

A NOVEL SOLUTION FOR AUTOMATIC SUMMARIZATION OF USER REVIEWS

Thesis Submitted For The Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Applications

By

Sonia Rani

Registration Number: 41900060

Supervised By

Dr. Tarandeep Singh Walia (25153)

Department of Computer Applications (Associate Professor)

LOVELY PROFESSIONAL UNIVERSITY



LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2025

DECLARATION

I, hereby declared that the presented work in the thesis entitled “A Novel Solution for Automatic Summarization of User Reviews” in fulfilment of degree of Doctor of Philosophy (Ph.D.) is outcome of research work carried out by me under the supervision of Dr. Tarandeep Singh Walia, working as Associate Professor, in the School of Computer Applications of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar):

Name of the scholar:	Sonia Rani
Registration No.:	41900060
Department/school:	Computer Applications
University:	Lovely Professional University

CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled “A Novel Solution for Automatic Summarization of User Reviews” submitted in fulfillment of the requirement for the award of degree of Doctor of Philosophy (Ph.D.) in Computer Applications is a research work carried out by Sonia Rani (41900060), is bonafide record of her original work carried out under my supervision and that no part of the thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor):

Name of supervisor: Dr. Tarandeep Singh Walia

Designation: Associate Professor

Department/School: Computer Applications

University: Lovely Professional University

Abstract

Automatic review summarisation is crucial to extract information from a large set of product reviews without losing the core meaning. Aspect-based product review summarisation uses only sentences related to product features and disregards irrelevant sentences. This thesis aims to bridge the gap by presenting a novel framework for aspect-based product review summarisation.

This thesis investigates English and Hinglish reviews on e-commerce websites. We curate novel review datasets based on actual reviews from popular e-commerce websites. We analysed the effects of various pre-processing and feature extraction techniques on the reviews. We explore traditional and existing feature extraction techniques. We analyse that traditional feature extraction techniques are insufficient to extract features from abbreviated and out-of-vocabulary words in reviews. The experimental analysis shows that the BERT model can provide embeddings for non-vocabulary and abbreviated words.

In this thesis, we aim to bridge the gap by proposing an aspect-based product review summarisation using a hybrid approach of rule-based and transfer learning methods. The proposed framework has developed a synthetic dataset of aspect-based summaries for multiple categories of product reviews, including sentiment polarity and aspects. We have also experimented with labelled and unlabelled topic modelling techniques on Hinglish code-mixed product reviews to extract important topics related to the products, and found that the labelled topic modelling methods can extract relevant topics more effectively related to products than unlabelled topic modelling methods. This can help to develop solutions for NLP applications in Indian languages.

We have evaluated summaries using several parameters: ROUGE, BLEU, BERT-F1 score, consistency, relevance, coherence, fluency, and human scores. We have also evaluated the predicted summaries using annotated summaries. Our results demonstrate that our hybrid approach, combining rule-based and transfer learning methods, yields state-of-the-art results compared to existing work. Our research has the potential to benefit e-commerce websites, manufacturers, and consumers by providing a comprehensive understanding of user intentions through an aspect-based summary of product reviews. Furthermore, our work contributes to the NLP community by advancing research in developing applications in the English language and Hinglish language.

Acknowledgments

First of all, I thank the Divine force Maa Chintpurni, who has led me to the path of pursuing my PhD and the countless blessings and mercies bestowed upon me. This endeavour would never have been possible without divine grace. I am always debited for her unwavering faithfulness in all conditions and for providing me with the extraordinary dedication level and courage to complete my PhD Journey. I would like to express my deepest appreciation to my Supervisor, Dr Tarandeep Singh Walia, for supporting and motivating me throughout my PhD journey. I'm grateful to him for their constant support and for giving me valuable feedback during interactive sessions, which have helped me to complete PhD degree and to identify my research plans and hurdles during data preparation, discuss ideas, and draft corrections.

I felt honoured to interact with Prof. Radhika Mamidi at IIIT Hyderabad during the Anveshan Setu fellowship program, where I completed a one-month internship under the supervision of Prof. Radhika Mamidi. I would like to express my deepest gratitude to Prof. Radhika Mamidi for encouraging and motivating me on every step, which I achieved in my PhD journey. I am always gratified to her for all the research discussions and guidance for my further career plans and objectives in the PhD clinic mentorship sessions. She has always given me the confidence that I am capable of doing research and can do better in exploring ideas for research direction and, of course, completing the PhD journey.

I'd like to acknowledge my husband, Upkar Singh, for his valuable support during the initial days of my research. He has always supported me in my entire journey toward my PhD. Words cannot express my gratitude to my lovely husband. Special thanks to my daughter Taranpreet Kaur for always being intimate and reminding me to achieve my aim in life to become a doctor. Thanks to each one of you for your support and motivation to finish my PhD. Many thanks to my son for supporting me in my entire PhD journey. I am also highly thankful to my mother for all types of genuine support in all circumstances. I am also thankful to my father, sisters, brother and his wife, who have always supported and motivated me throughout the PhD journey.

Furthermore, I would like to extend my sincere thanks to my department. I am fortunate to be a part of the Computer Applications Department at Lovely Professional University, Phagwara, Punjab. Here, I gained the confidence to learn new topics, which helped me do research. I am also grateful to Manikant Roy, an ex-faculty member at Lovely Professional University and currently a research scholar at IIT Delhi. He

always motivated and encouraged me throughout my PhD Journey. Thank you to my friend Angelina Gill for motivating me throughout my PhD life.

I would like to express my gratitude to Prof. Bhaveshkumar Dharmani for their great inspiration and guidance. I am also thankful to Prof. Gurdas Singh, who inspired me to start the PhD journey. His constant motivation and confidence in my abilities have been a driving force in my academic journey. This achievement wouldn't have been possible without the support of Prof. Ponnurangam Kumaraguru and Prof. Radhika Mamidi. I am grateful to ACM India for providing me with travel grants, accommodation, and delicious food for attending the ACM India annual events held in Bhopal and Bhubaneswar in 2023 and 2024. I felt honoured to have been selected by the ACM India team for the Anveshan Setu Fellowship program.

I would like to thank Dr. Lovi Raj Gupta, the Head of the faculty in the Lovely Faculty of Technology and Sciences at LPU, for showing me the right direction in my PhD Journey. Furthermore, I would like to thank Dr Apash Roy, who has guided me in starting my PhD journey. I am also thankful to Jaskirat Singh, the instructor at Axis India Machine Learning Company, for teaching me practical skills. Last but not least, I am so grateful to my pet Bruno, who is a loyal companion and whose presence has always given me happiness and enthusiasm. His emotional support has always made me emotionally and mentally strong throughout my PhD Journey.

Contents

Chapter	Page
1 Introduction	1
1.1 Types of Summarisation	2
1.1.1 Extractive summary	3
1.1.2 Abstractive Summary	3
1.2 Motivation for the Study	3
1.3 Problem Statement	4
1.4 Types of Reviews	4
1.4.1 Generic Reviews and Aspect-based Reviews	4
1.4.2 Single Product and Multi-Product Reviews Summarisation	5
1.5 Purpose of Summarisation	6
1.5.1 Domain-Specific	6
1.5.2 Generic Summary	7
1.5.3 Query-focused Summarization	7
1.5.4 Aspect-based Summarisation	7
1.6 Nutshell of Purpose of Summarization	7
1.7 Importance of Aspect-based Summarisation	8
1.7.1 Applications and Benefits of Aspect-based Summarisation	9
1.8 Approaches of Aspect Extraction	10
1.9 Research Challenges	11
1.10 Objectives of the Proposed Work	13
1.11 Scope of Proposed Work	13
1.12 Visualisation of Existing Summarization Datasets	14
1.12.1 Home and Kitchen Review Dataset	14
1.13 Findings of Existing Review Summarization Datasets	15
1.14 Motivation for Data Collection	15
1.14.1 Examples of Reviews on Amazon	16

1.14.2	Examples of Ladies' Clothing Reviews on Meesho	16
1.14.3	Mismatch Rating of Reviews on E-commerce Websites	17
1.15	Novelty Contributions of Thesis	17
1.16	Workflow of Proposed Work	18
1.17	Structure of the Thesis	19
2	Literature Review	21
2.1	Contributions of this Chapter	21
2.2	Approaches of Multi-review Summarisation	22
2.3	Aspect-based Extractive Review Summarisation	22
2.3.1	Aspect-Based Review Summarisation Using Rule-based	22
2.4	Reviews Summarisation Using Unsupervised Methods	25
2.4.1	Extractive Reviews Summarisation using Statistical Methods	25
2.4.2	Aspect-Based Review Summarisation Using Graph Methods	25
2.4.3	Aspect-Based Review Summarisation Using Clustering	26
2.4.4	Aspect-based Extractive Summary Using Topic Modelling	28
2.5	Review Summary Using Supervised Machine Learning	29
2.5.1	Strengths and Weaknesses of Machine Learning Approach	29
2.6	Aspect-based Abstractive Reviews Summarization	30
2.6.1	Aspect-based Reviews Summarization Using Deep Learning	30
2.6.2	Aspect-based Reviews Summarization Using Transfer Learning	31
2.7	Hybrid Aspect-based Extractive-Abstractive Summary	34
2.7.1	Strengths and Weaknesses of Hybrid Approach	36
2.8	A Nutshell of Existing Work on Summarization	37
2.9	Comparative Analysis of Aspect-based Summarisation	37
2.10	Other Related Work	40
2.11	Research Gaps	41
2.12	Analysis of Available Reviews Summarisation Datasets	42
2.12.1	Home and Kitchen Reviews Dataset Analysis	43
2.12.2	Clothing Shoes and Jewellery Dataset Analysis	43

2.12.3	Analysis of Flipkart Multiple Product Reviews Dataset	44
2.12.4	Women’s E-commerce Clothing Reviews	45
2.13	Pros and Cons of Available Datasets of Reviews	45
2.14	Conclusion	48
3	Text Preprocessing and Feature Extraction Techniques’ Challenges on Domain-Specific Data	49
3.1	Contributions of this Chapter	49
3.2	Introduction	49
3.3	Data Collection Challenges in While Scraping	50
3.4	Preprocessing Techniques	50
3.4.1	Data Cleaning	50
3.4.2	Data Normalisation	51
3.4.3	Contractions	51
3.4.4	Data Filtering of English and Hinglish Reviews	51
3.4.5	Spell Checking Using Pyenchant	51
3.4.6	Spell Correction Using TextBlob and Autocorrect	52
3.4.7	Experimentation with Tokenization	52
3.4.8	Experimentation with Stemming and Lemmatisation	53
3.4.9	Experimentation with Stop words removal	54
3.5	Experimentation with Feature Extraction Techniques	54
3.5.1	Features Extraction Using Rule-based Methods	55
3.5.2	Feature Extraction Using Bag-of-Words	55
3.5.3	Feature Extraction Using TF-IDF	56
3.5.4	Feature Extraction Using Word2vec Method	57
3.5.5	Feature Extraction Using FastText Embeddings	57
3.5.6	Feature Extraction Using BERT	58
3.6	Pros and Cons of Feature Extraction Techniques	59
3.7	Conclusion	61

4	Experimentation With Topic Modelling Techniques on Hinglish Code-Mixed Product Reviews	62
4.1	Contributions of this Chapter	62
4.2	Introduction	63
4.3	Examples of Hinglish Mobile Reviews	64
4.4	Topic Modelling Techniques	64
4.5	Experiments and Methodology	65
4.5.1	Data Collection	65
4.5.2	Exploratory Analysis of Hinglish Reviews	66
4.5.3	Preprocessing	66
4.5.4	Most Frequent Words Including and Excluding Stop Words . .	67
4.5.5	Unigram, Bigram and Trigram Count Analysis of Reviews . . .	68
4.5.6	Experimentation with Unlabeled and Labelled Topic Modelling	68
4.5.7	Experimentation Using LDA on Unlabelled Data	69
4.5.8	Topic Modeling using BERTopic	71
4.5.9	Topics Extracted Using Labelled Product-aware Topic Modelling	73
4.5.10	Sentiment Analysis and Labelled LDA Topics	76
4.6	Summary of Recent Studies on Topic Modelling	79
4.7	Comparison of Unlabelled/Labelled Topic Models	80
4.8	Conclusion	82
5	Aspect-based Product Reviews Summarization Using Hybrid Approach . . .	83
5.1	Contribution of this Chapter	83
5.2	Introduction	84
5.3	Proposed Methodology of Aspect-Based Reviews Summarisation	85
5.3.1	Data Collection	85
5.3.2	Text Preprocessing	85
5.4	Data Statistics and Exploratory Analysis	87
5.4.1	Reviews Categorical Distribution	87
5.4.2	Reviews Words' and Characters Length Distribution	87

5.4.3	Distribution of Reviews Word Length Grouped By Products and Likert Label	88
5.4.4	Distribution of Top Frequent Words	88
5.4.5	Visualisation of Combined English and Hinglish Reviews' Words	89
5.5	Data Filtering of English and Hinglish Reviews	89
5.5.1	Filtering of English Reviews	90
5.5.2	Filtering of Hinglish Reviews	90
5.5.3	Visualization of English Reviews' Words	90
5.6	Hybrid Approach for Summarization	91
5.6.1	Feature Extraction Using DistilBERT and Extractive Summary Generation	91
5.6.2	Sentiment Analysis of Reference Summary using the BERT . .	92
5.6.3	BART and Pegasus for Extractive to Abstractive Summary Generation	92
5.7	Synthetic Corpus of Abstractive Summaries With Features and Products	93
5.7.1	Trained the Data using Fine-tuning T5	93
5.7.2	Visualisation of Summaries Using Word Cloud	95
5.7.3	Distribution of Reference and Summaries Length	95
5.7.4	Sentiment Distribution of Summaries With Features of Products	96
5.7.5	Distribution of Features and Sentiment for Per Product	96
5.8	Summary Evaluation using Rouge Scores	97
5.9	Examples of Predicted Summaries	97
5.10	Conclusion	98
6	Results and Analysis	99
6.1	Contribution of this Chapter	99
6.2	Preprocessing, and Feature Extraction Results	100
6.2.1	Preprocessing Results	100
6.2.2	Feature Extraction Results	101
6.3	Topic Modelling Results on Hinglish Reviews	101
6.3.1	Limitations of Existing Work on Topic Modelling Results . . .	101

6.3.2	Impact of Including and Excluding Hinglish Stop words	102
6.3.3	Distribution of Coherence Score for Labelled Vs Unlabelled Topic Models	102
6.3.4	Extracted Topics Similarity Score and Coherence Mean Scores .	103
6.3.5	Comparison of Coherence Score of Topics Using Labelled and Unlabeled LDA, NMF and BERT	103
6.3.6	Jaccard Similarity Score and Mean Diversity Vs Coherence Scores	104
6.3.7	Comparison of Topics Coherence Score Using Labelled and Unlabeled Topic Models	104
6.4	Aspect-based Multi-product Reviews Summary Results	105
6.4.1	Distribution of Rouge, BERT and BLEU Scores of Predicted and Human Summary	105
6.4.2	Visualisation of Predicted Summaries With Human Summary .	106
6.4.3	Similarity of Rouge Score with BERT Score	106
6.4.4	Visualisation of all Metrics Scores With Summaries	107
6.4.5	Distribution of All Metrics Scores of Summaries	108
6.4.6	Visualisation of Evaluation Parameters Scores	109
6.4.7	Correlation of Rouge and BERT and All Metrics Score	109
6.4.8	Inter-agreement Analysis of Annotators	110
6.4.9	Meaningful Summaries Based on Threshold Value	111
6.5	Bench Marking	111
6.5.1	Comparative Analysis of Baseline with Proposed Work	112
6.5.2	Comparative Analysis of Proposed Work with Existing Work . .	112
6.6	Conclusion	113
7	Summary, Conclusion and Future Work	114
7.1	Summary	114
7.2	Conclusion	115
7.3	Future Work	115
	Bibliography	116

Appendices	131
A.1 Existing Datasets of Product Reviews Summarisation	132
A.1.1 Clothing, Shoes and Jewellery Dataset	132
A.1.2 Home and Kitchen Reviews Labelled Dataset	132
A.1.3 Flipkart Product Reviews Labeled Dataset with Sentiment	133
A.1.4 Amazon Food Reviews Summarization Dataset	133
A.1.5 Some Summaries with Evaluation Parameter's Scores	134
List of Publications	143
List of Conferences and Workshops	144

List of Tables

Table	Page
1.1 Nutshell of Purpose of Summarization	7
1.2 Importance of Aspect-based Summarization	8
1.3 Applications and Benefits of Aspect-based Summarization	9
1.4 Comparative Analysis of Aspects Extraction Approaches	10
1.5 Contributions of the Thesis	17
2.1 Contributions of this Chapter	21
2.2 Strengths and Weaknesses of Rule-based Approach	24
2.3 Strengths and Weaknesses of Traditional ML Approach	29
2.4 Strengths and Weaknesses of Deep Learning Approach	31
2.5 Strengths and Weaknesses of Transfer Learning Approach	34
2.6 Strengths and Weaknesses of Hybrid Approach	36
2.7 Researchers' Work Approach Wise	37
2.8 Comparative Analysis of Existing Work on Aspect-based Summarisation	37
2.9 Examples of Home and Kitchen Review Dataset	43
2.10 Reviews of Clothing Shoes and Jewellery Dataset	43
2.11 Some Reviews of Flipkart Reviews Dataset	44
2.12 Women E-commerce Clothing Reviews	45
2.13 The Pros and Cons of Standard Datasets	45
3.1 Contributions of Data Collection, Preprocessing and Feature Extraction	49
3.2 Pros and Cons of Feature Extraction Techniques	59
4.1 Contributions of this Chapter	62
4.2 Unigram, Bigram and Trigram Count Analysis of Reviews	68
4.3 Sentiment Analysis of Reviews and d Labelled LDA Topics	76
4.4 Coherence score of LDA NMF and BERT	79

4.5	Some Recent Studies of Topic Modelling Methods	80
4.6	Comparison of Topic Modeling Methods	80
5.1	Contribution For Aspect-based Summarisation	83
5.2	Examples of Predicted Summaries	97
6.1	Contribution of Each Experimentation Results	99
6.2	Analysis of Baseline with Proposed Work	112
A.1	Summaries with Evaluation Parameter Scores	134

List of Figures

Figure	Page
1.1 Types of Summarisation	2
1.2 Purpose of Summarisation	6
1.3 (a)Amazon Food Review Dataset (b) Clothing and Jewellery Dataset . .	14
1.4 home and kitchen reviews dataset	14
1.5 Examples of English and Hinglish Reviews	16
1.6 Examples of Reviews on Meesho	16
1.7 Mismatch Rating of Reviews	17
1.8 Block Diagram of Proposed Research Work	18
2.1 Approaches of Summarization	22
3.1 Spell Checking Using Pyenchant	51
3.2 Spell and Abbreviation Correction Using TextBlob and Autocorrect . .	52
3.3 Results of Tokenisation on Reviews	52
3.4 Stemming and Lemmatisation Results	53
3.5 Excluding Stop Words Results	54
3.6 (a) Pos tagging to extract features (b) Distribution of top 20 noun words	55
3.7 (a) Bag of words to get word vectors (b) Bag of words to extract feature words	56
3.8 (a) Feature Vector using TF-IDF (b)Feature Words using TF-IDF	56
3.9 (a) Word2Vec embeddings of nyc word (b) Word2vec can't give the OOV word embeddings	57
3.10 (a) Implementation of FastText (b) FastText didn't give embeddings of abbreviated words	58
3.11 (a) Feature words including stop words (b) Features excluding stop words	58
3.12 BERT Working	59
4.1 Some Examples of Hinglish Reviews	64

4.2	Techniques of Topic Modelling	64
4.3	Pipeline of the methodology of unlabelled and labelled topic modelling	65
4.4	Hinglish Reviews' Products Categories Analysis	66
4.5	(a)Word length distribution (b) Character Length distribution	66
4.6	(a) Word Length Distribution Excluding Stopwords (b) Word Length Distribution Excluding Stopwords	67
4.7	(a)Word Length distribution Excluding Stopwords (b) Most Frequent Words	67
4.8	(a)Sentiment Distribution of Hinglish Product Reviews	68
4.9	Document-term matrix using Unlabelled LDA	69
4.10	Visualisation of topics' words using pyLDAvis	70
4.11	Document-term matrix of extracted topics using unlabeled NMF	71
4.12	Score of topic's words	72
4.13	Hierarchical clustering of topics	72
4.14	Topics Words Extracted using Labelled LDA	73
4.15	Extracted Topics Using Labelled NMF	73
4.16	(a) Topics words extracted using LDA (b) Topic words extracted using NMF (c) Extracted Topic Words using BERT	74
4.17	Topic Words using Unlabeled LDA, NMF and BERT	74
4.18	(a) Saree, palazzo, duppata topic words using LDA, NMF and BERT (b) Ladies suit,lehenga choli, mobile topic Words extracted using LDA, NMF and BERT	75
4.19	Topics Words Extracted using BERTopic	75
4.20	Distribution of BERT Topics Word	76
4.21	Trending Line of Coherence Topics Scores using Labelled Topic Mod- elling	78
5.1	Block Diagram of Proposed Methodology	85
5.2	Cleaned Data	86
5.3	(a) Product reviews' percentage distribution (b) Reviews categories dis- tribution	87
5.4	(a) Words Count Distribution (b) Characters Count Distribution in reviews	87

5.5	(a)Word length of reviews grouped by products (b) Likert label of reviews distribution	88
5.6	Distribution of top 50 frequent words	89
5.7	Visualization of Hinglish and English words	89
5.8	Filtering of English Reviews	90
5.9	samples of Filtered Hinglish Reviews	90
5.10	Word Cloud of Multi Products English Reviews' Words	90
5.11	Reference Summary Generated using DistilBERT	91
5.12	Abstractive Summary generated using BERT, BART and Pegasus	92
5.13	Best summary with rouge score, product, feature and reference	93
5.14	(a)T5 Model Architecture (b) Trained Summaries using T5	94
5.15	(a) Training and Testing loss using epochs (b) Sentiment Distribution of Summaries	94
5.16	Word cloud of Individual Summaries	95
5.17	(a) Reference Word Length (b) Summaries Word Length	95
5.18	Sentiment Distribution of Summaries features with products	96
5.19	(a)Features distribution per product (b) Likert label feature distribution .	96
5.20	(a)Rouge Score of Individual Summaries (b) Mean Rouge Scores	97
6.1	Results of Preprocessing techniques	100
6.2	Visualisation of Embeddings using BERT	101
6.3	Reviews' Length with Including and Excluding Hinglish and English Stop words	102
6.4	Coherence Score of Labelled vs Unlabelled LDA, NMF and BERTopic for all Products	102
6.5	(a) Heat map of LDA NMF and BERT score (b) Mean Coherence Score of Labelled and Unlabelled Topic Models	103
6.6	Jaccard Similarity Score of Product-aware Topic Models	103
6.7	(a)Jaccard Similarity Score for Individual Products(b) Diversity and Coherence Mean Score of LDA, NMF, BERT Models	104
6.8	Coherence Scores of Labelled and Unlabeled LDA, NMF, BERT based on the Product Topics	104

6.9	(a) Distribution of Rouge, BLEU and BERT Score (b) Distribution of Rouge and BERT Score for Human Summary	105
6.10	Predicted Summary With Reference, Product and Features	106
6.11	(a) Similarity of Rouge1 and BERT score (b) Similarity of RougeL and BERT score (c) Similarity of Rouge2 and BERT score	107
6.12	Visualisation of all Metrics Scores	107
6.13	Distribution of all metrics scores	108
6.14	(a) Bigram-diversity score (b) Coherence score (c) Informative score . .	109
6.15	(a) Conciseness score (b) Fluency score with summary number	109
6.16	(a)Heat map of Rouge, BERT Metric Scores (b) Heat map of All Metric Scores	110
6.17	(a) Mismatch Sentiment Distribution of Reference Summary vs Predicted Summary (b) Reference Vs Summary Sentiment (c) Rater 1 and Rater 2 Scores	110
6.18	(a) Meaningful Summaries based on 55 Threshold Value (b) Meaningful Summaries based on 50 Threshold Value	111
6.19	(a)Comparative Analysis of Proposed Work (b) Comparison of Average score and Human Scores	112
A.1	Clothing, Shoes and Jewellery Dataset	132
A.2	Some reviews of Home and Kitchen Reviews Dataset	132
A.3	Flipkart Multiple Product Reviews Labeled Dataset	133
A.4	Multiple Product Reviews Labeled Dataset on Hugging Face	133

Abbreviations

ABRS	Aspect based Review Summarisation
ABS	Aspect based Summarisation
BART	Bidirectional and Auto-Regressive Transformer
BERT	Bidirectional Encoder Representations from Transformers
BiLSTM	Bidirectional Long Short-Term Memory Network
CNN	Convolution Neural Network
DL	Deep Learning
DT	Decision Tree
GPT	Generative Pre-trained Transformer
ILP	Integer Linear Programming
KNN	K-Nearest Neighbour
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
LSTM	Long Short-term Memory
ML	Machine Learning
NB	Naive Bayes
NLP	Natural Language Processing
NMF	Non-Negative Matrix Factorisation
PGN	Pointer Generator Network
PMI	Pointwise Mutual Information
PSO	Particle Swarn Optimisation
RNN	Recurrent Neural Network
RNTN	Recursive Neural Tensor Network
SVM	Support Vector Machine
T5	Text-To-Text Transfer Transformer
TF-IDF	Term Frequency Inverse Document Frequency
TF-ISF	Term Frequency Inverse Sentence Frequency
UCI	University of California Irvine machine learning repository

Chapter 1

Introduction

This chapter begins with an overview of the summarisation process, online shopping and product reviews on e-commerce websites. It describes the role of natural language processing techniques and machine learning methods in developing AI-based summarisation systems. Then, it discusses the background of the problem statement and the motivation of the study, data collection, and the research issues related to summarisation systems. Furthermore, the scope and objectives of the proposed work, as well as the contribution of the thesis, have been discussed. Lastly, the thesis structure is concluded.

Automatic summarisation is crucial for summarising large text content on e-commerce websites, scientific articles, blogs, and social media comments. Text summarisation is the process of extracting relevant and important information in a concise way without changing the core meaning of the text [1]. It plays a vital role in maintaining the escalating volume of textual data in digital technology. In fact, due to the development of vast AI-based technologies, both professionals and non-professionals utilise the technology to communicate and conduct online shopping, enabling time-saving and cost-effective transactions. We live in a digital era. Nowadays, most people, whether professionals or not, use gadgets constantly due to the exponential rise of digital devices and reasonably priced internet. A significant amount of unstructured text is generated online through product reviews, student feedback, and comments on social media posts.

Electronic gadgets and high-speed internet make online shopping more trending now, and it engages a large number of consumers. There are more reasons for online shopping; it saves buyers time and money, and it also gives discount offers. Returning products and refund policies are also increasing the number of online customers on e-commerce websites. Online shopping is more beneficial to buyers, manufacturers and e-commerce companies. After purchasing products, consumers write excellent or bad feedback in the form of reviews. There are an enormous number of reviews available on e-commerce websites (Amazon, Flipkart, Myntra, etc.) for a single product. Professionals write reviews in English, and those who are not professionals write the reviews in their local language. However, sometimes people also use code-mixing, like mixing one language into another language and writing reviews in Hindi using English characters. These reviews are called Hinglish reviews. All reviews, either in English or Hinglish, are beneficial for consumers as they allow them to spend their money buying good products at low prices and are also important for e-commerce companies and manufacturers.

Through these reviews, manufacturers and e-commerce companies can comfortably determine consumers' intentions and increase sales by improving the cons of the products. The problem is that if users want to buy more than one product, they must have to read many reviews. Reading these short or lengthy reviews is a typical and cumbersome task, and consumers and e-commerce companies can't read these reviews [2]. That's why automatic text summarisation systems are required here, and NLP techniques and machine learning methods are used to design the AI-based systems. The overarching goal is to highlight the crucial role of NLP and machine learning methods in e-commerce and how they can be leveraged to enhance the understanding of consumer intentions about products. While existing research work focuses on traditional methods to generate generic or domain-specific product review summaries, they do not succeed in providing feature-specific insights in the summary. E.g., a lady customer who wants to read reviews for a kurti might be interested in summaries which specifically about design, size or fabric quality. Existing summarisation methods do not present the granularity and personalisation required for summarisation based on such queries. Two types of summaries can be generated from single or multiple documents, which are described below.

1.1 Types of Summarisation

Automatic Summarisation can be performed from 2 types of text documents: Single document and multi-document [3]. Single or multi-document text can be summarised into 2 types: extractive or abstractive. Product reviews summary can be generated in both types. But the main point is to consider the product features that must be included in the summary. We have described the types of summaries in the figure below:



Figure 1.1: Types of Summarisation

Figure 1.1 presents the extractive and abstractive types of summaries that can be generated from single or multiple documents.

Single document Summarisation: It summarises a single article, story, news or blog. It is mostly common to summarise the news articles and blog summaries. The study performed the single-document summarisation on the DUC02 dataset [4, 5].

Multi-document Summarisation: Multi-document summarisation summarises multiple documents about several contents, and multiple documents on the same topic, like

multiple reviews about one or more products, movies, or social media posts. It helps to reduce redundancy and fuse multiple perspectives. Product review summarisation comes under the multi-document summarisation.

1.1.1 Extractive summary

This type of summary extracts the text's crucial sentences and generates the original text's gist [6]. It generates a fluent and concise summary from a large text, preserving important information. An extractive summary is generated without rephrasing the original text. This type of summary extracts the important sentences from the text by measuring the frequency using statistical machine learning or NLP methods. It extracts the frequency-based sub-parts of the sentences from the original text while containing no new words in the summary. An extractive summary is generated without rephrasing the sentences of the original text. It selects the frequency-based sub-parts of sentences from the original text while retaining the important information of the text. However, it finds the important sentences from the text, but it does not contain any new words in the summary. Moratanch and Chitrakala [7] described how important sentences and paragraphs are selected and summarised in the original documents in a concise form, and those sentences are selected based on statistical and linguistic features. The extractive summarisation extracted only very important sentences according to high scores and passages [8]. There are many approaches and methods of generating extractive summaries, which are described in the next chapter.

1.1.2 Abstractive Summary

An abstractive summary is created by understanding the text's semantics. An abstractive summary extracts the introductory sentences and paraphrases the original text. This summary type also generates new words that are not present in the original text [9]. Product review summaries can be generated in both types. However, the main point is to consider the product features must be included in the summary. However, both types of summaries contain important information from the text. But an Abstractive summary is more important than an extractive summary because it contains the semantics of the text. A detailed description of the abstractive summary techniques is presented in Chapter 2.

1.2 Motivation for the Study

Due to the popularity of online shopping and the exponential growth of product reviews, there is a need for automatic summarisation of product reviews, providing important insights for informed decision-making and business growth in the e-commerce industry. After exploring the review content on online e-commerce websites, analysing

and summarising a large amount of product reviews has become very important. The evolution of high computational power and the advancement of artificial intelligence technologies have a significant impact on the development of AI-based summarisation systems. These systems can save time for manufacturers, e-commerce companies, and consumers and provide important insights for decision-making and growth in business. To develop AI-based summarisation systems, natural language processing methods and machine learning algorithms are used to extract important information from user reviews and automatically summarise reviews into concise and meaningful sentences.

1.3 Problem Statement

Although researchers have been more focused on the product reviews summarisation problem in recent years, it has still been challenging to encompass the aspect-based summarisation of product reviews. There are several review summarisation datasets available, but they only cover limited solutions. This relies on a paradigm that builds model architecture based on English, Spanish, and German language reviews for generic summarisation only. Moreover, there is a need to develop aspect-based summarisation datasets and novel solutions to summarise product reviews based on the product's features. By gaining insights through the exploration of reviews on e-commerce websites and existing review summarisation datasets, we are striving towards collecting data from e-commerce websites. This thesis addresses the limitations of existing datasets by applying preprocessing and feature extraction techniques to real-world data. The proposed work aims to fill the gap in existing research by presenting a hybrid approach for aspect-based summarisation of product reviews in the English language and experimenting with unsupervised and product-aware topic modelling on Hinglish code-mixed product reviews.

1.4 Types of Reviews

There are two types of reviews: generic and aspect-based reviews. A detailed explanation of the review types is provided below.

1.4.1 Generic Reviews and Aspect-based Reviews

Generic Reviews: Generic reviews define the overall general feedback without focusing on specific features of the product. These reviews do not describe any specific aspects of the product; they only describe the overall feedback about the product or satisfaction. These types of reviews are only helpful in estimating the user's satisfaction with the product. These types of reviews do not describe what exactly the user liked or

disliked. These reviews are not particularly helpful in improving the product features. The examples of generic reviews are described below.

1. “This dress is fine. I like it”
2. “I didn’t like this dress at all.”

Aspect-Based Reviews: These reviews describe the specific features of the product, like size, design, fabric, stitching, etc. They are crucial for both purchasers and manufacturers in understanding what customers like and dislike about features. The reviewers discuss more than one aspect in the reviews. These types of reviews help identify specific features. Some examples of aspect-based reviews are given below:

1. “The size of the dress is correct, but the fabric quality is poor.”
Aspects: Size(positive), fabric quality (negative)
2. “Stitching is great, but the design is disappointing.”
Aspects: Stitching (positive), design (negative)

1.4.2 Single Product and Multi-Product Reviews Summarisation

Multi-review summarisation can be performed on multiple reviews about a single product or multiple products.

1. Single product Reviews summarisation: When a product has hundreds or thousands of reviews, it’s hard for users to read them all. Single product reviews summarisation is the process of generating an informative and concise summary from multiple user reviews of a product, service, or social media post. Instead of summarising a single document, story or article, it combines the information of many reviews to highlight the important topics, opinions, and sentiments to make a decision for purchasing goods. The example of multiple reviews about a single product is given below:

Review	Product
“The saree quality is amazing! but fitting is not accurate”	Saree
“The design is disappointing worst fabric and blouse fitting.”	Saree
“The saree length is okay, but the stitching quality of astar is poor.”	Saree
“The saree is very nice but design is outdated ”	Saree
“The phone battery is damaged and camera is poor”	Mobile
“The adult kurti stitching quality is poor but fitting is okay”	Adult kurti
“The kids dress zip is very loose and size is small”	Kids dress
“The kids kurti fabric is transparent but materiel quality is awesome”	Kids kurti

2. Multi-product Reviews summarisation: It is more complex and less common. It summarises the multi-product reviews of different categories, like clothing and elec-

tronics. These review summaries can be generated for each product individually. For E.g., saree, adult or kids kurti, kids dress, mobile, etc. The below table describes some examples of multiple product reviews.

Review	Product
“The suit is very nice but design is outdated ”	Lady suit
“The phone battery is damaged and camera is poor”	Mobile
“The adult kurti stitching quality is poor but fitting is okay”	Adult kurti
“The kids dress zip is very loose and size is small”	Kids dress
“The kids kurti fabric is transparent but materiel quality is awesome”	Kids kurti

1.5 Purpose of Summarisation

The primary context of summarisation is extracting important insights from text in various areas, including e-commerce, healthcare, and education. There are different purposes of summarisation, such as domain-specific, general, query-focused, and aspect-based. These types are described in detail as follows.



Figure 1.2: Purpose of Summarisation

Figure 1.2 presents the 4 purposes of summarisation: 1. Domain-specific 2. Generic Summary 3. Query-focused 4. Aspect-based Summary.

1.5.1 Domain-Specific

Domain-specific summarising is used when we want to summarise documents related to specific domains, like product reviews, movie reviews, medical, news, and hotels. However, product review summarisation is also called domain-specific summarisation. Zhang et al. [10] proposed DSGPT for summarising the domain-specific reviews and product titles. Nowadays, more important challenges are to summarise domain-specific summarisation due to unlabeled domain-specific datasets, such as those in the medical domain and product reviews. It also requires domain knowledge of those domains. Because, with the help of that domain knowledge, we can predefine some product features and, based on those attributes, summarise the product reviews within the context of that domain. Tailor the summarisation process to a specific domain,

E.g., electronics, fashion, restaurants. It focuses on vocabulary, structure, and aspects specific to the domain.

1.5.2 Generic Summary

The generic summarisation generates a generic summary based on important points in the overall text. This type does not focus on any query or any specific features and points. It summarises the overall document into essential content, only containing the main theme and idea from the overall document [11]. It is only useful for the general audience. It provides a generic summary without focusing on any query or specific aspects, and focuses only entire review text.

1.5.3 Query-focused Summarization

Query-focused summarisation is mainly used for question-answering summarisation. E.g., query on specific questions related to any topic and summarise only the information relevant to a specific query. It focuses only on the user query or interest. finding words related to the query using a query-attentive semantic graph. The query-focused summarisation is proposed by [12, 13], for specific domains.

1.5.4 Aspect-based Summarisation

Aspect-based Summarisation summarise the documents related to aspects rather than summarising the generic documents [14]. The purpose and use case example is that reviewers want to know what people say specifically about the size of the dress. It summarises the product reviews based on the specific aspects of a product mentioned in the reviews. E.g., in clothing product reviews, size, design, fitting, quality and fabric are the features, and reviews related to a mobile phone contain info about battery life, and camera quality, which are more important than the person about whom and what occasion this product is purchased [15].

1.6 Nutshell of Purpose of Summarization

Table 1.1: Nutshell of Purpose of Summarization

Domain-Specific	Product domain	Specific domain reviews	Domain-specific summary	Summarize reviews of dresses or TVs.
-----------------	----------------	-------------------------	-------------------------	--------------------------------------

Continued on next page

Table 1.1 – continued from previous page

Type	Focus	Input Format	Output Format	Use Case
Generic	Reviews	Multiple reviews	Condensed overall summary	Summarise the multiple reviews in short.
Query-Focused	User query	Reviews, Query	Query-answer summary	What people say about size and design of dress.
Aspect-Based	Specific product aspects	Multiple reviews	Aspect-wise summaries	Summarize the opinions related to any aspect like design,size or quality, battery.

Table 1.1 describes the purpose of summarization types, focus, input with output format of summary and use cases.

1.7 Importance of Aspect-based Summarisation

Aspect-Based Review Summarization is a natural language processing technique that extracts and summarises opinions about specific aspects or features of a product or service from user reviews. Instead of generating a general summary of a review, Aspect based Summarisation (ABS) breaks down the review into feature-specific insights, helping users understand exactly what people liked or disliked about each aspect.

Table 1.2: Importance of Aspect-based Summarization

Goal	Explanation
Extract Specific Information	Aspect based Review Summarisation (ABRS) identifies distinct features or aspects (e.g., "battery", "camera", "design") mentioned in reviews.
Summarize Per Aspect	It summarizes opinions for each aspect individually, instead of the whole document or review.
Provide Structured Insights	Helps convert unstructured, lengthy reviews into structured, digestible summaries.
Aid in Decision Making	Users can quickly understand what people like or dislike about a specific feature of a product.

Continued on next page

Table 1.2 – continued from previous page

Goal	Explanation
Enable Fine-Grained Sentiment Analysis	It links positive or negative sentiment to specific aspects rather than the whole review.

Table 1.2 defines the needs of aspect-based summarization with a detailed explanation.

1.7.1 Applications and Benefits of Aspect-based Summarisation

Table 1.3: Applications and Benefits of Aspect-based Summarization

Domain	Application	Benefits
E-commerce	Summarising product reviews by aspects (E.g., camera, battery, display for phones)	Helps buyers make informed decisions; improves trust and conversions
Hospitality	Summarizing hotel reviews (E.g., location, cleanliness, staff, amenities)	Tourists can compare hotels by features; helps in hotel selection
Restaurants	Summarizing food delivery or restaurant reviews by food, service, delivery, price	Improves customer satisfaction and restaurant discoverability.
Recommender Systems	Enhancing recommendations based on users' preferences	Suggests items based on what the user values most (E.g., comfort vs. price)
App Stores	Summarising user feedback for mobile apps by UI, performance, features	Developers identify pinpoints, users choose better apps.
Brand/Product Monitoring	Monitoring customer feedback on products/services over time	Companies can detect emerging issues or trends per feature.
Social Media Monitoring	Summarising public opinion on products/events by different aspects	Track public sentiment on issues (e.g., policy, launch events)
Market Research	Analysis, survey or review data across different products/features	Better product planning and customer segmentation.

Continued on next page

Table 1.3 – continued from previous page

Domain	Application	Benefits
Educational Feedback	Summarising course/instructor reviews by content, teaching style, assignments	Help to improve curriculum and teaching methods
Automotive Reviews	Summarising car reviews by mileage, comfort, safety, design	Assists car buyers with targeted comparisons.

Table 1.3 describes the applications and benefits of aspect-based summarization.

1.8 Approaches of Aspect Extraction

Aspect extraction is the vital step for Aspect-based summarization. It identifies the relevant aspect terms from the review text. However, there are several approaches to extracting the aspects from the reviews, which are categorised as follows:

Table 1.4: Comparative Analysis of Aspects Extraction Approaches

Approach	Type/Name	Strengths	Weaknesses
Rule-Based Methods	Unsupervised (Handcrafted rules or dictionary approach)	It can handle the domain-specific control, and it is simple, interpretable, and no labelled data is needed.	Not scalable, hard to apply on multiple domains.
Statistical Methods	Unsupervised (TF-IDF, TF-ISF)	Lightweight, interpretable, and can work on large data	It may extract non-informative frequent terms and ignores context.
Machine Learning Methods	Supervised (SVM, NB, KNN)	It learns patterns from labeled data, better than rules	It is less interpretable because requires labeled data and feature engineering.
Deep Learning Methods	Supervised (CNN, RNN, LSTM)	Captures context, semantic meaning, less manual feature engineering	Requires large labeled data, high computational cost.

Continued on next page

Table 1.4 – continued from previous page

Approach	Type/Name	Strengths	Weaknesses
Topic Modelling	Unsupervised (LDA, NMF, BERT)	No labelled data required, identifies latent hidden semantic topics (aspects)	It is hard to control granularity because topics may not always align with exact aspects.
Hybrid Approach	Mixed (Rule-based + Transfer learning)	Combines strengths of multiple methods, handles multiple domain rules and context	Complex integration, High cost, GPU, higher implementation effort, needs finetuning the model.

Table 1.4 explains the several approaches of aspects extraction with the strengths and weaknesses.

1.9 Research Challenges

In real-life scenarios, most professional and non-professional people do online shopping, and professional people write good or bad feedback about product features in English. On the other side, non-professionals write reviews in Hinglish. We have analysed that both English and Hinglish reviews are important for e-commerce companies and manufacturers to increase sales, and also beneficial for consumers to purchase goods. We have observed many research issues in existing research work, which are described as follows.

1. **Lack of Aspect-based Summarisation Datasets:** To our knowledge, there is a lack of domain-specific, especially women’s clothing, labelled datasets. Although there are several product review summarisation datasets available, but these are not sufficient to solve summarisation issues in real life. The summaries of those datasets are written in a generic manner and do not include sentences related to product features, which cannot solve the aspect-based summarisation issue. To solve the aspect-based summarisation problem, it is necessary to develop or annotate the aspect-based summarisation datasets [16, 17].
2. **Hinglish Reviews:** Another area that requires attention is the summarisation of Hinglish reviews. While there has been significant research on sentiment analysis of code-mixed Hinglish reviews, the field of Hinglish reviews summarisation remains largely unexplored. It presents a unique opportunity for future research and development in the field of natural language processing [18].

3. **Pre-processing techniques for summarisation:** Although a great deal of work has been done on pre-processing techniques on existing datasets. However, there are many challenges to applying preprocessing techniques to real-life data. Because real data often contains numerous abbreviations, spelling errors, contradictions, noise, and Hinglish reviews, it is very challenging to apply preprocessing techniques to noisy data. There are pros and cons of several preprocessing techniques, and they can be applied based on the application’s development stage [19–22].
4. **Feature Engineering and Extraction Techniques:** A key research issue that has emerged is the crucial role of domain knowledge in feature engineering and extraction techniques. The absence of such knowledge, poses a significant challenge in extracting features from different categories of product reviews, and it underscores the importance of understanding the product domain. Although extracting features from unstructured reviews is challenging, especially since these reviews are often written in native languages and contain abbreviations, grammatical and spelling errors [23, 24].
5. **Aspect-based summarisation:** There is admirable research done on product review summarisation, but we found that there are some bottlenecks in aspect-based summarisation techniques and aspect-based domain-specific datasets. We analysed that in reviews, many reviewers often include irrelevant information that is not important for the summary. For ex. I purchased it for my son’s birthday. Although it does not define any aspects so, these types of sentences should not be included in the summary. Existing techniques have some loopholes to handle these reviews and make it difficult to generate clear aspect-based summaries [25–28].
6. **Coherent, Fluent and Accurate Summary:** Existing approaches have faced the challenges of being semantic, informative, relevant, fluent, coherent and accurate predicted summaries [29, 30].
7. **Multiple Domain-Specific Needs:** Although profound work has been done on summarisation but there are still open research gaps to summarise the domain-specific datasets’ reviews. E.g., in the fashion category, there are domains for ladies’ clothing, kids’ clothing, and men’s clothing. In the electronics category, mobile reviews can also be found. The features of all these reviews are always different, so it is very challenging to include all the products’ features in the summary according to product type, and it requires a novel summarisation approach. [31, 32].

8. **Summarisation with Sentiment:** There are some bottlenecks in the existing work, and models sometimes struggle to include sentiment in the summary properly [33].
9. **Evaluation of Summary Quality:** Traditional metrics like ROUGE and BLEU do not fully capture the quality of aspect-based summaries. Human Evaluation is also required to evaluate the summary. More research is needed on human-centric and explainable evaluation metrics. [34, 35].

1.10 Objectives of the Proposed Work

The objectives of the proposed research work are listed as follows:

- To study and analyse existing multi-review summarisation methods.
- To extract and preprocess the data for text summarisation of multi-documents.
- To propose a novel solution for an automatic user review summarisation.
- To validate the robustness of the proposed solution on various performance parameters.

1.11 Scope of Proposed Work

The proposed research aims to develop an effective approach for aspect-based review summarisation for the English language reviews by combining rule-based, and transfer learning techniques. The framework is designed to extract fine-grained aspects and generate informative summaries across various product domains, such as fashion and electronics. This research work introduces a hybrid approach for aspect-based abstractive summarisation, capturing associated sentiments and product features to develop a synthetic corpus of summaries. Evaluation of summarisation quality is based on coherence, diversity, fluency, relevance, informativeness, ROUGE, BLEU metrics, BERT F1, and Human score, as well as inter-rater agreement analysis. This research work also focuses on Hinglish reviews and extracting important information to mine the Hinglish code-mixed product reviews. It utilises both weakly supervised and unsupervised LDA, NMF, and BERT topic modelling techniques to extract the meaningful topics. The evaluation of extracted topics is performed using coherence, diversity and Jaccard similarity scores. However, the scope is limited to the English and Hinglish languages, and multi-domain, especially clothing and mobile product reviews, and does not extend to multilingual product reviews.

1.12 Visualisation of Existing Summarization Datasets

There are many summarisation datasets publicly available. But through these datasets, real-life issues can't be solved because the real data contains so much noise and abbreviated words. The visualisation of the standard datasets is presented below.

Amazon Food Reviews Dataset Analysis			Clothing and Jewellery Reviews Dataset Analysis		
ReviewText	Rating	Summary	ReviewText	Rating	Summary
Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".	1	Not as Advertised	Just finished two tiers 6 inch and 10 inch cake for my daughter's birthday. I can't say enough thank you for this cake cutter. No hustle, easy to use and works like a magic. I love it and I would recommend it to everyone I know.	5	I love it.
If you are looking for the secret ingredient in Robitussin I believe I have found it. I got this in addition to the Root Beer Extract I ordered which was good and made some cherry soda. The flavor is very medicinal.	2	Cough Medicine	I'm not big on unique gadgets since running out of storage space in the kitchen, but this one beats all the workarounds for slicing cakes horizontally.	4	what can I say
My cats have been happily eating Felidae Platinum for more than two years. I just got a new bag and the shape of the food is different. They tried the new food when I first put it in their bowls and now the bowls sit full and the kitties will not touch the food. I've noticed similar reviews related to formula changes in the past. Unfortunately, I now need to find a new food that my cats will eat.	1	My Cats Are Not Fans of the New Food	I am also one of those reviews that found this unit need more work. second time i used it the receiver did not work. I remembered one of the reviews wrote there is a reset button on the back. well tried that and came back alive. also you have to turn on the receiver before the sending unit. otherwise it does not connect. for me i just need to know what the meat is doing in my bbq pit when im inside the house. we it works for me. accuracy and distance i still have to check out but im pretty sure its close. not on the nail....	3	my son loves it
Right now I'm mostly just sprouting this so my cats can eat the grass. They love it. I rotate it around with Wheatgrass and Rye too.	5	Yay Barley	My husband wanted this for Christmas to go along with the new grill he bought a month ago. It works great because we can sit inside and watch tv while the food cooks outside on the grill and check on the internal temperature from the comfort of our couch.	5	Christmas gift
			What can I say.. my daughters have it in orange, black, white and pink and I am thinking to buy for they the fuccia one. It is a very good way for exalt a dancer outfit great colors, comfortable, looks great, easy to wear, durables and little girls love it. I think it is a great buy for costumer and play too.	5	I have buy more than one

(a)

(b)

Figure 1.3: (a)Amazon Food Review Dataset (b) Clothing and Jewellery Dataset

Figure 1.3 (a) presents some review examples from the Amazon Food Reviews dataset, which includes various product summaries. The summaries only describe the product names, but no summary including the features of the product (b) shows some review examples from the Clothing, Jewellery and Shoes Dataset. This dataset summary is written for general purposes only and describes only the unimportant information in the summaries, excluding any product features.

1.12.1 Home and Kitchen Review Dataset

Home Kitchen Reviews Dataset Analysis		
ReviewText	Rating	Summary
My last toaster stopped working and, after research, I thought this a good choice. I am disappointed that it does not toast evenly on both sides.	2	new toaster
Got my mom this when she requested a new toaster. What can I say...it's a toaster...it toasts toast. Nothing has blown up or caught fire. It does what it is supposed to. She likes it. Would purchase again.	4	its a toaster
I wanted a toaster that was simple and this does the trick. If things start to burn it has a stop button and we like that. We have set it at a 21/2 temp and that toasts just fine.....Easy to clean just pull out the bottom tray and then I shake it over the sink for lose crumbs.	4	New toaster
Pros: Fit and finish are nice. Does not look cheap. Slots are adequately big enough. Multifunction if you want to do bagels or defrost frozen waffles. Easy to clean.Cons: cord is a little short, to brown toast. it seems like a very high indicator it takes longer than other toasters not pay more than \$25	3	Okay toaster
Toaster is compact and very basic. We use it for English muffins, which use the bagel setting. Seems to do a fine job. Has non-skid feet, if that is a plus for anyone.		Basic toaster
I had a more expensive toaster before. I hated it. This one is so much better and I like the white.	5	Toaster

Figure 1.4: home and kitchen reviews dataset

Figure 1.4 explains the sample reviews from the Home and Kitchen Review dataset. In this dataset summary with rating is explained. But in the summary is only written about the product like New toaster, Its a toaster, Okay toaster, Basic toaster and last summary is Toaster, Which does not describe any important information or not about any features of the products.

Although these datasets are publicly benchmarked standard datasets but they are not sufficient to solve real-life issues.

1.13 Findings of Existing Review Summarization Datasets

- Do the reviews in existing datasets have the same pattern as real-world reviews?
- Do existing datasets represent the abbreviations and noise, such as present in real reviews?
- Do existing datasets contain reviews in code-mixed formats like Hinglish?
- Are reviews in existing datasets written by real shoppers and hence represent their non-professional nature?
- Are the review summaries of existing datasets generic-based or feature-based?
- Do summaries of those datasets include the product names and product features? Which can help to solve the challenge of multi-product review summarisation.
- Does sentiment analysis of summaries described in those datasets?

1.14 Motivation for Data Collection

Despite the fact that there are many unlabeled and labelled review summarisation datasets available. However, some loopholes still exist in those datasets. Those datasets do not contain any updated reviews, especially summaries of reviews that are not written in accordance with product features. Existing datasets don't contain indian reviews, mainly in the women's clothing domain. It is the novelty of the research, which analyses English and Hinglish reviews on e-commerce websites. The main aim of data collection is to analyse the reviews and how professional and non-professional people write feedback about products. We have analysed and collected various product reviews from Amazon, Meesho and Flipkart, from which some reviews are in the English language and some are in the Hinglish language. Some examples of Hinglish and English reviews are shown below.

1.14.1 Examples of Reviews on Amazon

There are numerous reviews present for each product on e-commerce websites like Amazon and Meesho from which some samples of reviews are given below.



Figure 1.5: Examples of English and Hinglish Reviews

Figure 1.5 shows the some English reviews for clothing and Hinglish reviews for mobile phones present on the Amazon e-commerce website. It also presents the reviews as are latest up-to-date from 2020 to 2024.

1.14.2 Examples of Ladies' Clothing Reviews on Meesho

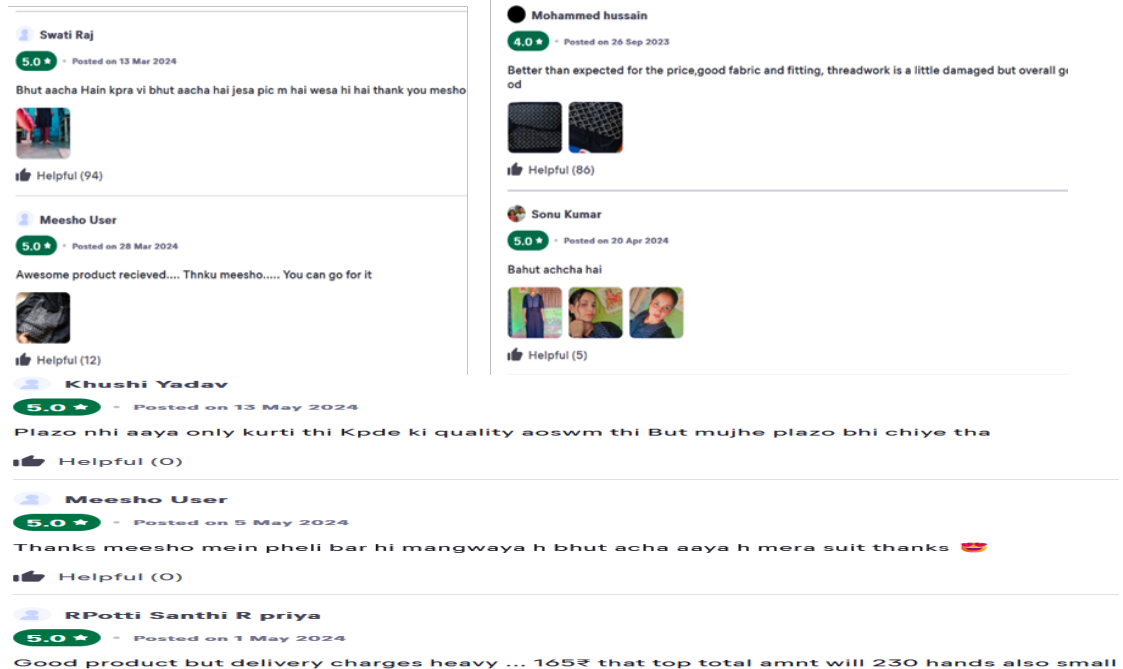


Figure 1.6: Examples of Reviews on Meesho

Figure 1.6 shows some review of ladies' clothing available on Meesho. These women's clothing reviews are mixed in English and Hinglish. To analyze the real-life issues for review summarisation, we have collected the latest reviews from the 2019 to 2024 time period.

1.14.3 Mismatch Rating of Reviews on E-commerce Websites

There are many reviews available on e-commerce websites whose rating is the opposite of the reviews. The following figure presents an example of an opposite rating from the reviews.



Figure 1.7: Mismatch Rating of Reviews

Figure 1.7 shows that the ratings are not matched with the reviews. In real-life data, many reviewers are not experts in writing reviews, and sometimes they give incorrect ratings based on their reviews. The figure shows that the first review title is 'theak hai' and the review text is 'sari theak hai zyada kuch khas nhi'. The reviewer has given a rating of 4 stars. In the 2nd review user has written 'Damage cloth is very poor'. However, the rating is 4 stars, which contradicts the review. To address this issue, we employed Likert scaling, ranging from 1 to 5, to categorise the most positive and negative reviews, along with their corresponding ratings.

1.15 Novelty Contributions of Thesis

Table 1.5: Contributions of the Thesis

Contribution Area	Novelty Description
Product-aware Topic Modelling on Hinglish Code-Mixed Reviews	Used unsupervised and supervised topic modelling (LDA, NMF, BERTopic) to extract coherent and interpretable topics from reviews and grouped topics based on specific product names with the sentiments. The evaluation is performed using coherence, diversity and Jaccard similarity scores.
Likert Scale of Labelling Reviews	The proposed work has used a Likert scale to identify sentiment polarity of the reviews.

Continued on next page

Table 1.5 – continued from previous page

Contribution Area	Contribution Description
Aspect-based Sentiment-oriented Reviews Summarisation	The proposed work has developed a framework for aspect-based summaries of multi-product reviews and created a synthetic corpus of summaries that include sentiments and features. The evaluation of summaries is performed using the Rouge score, BLEU, BERT score, and human scores.

The table 1.5 describes the novelty contribution of the thesis for aspect-based summarisation and product-aware labelled topic modelling on Hinglish code-mixed reviews.

1.16 Workflow of Proposed Work

The workflow of the thesis framework and proposed architecture is as follows:

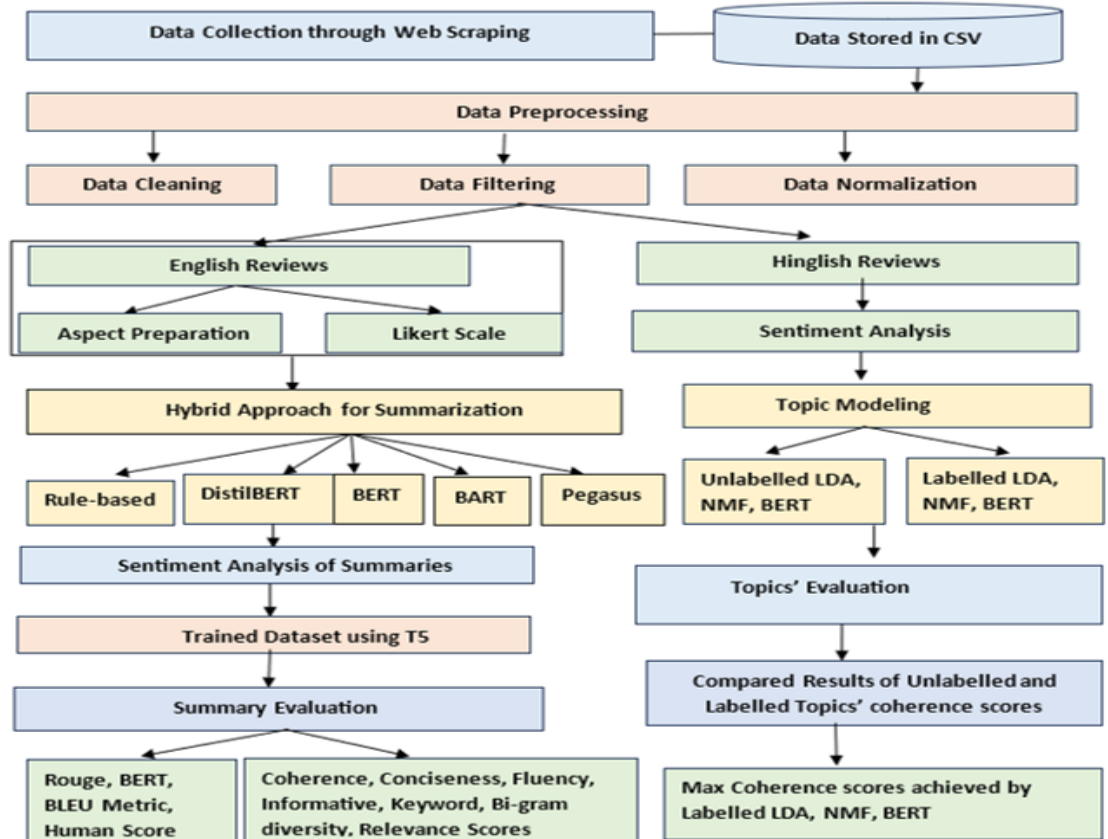


Figure 1.8: Block Diagram of Proposed Research Work

Figure 1.8 shows the steps of the proposed research work of aspect-based summarisation of English language reviews and topic modelling on Hinglish code-mixed prod-

uct reviews, which are described as follows:. The English language reviews are summarised through a hybrid approach, and important topics are extracted from Hinglish code-mixed product reviews using topic modelling techniques.

1. The 1st step of this research is to collect the latest multi-domain product reviews from the Meesho, Amazon, and Flipkart e-commerce websites.
2. Load unstructured data into a structured format in a CSV file, and experiment with the various preprocessing techniques and filter the English and Hinglish reviews and save into different data frames.
3. The Likert scale is applied to the English reviews to determine sentiments, and a list of aspects for different products is prepared.
4. Apply the unsupervised and product-aware topic modelling techniques on Hinglish code-mixed product reviews. To remove hinglish stop words, we have created a Hinglish domain-specific stopwords list. Coherence scores are used to evaluate the topics.
5. Perform aspect-based summarisation using a hybrid approach to summarise the product reviews in the English language.
6. Evaluation of predicted summaries is made using the Rouge, BERT and BLEU, Human score and coherence, fluency, informativeness, keyword, and fluency parameters.

1.17 Structure of the Thesis

The remaining thesis chapters, along with a brief overview, are presented below.

1. **Chapter 2. Literature Review:** This chapter discusses the literature survey of existing summarization approaches, methods and datasets of review summarisation. This chapter also analyses the limitations of existing review summarization datasets.
2. **Chapter 3: Text Preprocessing and Feature Extraction Techniques’ Challenges on Domain Specific Data:** This chapter analyses the challenges of collecting reviews from e-commerce websites and applying various preprocessing and feature extraction techniques.
3. **Chapter 4: Experimentation With Topic Modelling Techniques on Hinglish Code-Mixed Product Reviews:** In this chapter, we present the experimentation of unsupervised and labelled topic modelling techniques on Hinglish Code-Mixed product reviews.

4. **Chapter 5: Aspect-based Product Reviews Summarization Using Hybrid Approach:** In this chapter, we propose the hybrid approach for aspect-based product review summarisation, and we have experimented with the rule-based and transfer learning methods.
5. **Chapter 6: Results and Analysis:** This chapter describes the evaluation parameters, including the Rouge metric, coherence, conciseness, relevance, informativeness, keyword coverage, and Bi-gram diversity, for evaluating summaries.
6. **Chapter 7: Summary, Conclusion and Future Work:** In this chapter, we present the summary of our proposed work with a conclusion and future scope for improving the results in this research area.

Chapter 2

Literature Review

This chapter elucidates the role of various natural language processing and machine learning methods in developing automatic summarisation systems. The main aim of this chapter is to explore the techniques of multi-document extractive and abstractive summarisation and datasets of review summarisation. This chapter focuses on identifying the best approach from all summarising techniques and also examines the summaries of benchmark data sets. It also provides a comparative analysis of rule-based and machine learning, deep learning and transfer-learning methods. The detailed analysis of existing review summarisation datasets, with the pros and cons, is also elaborated in this chapter.

2.1 Contributions of this Chapter

Table 2.1: Contributions of this Chapter

Terms	Contribution
Review	Finding the techniques and methods of multi-document summarisation, and found that the latest transfer learning methods are most suitable for summarisation.
Analysis of Reviews in Data Sets	A detailed analysis of product review summarisation data sets and found the difference b/w those datasets' reviews style and grammar and real-world reviews' style.
Style of Summary	We have analysed publicly available datasets, and the summaries of those datasets are written in a generic style instead of a feature-based style, which is not suitable for aspect-based summarisation.
Multi-domain product	We found that the existing datasets are mostly focused on electronics products, and we contribute to collecting data from multiple domains, including traditional ladies' dress reviews in the fashion category and mobile reviews.

Table 2.1 describes our contributions in this chapter. It describes the review of all techniques for summarisation, and then a deep analysis of publicly available summarisation datasets.

2.2 Approaches of Multi-review Summarisation

There are many approaches and methods of generating extractive and abstractive summaries. Unsupervised techniques, such as graph-based and statistical methods, are used to generate an extractive summary. Lin and NG [36] surveyed existing approaches to generating abstractive summaries, with a focus on recent neural approaches. Rani and Walia [37] explained several summarisation techniques of automatic abstractive and extractive summarisation. Shakil et al. [38] highlighted the challenges of controllable abstractive text summarisation, including inadequate meaning, cross-lingual summarisation, factual consistency, long-document summarisation, evaluation metrics, and handling noisy data.

There are many approaches and methods of generating extractive and abstractive summaries that are described as follows:



Figure 2.1: Approaches of Summarization

Figure 2.1 shows that several approaches like Rule-based, graph-based, machine learning, deep learning, transfer learning and hybrid techniques are used to summarise user reviews.

2.3 Aspect-based Extractive Review Summarisation

Aspect-based Summarisation summarise the documents related to product aspects rather than summarising the documents according to frequency or a generic summary. To generate meaningful summaries, the aspects must be included in the Aspect-based summary.

2.3.1 Aspect-Based Review Summarisation Using Rule-based

The Natural Language Processing (NLP) rule-based approach is the primary and traditional approach. This approach is used for extracting features using predefined linguistics rules or part-of-speech tagging, such as nouns, verbs, or adjectives. This technique is mainly used for sentiment analysis and the classification of positive and negative reviews. Patkar et al. [16] applied a semi-supervised approach to generate a feature-based summary for product reviews. They used the Link Grammar parser to extract frequent nouns and noun phrases. Proper nouns and named entities present good

features involved in the summary. A pos tagger was used to identify the proper noun feature. Stanford NER was used to identify named entity features. Htay et al. [24] extracted explicit features of 5 electronic products using adjectives, adverbs, nouns, words and the reviews collected from Amazon. Hou et al. [25] used linguistic patterns to select the highly correlated noun features. They annotated the Kindle e-reader 265 review labels. Fleiss' Kappa metric was used for the evaluation. Their proposed framework summarised the reviews based on customers' aspects and preferences. They extracted feature-based relationships between sentiment and feature words and assigned the scores according to dependency relationships. The shortcoming of their approach is that a few labelled reviews and implicit aspects are not considered. They suggested that extracting the product's features, such as affordance, price, validity, linguistic patterns, and ground truth results, can be used for further study.

Sheikh et al. [32] proposed a rule-based lexicon approach for extracting features from mobile Samsung reviews, classifying them as positive, negative, and neutral using Sentiwordnet. Their approach can be improved with a supervised approach and an attempt to build a predictive model. Hu and Liu [39] proposed a rule-based part-of-speech tagging approach with an a priori algorithm to extract the features of products and summarised the opinions related to features into positive and negative categories using bootstrapping, association mining, and Word Net. They manually annotated the features of 5 products and evaluated the summary by reading all the reviews. Konjengbam et al. [40] proposed the aspect-based ontology tree model to summarise the product reviews. Part-of-speech tagging, combined with grammatical rules, was used to extract opinions, words, and aspects. Still, one limitation of their approach is that it overlooks the identification of synonyms and hierarchical relationships between different aspects.

Hanni, Patil, and Patil [41] proposed a dynamic system for a feature-based extractive summary of the web link of a product, using Pos-tagged words that are more related to that product's features. Reviews' polarity and rating of reviews are included in the summary. Chung and Tseng [42] developed a business intelligence system with a rule-based technique for information retrieval. They employed association rule mining, specifically the A priori algorithm, to extract features that revealed the relationship between customer reviews and ratings. Siautama et al. [43] applied the pos tagging with NLTK and extracted features by combining the adjective-noun words. Hong and Wang [44] utilised POS rules to generate extract phrases from English and Chinese reviews. 3 product reviews (hotels, Apple MacBook, Movies) are in Chinese, and another 3 (Cars, Jiudian, office equipment) are in English. Chigateri and Bhandarkar [45] proposed the rules to extract and classify the necessities and preferences from reviews. Modal verbs 'has to' define the necessity and 'would' defines the preference. They classified preferences with 99.78% accuracy and necessity with 91% accuracy based on reviews. Amarouche et al. [46] applied the pos tagger to extract the features using

nouns and converted plural nouns to singular nouns. One important key strength of the system proposed by them is that it generates a summary that includes product features, sentiment, and the period. Singh et al. [18] applied Pos tagging to extract the nouns, verbs, adjectives, and adverbs stored in a frequent feature database. They experimented on 1100 reviews of Android phones, crawled from Snap deal, Flipkart, Shop Clues, and Amazon and classified the positive reviews with 62.60% and negative reviews-32.68 accuracy.

2.3.1.1 Strengths and Weaknesses of Rule-based Approach

Table 2.2: Strengths and Weaknesses of Rule-based Approach

Intro/Work	Strengths	Weaknesses
Rule-based systems identify important information from reviews using POS tagging and linguistic patterns.	• It is easy to design rules for specific domain data, and this approach yields maximum accuracy. • It learns from data and can handle a wide range of language patterns of summary styles. • It can generate summaries with high accuracy and can handle large, complex datasets. • It can be trained on small labelled data of the specific domain and can easily adapt to new domains with high performance. • Easy to interpret and no training data required and good for controlled domains.	• To design the rules, domain and linguistics knowledge is required for data training. • Requires quality of labelled training data. • Requires high cost, max computational resources and large amounts of data. • It requires domain-specific data for fine-tuning and is complex to implement and fine-tune effectively. • Not scalable and fails with noisy or informal text and Difficult to adapt.

Table 2.2 defines the strengths of the rule-based method is its simplicity and interpretability, which make it suitable for small, well-structured datasets. However, the weakness of approach it is not adaptable to informal language or large-scale data, limiting its utility in dynamic environments.

2.4 Reviews Summarisation Using Unsupervised Methods

2.4.1 Extractive Reviews Summarisation using Statistical Methods

The statistical approach is mainly used for generating extractive summaries. This approach extracts sentences based on the frequency of words and probabilistic distributions to identify importance sentences and generate summaries. These methods do not require labelled training data and are applied to calculate the frequency of sentences. Siautama et al. [43] proposed a Term Frequency Inverse Document Frequency (TF-IDF) method to summarise the hotel reviews in an extractive format and phrase summary style by pairing adjectives with nearest nouns and considering polarity. TextBlob was used to identify sentiments, and evaluation was performed through BLEU, ROUGE-1, and ROUGE-2 recall scores. Li et al. [47] introduced TF-IDF with a weighted directed graph to generate an extractive summary of movie and product reviews without losing the sentiments by considering the sentiments' scores. Meena and Gopalani [48] proposed Term Frequency Inverse Sentence Frequency (TF-ISF) method to generate an extractive summary of the DUC2002 corpus. They adopted the priority-based feature filtering method to generate a multi-document summary.

Dalal and Zaveri [49] utilised a semi-supervised approach for mining user reviews to generate feature-based statistical summaries. Cataltas et al. [50] used the TF-IDF method to convert words into vectors. Reviews collected from the Amazon Review dataset. The polarity of reviews is determined by the sentiment of scores using a dictionary provided by VADER. Experiments were conducted on two Amazon review datasets: shoes and electronics. They suggested that in future, find out the hidden sentiments of consumers.

2.4.2 Aspect-Based Review Summarisation Using Graph Methods

Graph-based methods represent textual elements such as words, sentences, or aspects as nodes, and their semantic or syntactic relationships as edges in a graph structure. Text Rank, inspired by the PageRank algorithm, is one of the most well-known graph-based algorithms developed by Google, which ranks sentences or phrases based on their centrality in a similarity graph. Konjengbam et al. [40] employed the Latent Semantic Analysis (LSA) method to identify hidden topics within the text and reduce the matrix dimensions. Mobile, camera, DVD player and jukebox reviews were collected for experiment purposes. One limitation of their approach was that it neglected to identify the synonyms and hierarchical relationships between different aspects. Al-Dhelaan and Al-Suhaim [51] proposed the topic signatureRecursive Neural Tensor Network (RNTN) method to generate an extractive summary containing the sentiments.

The Experimentation was performed on Opinions, Edmunds corpus, and hotel reviews collected from TripAdvisor.com. An Evaluation was performed using Rouge scores and Amazon Mechanical Turk. They also compared the results of their proposed graph-based topic signature ranking algorithm with Lex Rank, Text Rank, Sum Basic, and TF-IDF algorithms.

Modani et al. [15] proposed a variant of the Page Rank method to generate an extractive summary. They employed a greedy algorithm to select the maximum value of sentences through the graph nodes and extracted the imp sentences and assigned scores to them. The highest-weighted scores of words are added to the summary. Evaluation was done through KL-divergence and compression rate, and retention methods. Ansari et al. [52] proposed the graph-based semi-supervised algorithm for extracting aspects from reviews of restaurants and laptops' domain-specific datasets. Thessrimuang and Chaowalit [53] used the PageRank and hop field method to extract the essential sentences from 459 cosmetic product reviews collected from Twitter through the Twitter API. The cosine similarity method is used to find the most important phrases and rank these phrases, and find the pros and cons of each product through the phrases. They evaluated system results with Rouge-1,2, Rouge-SU4 and KL-Divergence scores. The limitation of their approach was the presence of grammatical errors in the unstructured social media data. Another limitation is the less frequent use of synonyms.

2.4.3 Aspect-Based Review Summarisation Using Clustering

Clustering is an unsupervised machine learning technique that groups similar items together based on their features. Clustering involves grouping similar sentences and is widely used to group semantically related opinion sentences or aspect terms from the unlabelled data. This helps in organising unstructured review data into coherent, aspect-specific summaries. Cataltas et al. [50] proposed the DBSCAN clustering algorithm to identify negative features through product defects, generating a summary of product defects. Manjupriya et al. [54] proposed a parallel approach using the Hadoop cluster to extract the features and opinion summarisation. The strength of their approach is that they focused on scalability over summary quality, but it lacks semantic understanding. They suggested that accuracy can be improved by extracting more accurate features, including opinion words from adverbs and nouns.

Meng and Wang [55] introduced the clustering method to tag products and extract explicit and implicit features appearance and price through product specifications fetched from Amazon. Angelidis and Lapata [56] introduced two weakly supervised approaches to identify salient opinions and generate extractive summaries across six diverse domains. They developed a novel dataset annotated by a human annotator with gold-standard aspect annotations. They created the new dataset named OPOSUM to

evaluate the summary of Amazon reviews from the Amazon product dataset. Their proposed framework combined the seeded aspect extractor based on a multi-task objective without directly supervision. There is a need to develop a more integrated approach that can jointly identify sentiments and aspects across multiple languages and domains, and to develop methods using weak supervision to generate abstractive summaries.

Li et al. [57] proposed Aspect-based summarisation by utilising the token cluster method and BERT model to extract aspect-related words to generate an unsupervised extractive summary. They extracted the root of noun phrases and adjectives related to aspect word features from reviews. E.g., the noun phrase 'front desk staff' the noun is considered its root- 'staff'. The limitation of the approach is that noun phrases can't extract all product aspects, and the a lack of domain knowledge implementation for extracting meaningful aspects. Tsai et al. [58] investigated the systematic approach to summarising hotel reviews according to features. They created a classifier that identified helpful reviews and classified sentiment polarities. Their approach filtered the reviews by hotel quality, such as cleanliness, service, traffic, and indoor and outdoor facilities.

Bafna and Toshniwal [59] proposed the feature opinion mining approach based on the product domain. Their system extracted the relevant reviews according to the clustered features of a product. They proposed a system that automatically extracted features based on the polarity of reviews. They clustered the aspect-related words. They applied the token length, both with and without sentence order, along with a word length penalty. Krishnakumari and Sivasankar [60] handled large datasets of reviews, used Hadoop to cluster features and opinions and generated a feature-based summary. Phong et al. [61] considered the MARK web-based online tool, along with a clustering keyword-based framework, to analyse semi-automated reviews. Their proposed framework automatically extracted the keywords from 300,000 mobile reviews gathered from Google Play. Their proposed system ranked the associated negative reviews using a k-means clustering algorithm. The evaluation was conducted to determine the ratings using Pearson correlation and assess the distribution of ratings for skewness.

Coavoux et al. [62] proposed the k-means clustering algorithm for clustering sentences based on aspect and summary generation. Gautam et al. [63] proposed clustering approach to extract features. They employed the cosine similarity method to identify the most similar sentences and rank them accordingly. Their approach clustered the sentences according to features, and top-ranked score sentences were selected to create an extractive summary. Souza and Manzato [64] proposed Aspect-based extractive summarisation using POS tagging and the hierarchical agglomerative clustering technique. They used Euclidean distance, and single linkage to extract aspect-based sentences from reviews. Their approach clustered the labels of aspects using TF-IDF and centroid methods, and selected sentences related to at least one aspect. Experimentation

was performed on the Opinions, Space, and Amazon datasets. An evaluation was done using Rouge scores. The limitation of their proposed work is that they considered only a few aspects to generate summaries of different domains.

2.4.4 Aspect-based Extractive Summary Using Topic Modelling

Topic modelling is an unsupervised machine learning technique used to discover hidden topics, such as price, design, battery, or thematic structure, from unlabeled reviews. Unlike supervised approaches that require labelled aspect terms, topic modelling automatically learns the distribution of words across topics and the distribution of topics across documents. Models such as Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorisation (NMF), and more recently, BERTopic and Top2vec, have been widely used for extracting interpretable and coherent topics from customer reviews. Yang et al. [27] proposed the two-layer categorical topic modelling (CAT-LDA) Method to extract the aspects of the category. Experiments were conducted using the Amazon Review dataset. Wang, Zhu and Li [65] proposed the NMF method for generating feature-based extractive summaries based on time, price, size and quality without losing any information. They used the centroid method to find the frequent adjective features. The experiment was conducted with three types of rice cookers. The evaluation was conducted using DUC05, DUC06, Rouge-1, Rouge-2, W, and Rouge-SU. Belwal, Rai and Gupta [66] employed the LDA topic modelling method and semantic measure to generate the extractive summary. Experimentation was performed on Opinions, Cars and the CNN/Daily Mail dataset. Cosine similarity, Jaccard coefficient, and Euclidean distance were used to measure the similarity b/w sentences. In future, their proposed method can be implemented on other language datasets also, and topic modelling (graph-based method) can be incorporated to extractive summarisation.

Jin, Huang and Zhu [67] employed the labelled LDA method to identify critical explicit aspect topics from the Edmunds dataset of car reviews. Mukherjee et al. [68] used an unsupervised topic model and Integer Linear Programming (ILP) method to extract coherent sentences around certain aspects from tourist reviews. They manually labelled the identified aspects with fine-grained labels and categorised the topics into aspect classes. Chin, Bhowmick and Jatowt [69] proposed the LDA topic modelling to generate the topics of tweets and ranked the tweets. Akhtar et al. [70] proposed the LDA topic modelling method to find hidden topics and aspects for hotel review summarisation and classified the predefined classes into some common aspects. They crawled hotel reviews from TripAdvisor, GoIbibo, Yatra, and other similar websites. Their approach summarised what users expect from hotels when they visit, or do not, according to their budget. The main gap in their approach is that aspects are not extracted based on users' preferences, and they suggested that aspects can be extracted according to users' preferences.

2.5 Review Summary Using Supervised Machine Learning

Machine Learning (ML) plays a vital role in the development of summarization systems. It enables the learning of patterns from annotated review data and makes predictions about relevant aspects, classifying the polarity of reviews, sentiments, and summary-worthy content, without relying solely on manually crafted rules. ML methods such as Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT) are used to classify the aspects based on the predefined aspects and for opinion mining.

Hanni, Patil and Patil [41] developed an Android app called Help Buy, which can take a link to the web page of an Amazon product and display a summary of all reviews on that page. They used NB, K-Nearest Neighbour (KNN) classifiers to classify reviews into positive, negative and neutral and summarise customer reviews based on the features of the product. Thessrimuang and Chaowalit [53] employed the Naive Bayes algorithm to classify the polarity of cosmetic products. The cosine similarity method is used to find similar phrases. Tsai et al. [58] developed a framework to filter the reviews and grouped them by hotel quality, cleanliness, service, traffic, and indoor and outdoor facilities. They employed the Decision Tree, SVM, Random Forest and Logistic Regression methods to categorise the helpful sentences and classified sentiment polarities.

2.5.1 Strengths and Weaknesses of Machine Learning Approach

Table 2.3: Strengths and Weaknesses of Traditional ML Approach

Type of ML	Intro/Work	Strengths	Weaknesses
Supervised Learning	Input and output labels are available. Commonly used in ABS for aspects classification and sequence labeling tasks.	• High task-specific accuracy • Supports fine-grained sentiment analysis • Predictable performance	• Requires large labeled datasets • Not domain adaptable • Annotation is time-consuming

Continued on next page

Table 2.3 – continued from previous page

Type of ML	Intro/Work	Strengths	Weaknesses
Unsupervised Learning	No labeled data required. Often used in ABS via topic models (e.g., LDA, NMF) to discover latent aspects from reviews.	• No need for manual labels • Good for aspect discovery • Works on large datasets	• Output may be noisy or incoherent • Hard to evaluate automatically • Lacks sentiment polarity by default

Table 2.3 describes the Supervised and Unsupervised ML role, pros and cons for aspect-based summarisation.

2.6 Aspect-based Abstractive Reviews Summarization

2.6.1 Aspect-based Reviews Summarization Using Deep Learning

Deep learning has significantly advanced the field of aspect-based summarization by enabling models to learn hierarchical and semantic representations of text data automatically. Unlike traditional rule-based, statistical and machine learning methods, deep learning models do not rely on manual feature engineering and utilise architectures such as Recurrent Neural Network (RNN), Convolution Neural Network (CNN) to capture contextual, sequential, and aspect-specific patterns within reviews.

Mabrouk, Redondo and Kayed [17] employed RNN, CNN, SVM, and a template for hierarchical aspect-based opinion mining polarity. They crawled the reviews of 5 companies’ laptops using a web scraper, extracted aspects from product templates. A key strength of their work is that their proposed framework can be applied to movies and restaurant domains. However, a notable limitation is that their work only performed the aspect classifications using explicit and implicit aspects. Boorugu and Madhavi [30] proposed the Seq2seq model and Conceptnet embedding to generate an abstractive summary. They experimented on the standard Amazon food Review corpus, and evaluation was performed using the Rouge Metric. They recommended that, in the future, the BERT model can be used to get a more accurate summary rather than a Seq2seq summary.

Bravzinskas, Lapata, and Titov [71] proposed a Seq2seq with LSTM methods to generate an abstractive summary. They experimented with Amazon and Yelp datasets,

and evaluation was performed using the Flesch Kappa metric. Bravzinskas et al. [72] employed the copycat algorithm for generating an abstractive summary. Boorugu and Madhavi [30] proposed the Seq2seq model and Conceptnet embedding to generate an abstractive summary. They experimented on the standard Amazon food Review corpus, and evaluation was performed using the Rouge Metric. They recommended that, in the future, the BERT model can be used to get a more accurate summary rather than a Seq2seq summary.

2.6.1.1 Strengths and Weaknesses of Deep Learning Approach

Table 2.4: Strengths and Weaknesses of Deep Learning Approach

Intro/Work	Strengths	Weaknesses
Neural networks (RNN, CNN, LSTM) automatically learn features.	• End-to-end learning and can capture complex patterns. • Effective with large-scale text.	• Requires large training datasets and low interpretability. • Slow training and tuning.

Table 2.4 describes the deep learning approach benefits from its ability to capture complex contextual relationships, leading to more accurate summaries. The major drawback, however, is the requirement for large labelled data for training and computational resources, which can hinder real-time deployment.

2.6.2 Aspect-based Reviews Summarization Using Transfer Learning

Transfer learning is an advanced version of deep learning methods with an attention mechanism that can be used with small annotated datasets by fine-tuning a model on a target task. Instead of training a model from scratch, knowledge acquired from solving a general problem is transferred to a domain-specific or downstream task. In the field of Natural Language Processing (NLP), transfer learning has gained tremendous popularity with the emergence of pretrained language models such as BERT, RoBERTa, T5, and PEGASUS models are pretrained on massive unlabeled text corpora to learn deep contextual representations, and then fine-tuned on specific tasks such as summarization, sentiment analysis, or aspect-based opinion mining. In the context of aspect-based summarisation, transfer learning methods enable the understanding of semantic nuances and aspect-specific opinions, even with limited domain-specific data. This makes it a crucial strategy for developing robust and scalable NLP systems in real-world applications.

Amplayo, Angelidis and Lapata [14] proposed an Aspect-based abstractive summarisation based on the queries about aspects. They created a synthetic review corpus paired with aspects and a controller that predicted the aspects in a document and developed the pseudo summary. Sileidar, Makwana and Bhattacharyya. [29] proposed the aspect-sentiment-based opinion summarisation by fine-tuning a transfer learning model. They annotated the AMASUM synthetic dataset by collecting information from product descriptions, specifications, reviews, and question-answer pairs. They employed an extractive-abstractive approach that extracted salient opinion sentences based on an aspect-sentiment combination. The limitation of their proposed approach is that the summaries are not factually correct. Zhang et al. [33] proposed the Roberta transformer-based model for aspect-based opinion summarisation. Their proposed system also performed aspect-based sentiment analysis to extract opinion phrases from reviews. They performed experiments on Yelp, Amazon, and Rotten Tomatoes datasets and conducted an evaluation using ROUGE metrics. The limitation of their proposed work is summaries are not informative, and domain knowledge is not considered. They suggested that in the future, better strategies can be applied to model the extraction of aspect and sentiment information explicitly.

Tang, Zhang and Dinh [73] proposed the aspect-based key point analysis for quantitative review summarisation. Luo [74] highlighted the significance of BERT and BART, two contextual pre-trained embedding models that have demonstrated considerable potential in NLP applications. Takeshita et al. [75] developed a novel summarisation ACLSum dataset with the input of domain experts. Anker et al. [76] proposed a zero-shot text classification and generic summarization approach for aspect-based summarization without any supervision. Soleimani et al. [77] experimented with the zero-shot self-supervised pre-training approach for biomedical aspect-based summarisation by leveraging the PubMed structured abstracts over unseen aspects. Wang et al. [78] employed the two-stage model to generate abstractive Multi-document summaries using the AspSumm dataset. They experimented with the WikiAsp dataset. They performed zero-shot, few-shot, and fine-tuning on seven downstream datasets. Zhang et al. [79] developed a framework using the Roberta transformer-based model to create an aspect-based abstractive sentiment-oriented summary of unlabeled reviews, denoted as ASU-OSum, an aspect-based opinion summarisation. They experimented with Yelp, Amazon, and Rotten Tomatoes datasets, evaluated the results using ROUGE metrics and human evaluation. The limitation of their proposed work is summaries are not informative, and domain knowledge is not considered. They recommended that, in the future, better strategies can be applied to model the extraction of aspects and sentiment information explicitly.

Pan et al. [80] proposed a large-scale aspect-based abstractive summarisation dataset by using aspect alignments. The abstractive summary is based on the aspect related to

phrases. They used the Pointwise Mutual Information (PMI) method and also utilised an attention-based Pointer Generator Network (PGN) recurrent neural network to generate an abstractive summary. Evaluation scores were achieved by BLEU-3.20, Rouge1-27.72, Rouge2-26.41, and Rouge SU-7.97 for the word-level and aspect-level evaluation, which were done using Precision-93.05, Recall-88.75, and F1-90.83 scores. The limitation of their proposed framework is that it is specific to the fashion domain, but it can be extended to other domains, such as books and movies. Korkankar et al. [81] explored the use of large language models, such as Llama, Generative Pre-trained Transformer (GPT), Gemma2, Mixtral, and Qwen2 Mistral, on the publicly available Amazon reviews dataset. An experiment was conducted on three different domains, and evaluation was performed using Rouge, Meteor, and BERTscore F1 and GPT4 to assess the quality of aspect-based abstractive summaries. They suggested that in the future, large language models can be fine-tuned to analyse and improve the performance and efficiency of large language models.

Xu, Meng and Cheng [82] proposed an aspect-based abstractive summarisation that employed the aspect-sensitive Markov random model and a greedy removal method to represent and meet diversity. They focused on the dependency between extracted sentences and intrinsic relationships for generating summaries. They used the MG-LDA model, a language model, and a multinomial distribution of words to extract aspects. Higher probability of words being selected as explicit aspects and sub-aspects. Tyss, Aly and Grabmair [83] proposed aspect-based summarisation of legal documents and introduced the novel dataset LexAbSumm for aspect-specific summarisation of legal case files. They used handcrafted rules and regular expressions to parse the documents. They proposed BERT embeddings and the LED, PRIMERA, LongT5, SLEd-BERT, and Unlim-BART models to generate an abstractive summary. An Evaluation was performed using BLEU scores. They recommended that, in the future, the model can be developed to consider aspects and prevent the generation of non-aspect-related words in summaries.

Kansal and Toshniwal [84] proposed aspect-based sentiment-oriented summarisation by clustering subjective aspect words and detecting polarity ratings using the PMI (Pointwise mutual information) method, as well as mapping features to opinion words, which can handle context-dependent words. They used an online sentiment dictionary to classify the context-independent words. They used features and opinion words together. The weakness of their approach is limited to 500 reviews of DVD players, cameras, and Nokia phones, with non-feature words selected through nouns and adverbs. Tan et al. [85] introduced a novel approach that utilises weak supervision and aspect-based modelling, ingeniously integrating rich external knowledge from Concept Net and Wikipedia. An experimentation was done on the CNN/Daily Mail, 280k MA-News dataset using the Bidirectional and Auto-Regressive Transformer (BART) model.

A limitation of their work is that NER can only be applied to news articles to extract aspects, not to product reviews. Yang et al. [28] proposed query-based summarisation for specific domains. They used the OASum dataset, which contained an open-domain aspect-based summarisation. They performed zero-shot, few-shot, and fine-tuning on seven downstream datasets. The limitation of their approach is that summaries are not correlated with aspects.

2.6.2.1 Strengths and Weaknesses of Transfer Learning Approach

Table 2.5: Strengths and Weaknesses of Transfer Learning Approach

Intro/Work	Strengths	Weaknesses
Pre-trained large language models, such as BERT, BART, GPT, and T5, handle contextual understanding and abstractive summarisation.	• Excellent contextual understanding and reduces the need for task-specific training. • Generalizes well to new domains	• Requires significant computational resources and Risk of domain mismatch. • Hard to interpret predictions

Table 2.5 describes the working of transfer learning methods with their strengths and weaknesses for aspect-based summarization.

2.7 Hybrid Aspect-based Extractive-Abstractive Summary

A hybrid approach combines Rule-based linguistics patterns with ML, Deep Learning (DL), and Transfer learning-based methods to generate an extractive-abstractive summary. Di Fabrizio, Stent and Gaizauskas [26] developed a STARLET-H system that generated extractive and abstractive summaries. They manually labelled the reviews of hotel and restaurant domains and selected the most frequent adjective phrases through polarity and aspects. They also experimented on the DUC05 and DUC06 datasets. Siledar, Makwana and Bhattacharyya [29] proposed the aspect-sentiment-based opinion summarisation by fine-tuning a transfer learning Text-To-Text Transfer Transformer (T5) model. They annotated the AMASUM synthetic dataset by collecting information from product descriptions, specifications, reviews, and question-answer pairs. They used an extractive-abstractive approach that extracted the salient opinion sentences. The limitation of their work is summaries are not factually correct.

Hong and Wang [44] proposed the LDA model with Long Short-term Memory (LSTM) to classify and summarise car and hotel reviews, transitioning from an extractive to an abstractive summary, across 6 datasets about cars and hotels collected from University of California Irvine machine learning repository (UCI). Ye and Lee [86] utilised the BERT autoencoder model to convert reviews into vectors. They crawled the reviews of iPhone and Samsung mobiles and extracted domain-related features, such as performance, capability, and functionality. Positive and negative summaries are generated separately with the Bidirectional Encoder Representations from Transformers (BERT) encoder and LDA model. They also created a sentiment analyser to detect aspects and their polarity and group relevant sentences for all prominent aspects. Frermann and Klementiev [87] introduced abstractive summarisation and extractive aspect-based summaries. Gurusamy, Rengarajan, and Srinivasan [88] proposed LDA and sentence concept matching to fine-tune the transformer model and generate both extractive and abstractive summaries. Their proposed system generated coherent summaries, including intermediate extractive summaries, by covering semantic aspects. Gautam, Kaur, and Josan [63] utilised the clustering and BiLSTM deep neural network to generate an extractive summary of TV reviews collected from Amazon and Flipkart. The evaluation was performed using the human scores and BLUE scores. The limitation of their approach is summary contains grammatical errors, and polarity is not considered.

Liu and Wan [89] utilised deep learning algorithms AttrEnc, AttrDec, AttrEncDec and MemAttr encoder-decoder model. The MemAttr model captured product features and user information to generate a summary. They experimented with the ReviewSum data set, which consists of 167,000 reviews, 3,080 users, and 3,329 products, labelled with 142.8 million samples. An evaluation was performed using Rouge 1, 2, and L scores. Nikhil Padhi et al. [90] experimented with Logistic Regression, Random Forest, SVM, XGBoost, ELMO, CNN, BiLSTM, BART, XLNet, BERT-base, and Legal BERT methods proposed for classifying SEBI case files. They utilised the embeddings of Word2vec, TF-IDF, and Glove for aspect-based summarisation. Their proposed framework achieved Rouge1-41, ROUGE2-26 AND ROUGEL-35 scores using the BART method and using BRIO, Rouge1-45, Rouge2-27, RougeL-35. Their system achieved high accuracy by exploring length-controlled abstractive summarization. Moreover, the indian legal documents dataset developed by them can be used for further analysis of indian legal documents.

Jeong and Lee [91] proposed a hybrid approach using the BERTopic modelling technique and ChatGPT to analyse the aspect-based frequency of complaints and identify negative sentiments in hotel reviews. They categorised complaint sentences based on the hotel's aspects and analysed the frequency of complaints. Their proposed technique achieved 70% accuracy in classifying the complaints. The main strength of their ap-

proach is that it can offer valuable insights into improving hospitality for hotel reviews and potentially reveal trends in customer expectations for hotel service quality. They indicated that fine-tuning of ChatGPT can be used to enhance the accuracy of aspect-based summarisation, and additional evaluation parameters can be used to assess the performance of ChatGPT.

Hayashi et al. [92] developed an open-source dataset using Wikipedia articles in 20 different domains. A two-stage model is proposed to identify aspects and summarise using extractive and abstractive models. They handled the proper pronoun quoted sources and consistent explanations of time-sensitive documents. Although they developed a weakly supervised dataset, the limitation is that it depends on Wikipedia articles and lacks multilingual representation. Rana and Cheah [31] employed a rule-based and machine learning approach to explore sequential patterns and normalised Google distance to extract both explicit and implicit aspects. They used similarity distance with the Particle Swarn Optimisation (PSO) method to group the synonyms. The strength of their approach is that their proposed method can be applied to real-life reviews across multiple domains.

2.7.1 Strengths and Weaknesses of Hybrid Approach

Table 2.6: Strengths and Weaknesses of Hybrid Approach

Intro/Work	Strengths	Weaknesses
A hybrid approach, integration of rule-based and machine learning models. This approach aims to balance inter-pretability and generalisation	• It can improve accuracy by capturing the domain-specific linguistic rules and leverage the complementary strengths of each technique. • Hybrid systems can be tailored to any domains(fashion, electrical) with domain-specific rules without retraining the models from scratch.	• It may increase system complexity and may require careful tuning of rule priority or model flow and rules annotation for new domain is a time-consuming process. • Complex for domain-specific summarization tasks because of risk of overfitting of rules on domain-specific • It can slow down performances due to involving multiple stages and models inference. • It may conflict decisions and arise discrepancies

Table 2.6 presents the strengths and weaknesses of the hybrid approach.

2.8 A Nutshell of Existing Work on Summarization

Table 2.7: Researchers' Work Approach Wise

Ref	Approach
[93], [25], [42], [24], [16], [46], [44], [94], [59], [95]	Rule-based
[43], [48], [47], [8]	Statistical Approach
[96], [44], [69], [20], [67], [97], [66], [27], [65], [98]	Topic Modelling
[59], [52], [99], [100], [101], [53], [102]	Supervised Machine Learning
[103], [104], [15], [53], [105], [106], [107], [40], [52]	Graph-based
[108], [63], [51]	Unsupervised Deep Learning
[58], [109], [110], [88], [26], [31], [87], [111], [43], [16], [49]	Hybrid Approach
[30], [44], [34], [112], [87], [78], [14], [89]	Supervised Deep Learning
[54], [59], [84], [55], [61], [94]	Clustering
[86], [113], [114], [71], [33], [17], [29], [57], [58], [85], [70], [14], [78], [79], [33], [28]	Transfer Learning

The table 2.7 describes the approach wise existing work on summarisation with the references. It also shows that the most of researchers are using transfer learning methods for developing summarisation systems framework.

2.9 Comparative Analysis of Aspect-based Summarisation

Table 2.8: Comparative Analysis of Existing Work on Aspect-based Summarisation

Ref, Year	Method	Dataset	Evaluation	Limitations/Future Scope
-----------	--------	---------	------------	--------------------------

Continued on next page

Table 2.8 – continued from previous page

Ref, Year	Methodology	Dataset	Evaluation	Limitations/Future Scope
[14] 2021	DistilRoberta for feature extraction and T5 for training	A synthetic dataset generated from OPOSUM and Space datasets	AMT, Info, Coh, Con, Flu and R1,R2,RL-40.37, 11.51, 23.23 for Space and 32.98, 10.72, 20.27 for the OPOSUM.	Length of the summary and the sentiments are not considered.
[17] 2021	BERT for summarization and CNN, SVM for classification.	Laptop reviews from five EC websites	F1-score- 82.6% achieved for classifying opinions	Implicit and explicit aspects not considered.
[25] 2019	Linguistic patterns and highly correlated features	Manually annotated 265 Kindle Reader reviews	Precision, Recall, F1 score	Not contained implicit aspects.
[26] 2014	Hybrid extractive to abstractive summary using a linguistic and statistical approach	Manually labelled the aspects of hotel and restaurant reviews	Amazon Mechanical Turk	Negative reviews are not considered.
[28] 2023	A transformer model for extractive to abstractive summarization	Long texts of 20 domains from WikiAsp, OASUM	Rouge score R1, R2, RL	Summaries are not correlated with aspects.
[29] 2023	Extractive-abstractive approach and fine-tuned a T5 model	Annotated the AMA-SUM synthetic dataset.	R1-37.27, R2-19.22, RL-35.32-verdict, Pros-R1-24.36, 7.91, 23.11, cons R1-21.17, 7.07, 20.08	Summaries are not informative and factually correct, which difficult to resolve conflicting information.

Continued on next page

Table 2.8 – continued from previous page

Ref, Year	Methodology	Dataset	Evaluation	Limitations/Future Scope
[30] 2019	Seq2seq model, Conceptnet embedding, for abstractive summary	Amazon food Review	Rouge Metric	Summary quality is low.
[33] 2023	Roberta model with attention mechanism and sentiment analysis.	Yelp, Amazon, Rotten Tomatoes datasets.	ROUGE metrics and human evaluation	Low quality summary and domain-specific knowledge is not considered.
[57] 2023	Noun phrases and adjectives, and BERT to extract features.	Amazon, Space	R1-40.89, R2-10.83, R3-24.58, inf, coh, concise	External domain knowledge is not considered.
[63] 2021	Bidirectional Long Short-Term Memory Network (BiLSTM) deep neural network and clustering	TV reviews collected from Amazon and Flipkart	BLEU Score	Summary contains grammatical errors, and polarity is not considered.
[64] 2022	Hierarchical agglomerative clustering, SBERT model, and TF-IDF	Opinions, space, and Amazon datasets.	Rouge R1, R2, RL Scores	A few aspects are considered.
[70] 2017	LDA model and linguistics patterns	Hotels reviews of Trip Advisor, GoIbibo, and Yatra	Human Evaluation	Domain-specific model lacks generalizability and use of deep models.
[78] 2023	A two-stage model to generate abstractive summaries.	Wiki Asp dataset.	Rouge scores.	some aspects not observed in training.

Continued on next page

Table 2.8 – continued from previous page

Ref, Year	Methodology	Dataset	Evaluation	Limitations/Future Scope
[85] 2020	BART model, NER, Weak supervision, Conceptnet	CNN/Daily Mail and 280k MA-News dataset, Wikipedia	Rouge R1, R2, RL Scores	The weak supervision may cause noise for different domains.
[87] 2019	Encoder-decoder Pointer-generator model and weak supervision	CNN/Daily Mail and RCVI datasets.	Diversity- 0.133, flu - 1.667, and info-1.433, Rouge R1- 0.23 and R2- 0.22.	Abstractive summary limited to the news domain only, and content informativeness varies.
[91] 2024	BERTopic, HDBSCAN for aspect based classification and sentiments analysis	Hotel Reviews from TripAdvisor	ChatGPT	Fine-tuning of large language models and the annotated datasets can be considered.
[92] 2021	Fine-tuned Roberta model, Presumm model, Text rank	WikiAsp, Opossum, Amazon, Rotten Tomatoes, MA-News	R1, R2, and RL Text Rank- 20.47 3.59 18.45, Presumm- 22.94 5.74 21.02	Limited to Wikipedia’s articles and lacks multilingual representation.
[112] 2021	Jointly Neural Attention model for summarisation with ratings	Electronics toys, games, movies and TV, Home and kitchen	ROUGE Score	Rating prediction is not considered

The table 2.8 presents a comparative analysis of existing work on aspect-based summarisation, including the methods, datasets, and evaluation metrics. The limitations of the existing work is also explained.

2.10 Other Related Work

Bravzinskas et al. [115] developed the framework for the abstractive summarisation system using a few-shot transformer, and experimentation was done on Amazon and

Yelp datasets. An evaluation was done using the Rouge-L score. Their model generated a summary that was concise, coherent, fluent, informative, and conveyed the sentiments effectively. One key strength of their proposed model is its applicability to cross-domain applications. Amplayo and Lapata [20] proposed the unsupervised LDA method and deep learning model to summarise unlabeled product reviews. They created synthetic datasets and utilised the LDA method to extract explicit aspects, incorporating sentiments derived from the opinions. Muniraj, Sabarmathi and Leelavathi [111] used Text Rank to extract important sentences from each news document to generate an extractive summary and then applied the Seq2seq hybrid model to generate an abstractive summary of news documents. Meena and Gopalani [48] utilised a rule-based method to select proper nouns and named entities from text and TS-ISF for sentence filtering, based on sentence scoring at the word level, and sentences were selected using ranking. The experimentation was performed on DUC 2002. They suggested that the proposed method might be explored by filtering the semantic features and involving the features of discourse-based approaches.

Mani et al. [8] utilised the unsupervised centroid-based distributed bag of words and beam search method for multi-document summarisation. They experimented with it on the DUC2006 and DUC2007 datasets. Implementation was performed with Gensim. Goodwin, Savary and Demner-Fushman [116] compared the 3 models BART(BART-large), T5(T5-base), and PEGASUS (Pegasus base) on the summarisation datasets. They experimented with zero-shot and few-shot learning (10 labelled examples) and found no difference in summary quality. They summarised the documents using topic and question-driven multi-document abstractive summarisation. They experimented with DUC 2007, TAC 2009, TAC 2010, and MEDIQA datasets. Shah et al. [34] utilised a Seq2seq encoder-decoder-based model to summarise food reviews based on quality, price, and maximum food demand, converting the input vectors to a smaller sequence of output vectors. Their proposed system summarised the reviews based on the product's ID with sentiment. They suggested that the BERT transformer-based method can achieve good results in generating summaries as compared to Seq2seq. However, a notable limitation is that their proposed approach relies heavily on large labelled datasets, and performance may degrade on unseen domains.

2.11 Research Gaps

Although a profound work has been done on Aspect-based summarisation. We found some bottlenecks in aspect-based summarisation techniques on domain-specific datasets. During the literature survey, research gaps found for the aspect-based summarisation are as follows:

1. Existing summarisation methods struggle to capture sentiments in the summary [112].
2. Existing models struggle with distinguishing between overlapping multiple aspects in a single sentence [17,28].
3. Domain adaptation remains a challenge due to varying terminologies for aspect-based summarisation [33].
4. There is a need for datasets of aspect-based summaries of domain-specific and multiple product reviews [28].
5. Real-world reviews writing style is different due to containing noise, typos, abbreviations, emojis, and inconsistent grammar. [63].
6. The absence of domain knowledge underscores the development of effective pre-processing and feature extraction techniques on domain-specific datasets [19].
7. Most research is focused on English, and the remaining unexplored. Another area that requires attention is the summarisation of Hinglish reviews [91].
8. Existing approaches have faced the challenges of being semantic, informative, relevant, fluent, coherent and accurate predicted summaries [29].
9. Few-shot and zero-shot learning approaches remain underdeveloped for training the small domain-specific datasets for the aspect-based summarisation task [30].
10. The traditional metrics ROUGE and BLEU do not fully capture the quality of aspect-based summaries. [28].
11. There is a need to investigate the generalisation of guided topic modelling techniques on various domains and languages [117].

Despite extensive research efforts and advancements in this area, there remain gaps that necessitate further investigation. Although significant progress has been made by researchers using various approaches for aspect-based summarisation, challenges such as handling contextual ambiguity, semantic, multiple domain handling, and maintaining summary coherence remain. These identified gaps motivate the methodology proposed in the next chapter.

2.12 Analysis of Available Reviews Summarisation Datasets

A detailed description of the pros and cons of existing datasets, which are publicly available on Kaggle or Hugging Face, is given below.

2.12.1 Home and Kitchen Reviews Dataset Analysis

Table 2.9: Examples of Home and Kitchen Review Dataset

ReviewText	Rating	Summary
My last toaster stopped working and, after research, I thought this a good choice. I am disappointed that it does not toast evenly on both sides.	2	new toaster
Got my mom this when she requested a new toaster. What can I say...it's a toaster...it toasts toast. Nothing has blown up or caught fire. It does what it is supposed to. She likes it. Would purchase again.	4	Its a toaster
I had a more expensive toaster before. I hated it. This one is so much better and I like the white.	5	Toaster
I wanted a toaster that was simple and this does the trick. If things start to burn it has a stop button and we like that. We have set it at a 21/2 temp and that toasts just fine.....Easy to clean just pull out the bottom tray and then I shake it over the sink for lose crumbs.	4	new toaster
I'm not big on unique gadgets since running out of storage space in the kitchen, but this one beats all the workarounds for slicing cakes horizontally.	5	what can I say
My husband wanted this for Christmas to go along with the new grill he bought a month ago. It works great because we can sit inside and watch tv while the food cooks outside on the grill and check on the internal temperature from the comfort of our couch.	5	Christmas gift

Table 2.9 shows some samples of the Home and Kitchen reviews dataset with the ratings and summaries, which are publicly available on Kaggle.

2.12.2 Clothing Shoes and Jewellery Dataset Analysis

Table 2.10: Reviews of Clothing Shoes and Jewellery Dataset

ReviewText	Rating	Summary
My 4 year old daughter always wants to play dress up. I bought this for her for Christmas and she never wanted to take it off. It is almost a year later and she still has it. It was well worth the money and has held up this long with nothing wrong with it. I would recommend it	5	My daughter loved it.

Continued on next page

Table 2.10 – continued from previous page

ReviewText	Rating	Summary
Our granddaughters are all very girly, so when the youngest one received this for Christmas, they all wanted it. Great color and fit for a 2 year old as well as her aunt who is 30. I would recommend this tutu for all little girls.	5	Sassy
Have they lost their minds. Over 100 dollar for a costume. Amazon, you are an idiot. This is an inexpensive costume. It better include a meeting with Tom Hanks for that price.	1	WHAT

Table 2.10 shows the outdated reviews of clothing, Jewellery and shoes dataset, that contained the columns of reviews, rating and summaries. The summaries of these reviews do not describe any product or aspect. The summaries of this standard dataset are written generic-based, not feature-based summaries.

2.12.3 Analysis of Flipkart Multiple Product Reviews Dataset

Table 2.11: Some Reviews of Flipkart Reviews Dataset

Summary	Rating	Review Text	Sentiment
awesome	5	amazing sound bar mivi very good quality sound and bass very rich battery backup very good and build quality good	positive
fabulous!	5	it came out very good than the expected bass is good battery backup is best we almost use it 335 hrs everyday still its battery comes down to 80 only loved the product	positive
terrific	5	great buy best in this money segment love the sound quality and the bass overall must buy and also love the made in india tag over the packaging sound quality 55 battery 45 bass 45	positive
brilliant	5	amazing sound and deep bass great build quality speaker build design is also very good battery backup and bluetooth connectivity is also very good best soundbar at this price range	positive
awesome	5	excellent product from mivi i really loved this bluetooth speaker battery backup amazing also fast connecting bluetooth	positive

Continued on next page

Table 2.11 – continued from previous page

Summary	Rating	Review Text	Sentiment
fabulous!	5	best product quality is to much awesome really love it you can definitely but it thank you	positive
just okay	3	thin blanket soft and overall good	positive
fair	3	the quality is good but the power of air is decent	positive

Table 2.11 defines some samples of reviews of Flipkart multiple product reviews dataset that have columns: reviews with ratings, sentiment and summary. But summaries do not properly describe any product or any aspect.

2.12.4 Women’s E-commerce Clothing Reviews

Table 2.12: Women E-commerce Clothing Reviews

ReviewText	Rating	Division Name	Class Name
Absolutely wonderful silky, sexy and comfortable	4	Initmates	Intimates
Love this dress. it’s so pretty. i happened to find it in a store, and i’m glad i did because i never would have ordered it online	5	General	Dresses
I love this dress. i usually get an xs but it runs a little snug in bust so i ordered up a size. very flattering and feminine with the usual retailer flair for style.	5	General	Dresses

Table 2.12 shows the reviews description of the Women’s E-commerce unlabeled product reviews. It defines the category ClassName "Dresses", which can be used to classify the categories of women’s dresses.

2.13 Pros and Cons of Available Datasets of Reviews

Table 2.13: The Pros and Cons of Standard Datasets

Dataset, Format	Description	Attributes	Pros	Cons
-----------------	-------------	------------	------	------

Continued on next page

Table 2.13 – continued from previous page

Dataset, Format	Description	Attributes	Pros	Cons
Amazon Fine Food Reviews [118] Format-CSV	It contains reviews on several products like dog,cat food, medicine, etc. This dataset contains 568454 rows and 10 columns.	ProductId, UserId, ProfileName, Helpfulness, Time, Summary, Text	• Due to 568485 reviews, sufficient to train machine learning model. • summaries of the dataset are written by experts. .	• It does not contain updated reviews. • Summaries are written in generic based, not based on product aspects.
Flipkart Reviews [119] Format-CSV	It has 205053 rows and 6 columns. It contains reviews of clothing, electronics, Home decor.	Product name, Product price, Rate, Reviews, Summary and Sentiment	• This dataset contains updated reviews. • This dataset can be used for multi-label text classification.	• Summaries are written only for generic purposes.
Home and Kitchen [120] Format-JSON	It contains 551,682 reviews and 8 columns. Reviews are related to home and kitchen appliances.	Index, reviewerId, reviewer name, reviewText, summary, reviewTime	• It can be used for generic summarization and topic classification.	• It has not included summaries related to product aspects.

Continued on next page

Table 2.13 – continued from previous page

Dataset, Format	Description	Attributes	Pros	Cons
Clothing, Shoes and Jewellery [121] Format-JSON	It has 278,677 reviews and 8 columns. It contains reviews of clothing, Jewellery and fashion products.	ReviewerID, reviewerName, helpful, reviewText, rating, summary,reviewTime	• It can be used for sentiment analysis and summarization.	• Reviews contained from 2004-2013. • Only generic summaries are written.
Women's Clothing E-commerce Reviews [122] Format-CSV	This dataset includes 10 feature variables with 23486 rows.	ClothingID, Title, ReviewText, Rating, Recommended, Division, Class	• It can be used for dress categorisation and sentiment analysis.	• Summaries are not mentioned.
Flipkart Product reviews [119] Format - CSV	It has 205053 rows and contains electronics, clothing, and Decor reviews.	Product name, price, Rating, Review, Summary, Sentiment	• It can be used for sentiment purposes and generic summary purposes.	• Summaries do not include any product features.
Flipkart Mobile Reviews [123] Format-CSV	It contains 900 mobile reviews and 4 columns.	Mobile name, Heading and Reviews	• It can be used for mobile topics classification.	• Only 900 reviews are present in this dataset.

Continued on next page

Table 2.13 – continued from previous page

Dataset, Format	Description	Attributes	Pros	Cons
Hugging face product-reviews [124] Format-CSV	It has 20k rows and 8 columns. It contains multiple product reviews in different languages.	ReviewId, product_id, stars, review, review_title, language, product category	• It can be useful for sentiment classification and summarisation.	• Considered review titles as review summaries.

Table 2.13 describes the pros and cons of standard datasets which are publicly available on Kaggle and Hugging Face websites.

2.14 Conclusion

This chapter provides an elaborative study of various text summarisation techniques and presents recent efforts and advances in NLP with transfer learning models. A detailed analysis of publicly available product review summarisation datasets is done, and it concludes that the loopholes of those datasets are summary are written in a generic form, not aspect-based. After analysing the Rule-based, machine learning and deep learning algorithms for summary generation, it is found that the rule-based approach can give maximum accuracy. Still, it can only be useful for small and domain-specific data. The pros and cons of existing review summarisation datasets are also described in this chapter. This chapter also explains the approach-wise existing research work and comparative analysis of rule-based, machine learning, deep learning and transfer learning methods. While rule-based and traditional machine learning methods offer interpretability, their scalability is limited. In contrast, deep learning and transfer learning methods offer strong performance across domains, albeit at the expense of transparency and computational efficiency. In the next chapter, Data collection and experimentation with preprocessing techniques are described.

Chapter 3

Text Preprocessing and Feature Extraction Techniques’ Challenges on Domain-Specific Data

Analysis of Real-world data is crucial for solving real-life issues. The use of preprocessing and feature extraction techniques on real-world data is a more tailored and effective approach, which is presented in the subsequent chapter. Text preprocessing and feature extraction play a crucial role in the performance of natural language processing (NLP) systems. This chapter explores the challenges of data collection and then experiments with preprocessing and feature extraction techniques. In the last, it identifies the challenges of preprocessing and feature extraction techniques for real-world data.

3.1 Contributions of this Chapter

Table 3.1: Contributions of Data Collection, Preprocessing and Feature Extraction

Aspect	Contribution
Data Collection	Collected 40k reviews of the fashion and electronics category.
Preprocessing Techniques	Analysis, and experimented with preprocessing techniques and found the challenges of preprocessing techniques on real-world data.
Feature Extraction Techniques	Experimented with several feature extraction techniques and found that the BERT transfer learning based method can extract more accurate features.

3.2 Introduction

Text preprocessing is the major and vital step for extracting the most important information from unstructured text. Preprocessing techniques transform the raw data into a more meaningful format. However, product reviews contain so much noise, including abbreviations and spelling errors, making it very challenging to understand these informal reviews. Although several preprocessing techniques can be applied to text, all of these have some impact on the text [125]. Therefore, text preprocessing is a crucial step in removing noise from reviews and normalising unstructured reviews. [126–128].

Before applying feature extraction techniques, text pre-processing is crucial for cleaning the text [129, 130]. Preprocessing steps like tokenisation, stemming, lemmatisation and stop word removal must be used carefully [21, 131, 132]. Ramadhan et al. [105] proposed the extractive summary of headphone product reviews with including and excluding stop words. The rouge score achieved R1-42.29 with typos and excluding stop words, and R1-46.71 with no typos but including stop words. The proposed research develops a dataset by collecting reviews on e-commerce websites and investigates the impacts of preprocessing techniques. However, preprocessing techniques should be applied according to the NLP application development. We examine the existing research work that has applied some preprocessing techniques like stop word removal and stemming, which are not suitable for summarisation purposes [19].

3.3 Data Collection Challenges in While Scraping

We have faced many challenges in data collection, which are mentioned as follows:

1. The challenge of blocking the IP address by Amazon due to scraping the maximum entries for data extraction.
2. Sometimes, duplicate entries are received after every 10 reviews due to the maximum time scraping.
3. We also received many blank entries in reviews after continuously scraping the same e-commerce website.

3.4 Preprocessing Techniques

Preprocessing is a crucial task to normalise the data. Most of the time is spent on applying preprocessing techniques to clean the noisy reviews. Online product reviews contain null values, noise, irrelevant information, emojis, abbreviations, punctuations, extra spaces, non-vocabulary words, spelling errors, informal words, etc. Several preprocessing steps are explained below.

3.4.1 Data Cleaning

Due to data extraction methods, the data contains Null and Nan values. The data cleaning process cleans Null, Nan, and duplicate values to contain the unique reviews and clean the noise, special characters, and punctuations from the data. The regular expressions (re) NLP library is used to remove the numbers, special characters, punctuation and alphanumeric characters from reviews.

3.4.1.1 Emojis Removal

Emojis are used to express emotions in the form of small digital icons. Nowadays, many people use emojis in reviews to express their feedback; however, machine learning algorithms often struggle to understand these emojis. The emoticons library used to detect the emojis.

3.4.2 Data Normalisation

We have applied the case conversion technique to convert uppercase to lowercase letters for normalising the reviews. Because "Ram and ram" have different vectors, and this can change the meaning of the sentence.

3.4.3 Contractions

Many times, users write the short form in reviews, as I've done. However, when converted into a numerical vector form, the text changes its meaning. The contraction library expands the contractions into the correct form.

3.4.4 Data Filtering of English and Hinglish Reviews

We filtered the reviews in English and Hinglish using the NLTK English words set and stored the English reviews in a CSV file. After filtering, we have stored the English and Hinglish reviews in two separate data frames.

3.4.5 Spell Checking Using Pyenchant

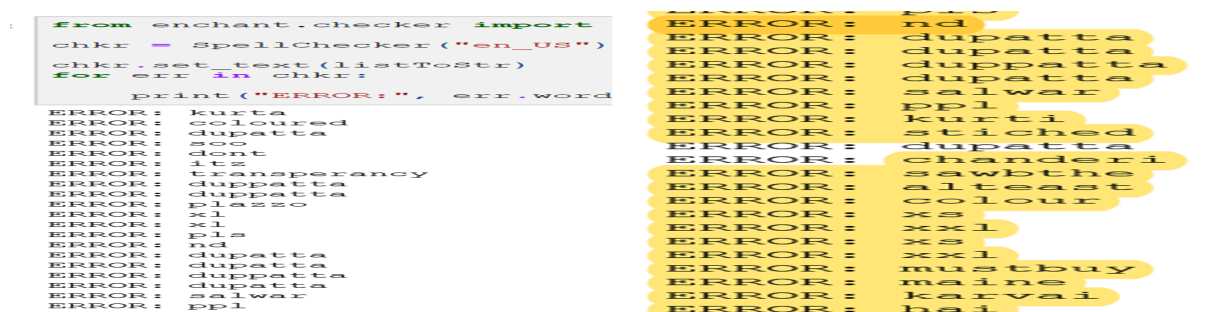


Figure 3.1: Spell Checking Using Pyenchant

Figure 3.1 shows errors in the out-of-vocabulary words in the reviews. The real-life reviews contain numerous spelling errors, abbreviation issues, and out-of-vocabulary words.

3.4.6 Spell Correction Using TextBlob and Autocorrect

	Reviews	corrected_review	num_corrections	cleaned	textblob_corrected	autocorrect_corrected
0	Suprb drs 🍌, Nys 🍌 Looking Vry Smart 🍌, Gud Cloth 🍌, Received the product before expected date, Same as Displayed	Supra dry drs 🍌, Was 🍌 Looking Cry mary 🍌, Mud Cloth 🍌, Received the product before expected date, Same as Displayed	6	suprb drs nys looking vry smart gud cloth received the product before expected date same as displayed	supra dry nays looking very smart god cloth received the product before expected date same as displayed	superb drs nys looking vry smart gun cloth received the product before expected date same as displayed
1	Perfect fit,material ,color. Unlike other party wear stuff, the material is cotton(hoisery) material. The kids wear it comfortably.My kid is extremely fussy when it comes to wearing anything other than cotton tshirts and cotton pants, but he wore it happily for a family function without complaining even once.Also the fit makes it look really smart. Have washed it and the color hasn't faded nor was there any color leak.	Perfect fit,material ,color. Unlike other party wear stuff, the material is cotton(misery) material. The kids wear it comfortably.By kid is extremely fussy when it comes to wearing anything other than cotton shirts and cotton pants, but he wore it happily for a family function without complaining even once.Also the fit makes it look really smart. Have washed it and the color hasn't faded nor was there any color leak.	6	perfect fitmaterial color unlike other party wear stuff the material is cottonhoisery material the kids wear it comfortablymy kid is extremely fussy when it comes to wearing anything other than cotton tshirts and cotton pants but he wore it happily for a family function without complaining even oncealso the fit makes it look really smart have washed it and the color has not faded nor was there any color leak	perfect fitmaterial color unlike other party wear stuff the material is cottonhoisery material the kids wear it comfortablymy kid is extremely fussy when it comes to wearing anything other than cotton shirts and cotton pants but he wore it happily for a family function without complaining even oncealso the fit makes it look really smart have washed it and the color has not faded nor was there any color leak	perfect immaterial color unlike other party wear stuff the material is cottonhoisery material the kids wear it comfortablymy kid is extremely fussy when it comes to wearing anything other than cotton shirts and cotton pants but he wore it happily for a family function without complaining even oncealso the fit makes it look really smart have washed it and the color has not faded nor was there any color leak
2	Very very happy with this purchase .the product quality is really awesome with a very decent price...good color and perfect stitching...lasts long..	Very very happy with this purchase .the product quality is really awesome with a very decent price...good color and perfect sticking...lasts long..	1	very very happy with this purchase the product quality is really awesome with a very decent pricegood color and perfect stitchinglasts long	very very happy with this purchase the product quality is really awesome with a very decent pricegood color and perfect stitchinglasts long	very very happy with this purchase the product quality is really awesome with a very decent pricegood color and perfect stickinglasts long
	Very soft very comfortable .It's really worth of money. I get this dress in 800 rs but it's looking like 2000 dress very	Very soft very comfortable .It's really worth of money. I get this dress in 800 is but it's looking like 2000 dress very		very soft very comfortable it is really worth of money i get this dress in 800 rs but it is looking like 2000 dress very	very soft very comfortable It is really worth of money i get this dress in 800 is but it is looking like 2000 dress very	very soft very comfortable it is really worth of money i get this dress in 800 rs but it is looking like 2000 dress very

Figure 3.2: Spell and Abbreviation Correction Using TextBlob and Autocorrect

Figure 3.2 shows errors in the out-of-vocabulary words in the reviews. After experimenting with several libraries, we concluded that TextBlob, Autocorrect, and Pyenchant are not suitable for correcting out-of-vocabulary words and many incorrect spellings in reviews.

3.4.7 Experimentation with Tokenization

Tokenisation is used for splitting the paragraphs into words or sentences. A sentence tokeniser can give a more correct output than a single word or phrase. Because the single word (nice) does not contain any semantics. The sentence tokeniser can be used for summarising the text.

	Reviews	cleaned	tokenized_sentences	word_tokens	phrases
0	Suprb drs 🍌, Nys 🍌 Looking Vry Smart 🍌, Gud Cloth 🍌, Received the product before expected date, Same as Displayed	suprb drs nys looking vry smart gud cloth received the product before expected date same as displayed	[suprb drs nys looking vry smart gud cloth received the product before expected date same as displayed]	[suprb, drs, nys, looking, vry, smart, gud, cloth, received, the, product, before, expected, date, same, as, displayed]	[suprb drs, vry smart gud cloth, the product, date, displayed]
1	Perfect fit,material ,color. Unlike other party wear stuff, the material is cotton(hoisery) material. The kids wear it comfortably.My kid is extremely fussy when it comes to wearing anything other than cotton tshirts and cotton pants, but he wore it happily for a family function without complaining even once.Also the fit makes it look really smart. Have washed it and the color hasn't faded nor was there any color leak.	perfect fitmaterial color unlike other party wear stuff the material is cottonhoisery material the kids wear it comfortablymy kid is extremely fussy when it comes to wearing anything other than cotton tshirts and cotton pants but he wore it happily for a family function without complaining even oncealso the fit makes it look really smart have washed it and the color has not faded nor was there any color leak	[perfect fitmaterial color unlike other party wear stuff the material is cottonhoisery material the kids wear it comfortablymy kid is extremely fussy when it comes to wearing anything other than cotton tshirts and cotton pants but he wore it happily for a family function without complaining even oncealso the fit makes it look really smart have washed it and the color has not faded nor was there any color leak]	[perfect, fitmaterial, color, unlike, other, party, wear, stuff, the, material, is, cottonhoisery, material, the, kids, wear, it, comfortablymy, kid, is, extremely, fussy, when, it, comes, to, wearing, anything, other, than, cotton, tshirts, and, cotton, pants, but, he, wore, it, happily, for, a, family, function, without, complaining, even, oncealso, the, fit, makes, it, look, really, smart, have, washed, it, and, the, color, has, not, faded, nor, was, there, any, color, leak]	[perfect fitmaterial color, other party, the material, cottonhoisery material, the kids, comfortablymy kid, anything, cotton tshirts, cotton pants, a family function, the fit, the color, any color leak]
2	Very very happy with this purchase .the product quality is really awesome with a very decent price...good color and perfect stitching...lasts long..	very very happy with this purchase the product quality is really awesome with a very decent pricegood color and perfect stickinglasts long	[very very happy with this purchase the product quality is really awesome with a very decent pricegood color and perfect stickinglasts long]	[very, very, happy, with, this, purchase, the, product, quality, is, really, awesome, with, a, very, decent, pricegood, color, and, perfect, stickinglasts, long]	[this purchase, the product quality, decent pricegood color, perfect stickinglasts]
	Verv soft verv comfortable .It's really worth of money. I get this dress in 800 rs but it's looking like 2000 dress very	verv soft verv comfortable it is really worth of money i get this dress in 800 rs but it is looking like 2000 dress very	[verv soft verv comfortable it is really worth of money i get this dress in 800 rs but it is looking like 2000 dress very]	[very, soft, very, comfortable, it, is, really, worth, of, money, i, get, this, dress, in, 800, rs, but, it, is, looking, like, 2000, dress, very]	

Figure 3.3: Results of Tokenisation on Reviews

Figure 3.3 presents the sentence tokenisation and word tokenisation on the reviews. The word tokenisation is not useful for summarisation purposes. The sentence tokeniser can be used to extract sentences or to increase the accuracy of summarising the text.

3.4.8 Experimentation with Stemming and Lemmatisation

Stemming is performed to normalise the reviews by transforming the words into their root form. It can also help to find similarities between texts. However, the main issue with stemming is that it can change the meaning of a sentence. After experimentation, it is analysed that stemming is not a good choice for summarising reviews. Lemmatisation is used to overcome the limitation of stemming. Lemmatisation strips the words into their base form. But nowadays, many reviewers have written many abbreviated and incorrect words in their reviews. So, stemming and lemmatisation are not a good choice for summarisation.

	Reviews	cleaned	stemmed	lemmatized
770	Quality of Material is good, fitting for 2-3 years is proper for 3 year old. Everyone really loved the product.	quality of material is good fitting for 23 years is proper for 3 year oldeveryone really loved the product	qualiti of materi is good fit for year is proper for year oldeveryon realli love the product	quality of material be good fitting for year be proper for year oldeveryone really love the product
771	Poor quality ... Do not purchase.. sooo bad don't like this. Very bad design don't purchase... Please don't be get fool	poor quality do not purchase sooo bad do not like this very bad design do not purchase please do not be get fool	poor qualiti do not purchas sooo bad do not like thi veri bad design do not purchas pleas do not be get fool	poor quality do not purchase sooo bad do not like this very bad design do not purchase please do not be get fool
772	Not bad ... Not met the expectations... Bad packaging	not bad not met the expectations bad packaging	not bad not met the expect bad packag	not bad not meet the expectation bad packaging
773	Without delivery, how can you show as "item delivered". Worst experience. Where is the product...so so so so so bad	without delivery how can you show as item delivered worst experience where is the productso so so so so bad	without deliveri how can you show as item deliv worst experi where is the productso so so so so bad	without delivery how can you show as item deliver bad experience where be the productso so so so so bad
774	This product is well fitted and gives rich look , the fabric is soft and quality of material too good	this product is well fitted and gives rich look the fabric is soft and quality of material too good	thi product is well fit and give rich look the fabric is soft and qualiti of materi too good	this product be well fit and give rich look the fabric be soft and quality of material too good
775	It's very nice dress ..And material is very soft and pretty. I am very happy with this product thank you so much .	it is very nice dress and material is very soft and pretty i am very happy with this product thank you so much	it is veri nice dress and materi is veri soft and pretti i am veri happi with thi product thank you so much	it be very nice dress and material be very soft and pretty i be very happy with this product thank you so much
776	Nice 🍕 good full bcoz kicks mc click bcoz di vchi icha didi chock LLC cl ch CM oc cn fulfill	nice good full bcoz kicks mc click bcoz di vchi icha didi chock llc cl ch cm oc cn fulfill	nice good full bcoz kick mc click bcoz di vchi icha didi chock llc cl ch cm oc cn fulfill	nice good full bcoz kick mc click bcoz di vchi icha didi chock llc cl ch cm oc cn fulfill

Figure 3.4: Stemming and Lemmatisation Results

Figure 3.4 describes the stemming and lemmatisation results, including stop words. After stemming, 'quality' changed into 'qualiti', 'experience' changed into 'experi', 'delivery' changed into 'deliveri', and 'material' changed into 'materi', which altered the meaning of the words. However, lemmatisation gives correct words, but still, after performing lemmatisation, some words' meaning has also changed.

3.4.9 Experimentation with Stop words removal

The stop words do not contain any important information. In NLTK, the stop words list contains (a, an, the, it, my, from, and not) in the text. E.g., (I bought it from Amazon). If we remove the stop words from this sentence (I bought Amazon), it can also change the meaning of some sentences. So, removing stop words is not an appropriate choice for summarisation purposes.

	Reviews	cleaned	stemmed	lemmatized	stemmed_withoutstop	lemmatized_withoutstop
770	Quality of Material is good, fitting for 2-3 years is proper for 3 year old. Everyone really loved the product.	quality of material is good fitting for 23 years is proper for 3 year oldeveryone really loved the product	qualiti of materi is good fit for year is proper for year oldeveryon realli love the product	quality of material be good fitting for year be proper for year oldeveryone really love the product	qualiti materi good fit year proper year oldeveryon realli love product	quality material good fitting year proper year oldeveryone really love product
771	Poor quality ... Do not purchase.. sooo bad don't like this. Very bad design don't purchase... Please don't be get fool	poor quality do not purchase sooo bad do not like this very bad design do not purchase please do not be get fool	poor qualiti do not purchas sooo bad do not like thi veri bad design do not purchas pleas do not be get fool	poor quality do not purchase sooo bad do not like this very bad design do not purchase please do not be get fool	poor qualiti purchas sooo bad like bad design purchas pleas get fool	poor quality purchase sooo bad like bad design purchase please get fool
772	Not bad ... Not met the expectations... Bad packaging	not bad not met the expectations bad packaging	not bad not met the expect bad packag	not bad not meet the expectation bad packaging	bad met expect bad packag	bad meet expectation bad packaging
773	Without delivery, how can you show as "item delivered". Worst experience. Where is the product...so so so so so bad	without delivery how can you show as item delivered worst experience where is the productso so so so so bad	without deliveri how can you show as item delivly worst experi where is the productso so so so so bad	without delivery how can you show as item deliver bad experience where be the productso so so so so bad	without deliveri show item deliv worst experi productso bad	without delivery show item deliver bad experience productso bad
774	This product is well fitted and gives rich look , the fabric is soft and quality of material too good	this product is well fitted and gives rich look the fabric is soft and quality of material too good	thi product is well fit and give rich look the fabric is soft and qualiti of materi too good	this product be well fit and give rich look the fabric be soft and quality of material too good	product well fit give rich look fabric soft qualiti materi good	product well fit give rich look fabric soft quality material good
775	It's very nice dress .And material is very soft and pretty. I am very happy with this product thank you so much .	it is very nice dress and material is very soft and pretty i am very happy with this product thank you so much	it is veri nice dress and materi is veri soft and pretti i am veri happi with thi product thank you so much	it be very nice dress and material be very soft and pretty i be very happy with this product thank you so much	nice dress materi soft pretti happi product thank much	nice dress material soft pretty happy product thank much
776	Nice 🍌 good full booz kicks mc click booz di vchl icha didi choock llc cl ch cm oc cn fulfill	nice good full booz kicks mc click booz di vchl icha didi choock llc cl ch cm oc cn fulfill	nice good full booz kick mc click booz di vchl icha didi choock llc cl ch cm oc cn fulfill	nice good full booz kick mc click booz di vchl icha didi choock llc cl ch cm oc cn fulfill	nice good full booz kick mc click booz di vchl icha didi choock llc cl ch cm oc cn fulfill	nice good full booz kick mc click booz di vchl icha didi choock llc cl ch cm oc cn fulfill

Figure 3.5: Excluding Stop Words Results

Figure 3.5 presents the results of Stemming and Lemmatisation, excluding stop words. The highlighted lines show that the reviewer has written 'do not purchase' and 'not bad not met', and after removing stop words sentence's meaning has been changed.

3.5 Experimentation with Feature Extraction Techniques

Feature extraction possesses remarkable abilities to extract meaningful features from text, and it is an important step. Traditional feature extraction techniques often fail to produce vector embeddings for abbreviated and noisy text. When examining traditional techniques, a natural query arises: Which techniques should be preferred for extracting features, and what is the corresponding difference between traditional and existing embedding techniques for extracting features from text? To extract features, Text data should be converted into numerical vectors because the machines cannot understand the text and can only understand the vectors. Word embedding methods convert text into vectors to extract features from the text data. Through similar vectors, we can find the similarity between texts.

There are two types of text embedding techniques: contextual-independent and context-dependent. One-hot encoding, Bag-of-words, TF-IDF, and TS-ISF are traditional, context-independent statistical methods, whereas Word2vec and BERT are context-dependent semantic embedding methods. The limitation of non-contextual embeddings is that they can not extract meaningful information from the text. Rani and Walia [133] suggested that BERT is the state-of-the-art transfer learning pretrained embedding model that can give semantic contextual pretrained embeddings, as well as it can also give the embeddings of abbreviated and out-of-vocabulary words. For example, sometimes consumers write the words in abbreviated form, and other embedding methods can't provide the embeddings. The BERT model gives the embeddings based on the surrounding text. The next section describes the experimentation with feature extraction techniques.

3.5.1 Features Extraction Using Rule-based Methods

The Rule-based methods are used to extract features. For example, the Pos tags (Noun, Adjective, Verb, Adverb, and Proper Noun) or Adjectives or Adjectives with noun phrases extracted from Reviews. The Rule-based method is applied using NLTK and Spacy libraries, which is demonstrated in the figure below.

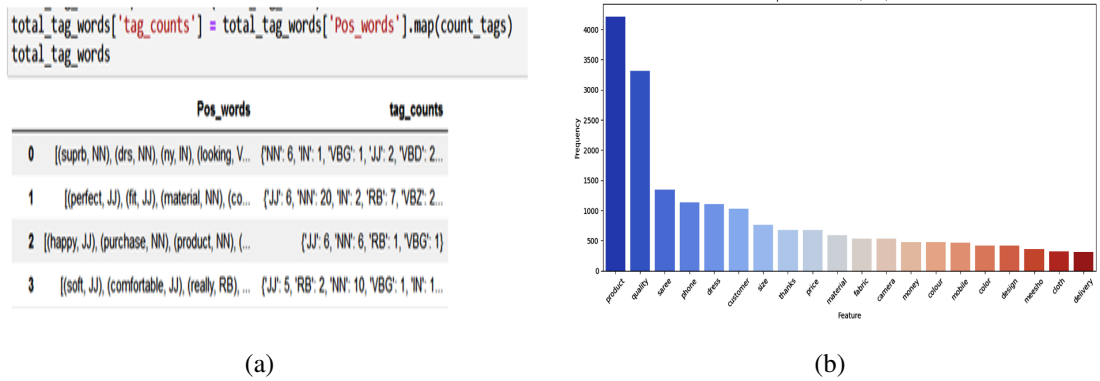
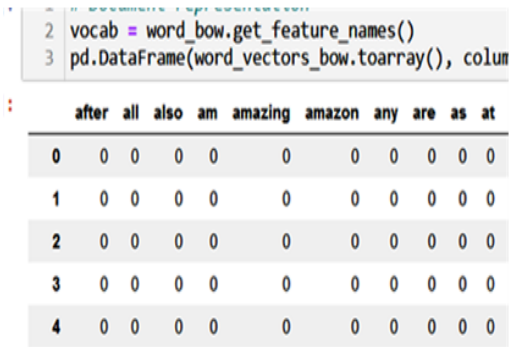


Figure 3.6: (a) Pos tagging to extract features (b) Distribution of top 20 noun words

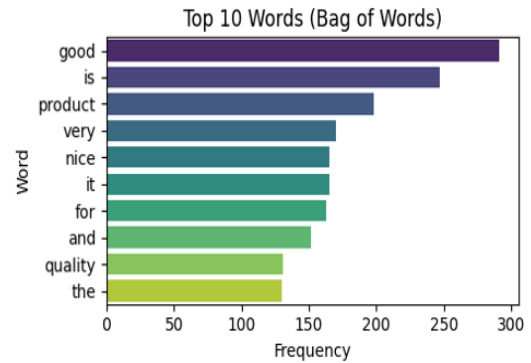
Figure 3.6 (a) shows the pos tags with the frequency of nouns, verbs, and adjectives. (b) presents the top 20 noun words distribution using pos tagging. The noun features play a crucial role in defining the aspects of a product's list of features.

3.5.2 Feature Extraction Using Bag-of-Words

The Bag-of-Words method is traditional, easy, and straightforward to implement for extracting a set of features. This method creates the features matrix of unique words and converts sentences' words into the same vectors, but the meanings are different according to the sentences.



(a)



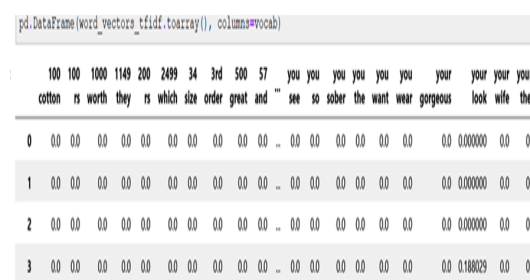
(b)

Figure 3.7: (a) Bag of words to get word vectors (b) Bag of words to extract feature words

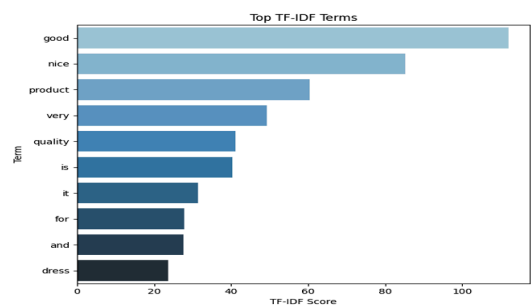
Figure 3.7(a) describes the Bag-of-words' work to create the feature vector matrix and places the frequency of words according to the presence in the documents, and (b) presents the distribution of features using bag of words and shows the top 10 words of extracted features using Bag of words. The main limitation of this method is that it makes a sparse matrix.

3.5.3 Feature Extraction Using TF-IDF

TF-IDF is used to find the important words, phrases and sentences based on frequency of the documents. The limitation of the TF-IDF method is that intuitive and essential words are not always captured through this technique; they can be captured for some amount. The figure below represents the experimentation with the TF-IDF method.



(a)



(b)

Figure 3.8: (a) Feature Vector using TF-IDF (b) Feature Words using TF-IDF

Figure 3.8 (a) shows the matrix of feature vectors extracted by the TF-IDF method, and (b) shows the extracted feature words using TF-IDF. The limitation of this method is that it cannot capture important words.

3.5.4 Feature Extraction Using Word2vec Method

Word2vec method is based on the Neural Network and generates the feature vectors; It helps to find the semantic similarity between words through feature vectors. This model can also detect synonyms and can also suggest different words for particular sentences. It places similar word vectors' embeddings nearby and places the different word vectors far away from other vector embeddings.

```
wv.most_similar('nyc')
```

```
[('brooklyn', 0.703615128993988),  
 ('atlanta', 0.6743649244308472),  
 ('boston', 0.6515336632728577),  
 ('chicago', 0.6503334641456604),  
 ('manhattan', 0.6443952322006226),  
 ('washington_dc', 0.6432445645332336),  
 ('springfield', 0.640512228012085),  
 ('baltimore', 0.6387253403663635),  
 ('houston', 0.638403594493806),  
 ('miami', 0.6332902908325195)]
```

(a)

```
vec_suprb = wv['suprb']
```

```
.....  
KeyError                                Traceback (most  
Input In [161], in <module>  
----> 1 vec_suprb = wv['suprb']  
  
File E:\python310\lib\site-packages\gensim\models\keyedvec  
Vectors._getitem_(self, key_or_keys)  
    390 """Get vector representation of 'key_or_keys'.
```

(b)

Figure 3.9: (a) Word2Vec embeddings of nyc word (b) Word2vec can't give the OOV word embeddings

Figure 3.9(a) shows the embeddings of the word2vec method and the similarity between embedding of the 'nyc' word and other words which have more similar embeddings to the 'nyc' country name. The main limitation of this method is that it gives similar embeddings for the same words but with different meanings and also gives the same embeddings for positive and negative sentences. Figure 3.9(b) shows that the word2vec model has no embeddings for the word 'suprb'.

3.5.5 Feature Extraction Using FastText Embeddings

FastText is an extension version of the Word2vec method developed by the Facebook Research team. This method can give a set of words, sub-words, character ngram or bigram embeddings. This method can also give the embeddings of wrong spell words. For ex, "sleping" is a misspelt word, but it is similar to sleeping. Although misspelt words are not present in the vocabulary, it can capture those words by understanding the prefixes or suffixes according to their semantic similarity.

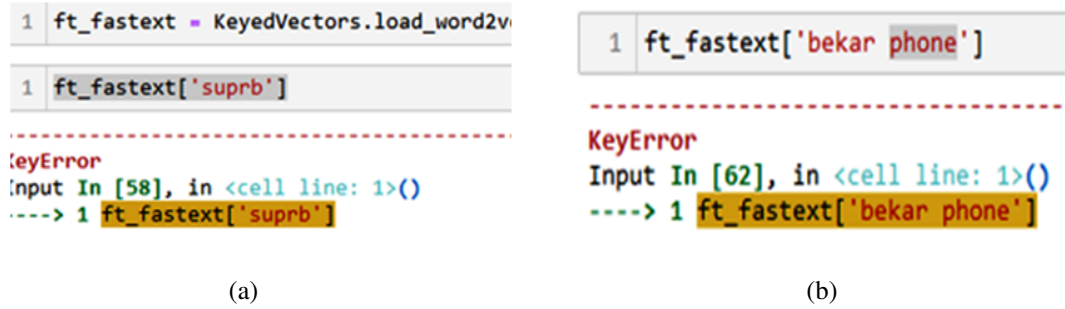


Figure 3.10: (a) Implementation of FastText (b) FastText didn't give embeddings of abbreviated words

Figure 3.10 (a) shows the error of FastText has no embedding of the 'suprb' word, and Figure 3.10(b) demonstrates that FastText didn't give the embedding of abbreviated and hinglish words.

3.5.6 Feature Extraction Using BERT

BERT is a transfer learning-based large language model, and it is an open-source framework for developing natural language processing applications. It is a state-of-the-art method for converting documents into contextual embedding vectors by using surrounding text to establish the contextual meaning of the sentence. The paper [134] introduced the working of the BERT model in semantic tasks. They described that the non-relevant words' position can affect the similarity relationships between sentences. The pros of the BERT model are that it can also give the embeddings of out-of-vocabulary words and Hinglish words because it is trained on a multilingual corpus and 3.3 Billion words.

3.5.6.1 Visualisation of Extracted Features Using BERT

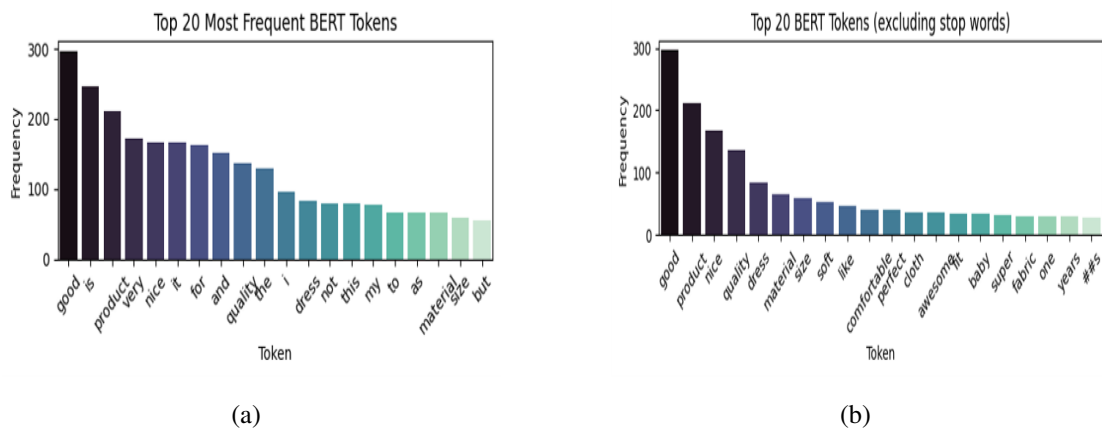


Figure 3.11: (a) Feature words including stop words (b) Features excluding stop words

Figure 3.11 (a) presents the top 20 extracted feature tokens using BERT including stop words, such as good, is, product, it, for, etc. Stop words are irrelevant to product

features. Figure (b) presents the top 20 extracted feature tokens, excluding the stop words like good, product, nice, quality, and size, which are the important features of products.

3.5.6.2 Working of BERT Model

I/p	[CLS]	I	buy	a	new	dress	[SEP]	It	is	very	beautiful	[SEP]
Token Embedding	E_{CLS}	E_I	E_{buy}	E_a	E_{new}	E_{dress}	$E_{[SEP]}$	E_{It}	E_{is}	E_{very}	$E_{beautiful}$	$E_{[SEP]}$
Sentence Embedding	E_A	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
Transformer positional Embedding	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}

Figure 3.12: BERT Working

Figure 3.12 shows the working of the BERT Model, and each input embedding combines the three embeddings:

1. **Token Embeddings:** A [CLS] Token is added at the beginning of the sentence, and the [SEP] token is inserted at the end of each sentence.
2. **Sentence Embeddings:** Sentence embeddings present the sentences added to each token. Each token receives a marking designating Sentence A or Sentence B.
3. **Positional Embeddings:** A positional embedding is added to each token and shows its position in the sequence.

3.6 Pros and Cons of Feature Extraction Techniques

Table 3.2: Pros and Cons of Feature Extraction Techniques

Technique	Pros	Cons
Bag-of-Words	• It is easy and intuitive to implement and better than one-hot encoding.	• It creates a sparse matrix and can't handle word order and can only be applied to small data.

Continued on next page

Table 3.2 — Continued from previous page

Technique	Pros	Cons
TF-IDF	• A few Amount of essential words can be captured using TF-IDF. • It reduces the sparsity issue.	• The semantic meaning of sentences is not captured. • It can't be applied to more repetitive words in the text. • Still, there is a sparsity issue. • TF-IDF does not consider out-of-vocabulary words.
Word2Vec	• It provides semantic embeddings and can identify the similarity between words. • It can also handle the sparsity issue.	• It can't measure the relationship between global words in contexts. • It can't create the embeddings of abbreviated words and can't capture the contextual order based on words position.
FastText	• It can generate character and sub-word embeddings. • It can handle some out-of-vocab words and semantic similarity between words.	• It is trained on fewer parameters. • It cannot generate the embeddings of abbreviated words. • It requires a large amount of memory to generate character embeddings.
BERT (Bidi-rectional Encoder Rep-resentations from Trans-formers)	• BERT can differentiate between polysemous words like ("cell, bank"),creates semantic contextual embeddings based on the surrounding words and can give the embeddings of non-vocabulary words. • It can be applied to large unlabeled data and fine-tuned on some labelled data.	• The high computational power, GPU and resources are required to run and train the BERT model. • The biased training data can lead to inaccurate results. • BERT is a large language model, so training is slow due to the Maximum token size of 512.

3.7 Conclusion

In this chapter, we have discussed the challenges of review collection and experimentation with preprocessing techniques on those reviews. After applying preprocessing techniques, it is concluded that stemming and stop word removal should not be performed on review summarisation because they can change the meaning of the sentence and can give incorrect output. Each preprocessing technique has its pros and cons, and some techniques can affect the NLP application. Therefore, preprocessing techniques should be applied in accordance with the development of the NLP application. After applying preprocessing techniques, this chapter has experimented with feature extraction techniques and found that traditional feature extraction methods, such as Bag-of-words, TF-IDF, Word2vec, and FastText, cannot provide semantic embeddings for out-of-vocabulary words and abbreviated words. However, the state-of-the-art transformer-based BERT method yields better results compared to other methods. This chapter also discusses the advantages and disadvantages of embedding techniques, as we have collected both English and Hinglish reviews. To extract the important information from the Hinglish code-mixed product reviews, we have experimented with topic modelling techniques, which are described in the next chapter.

Chapter 4

Experimentation With Topic Modelling Techniques on Hinglish Code-Mixed Product Reviews

The topic modelling techniques possess remarkable abilities in analysing important information by extracting important topics from reviews. The approximately 2,000 reviews of multiple-category products, such as clothing and electronics, are considered for applying topic modelling methods. This chapter analyses code-mixed Hinglish reviews using unlabeled and product-aware topic modelling techniques to extract important topics from reviews. The evaluation has been performed using coherence and diversity scores to compare the more relevant and coherent topics, and it concludes that the labelled topic modelling methods LDA, NMF, and BERT have outperformed the results.

4.1 Contributions of this Chapter

Table 4.1: Contributions of this Chapter

Terms	Contribution
Data collection	Collected 2k Hinglish reviews using scraping the e-commerce websites.
Preprocessing of Hinglish reviews	To stop words removal, we have created a domain-specific Hinglish stop word list.
Unsupervised and Labelled Product-aware Topic Modelling	We have experimented with LDA, NMF, and BERTopic modelling techniques on unlabelled and labelled topic documents with sentiments.
Evaluation	We have evaluated the topics using coherence and diversity scores of the words in the topics and compared the extracted topics' similarity with Jaccard similarity scores.

Table 4.1 has describes the contribution of experimenting with product-aware topic modeling techniques in this chapter.

4.2 Introduction

Topic modelling plays a vital role in extracting important information from reviews by finding important topics. Topic modelling techniques cannot comprehend the semantic context and meaning of the text. However, it can capture the hidden subjects using the context around the words and the probabilities assigned to each term in each document. The popularity of online shopping generates numerous product reviews on e-commerce websites. Most of the customers, especially non-professional Indians, write these reviews in Hinglish. It is very challenging to extract the topics from multi-domain categories of hinglish reviews [135]. Topic modelling is the best technique for mining text data by analysing and discovering relevant and important data to make decisions [136, 137]. It has been used to group related documents into coherent topics to mine the text [138]. Ponay [139] used Latent Dirichlet Allocation and BERTopic modelling techniques on customer feedback for an online ticketing system. Albalawi et al. [140] compared five topic modelling methods, including LDA, LSA, NMF, and PCA, as applied to short text data. There is a need for investigation of topic modelling techniques on various languages and domains. [117].

Extensive research on topic modelling has been done in the English language. However, little work has been done on the Indian languages. India is multilingual; most Indian people write feedback in their regional language. A few resources are available in Indian languages. In India, social media users use more than one language in writing. Nowadays, most people use Hinglish and code-mixed comments on social media and e-commerce websites. However, code-mixing is trending not only on social media but also on e-commerce websites. Indians mostly write feedback about products in Hinglish. Hinglish reviews are written in Hindi in English characters. We experiment with topic modelling on Hinglish reviews to extract important topics from reviews. These findings are significant in the context of e-commerce and data science, as they provide insights into the effectiveness of topic modelling techniques on Hinglish reviews, which can be used to improve product understanding and customer satisfaction.

In this chapter, we have experimented with the LDA, NMF and BERTopic modelling on Hinglish product reviews. In addition, it also compares the results, including hinglish stop words and excluding hinglish stop words, and found that the BERTopic modelling gives better results than LDA and NMF. The use cases, pros and cons of LDA, NMF, and BERTopic topic modelling techniques, have also been described in this chapter. The experimental results demonstrate that stop word removal is an important preprocessing technique for applying topic modelling. Especially including unimportant Hinglish stop words like 'hai', 'ho', 'hoo', 'ha', 'ho', 'wo', 'bhi' can reduce the coherence scores of topics.

4.3 Examples of Hinglish Mobile Reviews

Nowadays, most reviewers write their reviews in code-mixed Hinglish. Below are some examples of mobile phone reviews in Hinglish that are available on e-commerce websites, including Meesho, Amazon, Myntra and Flipkart.



Figure 4.1: Some Examples of Hinglish Reviews

Figure 4.1 demonstrates examples of mobile phone Hinglish reviews available on the Amazon e-commerce website. The reviews with the ratings are present.

4.4 Topic Modelling Techniques

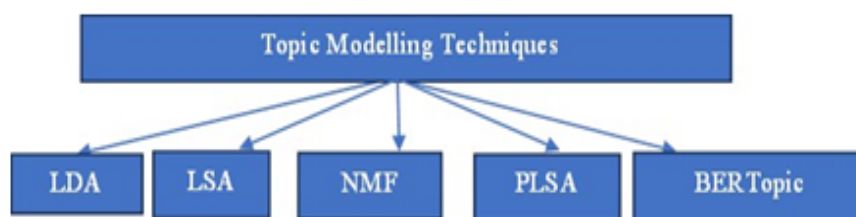


Figure 4.2: Techniques of Topic Modelling

Figure 4.2 shows several topic modelling algorithms, such as LDA, LSA, NMF, PLSA, and BERTopic. From which LDA and PLSA are probabilistic, LSA and NMF are Matrix Factorization, and BERTopic is Neural Network-Based. We have experimented with LDA, NMF, and BERTopic on Hinglish reviews and achieved better scores of BERTopic.

4.5 Experiments and Methodology

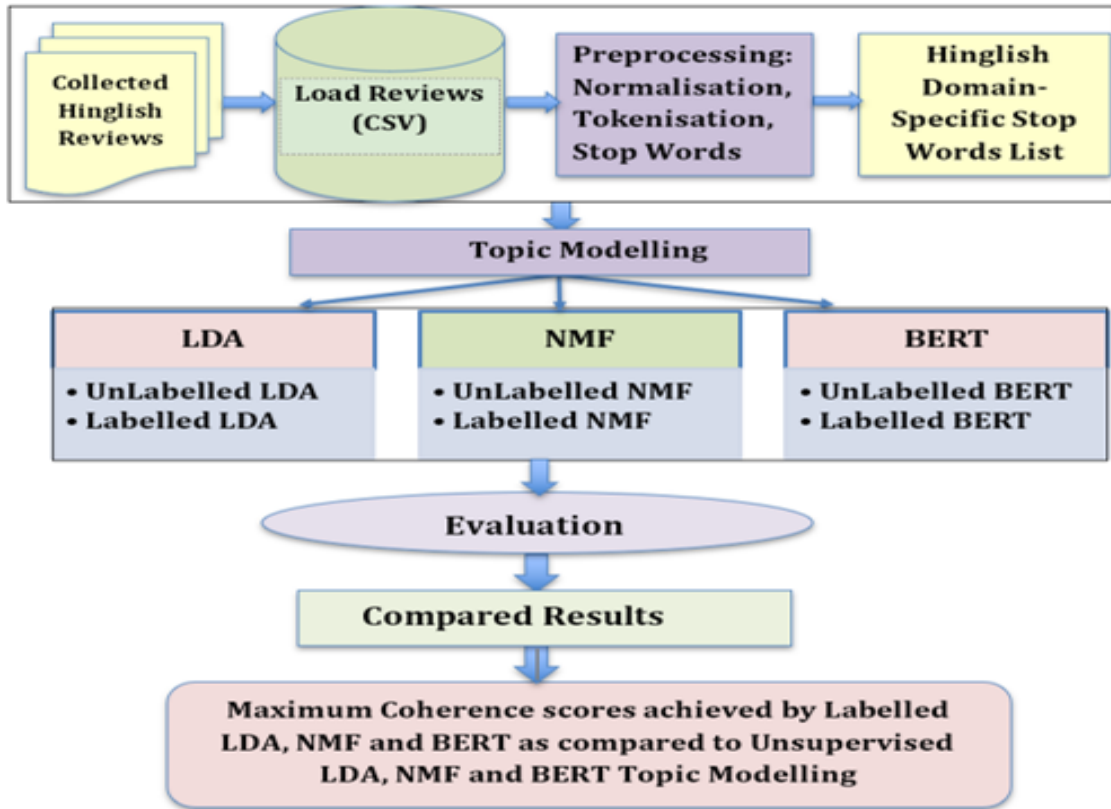


Figure 4.3: Pipeline of the methodology of unlabelled and labelled topic modelling

Figure 4.3 presents the methodology pipeline for extracting and visualising important topics from reviews. The first step is data collection and load data into a CSV file. In the second step, apply text pre-processing techniques, such as lowercase conversion, tokenisation and stop word removal. In the fourth step, we have experimented with topic modelling techniques on unlabelled and labelled Hinglish reviews' documents. The fifth step is to evaluate the topics and compare the results of extracted topics using both unlabelled and labelled topic modelling.

4.5.1 Data Collection

Data collection is the most important task of any machine learning pipeline. We have collected Hinglish product reviews of kids', ladies' dresses and mobiles from e-commerce websites.

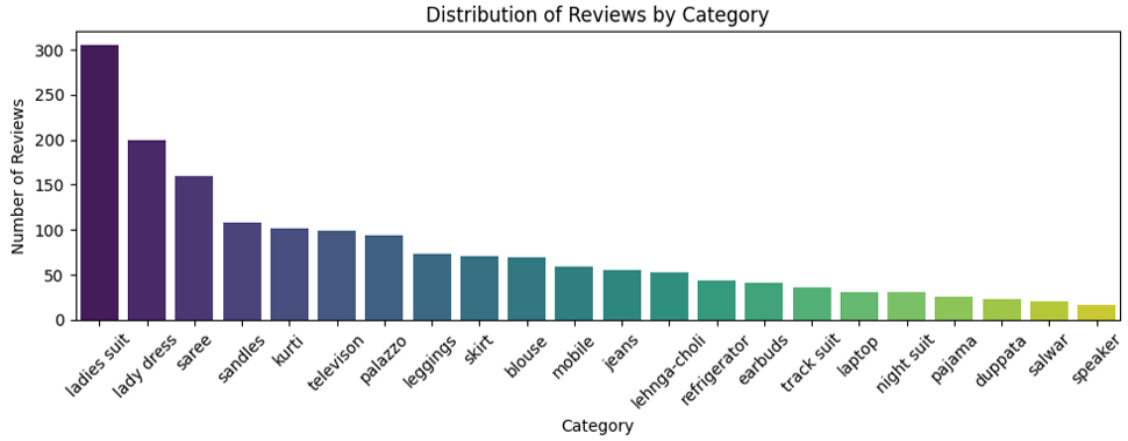


Figure 4.4: Hinglish Reviews' Products Categories Analysis

Figure 4.4 presents the distribution of product categories of Hinglish Reviews. There are 300 reviews of ladies' suits, 200 reviews of lady dress, approx.160 reviews of saree, 120 reviews of sandles, 100 reviews of television and palazzo and approx.50 reviews of duppata, pajama and so on.

4.5.2 Exploratory Analysis of Hinglish Reviews

4.5.2.1 Review Word and Character Length Distribution

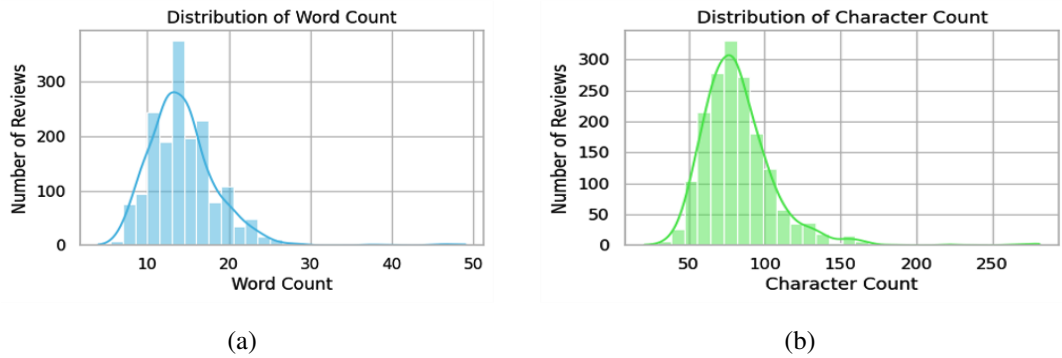


Figure 4.5: (a)Word length distribution (b) Character Length distribution

The figure 4.5(a) shows the distribution of word length distribution. It shows approx.180 reviews have an average length of 15 words, and rare reviews have a word length of 28, and (b) shows the character length distribution of reviews.

4.5.3 Preprocessing

After cleaning the Hinglish reviews, various preprocessing techniques are applied to the reviews. Case conversion is performed by converting reviews into lowercase using the regular expression library, and tokenisation and stop words removal are performed.

4.5.3.1 Review Length Distribution of Each Product Including and Excluding Stop Words

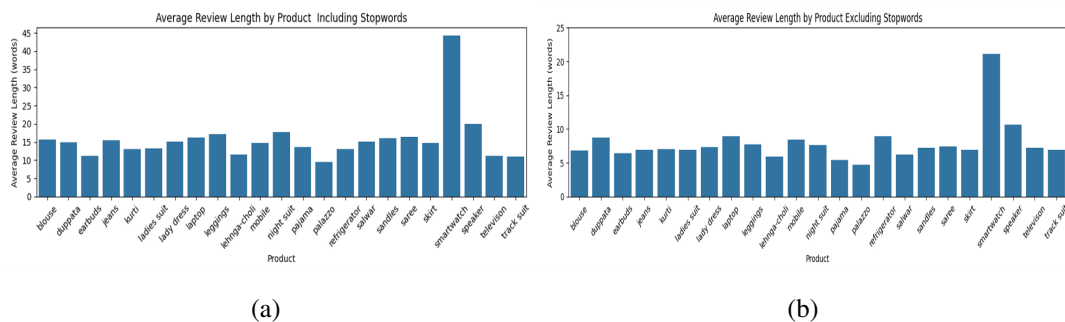


Figure 4.6: (a) Word Length Distribution Excluding Stopwords (b) Word Length Distribution Excluding Stopwords

Figure 4.6 (a) presents the word length distribution of individual products, including stopwords, and the average of review length is approx.45 words, maximum and minimum is 10 words. We can say that most of the time, reviewers write the stop words in reviews, which are irrelevant to extracting topic words, and (b) shows the word length after stop word removal. After removing the stop words maximum word length is approximately. 20 words, and a minimum is 5 words.

4.5.4 Most Frequent Words Including and Excluding Stop Words

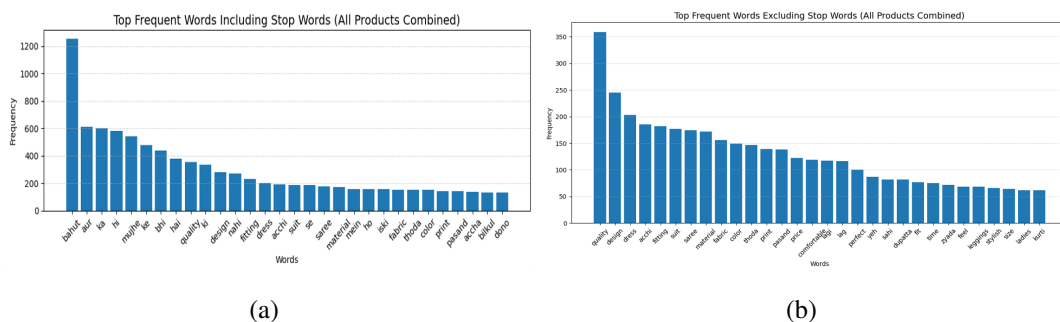


Figure 4.7: (a) Word Length distribution Excluding Stopwords (b) Most Frequent Words

Figure 4.7 (a) shows the most frequent words including hinglish stop words, are as bahut, ka, hai, etc., and (b) shows the most frequent words excluding hinglish stop words. To remove Hinglish stop words, we have created a Hinglish domain-specific stop words list.

4.5.4.1 Distribution of Positive and Negative Hinglish Product Reviews

The Hinglish reviews contain many negative and positive reviews about each product. We have performed sentiment analysis using VADER and TextBlob.

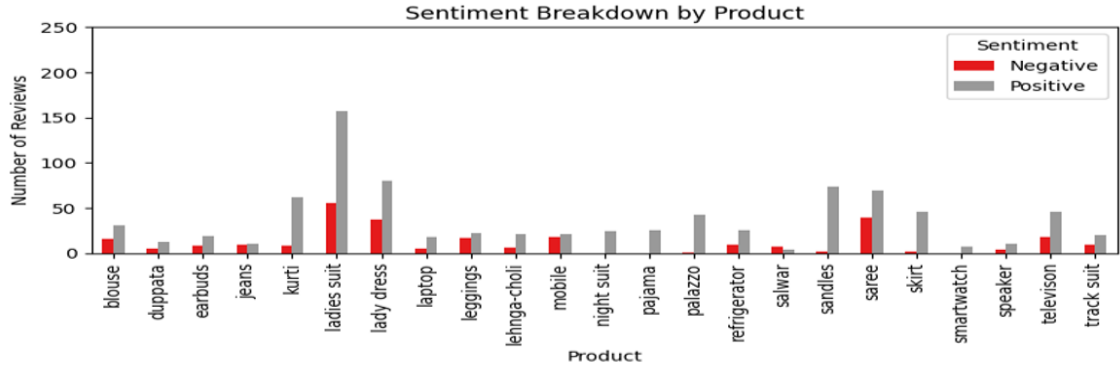


Figure 4.8: (a)Sentiment Distribution of Hinglish Product Reviews

Figure 4.8 shows the frequency distribution of positive and negative reviews based on the product categories. The x-axis represents the product names, and the y-axis shows the number of reviews. The red colour represents the negative reviews, and the grey colour represents the positive reviews. For the ladies’ suit, 150 reviews present negative opinions in the reviews and approximately sixty reviews discuss the positive aspects of the product.

4.5.5 Unigram, Bigram and Trigram Count Analysis of Reviews

Table 4.2: Unigram, Bigram and Trigram Count Analysis of Reviews

Unigram	Count	Bigram	Count	Trigram	Count
quality	361	dress material	61	saree print color	22
design	281	acchi lagi	55	chanderi work dupatta	20
fitting	233	flip flops	55	sidhidata womenâ heavy	19
saree	177	quality acchi	51	womenâ heavy cotton	19
fabric	156	blouse fitting	51	heavy cotton embroidery	19

Table 4.2 presents some examples of the unigram, bigram, and trigram counts in the reviews.

4.5.6 Experimentation with Unlabeled and Labelled Topic Modelling

4.5.6.1 Latent Dirichlet Allocation

LDA topic modelling is a natural language processing technique used to identify hidden topics in text documents and provide a high-level understanding of the main ideas by clustering the documents’ words, preserving their contextual meaning. It refers to the procedure that describes elements of an extensive dataset [141]. LDA Topic modelling

is a statistical and mathematical model for identifying hidden topics in multiple text documents LDA is a probabilistic model that discovers topics from documents [142], where each document is a mixture of topics, and each topic is represented by a distribution of words. The figure below represents the implementation of the LDA model.

4.5.6.2 Working of LDA model

$$P(w \mid d) = \sum_{k=1}^K P(w \mid z = k) \cdot P(z = k \mid d) \quad (4.1)$$

- $P(w \mid d)$: Probability of word w in document d
- z : Latent topic variable
- K : Number of topics
- $P(w \mid z = k)$: Word distribution for topic k
- $P(z = k \mid d)$: Topic distribution for document d

4.5.7 Experimentation Using LDA on Unlabelled Data

The stop words are irrelevant and unimportant in extracting the important words of the topics. However, the NLTK tool can only provide English stop words, not Hinglish, so we created a domain-specific list of Hinglish stop words and removed Hinglish stop words; then, we experimented with unlabeled LDA.

4.5.7.1 Document-term Matrix using Unlabelled LDA

1 get_lda_topics(ldamodel, num_topics)										
	Topics # 01	Topics # 02	Topics # 03	Topics # 04	Topics # 05	Topics # 06	Topics # 07	Topics # 08	Topics # 09	Topics # 10
0										
1	t	thoda	quality	is	pasand	quality	color	quality	fitting	design
2	design	tha	acchi	quality	color	design	quality	pasand	print	quality
3	kurs	fitting	design	acchi	fitting	fitting	Dupatta	design	color	time
4	fitting	color	fitting	fabric	acchi	Work	material	Blouse	material	acchi
5	perfect	material	Saree	comfortable	design	Blouse	print	Dress	Saree	acchi
6	quality	design	print	dress	Iski	the	Dress	color	dull	color
7	kharab	lagi	Bahut	Work	lagi	Suit	Suit	thoda	saree	service
8	color	quality	fabric	Dupatta	perfect	suit	tight	dress	price	TV
9	fit	comfortable	price	material	TV	lagi	pasand	Mughe	suit	Iski
10	stylish	time	lagi	Dress	quality	tha	Sabar	yeh	loo	Amazon
11	suit	perfect	sahi	Material	set	thoda	dull	set	design	dress
12	acchi	Blouse	dress	feel	Dress	print	fit	unique	uncomfortable	sahi
13	modern	suit	comfortable	suit	thoda	sahi	suit	aayi	sahi	pasand
14	Palazzo	uncomfortable	zyada	jeans	jaldi	time	Embroidery	Ye	zyada	wash
15	stitching	TV	Mughe	Suit	suit	Amazon	Cotton	the	chole	delivery
16	vibrant	tight	Blouse	soft	earbuds	average	Chanderi	Iski	earbuds	price
17	saree	saree	Quality	perfect	feel	price	Material	jate	Lehenga	fade
18	fabric	for	pasand	Kurs	pehna	Saree	Unstitched	acchi	premium	easy

Figure 4.9: Document-term matrix using Unlabelled LDA

Figure 4.9 shows the document term matrix of topics' words generated using the unlabeled LDA method. It shows the mixed topic words of all products. It encompasses many words across one or multiple topics. The total number of topics is 10, ranging from 0 to 9. The topic 0 words are: pasand, tight, design, quality, stitching, saree, smart, long and Topic 1 words are: comfortable, dress, quality, price, acchi, fabric, average,

sound, kapde, stitching, picture, reasonable. However, these are mixed topics across all products, which makes it very complex to identify specific words for particular products.

4.5.7.2 Visualisation using pyLDAvis

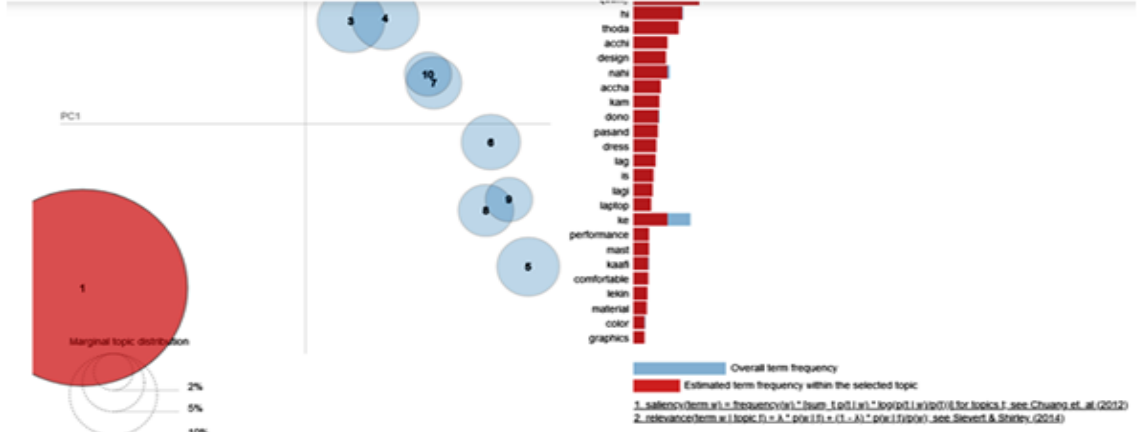


Figure 4.10: Visualisation of topics' words using pyLDAvis

Figure 4.10 shows the Visualization of topic word distributions and clusters of words that segregated the topics.

4.5.7.3 Non-Negative Factorization

NMF is a non-probabilistic, unsupervised matrix factorisation method, which includes a linear algebraic algorithm. It utilises the TF-IDF approach, where the semantic relationships between words determine their weight. To create a topic matrix, the word with the highest weightage in terms of frequency is considered the topic. NMF factorises a non-negative matrix into two smaller non-negative matrices. Mostly, it aligns with real data where negative values do not make any sense. According to [143], the NMF method projects data into low dimensions by reducing the number of features, breaking down a matrix into two matrices, and working with non-negative coefficients only, which results in a matrix of the product of two non-negative matrices by reducing the dimensions.

4.5.7.4 Working of NMF

NMF factorizes the term-document matrix V into two non-negative matrices W and H :

$$V \approx WH \quad (4.2)$$

- $V \in \mathbb{R}_{\geq 0}^{m \times n}$: Term-document matrix (terms $\times$ documents)
- $W \in \mathbb{R}_{\geq 0}^{m \times k}$: Term-topic matrix
- $H \in \mathbb{R}_{\geq 0}^{k \times n}$: Topic-document matrix
- k : Number of topics

4.5.7.5 Document-term Matrix of Unlabelled NMF

1 topic_words_nmf.head(20)

	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
0	fitting	dress	quality	design	se	lagl	suit	print	thoda
1	blouse	material	price	unique	tarah	acchi	perfect	saree	zyada
2	saree	work	acchi	pasand	wajah	pasand	size	color	phone
3	amazon	dupatta	fabric	stylish	flops	lag	track	dull	price
4	sahi	embroidery	bekar	modern	flip	skirt	set	fabric	features
5	time	chanderi	average	attractive	dress	aayl	night	bakwas	time
6	delivery	heavy	sound	lag	hisab	comfortable	comfortable	shipping	average
7	zyada	salwar	kapde	trendy	battery	yeh	fitting	sahi	cooling
8	service	cotton	stitching	yeh	waterproof	flops	chart	lamba	smart
9	price	unstitched	picture	elegant	backup	flip	fit	bright	size
10	perfect	sidhidata	product	kurti	hoon	feel	stylish	laga	sakta
11	tight	women	reasonable	dress	din	colors	pasand	unique	noise
12	uncomfortable	yeh	durable	laga	leggings	color	printed	boring	battery
13	acchi	fabric	thodi	saree	tv	perfect	achi	time	sound
14	brand	feel	brand	set	size	vibrant	color	pehanne	slow

Figure 4.11: Document-term matrix of extracted topics using unlabeled NMF

Figure 4.11 shows the document term matrix of 9 topics' words generated using the NMF method. The total topics is 9 from 0 to 8. The topic 0 words are: fitting, blouse, saree, amazon, sahi, time, delivery, zyada, service, price and Topic 1 words are: quality, price, acchi, fabric, average, sound, kapde, stitching, picture, reasonable. However, these are mixed topics across all products, which makes it very complex to identify specific words for particular products.

4.5.8 Topic Modeling using BERTopic

BERT is a transformer-based model developed by Google and is a state-of-the-art language model that has revolutionised various NLP tasks. BERTopic is a topic modelling technique that uses cTF-IDF to extract important, meaningful topics from unstructured text and create dense clusters. BERTopic combines transformer-based embeddings, dimensionality reduction, and clustering.

4.5.8.1 Visualization of Topics' Word Distribution using BERT

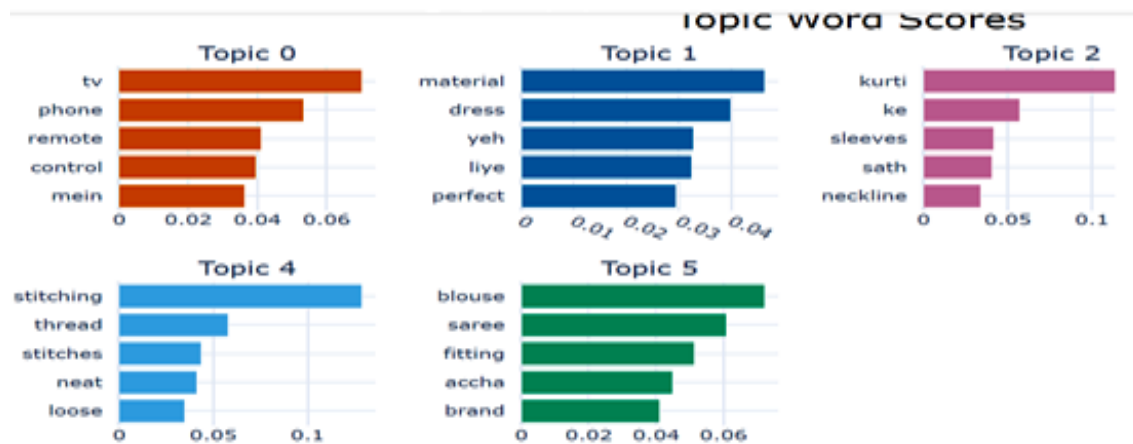


Figure 4.12: Score of topic's words

Figure 4.12 explains the words related to topics; topic 5 saree reviews contain the words 'blouse, saree, fitting, and brand', and the probability of the blouse word is higher than other words. BERTopic gives more accurate words related to topics. It can be easily examined through visualisation that reviews are of 5 categories. Topic 0 contains words related to TV. And phones. Topic 1 and 2, 3 about dresses and 4 about saree reviews.

4.5.8.2 Hierarchical Clustering using BERTopic

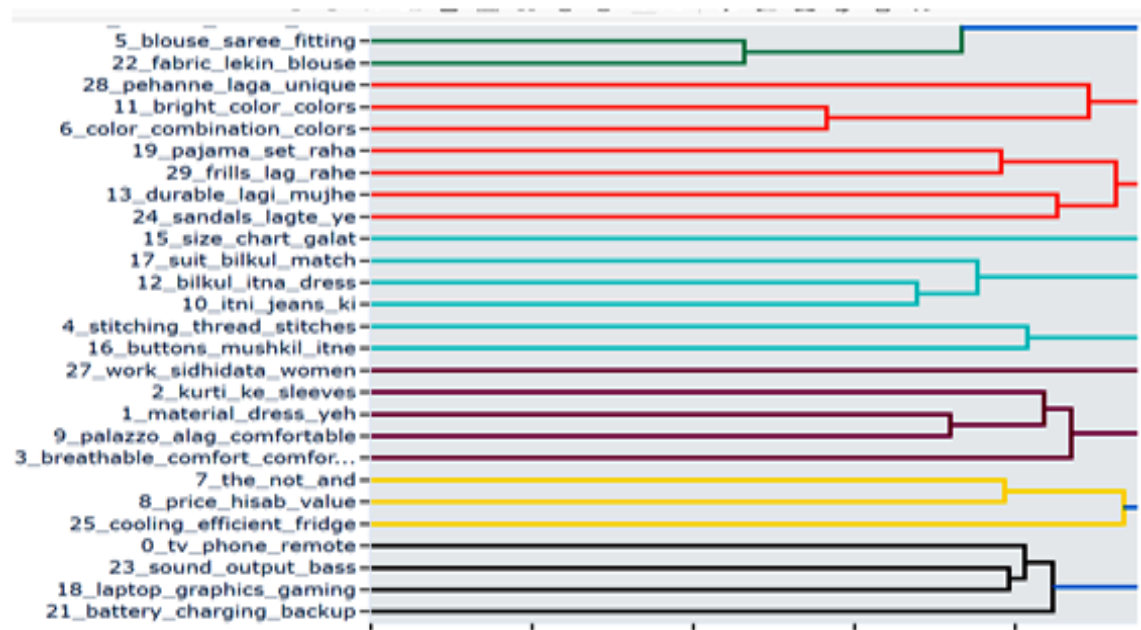


Figure 4.13: Hierarchical clustering of topics

Figure 4.13 demonstrates the hierarchical clustering of different topic words into different colours. Similar topic words are clustered in red, and similar words of other topic words are clustered in green colour.

4.5.9 Topics Extracted Using Labelled Product-aware Topic Modelling

4.5.9.1 Topics Words Extracted using Labelled LDA

[9] 1 df_lda.head(20)

	Product	LDA_Topic_1	LDA_Topic_2
0	blouse	bloufitting, print, quality, fitting, saree	acchi, thoda, shipping, time, quality
1	duppata	material, dress, chanderi, work, dupatta	poorly, dress, material, chanderi, dupatta
2	earbuds	bass, not, the, quality, thoda	noicancellation, samsung, design, bout, earbuds
3	jeans	size, sahi, dhone, kapda, jeans	pehnne, kamarband, purane, rang, jeans
4	kurti	sath, fabric, design, kurti, ladies	design, comfortable, quality, the, kurtat
5	ladies suit	price, the, design, quality, suit	tight, size, material, suit, fitting
6	lady dress	pasand, fabric, material, lag, dress	bekar, lagi, quality, yeh, dress
7	laptop	display, graphics, performance, quality, laptop	heat, average, experience, laptop, gaming
8	leggings	material, quality, jaate, pehnne, leggings	ya, in, size, fabric, leggings
9	lehnga-choli	design, acchi, choli, price, lehenga	bohot, dupatta, choli, lehenga, pasand
10	mobile	camera, zyada, software, thoda, phone	charging, quality, thodi, thoda, phone
11	night suit	hoon, suit, night, perfect, design	achi, fabric, quality, night, suit
12	pajama	pajamat, material, aayi, pasand, comfortable	trendy, stylish, colors, lag, pajamat
13	palazzo	material, waistband, quality, acchi, size	pasand, compliments, khushi, design, palazzo
14	refrigerator	zyada, door, storage, thoda, cooling	design, features, thoda, noilevel, smart
15	salwar	uncomfortable, kapdo, color, material, salwar	print, smudges, jaate, salwar, mei
16	sandles	pairon, acchi, laqi, flip, flops	laqi, design, flops, flip, sandals

Figure 4.14: Topics Words Extracted using Labelled LDA

Figure 4.14 shows the extracted num_2.topics using labelled LDA with product categories. This approach extracts the meaningful non-duplicate topic words and relevant topics related to the product.

4.5.9.2 Extracted Topics Using Labelled NMF

1 df_nmf.head(23)

	Product	NMF_Topic_1	NMF_Topic_2
0	blouse	bakwas, fabric, print, saree, bloufitting	amazon, price, delivery, zyada, fitting
1	duppata	rich, dupatta, chanderi, feel, texture	unstitched, women, material, dress, work
2	earbuds	samsung, bass, quality, bout, earbuds	works, awaaz, deti, noicancellation, sunai
3	jeans	unevenly, dhone, fade, rang, purane	chala, smell, ajeeb, bar, jeans
4	kurti	sath, design, fabric, kurti, ladies	material, color, quality, kapde, kurtat
5	ladies suit	perfect, price, suit, quality, fitting	salwar, dress, dupatta, work, material
6	lady dress	design, yeh, dress, lag, pasand	cotton, heavy, embroidery, material, work
7	laptop	acchi, backup, battery, display, quality	system, cooling, heat, impressive, graphics
8	leggings	fabric, quality, leggings, material, jaate	elastic, der, pehne, tight, waistband
9	lehnga-choli	aayi, color, pasand, bohot, dupatta	stitching, pasand, design, choli, lehenga
10	mobile	tak, software, zyada, phone, thoda	low, thodi, average, camera, quality
11	night suit	price, quality, pasand, night, suit	satisfied, comfortable, feel, fabric, hoon
12	pajama	pajamat, trendy, colors, stylish, lag	pajamat, fabric, comfortable, aayi, pasand
13	palazzo	size, fit, khushi, perfect, palazzo	lagi, stitching, fabric, quality, acchi
14	refrigerator	cooling, space, noilevel, design, thoda	efficient, installation, complicated, energy, ...
15	salwar	bleeding, transfer, kapdo, color, salwar	feel, sheer, tarahe, material, kapde
16	sandles	pairon, acchi, flops, flip, lagi	mujhpuchha, stylish, trendy, friends, sandals
17	saree	zyada, price, quality, sahi, fabric	saree, design, acchi, print, color

Figure 4.15: Extracted Topics Using Labelled NMF

Figure 4.15 shows the extracted coherent and meaningful unique topics nmf1_topic, nmf2_topic with the product's name using the labelled NMF method.

4.5.9.3 Visualisation of Topic Words using Unlabeled LDA, NMF and BERT

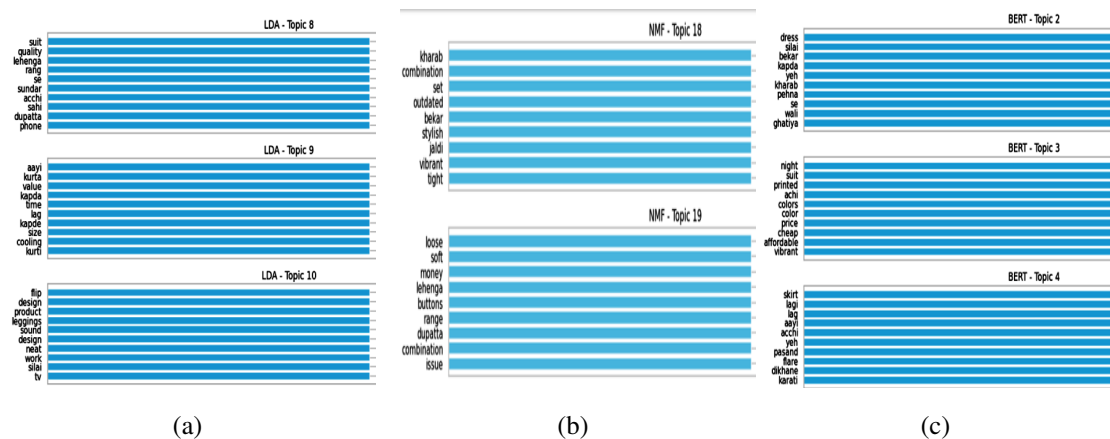


Figure 4.16: (a) Topics words extracted using LDA (b) Topic words extracted using NMF (c) Extracted Topic Words using BERT

Figure 4.16(a) shows the top words from 8,9,10 topics are extracted using unlabeled LDA. Topic 8 words are suit, quality, lehenga, sundar, sahi, dupatta, phone, which are a mixture of several product categories. (b) shows that the top words from 18 and 19 topics are extracted using unlabeled NMF. The top words are kharab, combination, outdated, bekar, stylish, vibrant, and tight, as demonstrated by topic 18 (c). The top words dress, silai, bekar, kapda, and ghatiya are extracted from topics 2, 3, and 4 using BERT.

4.5.9.4 Visualisation of Topic Words using Unlabeled LDA, NMF and BERT

```
1 df_topics.head(10)
```

	topic_id	unlabelled_lda	unlabelled_nmf	unlabelled_bertopic
0	0	tv, ladies, kurti, sakte, suit, palazzo, desig...	quality, price, fabric, bekar, sound, acchi, a...	ladies, kurti, sleeves, sath, neckline, kurta,...
1	1	display, battery, backup, din, suitt, smartwat...	work, suit, dupatta, material, embroidery, cha...	leggings, jaate, pehne, elastic, waistband, ch...
2	2	thoda, phone, cooling, features, smart, zyada,...	saree, print, color, blouffitting, bakwas, dull...	tv, remote, control, sakte, built, screen, eas...
3	3	saree, fabric, quality, print, acchi, bloufitt...	acchi, lagi, skirt, flip, flocs, lag, pairon, ...	flocs, flip, pairon, ache, lage, sakti, lagi, ...
4	4	size, suit, perfect, quality, fit, fade, dress...	fitting, time, sahi, price, zyada, delivery, b...	dress, bekar, kapda, ghatiya, pehna, yeh, sila...
5	5	comfortable, leggings, feel, lagi, flocs, flip...	design, stylish, unique, trendy, modern, attra...	skirt, lagi, lag, aayi, acchi, yeh, flare, kar...
6	6	design, pasand, lag, colors, stylish, dress, a...	dress, yeh, lag, material, bekar, buttons, fit...	breathable, smooth, feel, material, comfortabl...
7	7	quality, fitting, price, acchi, zyada, average...	comfortable, fabric, material, feel, soft, leg...	jeans, purane, rang, bar, kamarband, dhone, sa...
8	8	quality, length, dress, pasand, color, suit, l...	thoda, phone, zyada, price, features, battery...	price, hisab, reasonable, value, money, deal, ...
9	9	material, dress, dupatta, work, chanderi, unco...	pasand, aayi, color, lag, colors, vibrant, ski...	blouddesign, fitting, saree, amazonvice, shipp...

Figure 4.17: Topic Words using Unlabeled LDA, NMF and BERT

Figure 4.17 shows the extracted topics using unlabelled/unsupervised LDA, NMF and BERTopic modelling techniques.

4.5.9.5 Word Cloud of Topic Words using LDA NMF and BERTopic



Figure 4.18: (a) Saree, palazzo, duppata topic words using LDA, NMF and BERT (b) Ladies suit, lehenga choli, mobile topic Words extracted using LDA, NMF and BERT

Figure 4.18 (a) shows the word cloud images of products saree, palazzo, duppata topic words using LDA, NMF, and BERT and (b) shows the ladies suit, lehenga choli, mobile product topic Words extracted using LDA, NMF and BERT. It shows better results compared to unlabelled topic modelling methods. It displays the top extracted words based on the products.

4.5.9.6 Extracted Topics using Labelled BERTopic

```
1 df_bert.head(10)
```

	Product	BERTopic_Topic_1	BERTopic_Topic_2
0	blouse	saree, print, bloufitting, fabric, quality	delivery, fitting, amazon, time, price
1	duppata	dupatta, gir, poorly, easily, unkempt	work, dress, material, heavy, embroidery
2	earbuds	earbuds, boult, fit, samsung, comfortable	deti, awaaz, sunai, noicancellation, works
3	jeans	jagah, sahi, size, fit, chhoti	quality, uske, practically, patla, dikhta
4	kurti	kurtat, quality, fabric, kapde, the	print, attractive, fine, motifs, placement
5	ladies suit	formal, events, elegant, occasions, casual	breathable, hours, long, soft, fabric
6	lady dress	tight, fitting, thoda, lagi, fit	print, unique, boring, dull, lifeless
7	laptop	laptop, performance, gaming, quality, display	heat, cooling, graphics, laptop, impressive
8	leggings	uncomfortable, band, sweat, waist, transparency	stitching, khulne, stitches, durability, zero
9	lehnga-choli	dupatta, bohota, color, aayi, pasand	design, lehenga, colors, choli, laga

Figure 4.19: Topics Words Extracted using BERTopic

Figure 4.19 shows the extracted topics using BERTopic model and this model classifies and extracts the topics based on the products.

4.5.9.7 Distribution of BERT Topics Words

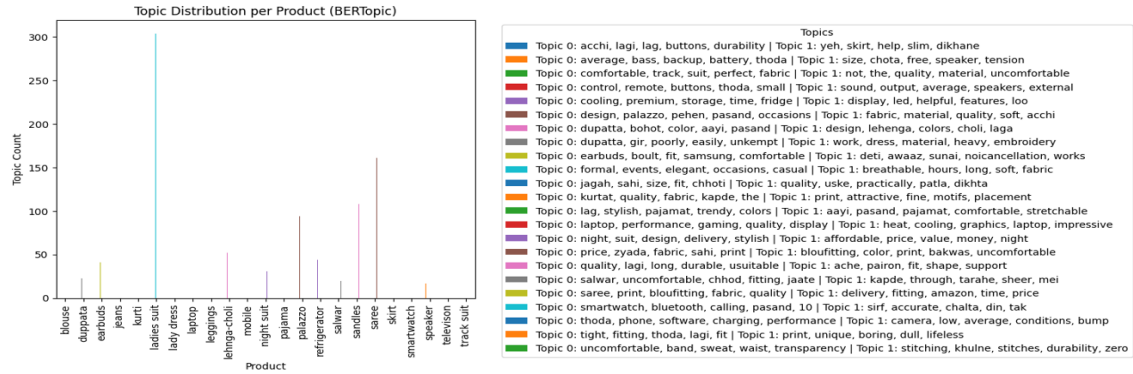


Figure 4.20: Distribution of BERT Topics Word

Figure 4.20 shows the topic names with the distribution of product categories. It describes that BERT has extracted max topics from ladies' suits and no topics from night suits, kurti, jeans, and mobile extracted from blouses.

4.5.10 Sentiment Analysis and Labelled LDA Topics

Table 4.3: Sentiment Analysis of Reviews and d Labelled LDA Topics

Product	Negative	Neutral	Positive	Labelled LDA Topics
blouse	13	36	19	Topic 1: blouffitting, print, quality, fitting, saree; Topic 2: accchi, thoda, shipping, time, quality
duppata	1	11	11	Topic 1: material, dress, chanderi, work, dupatta; Topic 2: poorly, dress, material, chanderi, dupatta
earbuds	6	17	18	Topic 1: bass, not, the, quality, thoda; Topic 2: noicancellation, samsung, design, bout,
jeans	8	36	11	Topic 1: size, sahi, dhone, kapda, jeans; Topic 2: pehnne, kamarband, purane, rang, jeans
kurti	7	44	50	Topic 1: sath, fabric, design, kurti, ladies; Topic 2: design, comfortable, quality, kurta
ladies suit	58	130	116	Topic 1: price, design, quality, suit; Topic 2: tight, size, material, suit, fitting

Continued on next page

Table 4.3 – continued from previous page

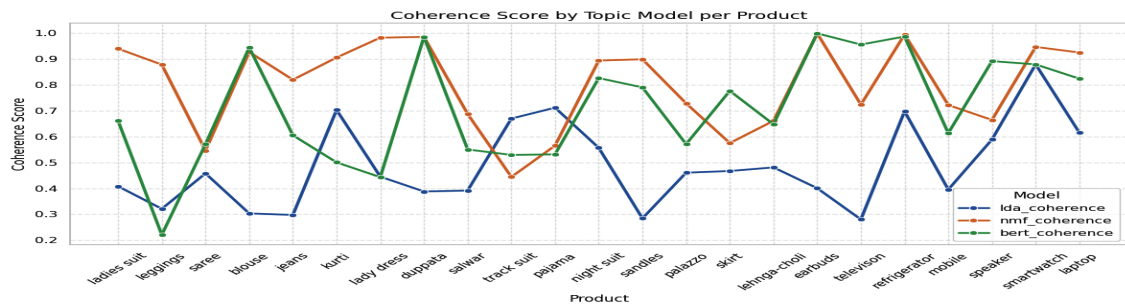
Product	Negative	Neutral	Positive	Labelled LDA Topics
lady dress	47	86	66	Topic 1: pasand, fabric, material, lag, dress; Topic 2: bekar, lagi, quality, dress
laptop	3	15	13	Topic 1: display, graphics, performance, quality, laptop; Topic 2: heat, average, experience, laptop, gaming
leggings	19	32	22	Topic 1: material, quality, jaate, pehne, leggings; Topic 2: size, fabric, leggings
lehenga-choli	4	25	23	Topic 1: design, acchi, choli, price, lehenga; Topic 2: bohot, dupatta, choli, lehenga, pasand
mobile	14	28	17	Topic 1: camera, zyada, software, thoda, phone; Topic 2: charging, quality, thodi, thoda, phone
night suit	1	8	22	Topic 1: suit, night, perfect, design; Topic 2: achi, fabric, quality, night, suit
pajama	3	5	18	Topic 1: pajamat, material, aayi, pasand, comfortable; Topic 2: trendy, stylish, colors, lag, pajamat
palazzo	3	52	39	Topic 1: material, waistband, quality, acchi, size; Topic 2: pasand, compliments, khushi, design, palazzo
refrigerator	0	13	31	Topic 1: zyada, door, storage, thoda, cooling; Topic 2: design, features, thoda, noilevel, smart
salwar	4	12	4	Topic 1: uncomfortable, kapdo, color, material, salwar; Topic 2: print, smudges, jaate, salwar, mei
sandles	18	27	63	Topic 1: pairon, acchi, lagi, flip, flops; Topic 2: lagi, design, flops, flip, sandals
saree	43	79	39	Topic 1: fabric, color, print, quality, saree; Topic 2: material, design, quality, bloufitting, saree
skirt	13	15	42	Topic 1: acchi, lagi, skirt; Topic 2: fabric, color, aayi, pasand, skirt

Continued on next page

Table 4.3 – continued from previous page

Product	Negative	Neutral	Positive	Labelled LDA Topics
smart watch	0	1	8	Topic 1: quality, bluetooth, calling, pasand, acchi; Topic 2: sirf, accurate, tak, din, chalta
speaker	0	6	11	Topic 1: sound, tension, free, wajah, waterproof; Topic 2: waterproof, wajah, treble, average, bass
television	7	38	54	Topic 1: built, in, features, smart, tv; Topic 2: thoda, remote, control, quality, tv
track suit	11	9	16	Topic 1: pasand, quality, suit, track; Topic 2: fitting, perfect, track, suit

4.5.10.1 Coherence Scores of Topics Using Labelled LDA, NMF and BERTopic



(a)

Figure 4.21: Trending Line of Coherence Topics Scores using Labelled Topic Modelling

Figure 4.21 defines the trends or changes in coherence scores of extracted topics based on products' reviews, achieved using LDA, NMF, and BERTopic. The x-axis represents the product categories, and the y-axis indicates the coherence score of the topics. The trending line indicates that the coherence score using LDA has yielded good results in extracting topics for certain products. But most of the topics' maximum coherence has been achieved by NMF and BERT.

4.5.10.2 Coherence Scores of Extracted topics using Labelled LDA, NMF and BERTopic

Table 4.4: Coherence score of LDA NMF and BERT

Product	LDA Co-herence	NMF Co-herence	BERT Coherence	Mean Co-herence	Best Model
blouse	0.250833622	0.662767452	0.56671668	0.493439251	NMF
duppata	0.725798404	0.702363096	0.557326175	0.661829225	LDA
earbuds	0.554756638	0.594805358	0.566697808	0.572086601	NMF
jeans	0.368192725	0.680400058	0.651039233	0.566544005	NMF
kurti	0.52973986	0.565867625	0.408340293	0.501315926	NMF
ladies suit	0.31512044	0.618882655	0.482416491	0.472139862	NMF
lady dress	0.390407798	0.633641949	0.294816566	0.439622104	NMF
laptop	0.477915233	0.635835525	0.493458502	0.53573642	NMF
leggings	0.405254867	0.57927555	0.288708127	0.424412848	NMF
lehnga-choli	0.495297572	0.713122433	0.7024061	0.636942035	NMF
mobile	0.428764304	0.575903823	0.390302393	0.464990173	NMF
night suit	0.500231433	0.590184897	0.569808382	0.553408237	NMF
pajama	0.438846804	0.560994911	0.523015409	0.507619041	NMF
palazzo	0.442846701	0.57961059	0.45702948	0.493162257	NMF
salwar	0.40660115	0.666361319	0.409749036	0.494237168	NMF
sandles	0.501823001	0.599356215	0.487417595	0.52953227	NMF
saree	0.384848824	0.536791589	0.596600797	0.506080403	BERTopic
skirt	0.523424785	0.572334277	0.572895612	0.556218224	BERTopic
smartwatch	0.737523288	0.669899011	0.745049636	0.717490645	BERTopic
speaker	0.630203848	0.738140178	0.595721001	0.654688342	NMF
television	0.280270373	0.703368702	0.684739901	0.556126325	NMF
track suit	0.606697476	0.747631616	0.494364153	0.616231082	NMF

Table 4.4 describes the product category wise labelled LDA, NMF, BERTopic coherence scores and mean coherence scores with the best model score according to maximum mean scores. It shows the most of the cases NMF and BERTopic is considered as best model as compared to LDA model.

4.6 Summary of Recent Studies on Topic Modelling

Table 4.5: Some Recent Studies of Topic Modelling Methods

Ref	Dataset	Language	Methods	Evaluation Metrics
[136]	Twitter posts	English	LDA, NMF, BERTopic	Coherence scores
[144]	Amar Ujala Hindi news articles	Hindi	LSI, NMF, LDA	Coherence, Perplexity Scores
[142]	14000 airline reviews collected from Asia.	English	LDA, TF-IDF	Recall, Precision
[145]	Cystic fibrosis comments from social media posts	English	LDA, Top2vec, Doc2vec,	Coherence score
[146]	1163 reviews were collected from 3 fast food restaurants.	English	LDA	Coherence score
[147]	IMDB dataset of movie reviews	English	LDA on Words, Topics and Documents	Precision, Recall, F1-score, support
[148]	111,728 documents from online newspapers	Arabic	BERTopic, LDA, NMF, AraBERT,	Normalised point-wise PMI.
[140]	20 newsgroup and short text data from Facebook	English	LSA, LDA, NMF, PCA, RP	Precision, Recall, F-Score, and coherence

Table 4.5 describes the existing research work of topic modelling techniques and it shows that most researchers have used unlabelled topic modelling on English, Arabic or Hindi. Still, there is a scope to experiment with the labelled topic modelling on Hinglish code-mixed product reviews.

4.7 Comparison of Unlabelled/Labelled Topic Models

Table 4.6: Comparison of Topic Modeling Methods

Techniques/Work	Pros	Cons
Unlabelled LDA Extracts a hidden mixture of topics based on probability and can be applied to any domain data.	Works on the unlabelled data, and hyperparameters can be customised.	Ignores word order, requires stop word removal, not find relationships in topics.

Continued on next page

Table 4.6 – continued from previous page

Techniques/Work	Pros	Cons
Unlabelled NMF • Decompose document matrix into topics according to weights. • Can capture latent topics from sparse, noisy, and heterogeneous data. • Can calculate each topic from each document.	• Faster than LDA and works better with brief texts. • Can handle missing values and outliers. • Can incorporate prior knowledge.	• Cannot measure topic importance. • It cannot determine an optimal number of features or clusters.
Unlabelled BERT • A state-of-the-art language model developed by Google. • Does not require prior knowledge.	• Considers context and semantic information.	• Takes one topic per document (not multiple topics). • Requires stop word removal preprocessing.
• It learns label-topic-word distributions. An extension of LDA Model where each document is associated with known labels (e.g., product names). • Each label corresponds to specific topics, restricting the topic model to only use the topics that are associated with the document.	• Work as supervised and topics align with labels; It Can be better for classification tasks. • Easier to interpret because topics are associated with known labels.	• It requires labelled data and slightly more complex implementation because it uses Guided LDA • It is less flexible if a document has a noisy or incorrect label, as it may degrade topic quality.

Continued on next page

Table 4.6 – continued from previous page

Techniques/Work	Pros	Cons
Labelled NMF - NMF approximates the document-term matrix with non-negative factors. - In labeled form, initial topic-word seeds are aligned with known labels to guide topic learning.	- It is most fast and scalable and more efficient than probabilistic models. - Can be guided by using seeded initialisation of labels for better control.	- It requires seed words for each label and produces redundant topics if the seed terms are not well-chosen. - It is harder to tune and lacks the probabilistic interpretability.
Labelled BERT - BERTopic uses clustering for topic modeling and allows fine-tuning and constraints to encourage label-specific topics. - Labeling is achieved by mapping documents to known labels and then grouping embeddings accordingly.	- It utilises context-aware embeddings to capture semantic meaning, thereby clustering topics within groups. - It generates the most accurate topics and works well for short texts.	- It is computationally intensive and requires a GPU or a long runtime for large datasets; it has a risk of over-clustering and may produce too many dynamic topics. - It is harder to interpret compared to LDA or NMF.

Table ?? describes the comparison of Unlabelled and Labelled LDA, NMF and BERT topic modelling techniques with their characteristics and pros and cons.

4.8 Conclusion

This chapter experiments with unsupervised and supervised product-aware topic modelling techniques on Hinglish reviews. For the Hinglish reviews, applying product-aware guided topic modelling techniques is a novel approach. We have compared the results of extracted topics using both unlabeled and labelled methods, evaluating coherence and diversity scores. In this chapter, we have also described a few recent studies on topic modelling techniques in several languages, along with their pros and cons. The next chapter presents the aspect-based extractive-to-abstractive summarisation of English language reviews using a hybrid approach.

Chapter 5

Aspect-based Product Reviews Summarization Using Hybrid Approach

Although admirable work has been done on review summarisation, despite the growing popularity of aspect-based summarisation, several unresolved challenges hinder the complete potential of ABS. Aspect-based summarisation focuses on generating summaries based on aspects of products or input text features. Despite its growing popularity, there are several unresolved research issues and challenges that hinder the full potential of ABS. Existing work on aspect-based summarisation is limited to specific domains. The aspects of different domains (E.g., restaurants, electronics, hotels and fashion domain, especially ladies' clothing) have distinct aspect sets that may not generalise across datasets. To resolve the issue of aspect-based summarisation for clothing and mobile domains, we have proposed a hybrid approach for aspect-based summarisation.

5.1 Contribution of this Chapter

Table 5.1: Contribution For Aspect-based Summarisation

Terms	Our Contribution
Data Collection	Collected the approx.40k reviews from e-commerce websites.
Aspects of Multi-domain products	Prepared the list of aspects of all products and generated a summary based on the aspects with product names.
Likert Scale	We have performed a Likert scale on reviews to find the highly positive and highly negative reviews.
Omitted Irrelevant Reviews	We have omitted irrelevant reviews using Aspect-based features extraction using BERT. E.g., "I purchased it for my son and daughter's birthday"
Hybrid Extractive to Abstractive summary	The proposed work has generated a hybrid summary from an extractive to an abstractive summary.
Sentiment Analysis	Proposed BERT for sentiment analysis and included only positive and negative sentiments for summaries, and omitted the neutral sentiments.

Continued on next page

Table 5.1 – continued from previous page

Terms	Our Contribution
Multiple Aspects Handling	Our proposed work has addressed multiple aspects in the summary. E.g, "The camera is awesome, but the battery is damaged."
Annotation of Summary	A human annotator has performed the annotation of the abstractive summaries from extractive summaries.
Evaluation is done using Rouge, BLEU, BERT, and Human scores, intrinsic parameters	The summaries' evaluation is performed using Rouge, BLEU, BERT, human scores, and informative, conciseness, relevancy, coherence, keyword coverage, bi-gram diversity, parameters, etc.

Table 5.1 describes the novelty contribution of this chapter, which we have covered in detail with experimentation and summary evaluation results.

5.2 Introduction

Aspect-based summarisation indicates that summarising the product reviews focuses on key product aspects (e.g., size, price, camera, battery) instead of summarising the entire review text. It finds and organises sentences based on the features of a product and captures the opinions related to those aspects. This helps users to get a deep and brief understanding of what customers like or dislike about specific features of a product. Multiple products are related to different domains and have distinct features; some domain knowledge is also essential for preparing for the aspects of these products. The most important step in summarising product reviews is to include product features in the summary. To summarise the product reviews, several key aspects must be considered. In this chapter, we first provide a comprehensive understanding of the aspect-based summarisation techniques and methods for generating extractive and abstractive summaries in the related work. This chapter aims to summarise the product reviews based on the products' aspects with polarity using a hybrid approach of Rule-based and Transfer learning methods. Following this, the chapter presents the practical implications of the proposed hybrid approach, which can also be applied in real-world scenarios for aspect-based product review summarisation.

5.3 Proposed Methodology of Aspect-Based Reviews Summarisation

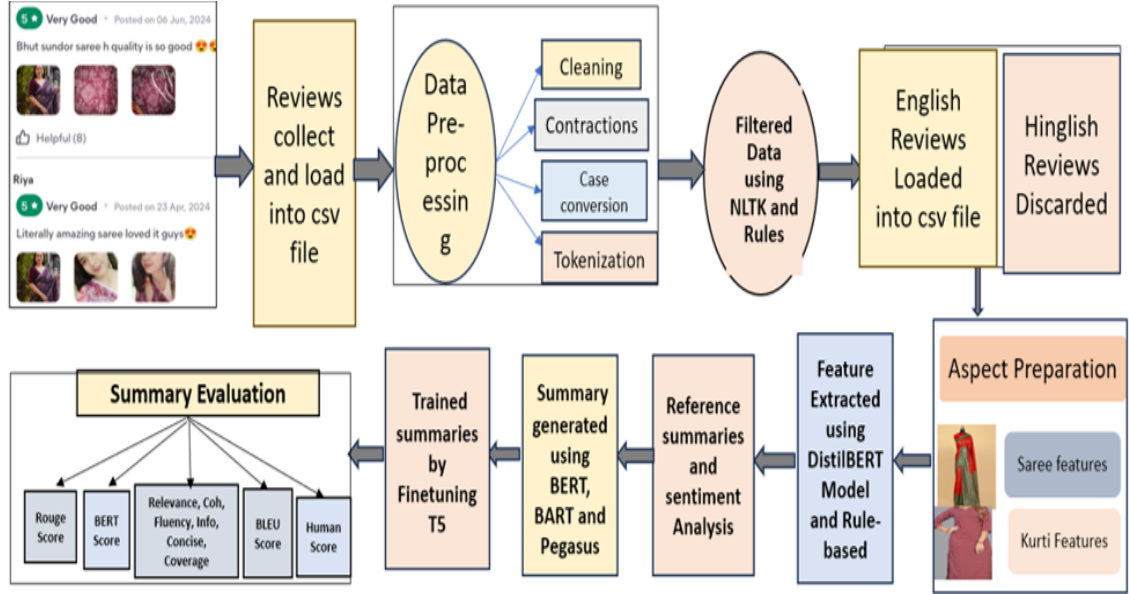


Figure 5.1: Block Diagram of Proposed Methodology

The Figure 5.1 shows the Block diagram of the proposed methodology for the Aspect-based summarisation. It describes all the steps from data collection to summarisation of reviews using Hybrid Rule-based and Transformer-based methods.

5.3.1 Data Collection

The data is collected through web scrapping e-commerce websites like Amazon and Meesho. We have scraped various categories of product reviews, including clothing for adults (Kurti, Sarees, Frock), kids' Kurtis, kids' dresses, and mobiles, using the Selenium and gspread libraries. We have crawled 10,180 saree reviews, 1,400 frock reviews, 7,560 kids' floral dresses, 1,205 reviews of kids' kurtis, and approximately 2600 reviews of adult kurtis. We have stored all these unstructured reviews, including columns for review text, customer name, and date, in CSV format.

5.3.2 Text Preprocessing

As we have mentioned in the 3rd chapter, stop word removal, stemming and lemmatisation techniques are not the best choice for summarization purposes [149]. These techniques can reduce the quality of the summary. So, we have not applied stemming, lemmatisation and stop word removal techniques. We used the q preprocessor and contractions libraries for text preprocessing.

5.3.2.1 Text Cleaning

We have cleaned the data by dropping null or NaN values in the first step. We cleaned the text using emojis, special symbols, numbers, unknown characters, and regular expressions.

5.3.2.2 Text Normalisation

We have applied the case conversion technique of upper case letters converted to lower case letters for normalising the reviews. Because "Ram and ram" have different vectors, this can change the entire meaning of the sentence.

5.3.2.3 Contractions

Many times, users write the short form in reviews, as I've done. But when converted into numerical vector form, the text changes meaning. The contraction library expands the contractions into the correct form.

5.3.2.4 Cleaned Reviews After Preprocessing

```
1 df.head()
```

	reviews_text	names	dates	help_f	Cleaned reviews
0	Best product 5/5 only @ 161. Thanks meesho	Mahalingayya	Posted on 5 Feb 2024	Helpful (32)	best product 55 only 161 thanks meesho
1	Bahut hi accha h thank you meesho	Jivani	Posted on 5 Feb 2024	Helpful (8)	bahut hi accha h thank you meesho
2	Nice frock , quality is good	Basubam Bairagi	Posted on 5 Feb 2024	Helpful (0)	nice frock quality is good
3	Very good quality as it is product thanks mesh...	Meesho User	Posted on 2 Feb 2024	Helpful (28)	very good quality as it is product thanks mesho
4	Nice dress, received so good in fabric. Colour...	Vinoth Kumar	Posted on 3 Feb 2024	Helpful (0)	nice dress received so good in fabric colour a...

Figure 5.2: Cleaned Data

Figure 5.2 demonstrates the data frame, including the cleaned reviews columns obtained after applying the pre-preprocessing techniques to the reviews using Regular Expressions and some NLP libraries.

5.4 Data Statistics and Exploratory Analysis

5.4.1 Reviews Categorical Distribution

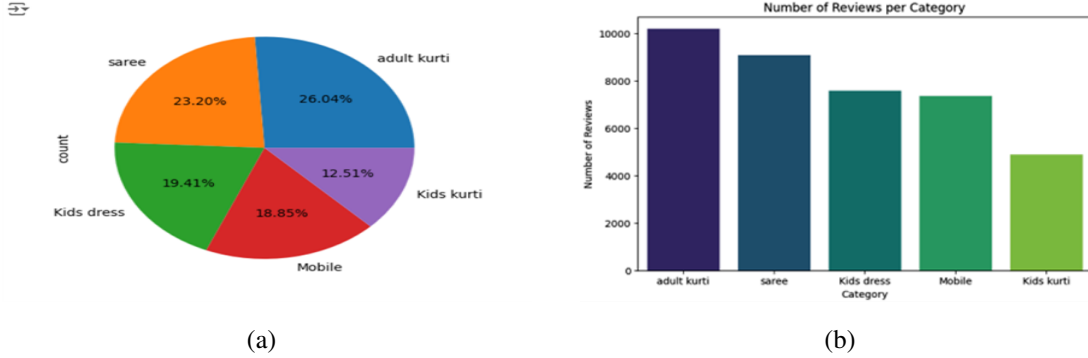


Figure 5.3: (a) Product reviews' percentage distribution (b) Reviews categories distribution

Figure 5.3(a) presents the percentage distribution of reviews according to product categories. It shows adult kurti- 26.04%, saree-23.20%, kids dress-19.41%, mobile-18.85%, kids kurti- 12.5% reviews. (b) shows the distribution of several categories of product reviews using the histogram. The x-axis defines the product's names, and the y-axis defines the number of reviews. The adult kurti has approx.10000 reviews, saree-8000, and kids' kurti have approx.5000 reviews.

5.4.2 Reviews Words' and Characters Length Distribution

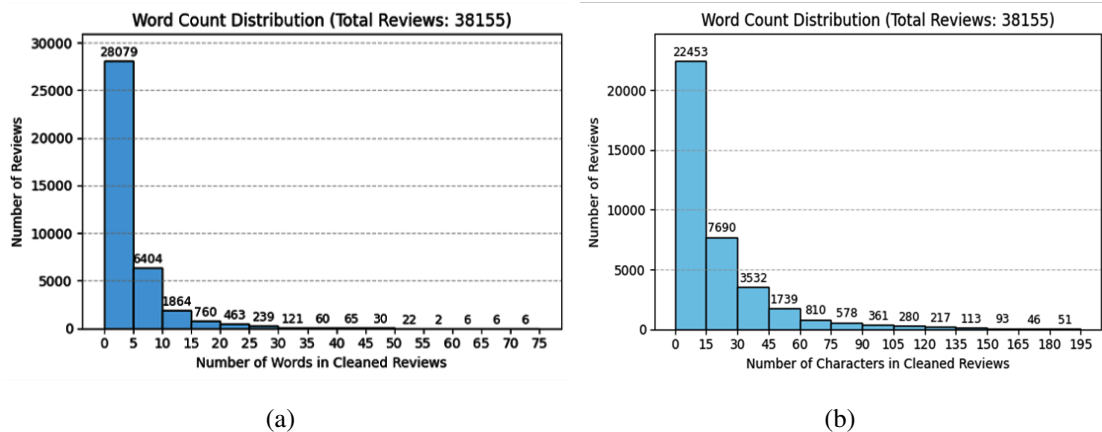


Figure 5.4: (a) Words Count Distribution (b) Characters Count Distribution in reviews

Figure 5.4 (a) shows the distribution of word count in reviews using a histplot. The x-axis shows the number of words, and the y-axis shows the number of reviews. The total number of reviews is 38155 after deleting null, nan and duplicate reviews. Most of the reviews are very short in length, and the figure shows that 28,079 reviews have 0 to 4 words, while 6,404 reviews have 5 to 10 words. The review length has a maximum of 4 words.

75 words and a minimum of 4 words. (b) explains the distribution of character counts from the reviews. The x-axis defines the number of characters, and the y-axis defines the number of reviews. The max characters in the reviews is up to 200, and the minimum number of characters in the reviews is between 0-15.

5.4.3 Distribution of Reviews Word Length Grouped By Products and Likert Label

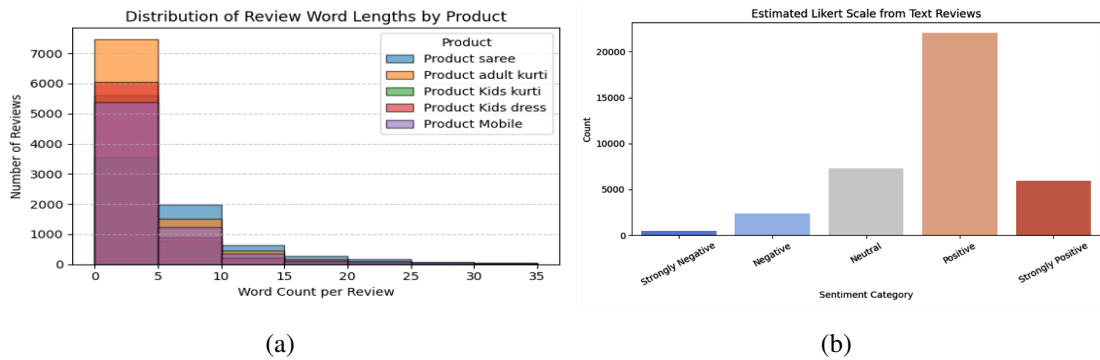


Figure 5.5: (a) Word length of reviews grouped by products (b) Likert label of reviews distribution

Figure 5.5 (a) shows the reviews word length distribution grouped by products. The x-axis defines the word count per review, y-axis presents the number of reviews. The maximum reviews are approximately 7000 reviews have only 5 words, and less than 1000 reviews have a word length of 15 words. Figure 5.5 (b) define the Likert label of reviews grouped by products. The x-axis presents the labels of sentiment categories, including strongly negative, negative, neutral, positive, and strongly positive. The y-axis represents the frequency of review counts and describes that there are 20k reviews are positive.

5.4.4 Distribution of Top Frequent Words

Our analysis also includes the top 50 frequently occurring words in the reviews. These words, extracted after counting the length of words in reviews, provide a snapshot of the most common sentiments expressed in the reviews. We have visualised these words as a word cloud, offering a quick and intuitive understanding of the most common themes in the reviews across different product categories.

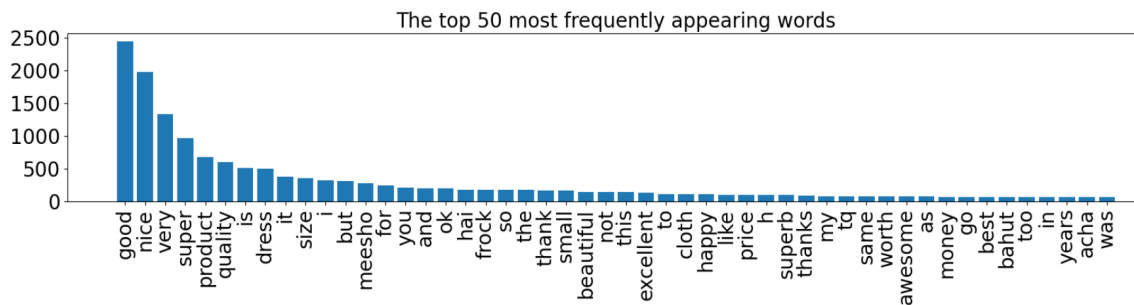


Figure 5.6: Distribution of top 50 frequent words

Figure 5.6 denotes the distribution of the 50 most common frequently appearing words like good, nice, very, super, product and quality. It demonstrates the distribution of review length, with a maximum length of 10 to 20 words. Figure 5(b)describes the maximum review length by characters and words for multiple categories of products.

5.4.5 Visualisation of Combined English and Hinglish Reviews' Words

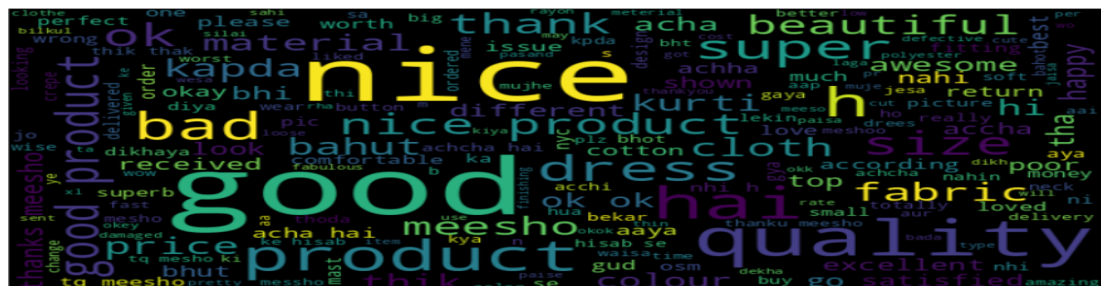


Figure 5.7: Visualization of Hinglish and English words

Figure 5.7 shows the visualisation of collected reviews combined in English and Hinglish reviews using the word cloud. It visualises English words from multiple categories of product reviews using a word cloud.

5.5 Data Filtering of English and Hinglish Reviews

We filtered the reviews in English and Hinglish using the NLTK English words set and stored the English reviews in a CSV file. After filtering, we have found the length of all product categories of English reviews, which is 235892.

5.5.1 Filtering of English Reviews

```
1 eng_sentences_kids_floral_dress
```

```
['best product 55 only 161 thanks meesho',  
'nice frock quality is good',  
'very good quality as it is product thanks mesho',  
'nice dress received so good in fabric colour and quality of dress so excited thanks for meesho',
```

Figure 5.8: Filtering of English Reviews

Figure 5.8 exposes the English reviews after filtering the reviews. We filtered the reviews in English and Hinglish using the english words set in the NLTK library and stored the English reviews in a CSV file. It demonstrates the English reviews after applying the filtering of reviews and shows the visualisation of English words of multiple categories of product reviews using the word cloud.

5.5.2 Filtering of Hinglish Reviews

```
1 non_eng_sentences_kids_floral_dress
```

```
['bahuut hi accha h thank you meesho',  
'bhot acha kapra hai quality bhi bhot acha hai thank you mesho',
```

Figure 5.9: samples of Filtered Hinglish Reviews

Figure 5.9 describes non-English sentences in reviews, such as Hinglish, after filtering the reviews.

5.5.3 Visualization of English Reviews' Words



Figure 5.10: Word Cloud of Multi Products English Reviews' Words

Figure 5.8 shows the visualization of English words of multiple categories' product reviews using the word cloud.

5.6 Hybrid Approach for Summarization

We have used a hybrid approach to extract the sentences from reviews related to product features. First, we created a list of features from various product review categories. For example, for mobile reviews, the key features include camera, RAM, memory, and battery, while for clothing reviews, the key features are design, size, quality, colour, fabric, and so on. We have stored all features in lists according to product categories. We have used Space, Benepar, NLTK, Tree, en_core_web_md, parse dependency tree and svglint to tag the reviews. We created the parse tree of sentences, for ex. The battery is good-(NP (NN battery) and (VP (VBZ is) (ADJP (JJ good))))'. We applied the POS tagging for 'JJ' or tree_pos == 'RB' (Adverb) or tr. label == 'NP'. Then, it formulates the list of phrases with the tagging and the Parse dependency tree.

5.6.1 Feature Extraction Using DistilBERT and Extractive Summary Generation

We have used the Question Answering DistilBERT model, a variant of BERT. In the Question, we input the feature of the product, and in the answer, it extracts only those sentences related to product features, while other sentences are neglected. Then, it formulates the list of feature-related sentences in the reference summary, including product features and product reviews. For E.g., if we provide the input of reviews and the 'camera' word to DistilBERT, it extracts all the sentences related to the camera based on similarity. These feature-related sentences are considered an extractive format reference summary.

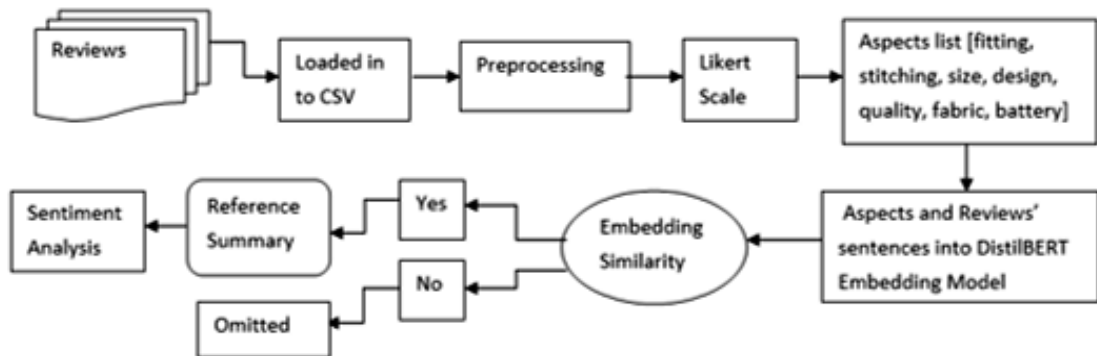


Figure 5.11: Reference Summary Generated using DistilBERT

Figure 5.11 shows the workflow of the feature-based extractive summary generation from the reviews. The extractive summary is generated based on the similarity between

feature embeddings. The sentences do not include any features that are omitted. BERT is used for sentiment analysis of extractive reference summaries.

5.6.2 Sentiment Analysis of Reference Summary using the BERT

Sentiment analysis of extracted sentences related to product features is performed using BERT Sentiment Analysis, and positive and negative sentiment is added in the reference summary. The neutral sentences are omitted in the reference summaries. Synset and WordNet are used to find synonyms and antonyms.

5.6.3 BART and Pegasus for Extractive to Abstractive Summary Generation

We have utilised the BERT, BART, and PEGASUS models, identical to those employed by [116], to generate the abstractive summary. We used the BERT, BART, and Pegasus models to generate the abstractive summaries. In BART, we have specified the maximum length as 150 and the minimum length as 30. Then, we selected the best abstractive summary by comparing the three generated summaries using Rouge scores and saved the best summaries along with their scores, features, products, and sentiment lists. The best generated summary's maximum token length is 128.

5.6.3.1 Extractive to Abstractive Summary Generation

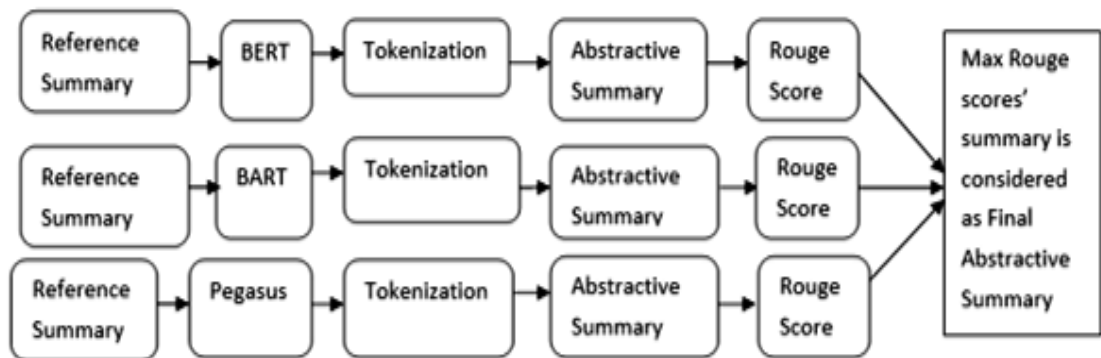
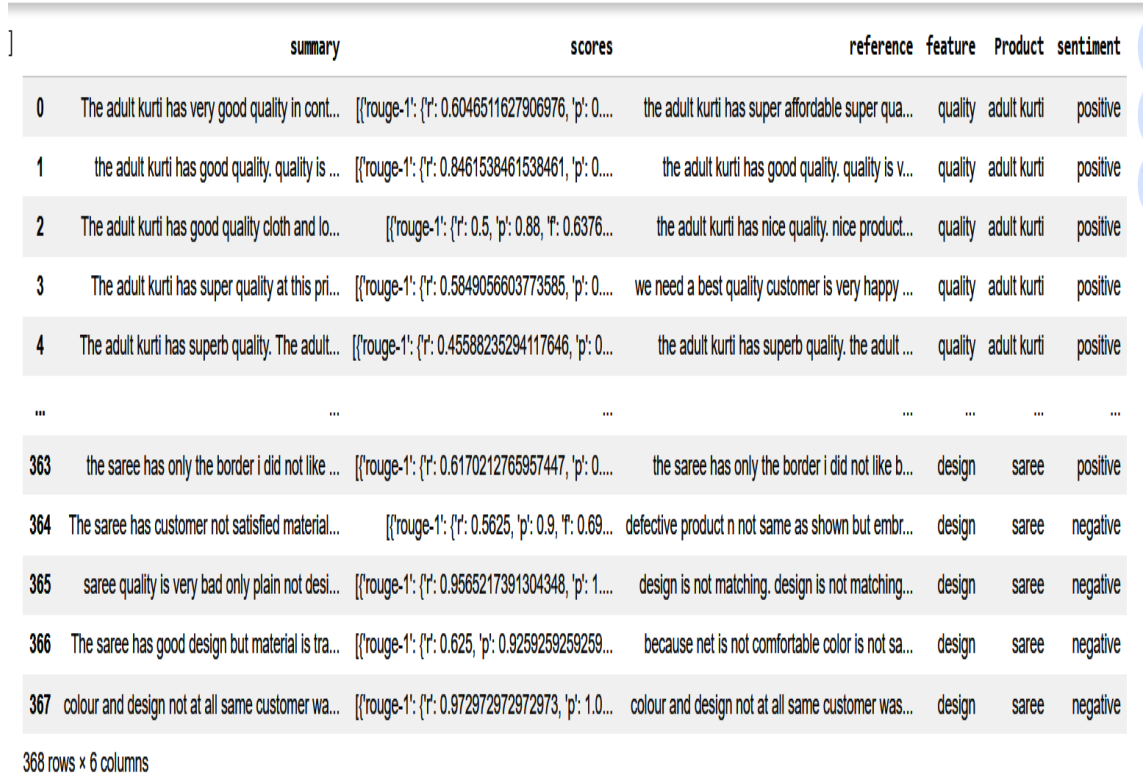


Figure 5.12: Abstractive Summary generated using BERT, BART and Pegasus

Figure 5.12 shows the abstractive summary generation workflow from the reference extractive summary. The max rouge scores' summary is considered as final abstractive summary.

5.7 Synthetic Corpus of Abstractive Summaries With Features and Products



	summary	scores	reference	feature	Product	sentiment
0	The adult kurti has very good quality in cont...	{{'rouge-1': {'r': 0.6046511627906976, 'p': 0....	the adult kurti has super affordable super qua...	quality	adult kurti	positive
1	the adult kurti has good quality. quality is ...	{{'rouge-1': {'r': 0.8461538461538461, 'p': 0....	the adult kurti has good quality. quality is v...	quality	adult kurti	positive
2	The adult kurti has good quality cloth and lo...	{{'rouge-1': {'r': 0.5, 'p': 0.88, 'f': 0.6376...	the adult kurti has nice quality. nice product...	quality	adult kurti	positive
3	The adult kurti has super quality at this pri...	{{'rouge-1': {'r': 0.5849056603773585, 'p': 0....	we need a best quality customer is very happy ...	quality	adult kurti	positive
4	The adult kurti has superb quality. The adult...	{{'rouge-1': {'r': 0.45588235294117646, 'p': 0....	the adult kurti has superb quality. the adult ...	quality	adult kurti	positive
...	...	...	...	...	...	...
363	the saree has only the border i did not like ...	{{'rouge-1': {'r': 0.6170212765957447, 'p': 0....	the saree has only the border i did not like b...	design	saree	positive
364	The saree has customer not satisfied material...	{{'rouge-1': {'r': 0.5625, 'p': 0.9, 'f': 0.69...	defective product n not same as shown but embr...	design	saree	negative
365	saree quality is very bad only plain not desi...	{{'rouge-1': {'r': 0.9565217391304348, 'p': 1....	design is not matching. design is not matching...	design	saree	negative
366	The saree has good design but material is tra...	{{'rouge-1': {'r': 0.625, 'p': 0.9259259259259...	because net is not comfortable color is not sa...	design	saree	negative
367	colour and design not at all same customer wa...	{{'rouge-1': {'r': 0.972972972972973, 'p': 1.0...	colour and design not at all same customer was...	design	saree	negative

368 rows x 6 columns

Figure 5.13: Best summary with rouge score, product, feature and reference

Figure 5.13 shows the best summary generated with rouge scores and reference summaries with feature, product, and sentiment columns. Best summary is selected by comparing the rouge score of 3 generated summaries by BART, BERT and Pegasus. The reference summary is an extractive format which is generated by DistilBERT.

5.7.1 Trained the Data using Fine-tuning T5

We split the data into 90% for training and 10% for testing. Then, we have fine-tuned the T5 model to train the predicted best summary from the reference summaries. The hyperparameters we have given max input length = 1024, min_target_length = 5, and max_target_length = 128, batch size = 8, learning rate = 2e-5, and max_epochs = 10, and model checkpoint = 'T5-small'. We give input of the reference summary with the prefix 'Summary_text' and output of the predicted summary,

5.7.1.1 T5 Architecture and Block Diagram for Trained Summaries

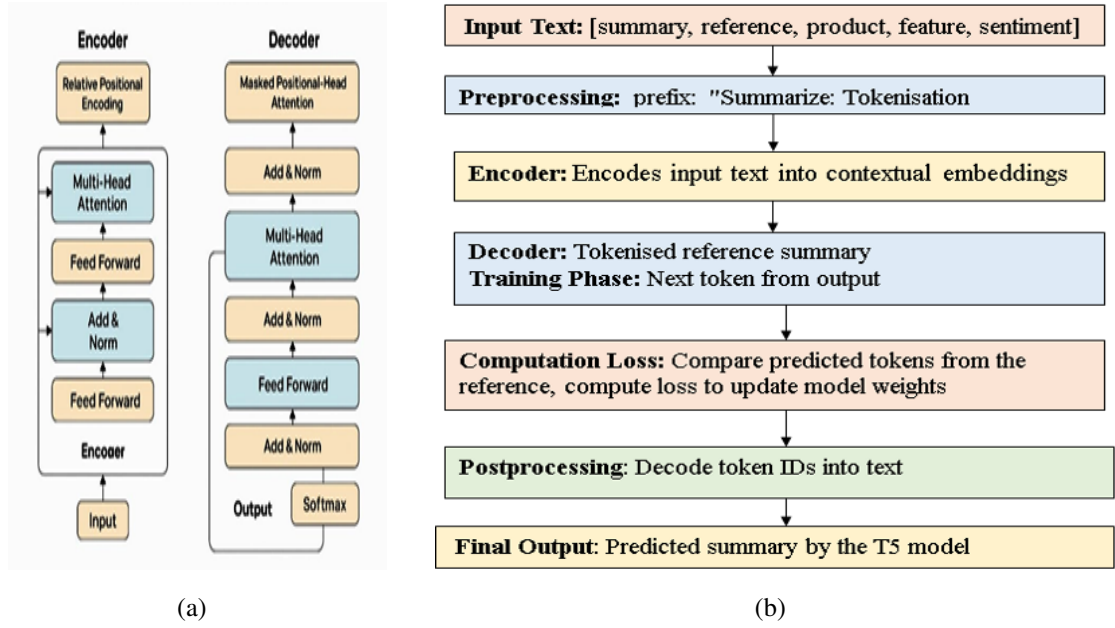


Figure 5.14: (a)T5 Model Architecture (b) Trained Summaries using T5

Figure 5.14 (a) shows the architecture of the T5 model, which contains the Encoder and Decoder. (b) shows the pipeline for training the data. First, input the synthetic corpus into the model and perform preprocessing, such as tokenisation. Then, input the text encoded into embeddings by the encoder and the tokenised reference summary into the decoder. Then, trained the next token from the output and computed the loss to update the weights by comparing the predicted tokens with the reference. The postprocessing step decodes the embedding IDs into text, and the T5 model predicts the final summary.

5.7.1.2 Training Testing Loss Using T5 and Sentiment Distribution of Summaries

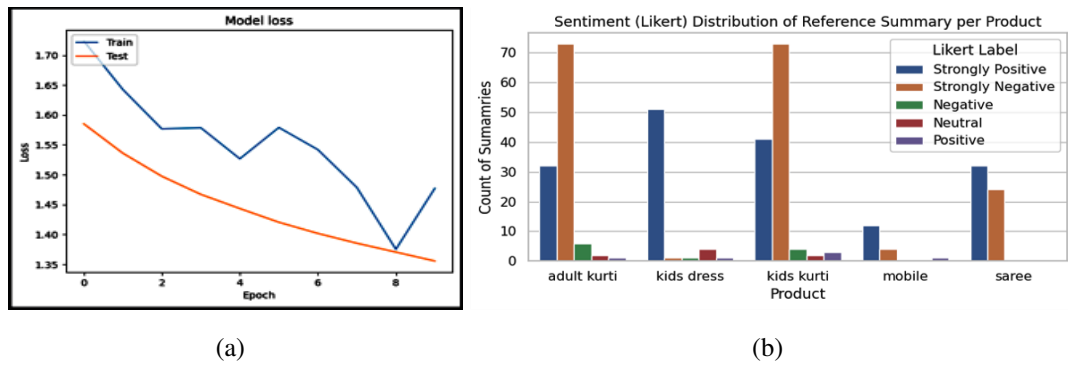


Figure 5.15: (a) Training and Testing loss using epochs (b) Sentiment Distribution of Summaries

Figure 5.15 defines the Train and Testing loss using 10 epochs. The orange colour line represents the Test, and the blue colour represents the Train. The x-axis represents the number of epochs, and the y-axis represents the loss rate of training and testing. (b) shows the sentiment distribution of summaries. The x-axis describes the names of products, and the y-axis defines the count of summaries. Most of the summaries are approx. 80 summaries of adult kurti, kids' kurti are negative, and approximately 30 summaries are positive.



Figure 5.16 visualises the word cloud images of summaries of some individual samples with the summary numbers.

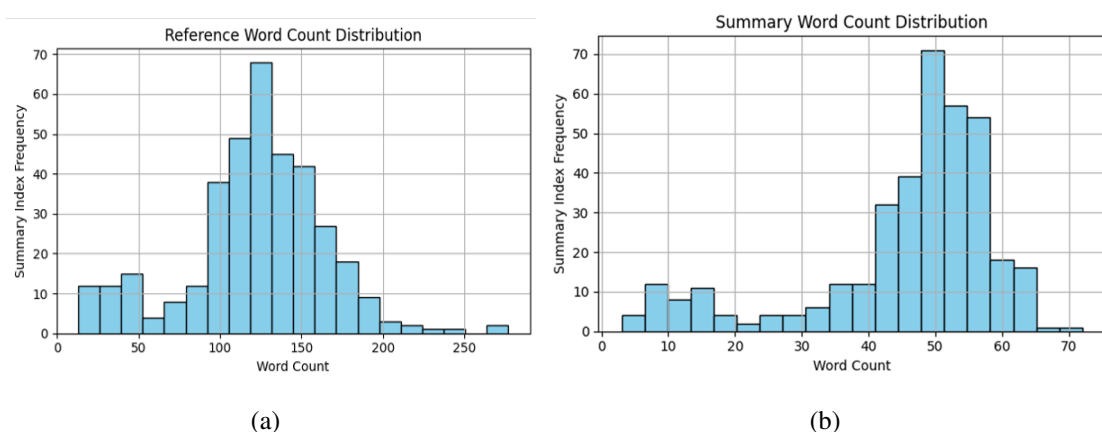


Figure 5.17 (a) shows the distribution of the length of reference summaries. The y-axis represents the number of review summaries, and the x-axis represents the length of words in the reference summary. The maximum length of the reference summaries is 250 words, and (b) presents the word length of summaries generated from the reference summaries. The x-axis presents the length of summaries using word count, and the y-axis represents the frequency of summaries.

5.7.4 Sentiment Distribution of Summaries With Features of Products

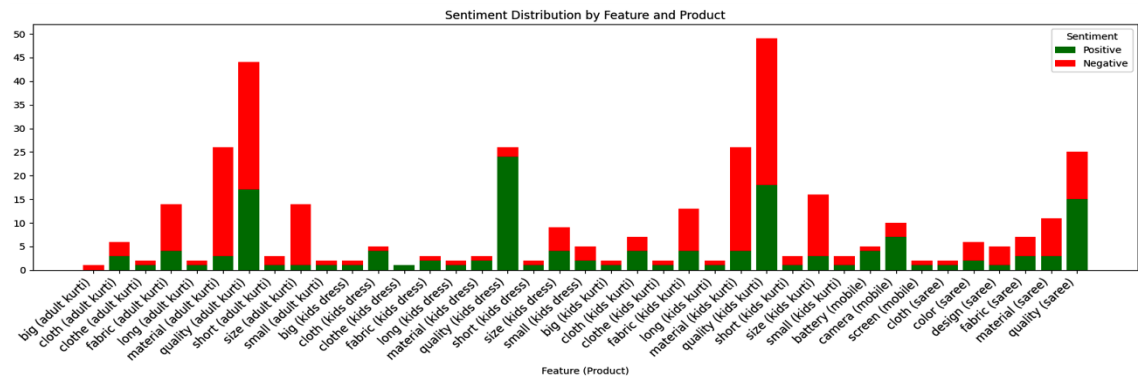


Figure 5.18: Sentiment Distribution of Summaries features with products

Figure 5.18 shows the sentiment distribution of features across several product categories. The green colour indicates that the reviewers are talking positively about those features, and the red colour indicates that they are talking negatively about those features.

5.7.5 Distribution of Features and Sentiment for Per Product

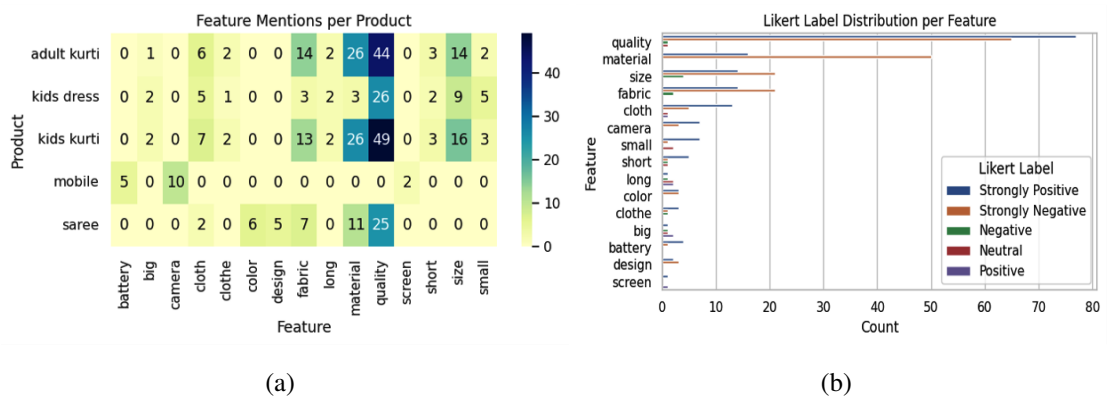


Figure 5.19: (a)Features distribution per product (b) Likert label feature distribution

Figure 5.19 (a) shows the number of features appearing in the summaries grouped by the products. The x-axis describes the feature names, and the y-axis shows the product names. The adult kurti has 26 times more material and 44 times more quality

features, as described by the reviewers. In the clothing domain, users most often discuss material, quality, fabric, and size. In the mobile reviews, most users discuss the battery and camera. Figure 5.19 (b) describes the sentiment analysis of features. The x-axis defines the count of summaries, and the y-axis shows the feature names. For the quality feature, most users are talking positively, but for the material feature, most users are talking strongly negatively.

5.8 Summary Evaluation using Rouge Scores

The Rouge metric scores, Rouge1, Rouge2, and RougeL scores are used to predict the summary's quality. The below visualisation shows the summary's accuracy of the predicted summary.

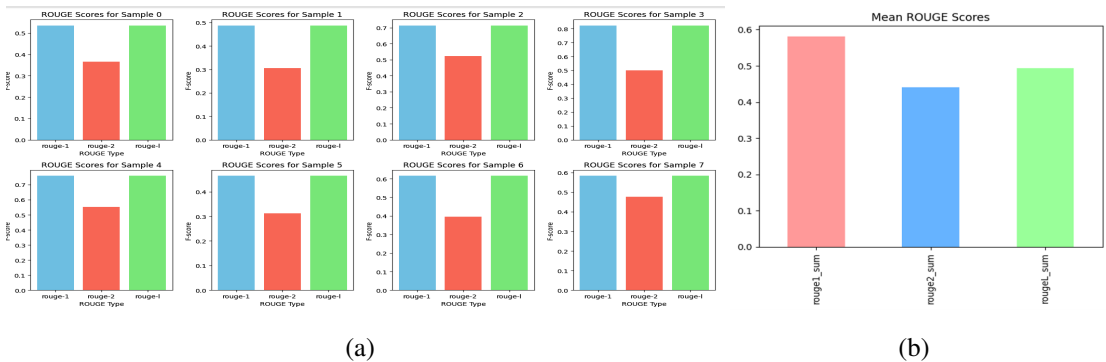


Figure 5.20: (a)Rouge Score of Individual Summaries (b) Mean Rouge Scores

Figure 5.20 (a) shows the R1, R2, and RL rouge scores of some individual samples. The x-axis defines the Rouge1,2,L scores, the y-axis describes the values of scores, and (b) presents the Mean scores of Rouge scores of all samples. The x-axis presents the Rouge1,2,L scores and the y-axis presents the score values. The Rouge1-51.9, Rouge2-49.2, RougeL-50.7 scores have been achieved.

5.9 Examples of Predicted Summaries

Table 5.2: Examples of Predicted Summaries

Summary_text	Predicted Summary
the adult kurti has very bad quality bad fabric. fabric material is very poor. fabric quality and size issue. cloth not good it is fabric am not satisfied do not repeat again meesho. fabric is not so good.	Product is nice but size is too small. Size and quality is not good. Neck size too small i have fitting issue. Fabric quality and size issue

Continued on next page

Table 5.2 – continued from previous page

the saree has only one the border I did not like but fabric is superb design is cute, quality is very good same color and design thanks, the saree has best quality same design, the saree has beautiful very good design. it is very good and design is also good. the saree has nice design and net also soft. nice working design customer like. the saree has good design. nice material and design. the saree has fabric quality is very nice design is also same. design not same as shown but border is totally different. the saree has good quality and good design. same color with design very nice. design is not matching quality is good. nice but design colour some dark. Nice colour and design weight less easy to wear. nice saree design. saree is good but design is not same.	The saree has only the border I did not like but fabric is superb design is cute. Saree is good but design is not same.
the kids dress has nice fabric. the kids dress has nice fabric quality.	The kids dress has nice fabric quality.

Table 5.2 represents the predicted summary from the reference summary. The `summary_text` column is the reference summary and predicted summary is generated by finetune the T5 model.

5.10 Conclusion

In this chapter, we have proposed a hybrid approach of rule-based and transfer learning methods to generate an aspect-based abstractive summary of product reviews. However, admirable work has already been done in summarisation. Still, aspect-based summarisation is important for research to focus on more. We have not applied the stop word removal, stemming and lemmatisation preprocessing techniques because these techniques can affect the summary's quality and can change the meaning of sentences. We have performed data cleaning, case conversion, filtering, and tokenisation preprocessing techniques on reviews. Then, 3 models, BERT, BART and Pegasus, are used to generate abstractive summaries. Then, all the summaries' evaluation is done using rouge scores, and the highest score of the predicted summary is considered the best summary. After this, we have fine-tuned the T5 model to train the predicted and reference summaries and achieved better accuracy than the baseline. In the next chapter, the results of predicted summaries are visualised using other intrinsic evaluation parameter scores.

Chapter 6

Results and Analysis

In this chapter, we present the results of each important step in the proposed pipeline: data collection, data preprocessing, feature extraction, topic modelling, and aspect-based sentiment-oriented summarisation. The experiments are conducted on 40k english language reviews, which we have scraped from e-commerce websites, Amazon, Flipkart, and Meesho.

Evaluating an automatic summary is a challenging and time-consuming task due to its subjective nature. Due to the lack of ground-truth summaries in aspect-based summaries across different domains, the annotation of ground-truth aspect-based summaries requires human resources and domain expertise. The standard Rouge and BLUE metrics are not always suitable for evaluating aspect-specific summaries. Bhandari et al. [150] concluded that the Rouge-2 metric can outperform all other metrics and is reliable for abstractive summaries, and Rouge-1 is reliable for extractive summaries. They also suggested the BERTscore metric for evaluating abstractive summaries. According to Lloret et al. [151], the content evaluation of summaries is assessed based on their relevance to the original documents. In this chapter, we propose evaluation parameters, including Rouge, BERT, BLEU, and intrinsic parameters such as coherence, information, relevance, fluency, and keyword coverage, for evaluating summaries. We have attempted to address the limitations of existing work by evaluating summaries using both intrinsic and human scores.

6.1 Contribution of this Chapter

Table 6.1: Contribution of Each Experimentation Results

Terms	Our Contribution
Real-world data analysis, preprocessing and feature extraction results	Found the noisy data and challenges to apply preprocessing and feature extraction techniques on the noisy and abbreviated reviews.
Topic Modelling results on Hinglish reviews	The extracted topics are evaluated using the coherence score and the diversity score, and the unlabelled topic models' results are compared with the labelled topic models.

Continued on next page

Table 6.1 – continued from previous page

Terms	Our Contribution
Aspect-based multi-product review summary results	The proposed work has generated a hybrid extractive and abstractive summaries, and evaluation is conducted using ROUGE, BLEU, BERT and Human scores.
Comparative analysis of proposed work with existing work	The proposed work has achieved good results as compared to existing models.

6.2 Preprocessing, and Feature Extraction Results

We found inconsistent formats, spelling mistakes, slang, emojis, abbreviated words, Hinglish reviews, and inconsistent capitalisation, as well as challenges in applying preprocessing to noisy data. We have experimented with several preprocessing techniques and found that stemming and stop word removal are not suitable choices for summarisation purposes.

6.2.1 Preprocessing Results

	reviews_text	names	dates	help_f	Label	Cleaned reviews	Review_Length	Length_after_removal_stopwords	Review_after_removing_stopwords	Tokens	Stemmed	Lemmaized
0	Bahut hi achha laga. I m so happy. Mere liye m...	Nandita Suklabaidya	Posted on 30 Dec 2019	Helpful (1937)	saree	bahut hi achha laga i m so happy mere liye man...	164	24	bahut hi achha laga happy mere liye mangwaya f...	[bahut, hi, achha, laga, i, m, so, happy, mere...]	[bahut, hi, achha, laga, i, m, so, happy, mere...]	[bahut, hi, achha, laga, i, m, so, happy, mere...]
1	Bahut Achcha hai amazing????????????	Meesho User	Posted on 5 Feb 2024	Helpful (2)	saree	bahut achcha hai amazing	36	4	bahut achcha hai amazing	[bahut, achcha, hai, amazing]	[bahut, achcha, hai, amaz]	[bahut, achcha, hai, amazing]
2	Quilting so good thanks so much mesho ?	Deepshikha	Posted on 2 Feb 2024	Helpful (4)	saree	quilting so good thanks so much mesho	39	5	quilting good thanks much mesho	[quilting, so, good, thanks, so, much, mesho]	[quilt, so, good, thank, so, much, mesho]	[quilt, so, good, thank, so, much, mesho]
3	Very Beautiful saree@toY@Very fast delivery...	Deepika Chavan	Posted on 7 Feb 2024	Helpful (0)	saree	very beautiful saree very fast delivery and pro...	100	7	beautiful saree very fast delivery product show...	[very, beautiful, saree, very, fast, delivery, a...]	[veri, beauti, saree, veri, fast, deliveri, and...]	[very, beautiful, saree, very, fast, delivery, a...]
4	It's worthy for money ??? you can go for it	Roselind.A	Posted on 25 Jan 2024	Helpful (9)	saree	it is worthy for money you can	44	3	worthy money go	[it, is, worthy, for, money, you, can, ...]	[it, is, worthi, for, money, you, can, ...]	[it, be, worthy, for, money, you, can, go, ...]

Figure 6.1: Results of Preprocessing techniques

Figure 6.1 shows the results of preprocessing techniques applied to the cleaned reviews after removing noise from the reviews. In real-life reviews, a lot of noise is present. It is so challenging to clean the noisy reviews. We have experimented with stemming, lemmatisation and stop words removal steps. However, after applying stemming, lemmatisation, and stop word removal, we found that these techniques can change the meaning of the sentence, as real-life reviews are already very brief. So, these techniques should not be applied for summarisation purposes.

6.2.2 Feature Extraction Results

We have utilised Bag-of-Words, TF-IDF, Word2vec, FastText, and BERT embeddings to extract features from reviews. Among these, BERT has yielded the best results, as it can generate embeddings for out-of-vocabulary and abbreviated words.

Reviews	cleaned	tokenized_sentences	bert_embedding
Suprb drs 🍌, Nys 🍌 Looking Vry SmarT 🍌, Gud Cloth 🍌, Received the product before expected date, Same as Displayed	suprb drs nys looking vry smart gud cloth received the product before expected date same as displayed	[suprb drs nys looking vry smart gud cloth received the product before expected date same as displayed]	[[[-0.10711852, -0.07393058, 1.357999, 0.6187935, 0.76441664, -0.4881686, -0.35069734, 0.7560985, 0.5974753, -0.42691132, 0.24055356, 0.4020504, 0.73893493, 0.2524736, -0.76901674, 0.48139283, -0.33484414, -0.4427012, 0.13072687, -1.140648, -0.6853546, -0.87566584, -0.2779493, 0.44552508, 0.6096051, -0.23711033, -0.16406435, -0.68894684, -0.2596769, 0.48076904, -0.097373225, -0.030604176, 0.16441347, -0.3568934, 0.13474369, 0.89558554, -0.0011144304, 0.7820595, -0.39665696, 0.21362285, -0.33256122, -0.020182805, 0.20361507, -0.33120766, -1.1871761, -0.33859497, 0.074383885, 0.3660675, 0.5447792, -0.90389144, 0.046757463, -0.21034488, 0.12722585, 0.44587964, -0.78243864, 0.3554299, 0.31436306, -0.80618685, -0.5953213, 0.38326237, -0.4146543, -0.2656914, -0.024398465, 0.24333423, 0.025877226, 0.3237921, 0.13919233, 0.32176194, -0.8484762, -0.44548047, -0.0029881764, -0.6437996, 0.0028526115, -0.29767066, -0.40597218, 0.32030442, -0.3402898, 0.08467974, 0.6291943, 0.48092636, -0.32676244, -0.41337562, 0.54084826, 0.26927707, 0.70651925, 0.6160942, 0.42682534, -1.0068495, -0.93833375, 0.8730394, 0.49520552, 0.42387953, 0.6299842, -0.50136286, -0.99548703, -0.3395359, -0.66722673, -0.35326356, 0.041096363, 0.29883683, ...]]
Perfect fit,material ,color. Unlike other party wear stuff, the material is cotton(hoisery material). The kids wear it comfortably.Mu	perfect fitmaterial color unlike other party wear stuff the material is cottonhoisery material the kids wear it comfortablymu kids le	[perfect fitmaterial color unlike other party wear stuff the material is cottonhoisery material the kids wear	[[[-0.55854344, 0.22788638, 0.12565504, 0.5540845, -0.30038407, -1.370391, 0.24660467, 0.14473861, 0.7473624, -0.46688157, -0.11265278, 0.63580006, -0.100148074, 0.030369733, -1.0521618, 0.7937555, 0.9466648, -0.47497523, 0.27058265, 0.23575364, -0.6640606, -1.1135948, -0.49410653, 0.039051324, 0.36367077, 0.52961236, -0.31396466, 0.48215798, -0.5087271, 0.5302194, -0.017673703, 1.3612194, 0.40689161, 0.7314069, 0.307518, -1.1678581, 0.8183705, ...

Figure 6.2: Visualisation of Embeddings using BERT

Figure 6.2 shows that the BERT Model can generate the embeddings of abbreviated and out-of-vocabulary words like suprb, drs, vry and nys. However, word2vec, flash text are unable to generate the embeddings of abbreviated and out-of-vocabulary words.

6.3 Topic Modelling Results on Hinglish Reviews

The Hinglish reviews are phonetically written by users, which leads to inconsistent spellings and variations. The hinglish stopwords such as accha, badiya, bariya, etc., increase the vocabulary sparsity and the number of topics generated by LDA and NMF. Hinglish reviews include the mixture of Hindi and English stop-words in Roman script. Removing stop-words is also very important in Hinglish code-mixed reviews.

6.3.1 Limitations of Existing Work on Topic Modelling Results

- **Lack of pre-trained models for Hinglish reviews:** There is a lack of NLP tools, such as NLTK, Spacy, and Gensim, to preprocess the Hinglish reviews, which impacts the lower quality of preprocessing, word representation, and topic assignment.
- **Short Reviews and Data Sparsity:** Most of the time, reviews are often very short, E.g., “acchi saree hai”. The topic modelling algorithm LDA can perform poorly on short reviews by generating repetitive topics.

6.3.2 Impact of Including and Excluding Hinglish Stop words

	Hinglish Reviews	Labels	word_count	char_count	tokens	filtered_tokens	filtered_length	filtered_without_hin_eng_stopwords	filtered_len_without_hin_eng
0	Is suit ka color aur cut bahut hi attractive hai.	ladies suit	10	49	[is, suit, ka, color, aur, cut, bahut, hi, att...]	[suit, color, cut, attractive]	4	[suit, color, cut, attractive]	4
1	Ye suit bahut hi comfortable hai aur me isse k...	ladies suit	17	86	[ye, suit, bahut, hi, comfortable, hai, aur, m...]	[suit, comfortable, kaam, parties, pehenti]	5	[suit, comfortable, kaam, parties, pehenti]	5
2	Ye suit bahut hi durable hai aur mujhe investm...	ladies suit	13	69	[ye, suit, bahut, hi, durable, hai, aur, mujhe...]	[suit, durable, investment, worth, lag]	5	[suit, durable, investment, worth, lag]	5
3	Suit ke button aur print bahut hi unique hai a...	ladies suit	17	83	[suit, ke, button, aur, print, bahut, hi, uniq...]	[suit, button, print, unique, stylish, lag]	6	[suit, button, print, unique, stylish, lag]	6
4	Is suit ka design bahut hi versatile hai, I ca...	ladies suit	23	127	[is, suit, ka, design, bahut, hi, versatile, h...]	[suit, design, versatile, i, can, wear, it, fo...]	14	[suit, design, versatile, wear, occasions, sui...]	9
5	Is suit ka quality bahut hi acha hai, and pric...	ladies suit	13	67	[is, suit, ka, quality, bahut, hi, acha, hai...]	[suit, quality, and, price, reasonable]	5	[suit, quality, price, reasonable]	4
6	This suit is my go-to option for any formal ev...	ladies suit	10	51	[this, suit, is, my, go-to, option, for, any, ...]	[this, suit, my, go, to, option, for, any, for...]	10	[suit, option, formal, events]	4
7	Ye suit mere husband ke liye perfect tha, wo i...	ladies suit	14	68	[ye, suit, mere, husband, ke, liye, perfect, t...]	[suit, husband, perfect, tha, wo, love]	6	[suit, husband, perfect, love]	4
8	Suit ke pocket me hole tha, mujhe bahut hi dis...	ladies suit	22	119	[suit, ke, pocket, me, hole, tha, mujhe, bahu...]	[suit, pocket, hole, tha, disappointment, suit...]	11	[suit, pocket, hole, disappointment, suit, fig...]	10
9	Suit ke pant bahut lambey hai, aur alteration...	ladies suit	11	60	[suit, ke, pant, bahut, lambey, hai, aur, alt...]	[suit, pant, lambey, alteration, sakta]	5	[suit, pant, lambey, alteration, sakta]	5
10	Suit ka quality bahut low hai, aur iske price...	ladies suit	13	64	[suit, ka, quality, bahut, low, hai, aur, isk...]	[suit, quality, low, price, se, worth]	6	[suit, quality, low, price, se, worth]	6
11	Suit ki color pehle hi wash ke baad hi fade ho...	ladies suit	13	56	[suit, ki, color, pehle, hi, wash, ke, baad, h...]	[suit, color, wash, fade, thi]	5	[suit, color, wash, fade, thi]	5

Figure 6.3: Reviews' Length with Including and Excluding Hinglish and English Stop words

Figure 6.3 shows the results of word length of reviews, including and excluding hinglish and english combined stop words. We have observed that, in most reviews, users often include stop words. For e.g., 4th Review length is 23 words and character length is 127 characters. After removing hinglish stop words, the filtered token length is 14 words, but after removing hinglish and English stop words in the column named 'filtered_without_hin_eng_stopwords' token length is only 9 words. Therefore, it is concluded that from code-mixed reviews, we must remove both Hinglish stop words and English stop words. Because in code-mixing, people write English and Hinglish mixed words in the reviews.

6.3.3 Distribution of Coherence Score for Labelled Vs Unlabelled Topic Models

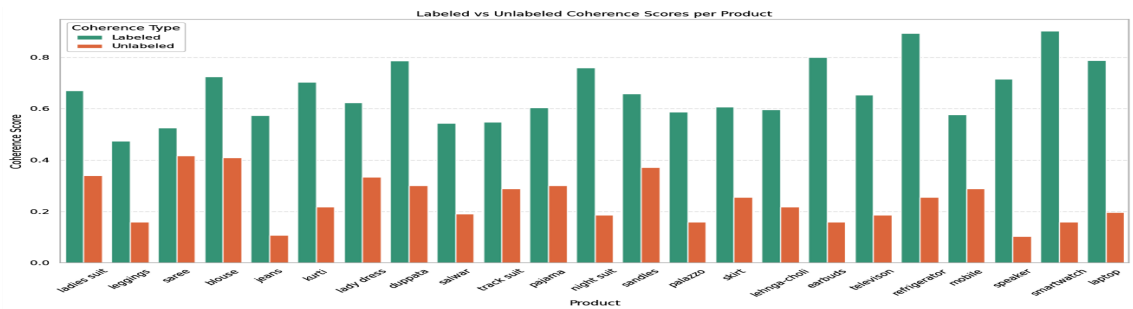


Figure 6.4: Coherence Score of Labelled vs Unlabelled LDA, NMF and BERTopic for all Products

Figure 6.4 defines the coherence score of labelled vs unlabelled topic models and shows that the labelled topic models have achieved greater coherence scores as com-

pared to unlabelled topic models. The x-axis presents the product names, and the y-axis shows the score values.

6.3.4 Extracted Topics Similarity Score and Coherence Mean Scores

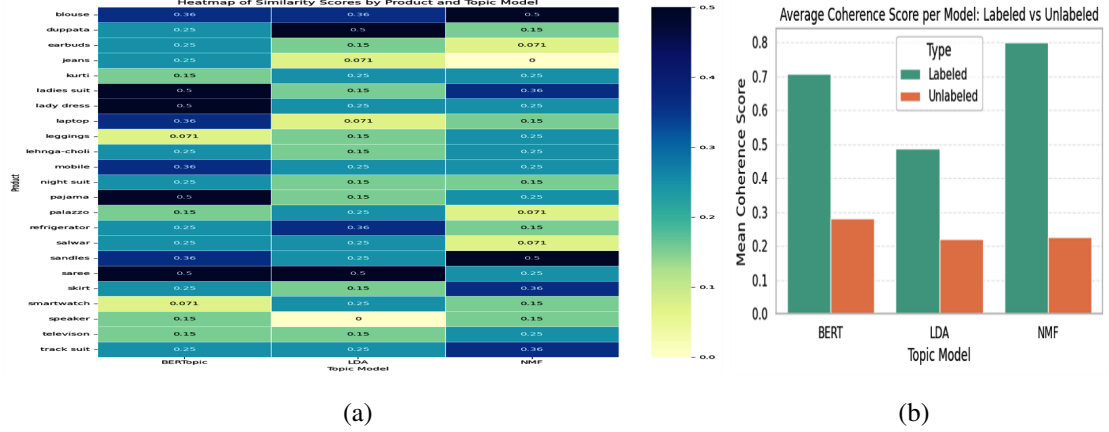


Figure 6.5: (a) Heat map of LDA NMF and BERT score (b) Mean Coherence Score of Labelled and Unlabelled Topic Models

Figure 6.5 (a) presents the heat map of the Labelled BERTopic, LDA and NMF models' scores grouped by the products. The dark blue colour represents the highest similarity. The light green colour shows the lower similarity between scores, and (b) shows the average coherence score of Labelled LDA, NMF and BERTopic Models.

6.3.5 Comparison of Coherence Score of Topics Using Labelled and Unlabeled LDA, NMF and BERT

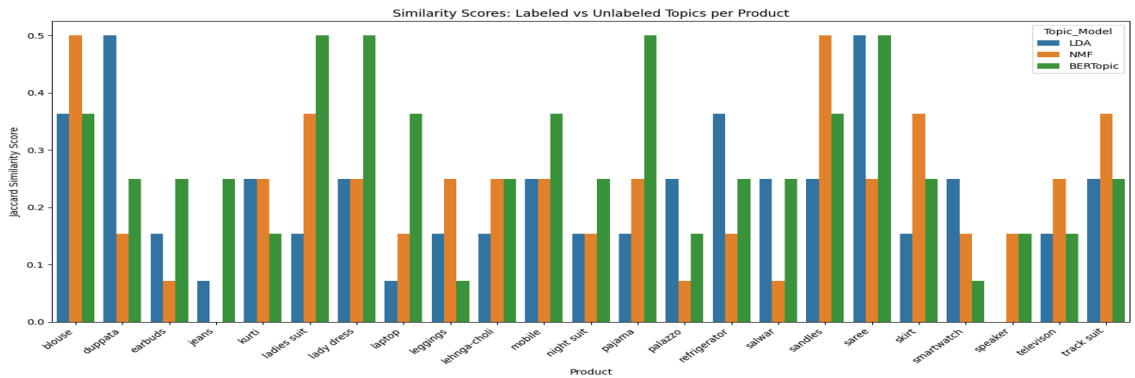


Figure 6.6: Jaccard Similarity Score of Product-aware Topic Models

Figure 6.6 The x-axis defines the model's name and the y-axis defines the Jaccard similarity scores. It describes the heat map of similarity scores of extracted topics using

LDA, NMF, and BERT. It demonstrates the similarity between the extracted topic scores obtained using topic models.

6.3.6 Jaccard Similarity Score and Mean Diversity Vs Coherence Scores

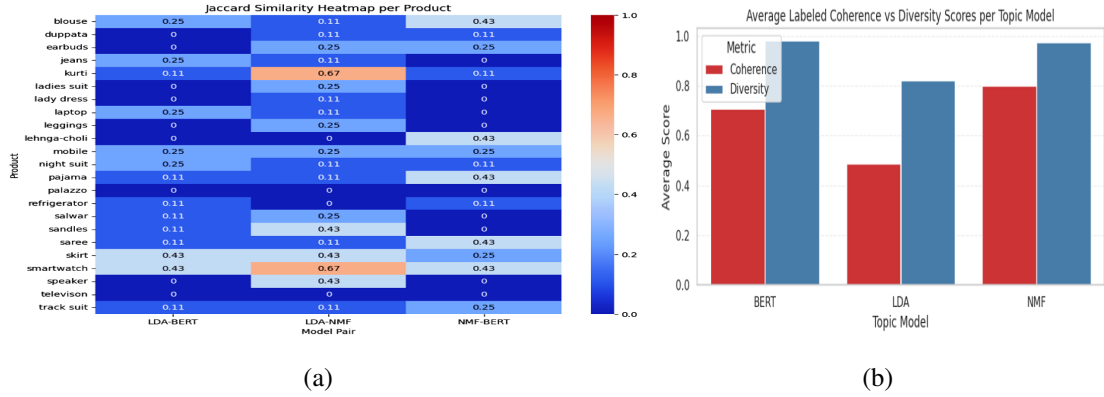


Figure 6.7: (a)Jaccard Similarity Score for Individual Products(b) Diversity and Coherence Mean Score of LDA, NMF, BERT Models

Figure 6.7 (a) presents the heat map of Jaccard similarity scores between extracted topics through labelled LDA, NMF and BERT using Jaccard similarity scores. The x-axis shows the model's name, and the y-axis shows the product's name. It describes the similarity scores of LDA-BERT and LDA-NMF, as well as NMF-BERT pairs. The maximum similarity score of LDA-NMF is 0.67. The dark blue colour describes no similarity, and the red colour defines the maximum similarity. (b) shows the mean of diversity vs coherence scores of using labelled topic models.

6.3.7 Comparison of Topics Coherence Score Using Labelled and Unlabelled Topic Models

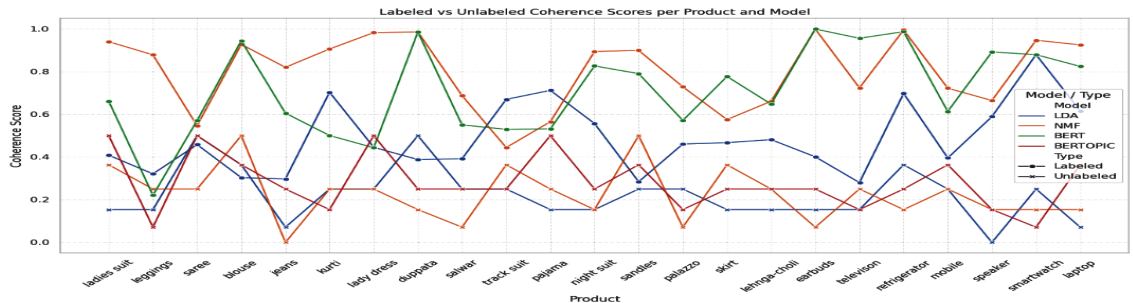


Figure 6.8: Coherence Scores of Labelled and Unlabelled LDA, NMF, BERT based on the Product Topics

Figure 6.8 defines the coherence score of extracted topics based on the products. The x-axis presents the product names, and the y-axis shows the score values. The

labelled product-aware topic models have achieved the maximum score as compared to the unlabeled topic models.

6.4 Aspect-based Multi-product Reviews Summary Results

6.4.1 Distribution of Rouge, BERT and BLEU Scores of Predicted and Human Summary

The results of the summaries have been evaluated using the Rouge, BLEU, BERT scores, and human scores and it has been found that the BERT score can better assess the quality of the summaries because the Rouge score cannot evaluate summaries based on semantics; instead, it checks summaries only through 1-gram, 2-gram, and n-gram approaches. However, the BERT score can evaluate summaries based on their semantics.

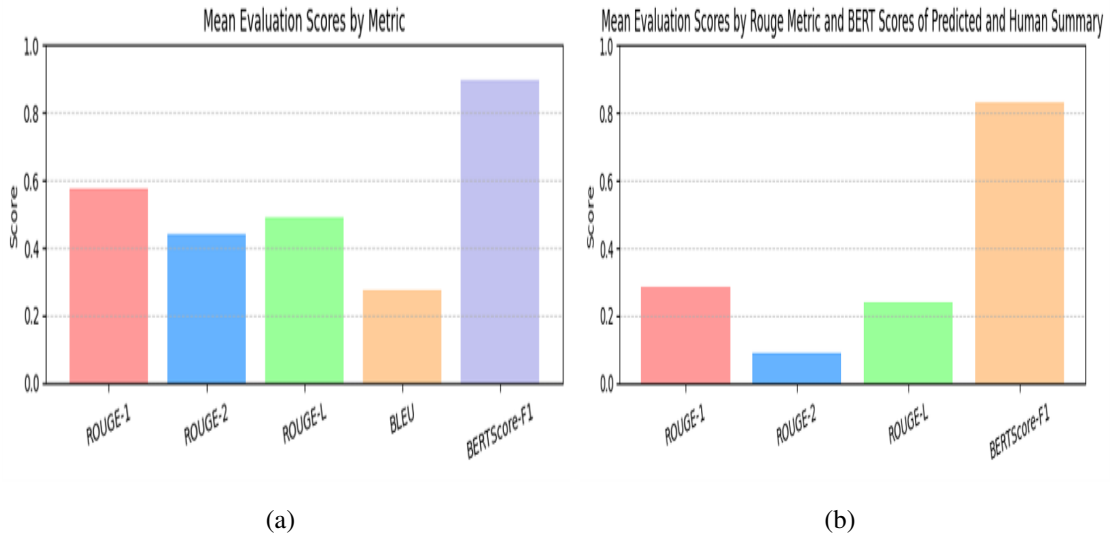


Figure 6.9: (a) Distribution of Rouge, BLEU and BERT Score (b) Distribution of Rouge and BERT Score for Human Summary

Figure 6.9 (a) shows the distribution of Rouge, BLEU and BERT scores of predicted vs reference summary generated by models. BERT score has achieved max scores as compared to Rouge, BLEU metric, because BERT works on the semantic embeddings, although the rouge score works on one-gram, n-gram and n-grams. (b) shows the distribution of Rouge1,2,L and BERT scores for the predicted summary versus the human summary. BERT has achieved maximum scores compared to the ROUGE score, but when compared to reference vs. predicted, it yielded lower scores.

6.4.2 Visualisation of Predicted Summaries With Human Summary

[12] 1 df.head(12)

	summary	reference	feature	Product	sentiment	Human Summary
0	the adult kurti has not good material quality...	material quality not good. material quality no...	material	adult kurti	negative	material quality is very poor and have issues ...
1	the kids kurti has not good size is not same ...	the kids kurti has not good size is not same a...	size	kids kurti	negative	colour and design are not same as shown in the...
2	the kids kurti has material quality not good ...	material quality not good received a defective...	fabric	kids kurti	negative	material quality is very poor, received defect...
3	The adult kurti has a very bad quality and it...	clothe quality is not good. material quality w...	quality	adult kurti	negative	fabric material quality is very poor and have ...
4	material is good but colour is black. materia...	design and shape is better but cloth material ...	material	kids kurti	negative	printing quality is poor, fabric is very slim ...
5	This is too short. border is very short and f...	very short. short. too short. this is too shor...	short	adult kurti	negative	fabric is very poor and had issues with size a...
6	the kids kurti has fabric is not good materia...	material quality not good. material quality no...	material	kids kurti	negative	material quality is very poor, colour and desi...
7	A defective product design is not matching si...	material quality not good. material quality no...	material	kids kurti	negative	material quality is very poor, have issues wit...
8	The kids kurti has nice clothe but size is no...	super but clothe quality is low. the kids kurt...	clothe	kids kurti	negative	product quality is very poor.
9	the adult kurti has loose material quality no...	material quality not good. material quality no...	material	adult kurti	negative	material quality is very poor and have issues ...
10	The kids dress has it is an awesome quality a...	the kids dress has good quality. the kids dres...	quality	kids dress	positive	quality of the product is very good with reson...

Figure 6.10: Predicted Summary With Reference, Product and Features

Figure 6.10 represents the predicted abstractive summaries by finetuning the T5 model, with reference summary and features, product, and human summaries. We have analysed that the predicted summary also covers the challenge of multi-domain products by adding the name of the product at the start of the summary, which makes it easier to understand the summary of that particular product.

6.4.3 Similarity of Rouge Score with BERT Score

A model might optimise for ROUGE (e.g., repeating phrases), yet produce summaries that are less semantically meaningful. Conversely, a model focusing on semantics (such as abstractive models) may achieve a lower ROUGE score but a higher BERT score. Relying on just ROUGE can misrepresent summary quality. So, both metrics should be considered to capture both lexical overlap and semantic accuracy. ROUGE is surface-based, rewarding exact word matches. BERT score is semantic, rewarding paraphrases and contextual matches. It highlights the misalignment between surface and semantic metrics. Bert Score is more meaning-aware, while ROUGE-2 is more token-aware.

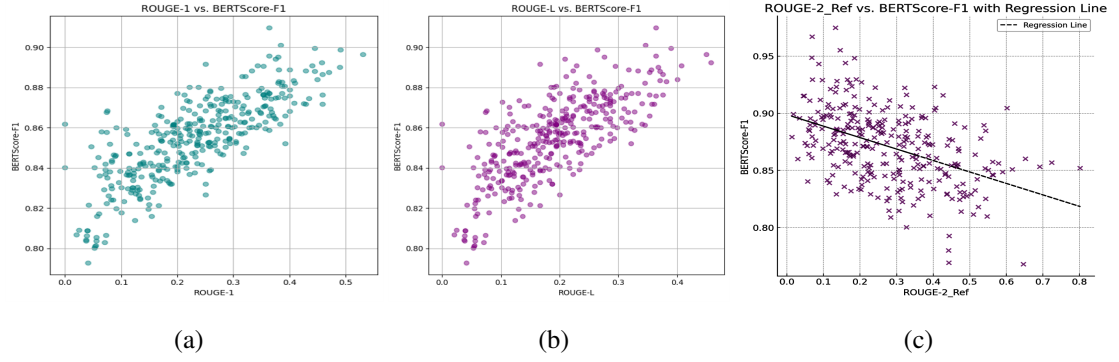


Figure 6.11: (a) Similarity of Rouge1 and BERT score (b) Similarity of RougeL and BERT score (c) Similarity of Rouge2 and BERT score

Figure 6.11 (a) presents the similarity between Rouge 1 and BERT-F1 score. The X-axis defines the ROUGE-1, and the y-axis BERT score. As ROUGE-1 increases, BERTscore-F1 also increases, which means more lexical overlap and more semantic closeness to reference summaries, and (b) shows the Rouge-L score is increasing, and the BERT F1 score is also increasing. There is a positive correlation between the Rouge score and the BERT score. Figure 6.9 (c) The X-axis shows the ROUGE-2 score, and the Y-axis presents the BERTScore-F1. It shows the weak relationship between Rouge2 and BERT. A regression line with a dashed black line indicates the overall trend. High Bert Score but low ROUGE-2, meaning the summary is semantically similar but uses different wording. This measures semantic similarity using contextual embeddings (like BERT). Higher values mean the summary is semantically closer to the reference, even if exact words differ.

6.4.4 Visualisation of all Metrics Scores With Summaries

We have also checked the summaries' quality through informative, conciseness, bi-gram diversity, coherence, fluency, relevance, and keyword coverage metrics scores.

	summary	coherence	informativeness	relevance	conciseness	coverage	fluency_score	bigram_diversity
1	the adult kurti has good quality. quality is ...	0.201119	0.857143	0.857143	0.362069	0.909091	0.738095	0.950625
7	the adult kurti has super quality. the adult ...	0.420761	0.923077	0.923077	0.373984	1.000000	0.652174	0.957037
8	The adult kurti has clour good quality not go...	0.533313	0.666667	0.666667	0.392308	0.600000	0.803922	0.964000
9	the adult kurti has super quality super. nice...	0.007815	0.933333	0.933333	0.344262	0.923077	0.571429	0.968471
10	the adult kurti has so good quality. quality ...	0.178006	0.823529	0.823529	0.361111	0.692308	0.794872	0.958449
...	...	...	...	...	...	...	...	...
362	the saree has material quality not good color...	0.268805	0.812500	0.812500	0.375000	0.700000	0.866667	0.928571
363	the saree has only the border i did not like ...	0.417347	0.625000	0.625000	0.394366	0.461538	0.821429	0.976529
365	saree quality is very bad only plain not desi...	0.368284	0.958333	0.958333	0.434211	0.866667	0.878788	0.966797
366	The saree has good design but material is tra...	0.273111	0.658537	0.658537	0.456140	0.500000	0.884615	0.974241
367	colour and design not at all same customer wa...	0.218860	0.973684	0.973684	0.379310	0.892857	0.854545	0.980796

Figure 6.12: Visualisation of all Metrics Scores

Figure 6.12 explains the summaries examples with coherence, informativeness, relevance, conciseness, coverage, fluency score and bigram diversity scores. The figure describes the coherence score as less as compared to other metric scores, and Informative and bigram_diversity and keyword coverage scores are high.

6.4.5 Distribution of All Metrics Scores of Summaries

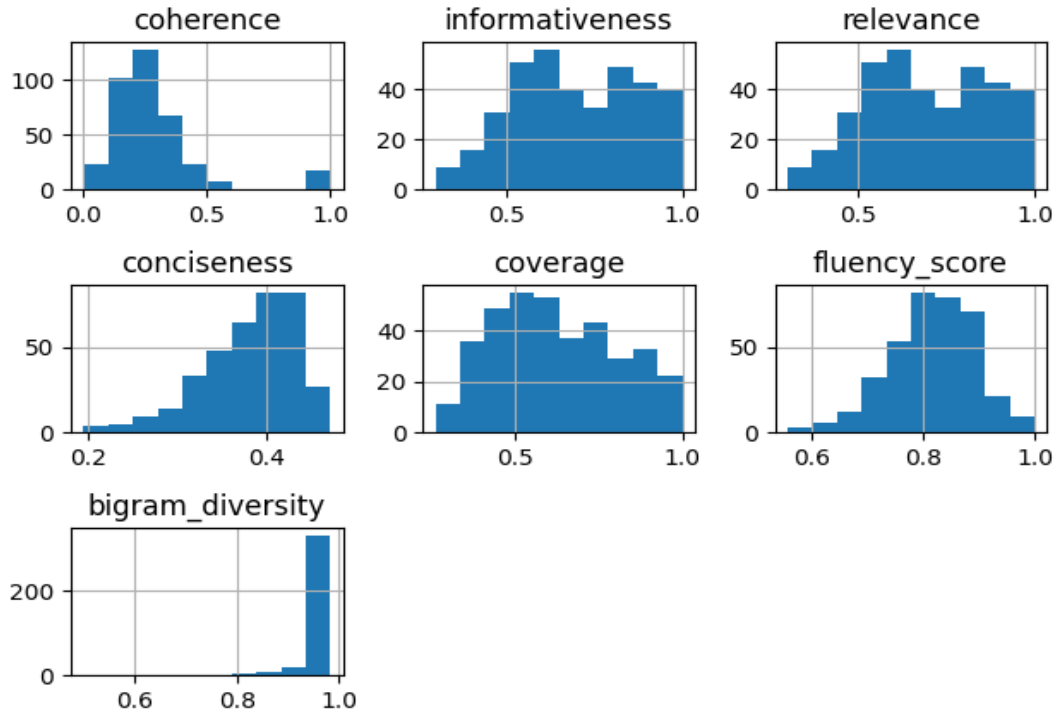


Figure 6.13: Distribution of all metrics scores

Figure 6.13 presents the histograms of all metrics. The x-axis represents the scores of all metrics, including coherence, informativeness, relevance, conciseness, coverage, fluency score, and bigram diversity scores. The y-axis presents the frequency of summaries. The diagrams show how many numbers of summaries are relied on between 0 to 1 metric scores. The first diagram shows that the coherence scores range between 0.0 and 0.5 for 100 summaries, and a very few summaries score is equal to 1.0. The informativeness scores range from 0.5 to 1.0 for 60 summaries, and the relevance score ranges from 0.5 to 1.0 for 20 to 50 summaries. The conciseness score is between 0.2 to 0.4 for 50 to 60 summaries, and the Coverage score is between 0.5 to 1.0 for 20 to 50 summaries. The fluency score is approximately between 0.6 and 1.0 for 100 summaries. The bigram_diversity scores are between 0.8 to 1.0 for 230 summaries. We found the maximum accuracy of scores achieved in Bigram diversity. The minimum accuracy achieved in coherence scores.

6.4.6 Visualisation of Evaluation Parameters Scores

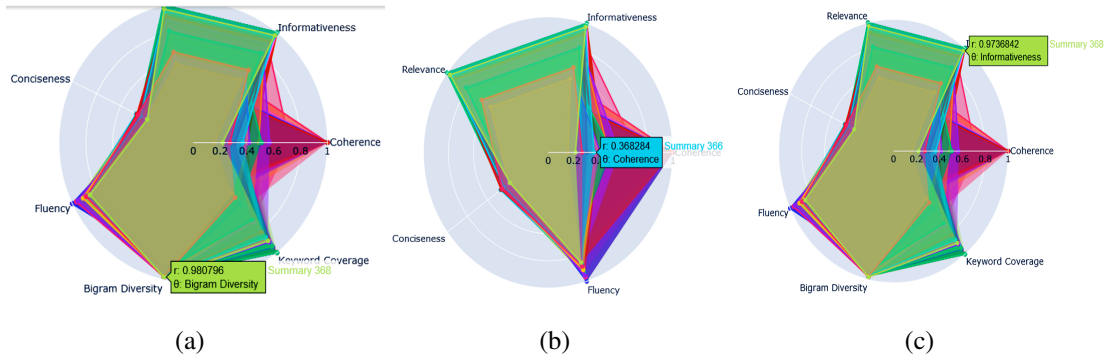


Figure 6.14: (a) Bigram-diversity score (b) Coherence score (c) Informative score

Figure 6.14 (a) presents the Bigramdiversity score of summary number 368 using a radar chart, and (b) shows the coherence score of summary number 366, and (c) explains the informative score for summary number 368.

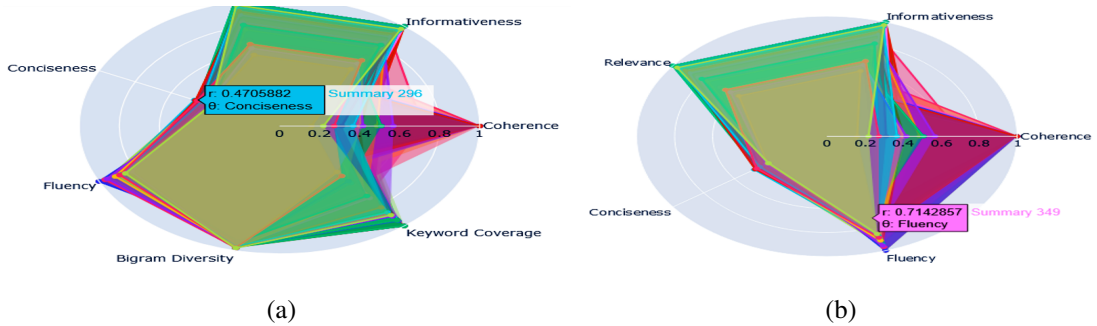


Figure 6.15: (a) Conciseness score (b) Fluency score with summary number

Figure 6.15(a) shows the conciseness score and (b) demonstrates the fluency score-0.7142857 of summary 349.

6.4.7 Correlation of Rouge and BERT and All Metrics Score

The BERT works on the semantic embeddings, although Rouge score work on one-gram, n-gram and n-grmas.

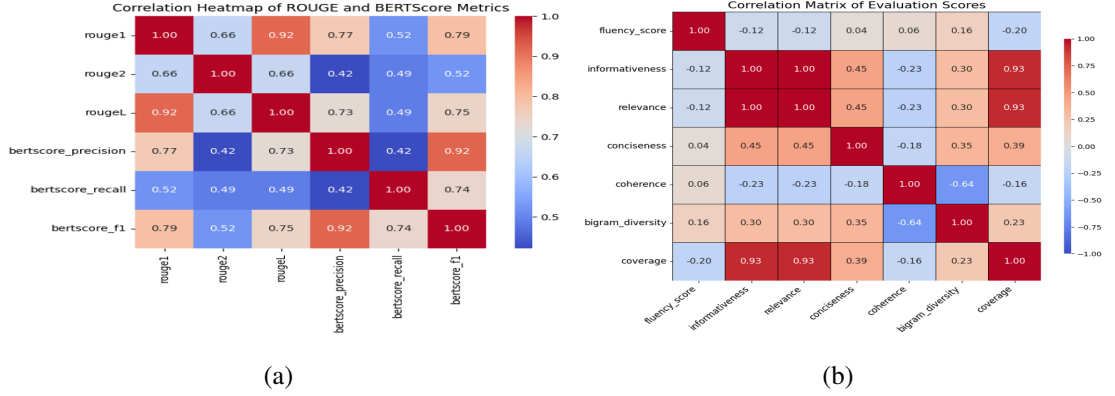


Figure 6.16: (a)Heat map of Rouge, BERT Metric Scores (b) Heat map of All Metric Scores

Figure 6.16 (a) shows the distribution of Rouge and BERT Scores. BERT score has achieved max scores as compared to the Rouge metric. (b) The correlation heatmap shows that ROUGE metrics (ROUGE-1, ROUGE-2, ROUGE-L) are strongly correlated with each other (b)presents the correlation between metric scores. Informative vs. Coverage metrics are highly positively correlated. It shows that if a summary scores a maximum of informativeness, then it has good coverage. Conciseness metric scores moderately correlate with relevance scores and informativeness. The red colour shows a high positive correlation. The blue colour indicates negative correlations (-0.12). Light colours show a weak or no correlation. The informative and coverage metrics scores (0.93) are highly positively correlated. Relevance and coverage scores are also highly positive correlated (0.93). But coherence vs. bigram Diversity has a negative correlation (-1.00). If the bigram diversity is high, then it can reduce the coherence score. No correlation is close to 0. Fluency and informative metrics (-0.12) have a negative correlation.

6.4.8 Inter-agreement Analysis of Annotators

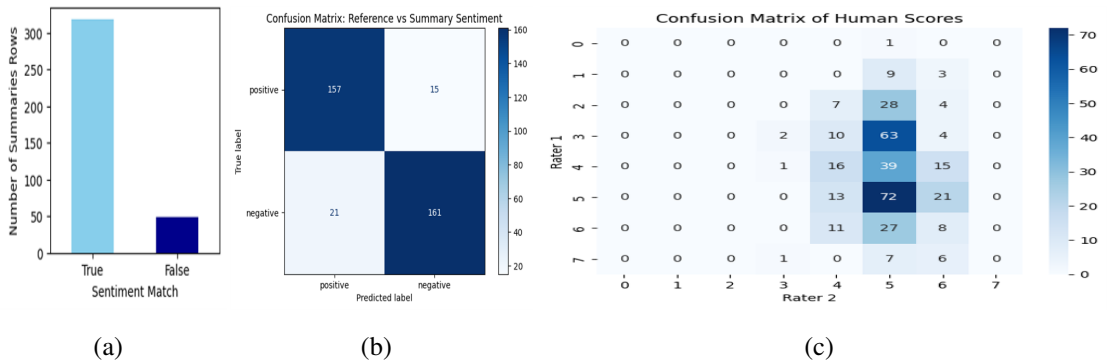


Figure 6.17: (a) Mismatch Sentiment Distribution of Reference Summary vs Predicted Summary (b) Reference Vs Summary Sentiment (c) Rater 1 and Rater 2 Scores

Figure 6.17 (a) shows the sentiment match and mismatch of reference vs predicted summary. The x-axis defines the True and False of sentiment match, and the y-axis de-

scribes the count of summary rows. The figure explains the approximate.310 predicted summaries' sentiment matches and 50 summaries' sentiment are a mismatch from the reference summaries' sentiment. Figure 6.17 (b) presents the confusion matrix of true positive and true negative values of reference vs summary sentiment, Figure 6.17 (c) presents the confusion matrix of the scores from the two experts' summary evaluations. The x-axis describes the evaluation scores given by Rater 2, and the y-axis describes the evaluation scores given by Rater 1. The main diagonal shows agreement cases, where both raters gave the same score; for example, a Score of 5 given by both raters occurs 72 times, and a Score of 3 given by both raters occurs 63 times. The off-diagonal value shows the disagreement. Score 4 is given by Rater 1, Score 5 is given by Rater 2 is 16 times, Score 5 by Rater 1 and Score 6 by Rater 2 is 21 times. This mainly shows the consistent agreement of scores from 3 to 6.

6.4.9 Meaningful Summaries Based on Threshold Value

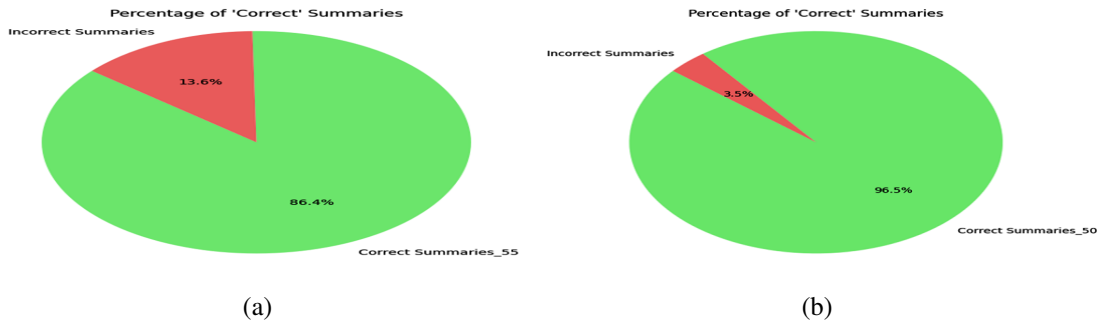


Figure 6.18: (a) Meaningful Summaries based on 55 Threshold Value (b) Meaningful Summaries based on 50 Threshold Value

Figure 6.18 (a) shows that 86.4% are meaningful summaries and 13.6% of summaries are not meaningful based on the if average score greater than 55 threshold value and Figure (b) presents the 96.5% summaries are correct, and only 3.5% summaries are not meaningful based on 50% threshold value.

6.5 Bench Marking

To evaluate the effectiveness of topic modelling methods, we compared the results using coherence and diversity scores for extracted topics from both unlabelled and labelled topic modelling techniques, achieving better results with product-aware labelled methods. We have proposed an aspect-based summarization for multi-product reviews and benchmarked it against several existing supervised and unsupervised methods. Standard evaluation metrics such as ROUGE 1,2,L, BLEU, BERT F1-score, and human score are used to evaluate the summaries' quality and the proposed work has achieved best scores as compared to existing work.

6.5.1 Comparative Analysis of Baseline with Proposed Work

Table 6.2: Analysis of Baseline with Proposed Work

Ref	Approach/Methods	Rouge1	Rouge2	RougeL	Summary/Dataset
[57]	Token cluster	33.4	6.71	17.95	Extractive (Amazon)
	method	44.48	12.84	26.61	Extractive (Space)
[64]	Kullkbaack	14.1	20.0	13.1	Extractive (Opinions)
	Leibler	16.9	20	16.3	(Space)
	Divergence Method, RNN	26.9	11.1	25	(Amazon)
[14]	ACESUM model	32.41	19.47	25.46	Abstractive (Space)
		29.53	6.79	21.06	(OPOSUM)
[85]	BART weak supervision	28.56	10.53	25.93	Abstractive
		35.62	15.8	33.01	(CNN/Daily Mail)
[29]	T5 Model	37.27	19.22	35.32	Abstractive (AMA-SUM)
Our	BERT, BART, Pegasus and T5	51.91	49.23	50.76	Hybrid Extractive to Abstractive (Multi-product Reviews)

The table 6.2 describes the comparison of the Rouge score of existing work with the proposed approach, and it demonstrates that the proposed approach has achieved maximum accuracy.

6.5.2 Comparative Analysis of Proposed Work with Existing Work

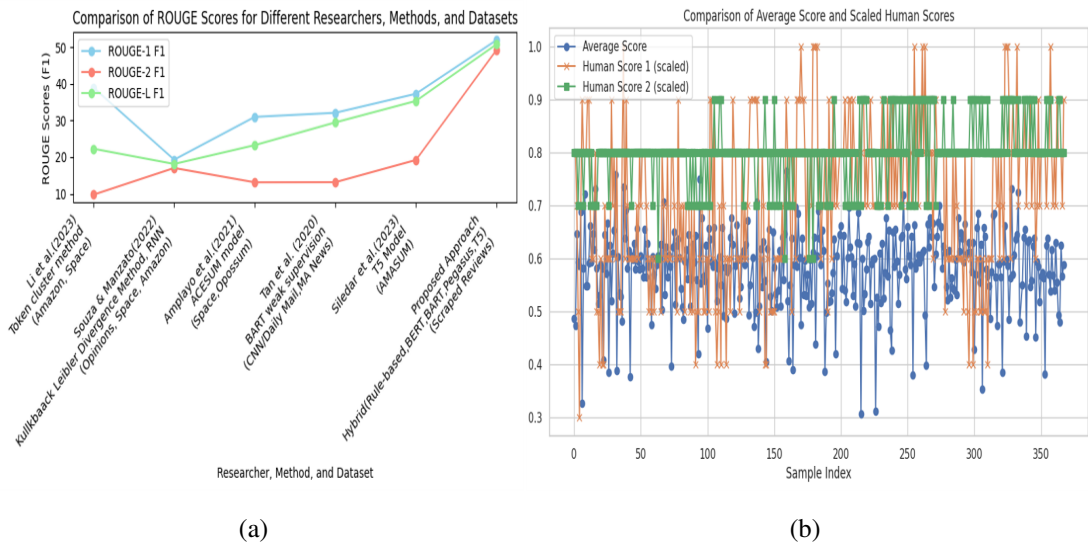


Figure 6.19: (a)Comparative Analysis of Proposed Work (b) Comparison of Average score and Human Scores

Figure 6.19 (a) presents the comparative analysis of proposed work with existing work. The x-axis defines the researcher's names and the dataset, as well as the method names used to develop a framework for summarisation. The proposed approach has achieved 51.9 Rouge1, 49.2 Rouge-2, and 50.7 RougeL scores, which are very good as compared to existing work. Figure 6.19 (b) describes the comparative analysis of proposed work with existing work. The x-axis represents the summary's index, and the y-axis displays the score values, ranging from 0.3 to 1.0. The blue-coloured lines define the average scores calculated by adding all metric scores, and the orange colour describes the Human 1 score; the Blue colour shows the Human 2 scores.

6.6 Conclusion

This chapter explains the use cases of summary evaluation parameter scores and experiments with several intrinsic parameter metric scores, including coherence, conciseness, informativeness, keyword coverage, relevance, fluency, bigram diversity, Rouge, BLEU, Human Evaluation, and BERT F1 scores. The best scores have been achieved using the BERT score because it evaluates summaries based on semantics, rather than n-grams or one-grams, unlike the Rouge score. The inter-rater agreement analysis is performed to calculate the similarity between human scores. At last, the percentage of correct or meaningful summaries is calculated by averaging all metric scores and based on the threshold value. The next chapter provides the conclusion and future work.

Chapter 7

Summary, Conclusion and Future Work

This chapter concludes the research work presented in this thesis. It provides a summary of the key findings, highlights the main contributions, outlines the limitations of the current work, and proposes directions for future research.

7.1 Summary

The proposed research presents a comprehensive study on Aspect-based sentiment-oriented review summarization extractive to abstractive summaries of unlabelled reviews. The proposed work has also employed unsupervised and supervised product-aware topic models, including LDA, NMF, and BERTopic, on Hinglish code-mixed product reviews to extract coherent and important topics from multi-product reviews. A summary of the main characteristics and results of each approach is presented in the following stages:

1. Pre-processing of English and Hinglish Reviews: We have considered both English and Hinglish Reviews. To filter the English reviews, regular expressions and natural language processing (NLP) tools are utilised. For the English language reviews, the contractions and sentence tokenisation are applied. However, for Hinglish code-mixed product reviews, stop word removal and word tokenisation preprocessing steps are used. To remove Hinglish stop words, we have created a Hinglish domain-specific stop words list.

2. Unsupervised and Product-aware Topic Modelling on Hinglish Reviews: Unsupervised topic models like LDA, NMF, and BERTopic are applied on Hinglish reviews without product names to extract topics. However, the extracted topics are not relevant or coherent, and they are a mixture of topics from multiple products. Product names are introduced as labels to guide the topic models, enabling the extraction of more coherent and relevant product-specific topics. Overall, the label-guided product-aware models have improved coherence and relevance of the extracted topics.

3. Aspect-based Sentiment Oriented Summarisation: Rule-based and Transfer learning methods are used to develop a framework for aspect-based sentiment-oriented summarisation. The product-related aspects list is created manually using domain knowledge. A reference summary is generated using DistilBERT, and abstractive sum-

maries are generated using BERT, BART, and Pegasus. The best score summary with sentiment, features, and product category is trained using a fine-tuned T5 model.

4. Evaluation Metrics: Coherence and diversity scores are used to compare the quality of topics. Human evaluation and ROUGE, BLEU, and BERT assessed the summary quality.

7.2 Conclusion

Aspect-based summarisation is important as it provides the correct context of the reviews. We collected the approx. 55k reviews from Indian based e-commerce websites, and applied preprocessing techniques on noisy, abbreviated Indian English and Hinglish code-mixed reviews. In this thesis, we attempted to address the lack of data and research on real-world English and Hinglish code-mixed product reviews. We have also made significant contributions to the field of Hinglish language processing. Product-aware, topic-guided, and unlabelled topic modelling techniques are applied to mine Hinglish code-mixed product reviews. An aspect-based sentiment-oriented summarization framework has been developed for English reviews to extract fine-grained opinions. The BERTscore and Rouge scores, as well as Human scores, have been utilised to evaluate summary quality. Our proposed hybrid approach, combining extractive and abstractive summarisation, has given the best results.

7.3 Future Work

There are several directions for future research to enhance this work.

1. Aspect-based summarisation datasets can be developed for different domains by labelling the aspect-based summaries through domain and linguistic experts.
2. To develop a hybrid summarisation system that can generate summaries in both English and Hinglish.
3. To extend the framework to other Indian and multilingual datasets.
4. To integrate with real-time review analysis systems for e-commerce.
5. To incorporate multimodal data (text, images, videos) for richer sentiment insights.
6. To develop a human–AI hybrid evaluation for a more meaningful summary assessment.

The insights from this thesis can be a stepping stone towards more explainable and user-centric review summarization systems across multi-product domains.

Bibliography

- [1] A. K. Yadav, A. Singh, M. Dhiman, Vineet, R. Kaundal, A. Verma, and D. Yadav, "Extractive text summarization using deep learning approach," *International Journal of Information Technology*, vol. 14, no. 5, pp. 2407–2415, 2022.
- [2] S. R. Rahimi, A. T. Mozhdehi, and M. Abdolahi, "An overview on extractive text summarization," in *2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI)*, pp. 0054–0062, IEEE, 2017.
- [3] A. Mahajani, V. Pandya, I. Maria, and D. Sharma, "A comprehensive survey on extractive and abstractive techniques for text summarization," *Ambient Communications and Computer Systems: RACCCS-2018*, pp. 339–351, 2019.
- [4] D. Patel and H. Chhinkaniwala, "Fuzzy logic-based single document summarisation with improved sentence scoring technique," *International Journal of Knowledge Engineering and Data Mining*, vol. 5, no. 1-2, pp. 125–138, 2018.
- [5] A. Barrera and R. Verma, "Combining syntax and semantics for automatic extractive single-document summarization," in *Computational Linguistics and Intelligent Text Processing: 13th International Conference, CICLing 2012, New Delhi, India, March 11-17, 2012, Proceedings, Part II 13*, pp. 366–377, Springer, 2012.
- [6] M. Zhang, G. Zhou, W. Yu, N. Huang, and W. Liu, "A comprehensive survey of abstractive text summarization based on deep learning," *Computational intelligence and neuroscience*, vol. 2022, no. 1, p. 7132226, 2022.
- [7] N. Moratanch and S. Chitrakala, "A survey on extractive text summarization," in *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, pp. 1–6, 2017.
- [8] K. Mani, I. Verma, H. Meisheri, and L. Dey, "Multi-document summarization using distributed bag-of-words model," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 672–675, IEEE, 2018.
- [9] S. Gupta and S. K. Gupta, "Abstractive summarization: An overview of the state of the art," *Expert Systems with Applications*, vol. 121, pp. 49–65, 2019.
- [10] X. Zhang, Y. Jiang, Y. Shang, Z. Cheng, C. Zhang, X. Fan, Y. Xiao, and B. Long, "Dsgpt: Domain-specific generative pre-training of transformers for text generation in e-commerce title and review summarization," in *Proceedings of the 44th*

International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2146–2150, 2021.

- [11] H. Bhandari, M. Shimbo, T. Ito, and Y. Matsumoto, “Generic text summarization using probabilistic latent semantic indexing,” in *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*, 2008.
- [12] C. Park and Y. Ko, “Query-focused summarization with the context-graph information fusion transformer,” *Expert Systems with Applications*, vol. 241, p. 122699, 2024.
- [13] M. Liu, Y. Fang, A. G. Choulos, D. H. Park, and X. Hu, “Product review summarization through question retrieval and diversification,” *Inf Retrieval J*, vol. 20, no. 6, p. 575–605, 2017.
- [14] R. K. Amplayo, S. Angelidis, and M. Lapata, “Aspect-controllable opinion summarization,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 6578–6593, 2021.
- [15] N. Modani, E. Khabiri, H. Srinivasan, and J. Caverlee, “Creating diverse product review summaries: a graph approach,” in *Web Information Systems Engineering—WISE 2015: 16th International Conference, Miami, FL, USA, November 1-3, 2015, Proceedings, Part I 16*, pp. 169–184, Springer, 2015.
- [16] M. Patkar, P. Pawar, M. Singh, and A. Save, “A new way for semi supervised learning based on data mining for product reviews,” in *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, pp. 819–824, IEEE, 2016.
- [17] A. Mabrouk, R. P. D. Redondo, and M. Kayed, “Seopinion: summarization and exploration of opinion from e-commerce websites,” *Sensors*, vol. 21, no. 2, p. 636, 2021.
- [18] J. P. Singh, N. P. Rana, and W. Alkhowaiter, “Sentiment analysis of products’ reviews containing english and hindi texts,” in *Open and Big Data Management and Innovation: 14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2015, Delft, The Netherlands, October 13-15, 2015, Proceedings 14*, pp. 416–422, Springer, 2015.
- [19] C. P. Chai, “Comparison of text preprocessing methods,” *Natural Language Engineering*, vol. 29, no. 3, pp. 509–553, 2023.
- [20] R. K. Amplayo and M. Lapata, “Unsupervised opinion summarization with noising and denoising,” in *Proceedings of the 58th Annual Meeting of the Association*

for Computational Linguistics, (Online), pp. 1934–1945, Association for Computational Linguistics, July 2020.

- [21] A. S. Nayak, A. P. Kanive, N. Chandavekar, and R. Balasubramani, “Survey on pre-processing techniques for text mining,” *International Journal of Engineering and Computer Science*, vol. 5, no. 6, pp. 16875–16879, 2016.
- [22] S. Vijayarani, M. J. Ilamathi, M. Nithya, *et al.*, “Preprocessing techniques for text mining-an overview,” *International Journal of Computer Science & Communication Networks*, vol. 5, no. 1, pp. 7–16, 2015.
- [23] R. Santhiran, K. D. Varathan, and Y. K. Chiam, “Feature extraction from customer reviews using enhanced rules,” *Peerj Computer Science*, vol. 10, p. e1821, 2024.
- [24] S. S. Htay, K. T. Lynn, *et al.*, “Extracting product features and opinion words using pattern knowledge in customer reviews,” *The Scientific World Journal*, vol. 2013, 2013.
- [25] T. Hou, B. Yannou, Y. Leroy, and E. Poirson, “Mining customer product reviews for product development: A summarization process,” *Expert Systems with Applications*, vol. 132, pp. 141–150, 2019.
- [26] G. Di Fabrizio, A. Stent, and R. Gaizauskas, “A hybrid approach to multi-document summarization of opinions in reviews,” in *Proceedings of the 8th International Natural Language Generation Conference (INLG)*, pp. 54–63, 2014.
- [27] Y. Yang, C. Chen, M. Qiu, and F. Bao, “Aspect extraction from product reviews using category hierarchy information,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pp. 675–680, 2017.
- [28] X. Yang, K. Song, S. Cho, X. Wang, X. Pan, L. Petzold, and D. Yu, “Oasum: Large-scale open domain aspect-based summarization,” in *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 4381–4401, 2023.
- [29] T. Siledar, J. Makwana, and P. Bhattacharyya, “Aspect-sentiment-based opinion summarization using multiple information sources,” in *Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD)*, pp. 55–61, 2023.
- [30] D. G. R. Ravali Boorugu and D. K. Madhavi, “Summarizing product reviews using nlp based text summarization,” *International Journal of Scientific & Technology Research*, vol. 8, no. 10, 2019.

- [31] T. A. Rana and Y.-N. Cheah, "Hybrid rule-based approach for aspect extraction and categorization from customer reviews," in *2015 9th International Conference on IT in Asia (CITA)*, pp. 1–5, IEEE, 2015.
- [32] A. Sheikh, T. Arif, M. Malik, and S. Bhat, "Extraction and summarization of reviews using lexicon based approach," in *IOP Conference Series: Materials Science and Engineering*, vol. 1022, p. 012117, IOP Publishing, 2021.
- [33] M. Zhang, G. Zhou, N. Huang, P. He, W. Yu, and W. Liu, "Asu-osum: aspect-augmented unsupervised opinion summarization," *Information Processing & Management*, vol. 60, no. 1, p. 103138, 2023.
- [34] J. Shah, M. Sagathiya, K. Redij, and V. Hole, "Natural language processing based abstractive text summarization of reviews," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 461–466, IEEE, 2020.
- [35] A. P. Wibawa, F. Kurniawan, *et al.*, "A survey of text summarization: Techniques, evaluation and challenges," *Natural Language Processing Journal*, p. 100070, 2024.
- [36] H. Lin and V. Ng, "Abstractive summarization: A survey of the state of the art," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 9815–9822, 2019.
- [37] S. Rani and T. S. Walia, "Automatic summarisation of product reviews using natural language processing and machine learning methods: a literature review," *International Journal of Data Mining and Bioinformatics*, vol. 27, no. 1-3, pp. 118–138, 2022.
- [38] H. Shakil, A. Farooq, and J. Kalita, "Abstractive text summarization: State of the art, challenges, and improvements," *Neurocomputing*, p. 128255, 2024.
- [39] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177, 2004.
- [40] A. Konjengbam, N. Dewangan, N. Kumar, and M. Singh, "Aspect ontology based review exploration," *Electronic Commerce Research and Applications*, vol. 30, pp. 62–71, 2018.
- [41] A. R. Hanni, M. M. Patil, and P. M. Patil, "Summarization of customer reviews for a product on a website using natural language processing," in *2016 International conference on advances in computing, communications and informatics (ICACCI)*, pp. 2280–2285, IEEE, 2016.

- [42] W. Chung and T.-L. B. Tseng, “Discovering business intelligence from online product reviews: A rule-induction framework,” *Expert systems with applications*, vol. 39, no. 15, pp. 11870–11879, 2012.
- [43] R. Siautama, A. C. IA, D. Suhartono, *et al.*, “Extractive hotel review summarization based on tf/idf and adjective-noun pairing by considering annual sentiment trends,” *Procedia Computer Science*, vol. 179, pp. 558–565, 2021.
- [44] M. Hong and H. Wang, “Research on customer opinion summarization using topic mining and deep neural network,” *Mathematics and Computers in Simulation*, vol. 185, pp. 88–114, 2021.
- [45] K. Chigateri and R. Bhandarkar, “Necessity and preference mining from text reviews: a non-bipolar assessment of text reviews,” *Indian Journal of Science and Technology*, vol. 14, no. 40, pp. 3064–3073, 2021.
- [46] K. Amarouche, H. Benbrahim, and I. Kassou, “Customer product review summarization over time for competitive intelligence,” *Journal of Automation, Mobile Robotics and Intelligent Systems*, pp. 70–82, 2018.
- [47] X. Li, P. Wu, C. Zou, H. Xie, and F. L. Wang, “Sentiment lossless summarization,” *Knowledge-Based Systems*, vol. 227, p. 107170, 2021.
- [48] Y. K. Meena and D. Gopalani, “Feature priority based sentence filtering method for extractive automatic text summarization,” *Procedia Computer Science*, vol. 48, pp. 728–734, 2015.
- [49] M. K. Dalal and M. A. Zaveri, “Semisupervised learning based opinion summarization and classification for online product reviews,” *Applied Computational Intelligence and Soft Computing*, vol. 2013, pp. 10–10, 2013.
- [50] M. ÇATALTAŞ, S. DOĞRAMACI, S. YUMUŞAK, and K. ÖZTOPRAK, “Extraction of product defects and opinions from customer reviews by using text clustering and sentiment analysis,” in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 4529–4534, IEEE, 2020.
- [51] M. Al-Dhelaan and A. Al-Suhaim, “Sentiment diversification for short review summarization,” in *Proceedings of the International Conference on Web Intelligence*, pp. 723–729, 2017.
- [52] G. Ansari, C. Saxena, T. Ahmad, and M. Doja, “Aspect term extraction using graph-based semi-supervised learning,” *Procedia Computer Science*, vol. 167, pp. 2080–2090, 2020.

- [53] N. Thessrimuang and O. Chaowalit, "Opinion representative of cosmetic products," in *2016 International Computer Science and Engineering Conference (IC-SEC)*, pp. 1–6, IEEE, 2016.
- [54] R. Manjupriya., T. Aarthi, and K. Krishnakumari, "A big data approach on aspect based summarization," *International journal of engineering research and technology*, vol. 4, 2018.
- [55] X. Meng and H. Wang, "Mining user reviews: from specification to summarization," in *Proceedings of the ACL-IJCNLP 2009 conference short papers*, pp. 177–180, 2009.
- [56] S. Angelidis and M. Lapata, "Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, eds.), pp. 3675–3686, 2018.
- [57] H. Li, S. B. R. Chowdhury, and S. Chaturvedi, "Aspect-aware unsupervised extractive opinion summarization," in *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 12662–12678, 2023.
- [58] C.-F. Tsai, K. Chen, Y.-H. Hu, and W.-K. Chen, "Improving text summarization of online hotel reviews with review helpfulness and sentiment," *Tourism Management*, vol. 80, p. 104122, 2020.
- [59] K. Bafna and D. Toshniwal, "Feature based summarization of customers' reviews of online products," *Procedia Computer Science*, vol. 22, pp. 142–151, 2013.
- [60] K. Krishnakumari and E. Sivasankar, "Aspect-based summarisation in the big data environment," *International Journal of Advanced Intelligence Paradigms*, vol. 25, no. 1-2, pp. 68–83, 2023.
- [61] M. V. Phong, T. T. Nguyen, H. V. Pham, and T. T. Nguyen, "Mining user opinions in mobile app reviews: A keyword-based approach (t)," in *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pp. 749–759, IEEE, 2015.
- [62] M. Coavoux, H. Elsahar, and M. Gallé, "Unsupervised aspect-based multi-document abstractive summarization," in *2nd Workshop on New Frontiers in Summarization*, pp. 42–47, Association for Computational Linguistics, 2019.
- [63] S. Gautam, J. Kaur, and G. S. Josan, "Deep neural network based multi-review summarization system.," *International Journal of Next-Generation Computing*, vol. 12, no. 3, 2021.

- [64] L. S. d. Souza and M. G. Manzano, “Aspect-based summarization: an approach with different levels of details to explain recommendations,” in *Proceedings of the Brazilian Symposium on Multimedia and the Web*, pp. 202–210, 2022.
- [65] D. Wang, S. Zhu, and T. Li, “Sumview: A web-based engine for summarizing product reviews and customer opinions,” *Expert Systems with Applications*, vol. 40, no. 1, pp. 27–33, 2013.
- [66] R. C. Belwal, S. Rai, and A. Gupta, “Text summarization using topic-based vector space model and semantic measure,” *Information Processing & Management*, vol. 58, no. 3, p. 102536, 2021.
- [67] F. Jin, M.-L. Huang, and X.-Y. Zhu, “Guided structure-aware review summarization,” *Journal of Computer Science and Technology*, vol. 26, no. 4, pp. 676–684, 2011.
- [68] R. Mukherjee, H. C. Peruri, U. Vishnu, P. Goyal, S. Bhattacharya, and N. Ganguly, “Read what you need: Controllable aspect-based opinion summarization of tourist reviews,” in *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pp. 1825–1828, 2020.
- [69] J. Y. Chin, S. S. Bhowmick, and A. Jatowt, “On-demand recent personal tweets summarization on mobile devices,” *Journal of the Association for Information Science and Technology*, vol. 70, no. 6, pp. 547–562, 2019.
- [70] N. Akhtar, N. Zubair, A. Kumar, and T. Ahmad, “Aspect based sentiment oriented summarization of hotel reviews,” *Procedia computer science*, vol. 115, pp. 563–571, 2017.
- [71] A. Bražinskas, M. Lapata, and I. Titov, “Few-shot learning for opinion summarization,” in *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, 2020.
- [72] A. Bražinskas, M. Lapata, and I. Titov, “Unsupervised opinion summarization as copycat-review generation,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, eds.), Association for Computational Linguistics, July 2020.
- [73] A. Tang, X. J. Zhang, and M. Dinh, “Aspect-based key point analysis for quantitative summarization of reviews,” in *Findings of the Association for Computational Linguistics: EACL 2024*, pp. 1419–1433, 2024.
- [74] Z. Luo, “Knowledge-guided aspect-based summarization,” in *2023 International conference on communications, computing and artificial intelligence (CCCAI)*, pp. 17–22, IEEE, 2023.

- [75] S. Takeshita, T. Green, I. Reinig, K. Eckert, and S. P. Ponzetto, “Aclsum: A new dataset for aspect-based summarization of scientific publications,” *arXiv preprint arXiv:2403.05303*, 2024.
- [76] Z. Ankner, P. Balaji, Y. Zhu, C. K. Hiew, P. Wang, and A. Gupta, “Entailsum: An entailment-based approach to aspect-based text summarization with automated aspect adaptation,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 36, no. 13, p. 2259017, 2022.
- [77] A. Soleimani, V. Nikoulina, B. Favre, and S. Ait-Mokhtar, “Zero-shot aspect-based scientific document summarization using self-supervised pre-training,” in *Proceedings of the 21st Workshop on Biomedical Language Processing*, pp. 49–62, 2022.
- [78] Y. Wang, Y. Zhou, M. Wang, Z. Chen, Z. Cai, J. Chen, and V. C. Leung, “Multidocument aspect classification for aspect-based abstractive summarization,” *IEEE Transactions on Computational Social Systems*, 2023.
- [79] Y. Zhang, Y. Liu, Z. Yang, Y. Fang, Y. Chen, D. Radev, C. Zhu, M. Zeng, and R. Zhang, “Macsum: Controllable summarization with mixed attributes,” *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 787–803, 2023.
- [80] H. Pan, R. Yang, X. Zhou, R. Wang, D. Cai, and X. Liu, “Large scale abstractive multi-review summarization (LSARS) via aspect alignment,” in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020*, pp. 2337–2346, 2020.
- [81] P. D. Korkankar, A. Abranches, P. Bhagat, and J. Pawar, “Aspect-based summaries from online product reviews: A comparative study using various llms,” in *Proceedings of the 21st International Conference on Natural Language Processing (ICON)*, pp. 562–568, 2024.
- [82] X. Xu, T. Meng, and X. Cheng, “Aspect-based extractive summarization of online reviews,” in *Proceedings of the 2011 ACM Symposium on Applied Computing*, pp. 968–975, 2011.
- [83] S. Tyss, M. Aly, and M. Grabmair, “Lexabsumm: Aspect-based summarization of legal decisions,” in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 10422–10431, 2024.
- [84] H. Kansal and D. Toshniwal, “Aspect based summarization of context dependent opinion words,” *Procedia Computer Science*, vol. 35, pp. 166–175, 2014.

- [85] B. Tan, L. Qin, E. Xing, and Z. Hu, “Summarizing text on any aspects: A knowledge-informed weakly-supervised approach,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Online), pp. 6301–6309, Association for Computational Linguistics, Nov. 2020.
- [86] W.-J. Ye and A. J. Lee, “Mining sentiment tendencies and summaries from consumer reviews,” *Information Systems and e-Business Management*, vol. 19, no. 1, pp. 107–135, 2021.
- [87] L. Frermann and A. Klementiev, “Inducing document structure for aspect-based summarization,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6263–6273, 2019.
- [88] B. M. Gurusamy, P. K. Rengarajan, and P. Srinivasan, “A hybrid approach for text summarization using semantic latent dirichlet allocation and sentence concept mapping with transformer,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, pp. 6663–6672, 2023.
- [89] H. Liu and X. Wan, “Neural review summarization leveraging user and product information,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2389–2392, 2019.
- [90] A. Nikhil Padhi, P. Parikh, S. Kanwal, K. Karlapalem, and N. Raman, “Aspect-based summarization of legal case files using sentence classification,” in *Companion Proceedings of the ACM Web Conference 2023*, p. 1249–1252, Association for Computing Machinery, 2023.
- [91] N. Jeong and J. Lee, “An aspect-based review analysis using chatgpt for the exploration of hotel service failures,” *Sustainability*, vol. 16, no. 4, p. 1640, 2024.
- [92] H. Hayashi, P. Budania, P. Wang, C. Ackerson, R. Neervannan, and G. Neubig, “Wikiasp: A dataset for multi-domain aspect-based summarization,” *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 211–225, 2021.
- [93] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177, 2004.
- [94] A. Bhatt, A. Patel, H. Chheda, and K. Gawande, “Amazon review classification and sentiment analysis,” *International Journal of Computer Science and Information Technologies*, vol. 6, no. 6, pp. 5107–5110, 2015.
- [95] N. Yu, M. Huang, Y. Shi, and X. Zhu, “Product review summarization by exploiting phrase properties,” in *Proceedings of COLING 2016, the 26th International*

Conference on Computational Linguistics: Technical Papers, pp. 1113–1124, 2016.

- [96] B. A. H. Murshed, S. Mallappa, J. Abawajy, M. A. N. Saif, H. D. E. Al-Ariki, and H. M. Abdulwahab, “Short text topic modelling approaches in the context of big data: taxonomy, survey, and analysis,” *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5133–5260, 2023.
- [97] M. Baimakhanbetov, “Determination of the optimal number of topics in the lda model when working with large arrays of text data,” in *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, pp. 332–336, IEEE, 2023.
- [98] H. Kitagawa, E.-S. Atlam, M. Fuketa, K. Morita, and J.-i. Aoe, “Document summarisation on mobile devices using non-negative matrix factorisation,” *International journal of computer applications in technology*, vol. 46, no. 1, pp. 13–23, 2013.
- [99] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, and M. Zhou, “Low-quality product review detection in opinion summarization,” in *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, pp. 334–342, 2007.
- [100] S. M. Basha and D. S. Rajput, “An innovative topic-based customer complaints sentiment classification system,” *International Journal of Business Innovation and Research*, vol. 20, no. 3, pp. 375–391, 2019.
- [101] A. Khan, M. A. Gul, M. I. Uddin, S. A. Ali Shah, S. Ahmad, M. D. Al Firdausi, and M. Zaindin, “Summarizing online movie reviews: a machine learning approach to big data analytics,” *Scientific Programming*, vol. 2020, pp. 1–14, 2020.
- [102] G. Anuradha and D. J. Varma, “Fuzzy based summarization of product reviews for better analysis,” *Indian journal of science and technology*, 2016.
- [103] L. S. Vakada, A. Ch, M. Marreddy, S. R. Oota, and R. Mamidi, “Gae-isumm: Unsupervised graph-based summarization for indian languages,” in *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2023.
- [104] A. Singh and S. Saha, “Graphic: A graph-based approach for identifying complaints from code-mixed product reviews,” *Expert Systems with Applications*, vol. 216, p. 119444, 2023.
- [105] M. R. Ramadhan, S. N. Endah, and A. B. J. Mantau, “Implementation of textrank algorithm in product review summarization,” in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, pp. 1–5, IEEE, 2020.

- [106] S. R. Basha, J. K. Rani, and J. Yadav, “A novel summarization-based approach for feature reduction enhancing text classification accuracy,” *Engineering, Technology & Applied Science Research*, vol. 9, no. 6, pp. 5001–5005, 2019.
- [107] X. Liu, Y. Li, F. Wei, and M. Zhou, “Graph-based multi-tweet summarization using social signals,” in *Proceedings of COLING 2012*, pp. 1699–1714, 2012.
- [108] N. S. Shirwandkar and S. Kulkarni, “Extractive text summarization using deep learning,” in *2018 fourth international conference on computing communication control and automation (ICCUBE)*, pp. 1–5, IEEE, 2018.
- [109] G. Bhagchandani, D. Bodra, A. Gangan, and N. Mulla, “A hybrid solution to abstractive multi-document summarization using supervised and unsupervised learning,” in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pp. 566–570, IEEE, 2019.
- [110] A. Abdi, S. Hasan, S. M. Shamsuddin, N. Idris, and J. Piran, “A hybrid deep learning architecture for opinion-oriented multi-document summarization based on multi-feature fusion,” *Knowledge-Based Systems*, vol. 213, p. 106658, 2021.
- [111] P. Muniraj, K. Sabarmathi, R. Leelavathi, *et al.*, “Hntsumm: Hybrid text summarization of transliterated news articles,” *International Journal Of Intelligent Networks*, vol. 4, pp. 53–61, 2023.
- [112] H. Xu, H. Liu, W. Zhang, P. Jiao, and W. Wang, “Rating-boosted abstractive review summarization with neural personalized generation,” *Knowledge-Based Systems*, vol. 218, p. 106858, 2021.
- [113] I. Gamzu, H. Gonen, G. Kutiel, R. Levy, and E. Agichtein, “Identifying helpful sentences in product reviews,” in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 678–691, Association for Computational Linguistics, 2021.
- [114] S. Abdel-Salam and A. Rafea, “Performance study on extractive text summarization using bert models,” *Information*, vol. 13, no. 2, p. 67, 2022.
- [115] A. Brazinskas, R. Nallapati, M. Bansal, and M. Dreyer, “Efficient few-shot fine-tuning for opinion summarization,” in *Findings of the Association for Computational Linguistics: NAACL 2022* (M. Carpuat, M.-C. de Marneffe, and I. V. Meza Ruiz, eds.), Association for Computational Linguistics, July 2022.
- [116] T. R. Goodwin, M. E. Savery, and D. Demner-Fushman, “Flight of the pegasus? comparing transformers on few-shot and zero-shot multi-document abstractive

- summarization,” in *Proceedings of COLING. International Conference on Computational Linguistics*, vol. 2020, p. 5640, 2020.
- [117] Y. Liu, “Comparison of lda and bertopic in news topic modeling: A case study of the new york times’ reports on china,” *Pacific International Journal*, vol. 7, no. 3, pp. 47–51, 2024.
- [118] A. Rumi, “Amazon fine food reviews.” <https://www.kaggle.com/datasets/jillanisofttech/amazon-product-reviews><https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews>, 2020.
- [119] N. Vaghani, “Flipkart product reviews.” <https://www.kaggle.com/datasets/niraliivaghani/flipkart-product-customer-reviews-dataset>, 2021.
- [120] J. McAul, “Home and kitchen.” <https://cseweb.ucsd.edu/~jmcauley/datasets/amazon/links.html>, 2014.
- [121] J. McAul, “Clothing shoes and jewellery.” <https://cseweb.ucsd.edu/~jmcauley/datasets/amazon/links.html>, 2014.
- [122] Nicapotato, “Women’s e-commerce clothing.” <https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews>, 2017.
- [123] V. Mane, “Flipkart mobile reviews.” <https://www.kaggle.com/datasets/vishalmanel0/flipkart-mobile-reviews>, 2021.
- [124] P. Keung, Y. Lu, G. Szarvas, and N. A. Smith, “The multilingual amazon reviews corpus,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 2020.
- [125] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, “Text preprocessing for text mining in organizational research: Review and recommendations,” *Organizational Research Methods*, vol. 25, no. 1, pp. 114–146, 2022.
- [126] R. Lourdasamy and S. Abraham, “A survey on text pre-processing techniques and tools,” *International Journal of Computer Sciences and Engineering*, vol. 6, no. 3, pp. 148–157, 2018.
- [127] A. Tabassum and R. R. Patil, “A survey on text pre-processing & feature extraction techniques in natural language processing,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 06, pp. 4864–4867, 2020.

- [128] J. Bektaş and A. Elsadig, “A unified workflow strategy for analysing large-scale tripadvisor reviews with bow model,” *International Journal of Business Intelligence and Data Mining*, vol. 21, no. 1, pp. 102–117, 2022.
- [129] S. A. Alasadi and W. S. Bhaya, “Review of data preprocessing techniques in data mining,” *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102–4107, 2017.
- [130] A. K. Uysal and S. Gunal, “The impact of preprocessing on text classification,” *Information processing & management*, vol. 50, no. 1, pp. 104–112, 2014.
- [131] W. Etaiwi and G. Naymat, “The impact of applying different preprocessing steps on review spam detection,” *Procedia computer science*, vol. 113, pp. 273–279, 2017.
- [132] Y. Ledeneva, “Effect of preprocessing on extractive summarization with maximal frequent sequences,” in *Mexican international conference on artificial intelligence*, pp. 123–132, Springer, 2008.
- [133] S. Rani and T. S. Walia, “An exploratory analysis of feature extraction techniques using nlp and text embedding methods,” in *Advances in Networks, Intelligence and Computing*, pp. 736–743, CRC Press, 2024.
- [134] T. Mickus, D. Paperno, M. Constant, and K. van Deemter, “What do you mean, bert?,” in *Proceedings of the Society for Computation in Linguistics 2020*, pp. 279–290, 2020.
- [135] V. Thakur, R. Sahu, and S. Omer, “Current state of hinglish text sentiment analysis,” in *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*, 2020.
- [136] R. Egger and J. Yu, “A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *frontiers in sociology*, 7, 886498,” 2022.
- [137] M. R. Maarif, “Summarizing online customer review using topic modeling and sentiment analysis,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 7, no. 3, pp. 177–191, 2022.
- [138] J. Qiang, Z. Qian, Y. Li, Y. Yuan, and X. Wu, “Short text topic modeling techniques, applications, and performance: a survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 3, pp. 1427–1445, 2020.
- [139] C. S. Ponay, “Topic modeling on customer feedback from an online ticketing system using latent dirichlet allocation and bertopic,” in *2022 2nd International*

Conference in Information and Computing Research (iCORE), pp. 1–6, IEEE, 2022.

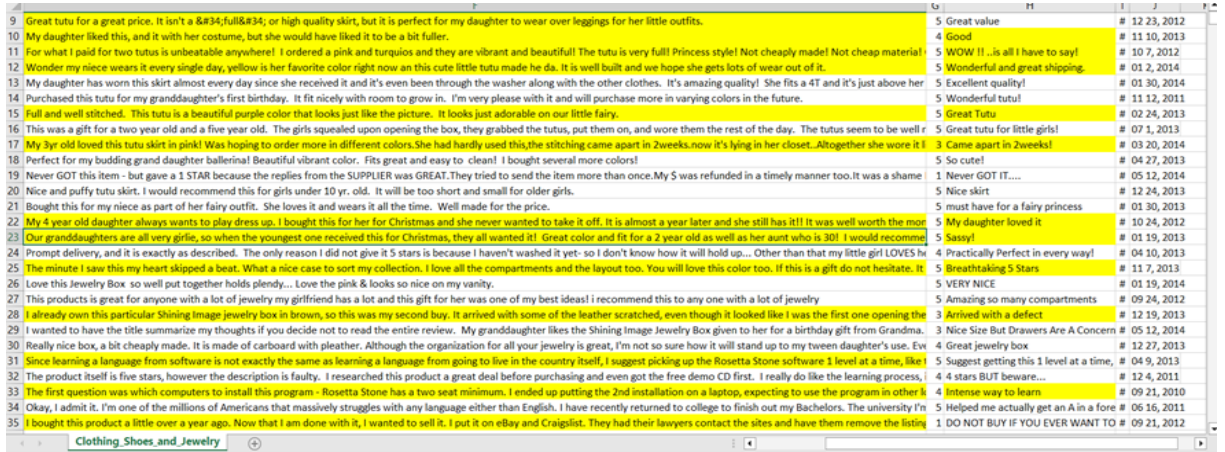
- [140] R. Albalawi, T. H. Yeap, and M. Benyoucef, “Using topic modeling methods for short-text data: A comparative analysis,” *Frontiers in artificial intelligence*, vol. 3, p. 42, 2020.
- [141] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [142] H.-J. Kwon, H.-J. Ban, J.-K. Jun, and H.-S. Kim, “Topic modeling and sentiment analysis of online review for airlines,” *Information*, vol. 12, no. 2, p. 78, 2021.
- [143] N. Lopes and B. Ribeiro, *Non-Negative Matrix Factorization (NMF)*, pp. 127–154. Cham: Springer International Publishing, 2015.
- [144] S. K. Ray, A. Ahmad, and C. A. Kumar, “Review and implementation of topic modeling in hindi,” *Applied Artificial Intelligence*, vol. 33, no. 11, pp. 979–1007, 2019.
- [145] B. Karas, S. Qu, Y. Xu, and Q. Zhu, “Experiments with lda and top2vec for embedded topic discovery on social media data—a case study of cystic fibrosis,” *Frontiers in Artificial Intelligence*, vol. 5, p. 948313, 2022.
- [146] S. F. Eletter, K. I. AlQeisi, and G. A. Elrefae, “The use of topic modeling in mining customers’ reviews,” in *2021 22nd International Arab Conference on Information Technology (ACIT)*, pp. 1–4, IEEE, 2021.
- [147] A. Farkhod, A. Abdusalomov, F. Makhmudov, and Y. I. Cho, “Lda-based topic modeling sentiment analysis using topic/document/sentence (tds) model,” *Applied Sciences*, vol. 11, no. 23, p. 11091, 2021.
- [148] A. Abuzayed and H. Al-Khalifa, “Bert for arabic topic modeling: An experimental study on bertopic technique,” *Procedia computer science*, vol. 189, pp. 191–194, 2021.
- [149] S. Rani and T. S. Walia, “An experimental study of text preprocessing techniques on user reviews,” in *Recent Advances in Computing Sciences*, pp. 226–230, CRC Press, 2023.
- [150] M. Bhandari, P. N. Gour, A. Ashfaq, P. Liu, and G. Neubig, “Re-evaluating evaluation in text summarization,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 9347–9359, Association for Computational Linguistics, 2020.

- [151] E. Lloret, L. Plaza, and A. Aker, “The challenging task of summary evaluation: an overview,” *Language Resources and Evaluation*, vol. 52, pp. 101–148, 2018.

Appendices

A.1 Existing Datasets of Product Reviews Summarisation

A.1.1 Clothing, Shoes and Jewellery Dataset

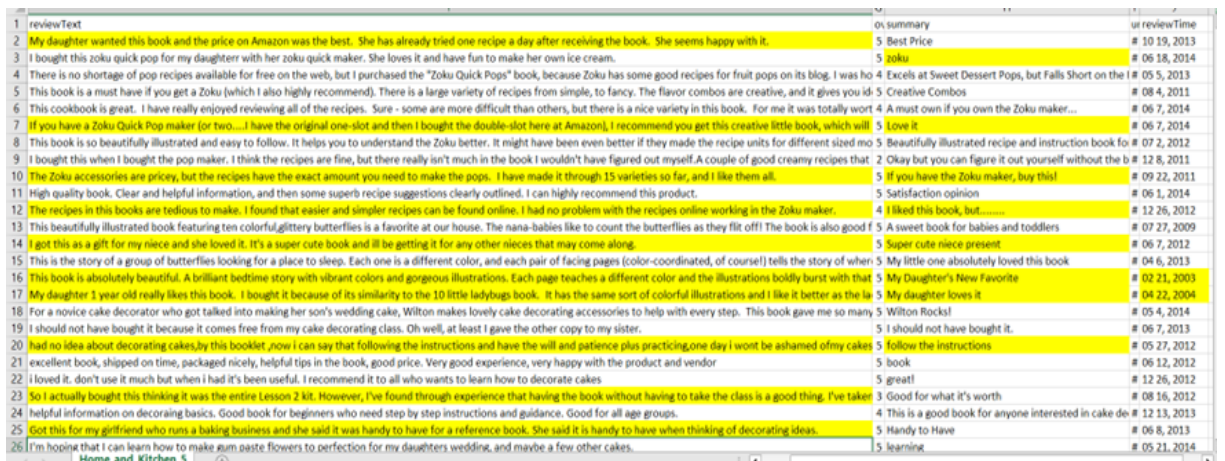


id	reviewText	summary	reviewTime
9	Great tutu for a great price. It isn't a \$34, full \$34, or high quality skirt, but it is perfect for my daughter to wear over leggings for her little outfits.	5 Great value	# 12 23, 2012
10	My daughter liked this, and it with her costume, but she would have liked it to be a bit fuller.	4 Good	# 11 10, 2013
11	For what I paid for two tutus is unbeatable anywhere! I ordered a pink and turquoise and they are vibrant and beautiful! The tutu is very full! Princess style! Not cheaply made! Not cheap material!	5 WOW!! ...is all I have to say!	# 10 7, 2012
12	Wonder my niece wears it every single day, yellow is her favorite color right now and this cute little tutu made her da. It is well built and we hope she gets lots of wear out of it.	5 Wonderful and great shipping	# 01 2, 2014
13	My daughter has worn this skirt almost every day since she received it and it's even been through the washer along with the other clothes. It's amazing quality! She fits a 4T and it's just above her	5 Excellent quality!	# 01 30, 2014
14	Purchased this tutu for my granddaughter's first birthday. It fit nicely with room to grow in. I'm very please with it and will purchase more in varying colors in the future.	5 Wonderful tutu!	# 11 12, 2011
15	Full and well stitched. This tutu is a beautiful purple color that looks just like the picture. It looks just adorable on our little fairy.	5 Great Tutu	# 02 24, 2013
16	This was a gift for a two year old and a five year old. The girls squealed upon opening the box, they grabbed the tutus, put them on, and wore them the rest of the day. The tutus seem to be well	5 Great tutu for little girls!	# 07 1, 2013
17	My 3yr old loved this tutu skirt in pink! Was hoping to order more in different colors, she had hardly used this, the stitching came apart in 2 weeks, now it's lying in her closet. Altogether she wore it	3 Came apart in 2 weeks!	# 03 20, 2014
18	Perfect for my budding grand daughter ballerina! Beautiful vibrant color. Fits great and easy to clean! I bought several more colors!	5 So cute!	# 04 27, 2013
19	Never GOT this item - but gave a 1 STAR because the replies from the SUPPLIER was GREAT. They tried to send the item more than once. My \$ was refunded in a timely manner too. It was a shame	1 Never GOT IT....	# 05 12, 2014
20	Nice and puffy tutu skirt. I would recommend this for girls under 10 yr. old. It will be too short and small for older girls.	5 Nice skirt	# 12 24, 2013
21	Bought this for my niece as part of her fairy outfit. She loves it and wears it all the time. Well made for the price.	5 must have for a fairy princess	# 01 30, 2013
22	My 4 year old daughter always wants to play dress up. I bought this for her for Christmas and she never wanted to take it off. It is almost a year later and she still has it!! It was well worth the	5 My daughter loved it	# 10 24, 2012
23	Our granddaughters are all very girls, so when the youngest one received this for Christmas, they all wanted it! Great color and fit for a 2 year old as well as her aunt who is 30! I would recom	5 Sassy!	# 01 19, 2013
24	Prompt delivery, and it is exactly as described. The only reason I did not give it 5 stars is because I haven't washed it yet- so I don't know how it will hold up... Other than that my little girl LOVES	4 Practically Perfect in every way!	# 04 10, 2013
25	The minute I saw this my heart skipped a beat. What a nice case to sort my collection. I love all the compartments and the layout too. You will love this color too. If this is a gift do not hesitate. It	5 Breathtaking 5 Stars	# 11 7, 2013
26	Love this Jewelry Box so well put together holds plenty... Love the pink & looks so nice on my vanity.	5 VERY NICE	# 01 19, 2014
27	This products is great for anyone with a lot of jewelry my girlfriend has a lot and this gift for her was one of my best ideas! I recommend this to any one with a lot of jewelry	5 Amazing so many compartments	# 09 24, 2012
28	I already own this particular Shining Image jewelry box in brown, so this was my second buy. It arrived with some of the leather scratched, even though it looked like I was the first one opening the	3 Arrived with a defect	# 12 19, 2013
29	I wanted to have the title summarize my thoughts if you decide not to read the entire review. My granddaughter likes the Shining Image Jewelry Box given to her for a birthday gift from Grandma.	3 Nice Size But Drawers Are A Concern	# 05 12, 2014
30	Really nice box, a bit cheaply made. It is made of cardboard with pleather. Although the organization for all your jewelry is great, I'm not so sure how it will stand up to my tween daughter's use. Ev	4 Great jewelry box	# 12 27, 2013
31	Since learning a language from software is not exactly the same as learning a language from going to live in the country itself, I suggest picking up the Rosetta Stone software 1 level at a time, like	5 Suggest getting this 1 level at a time,	# 04 9, 2013
32	The product itself is five stars, however the description is faulty. I researched this product a great deal before purchasing and even got the free demo CD first. I really do like the learning process,	4 4 stars BUT beware...	# 12 24, 2011
33	The first question was which computers to install this program - Rosetta Stone has a two seat minimum. I ended up putting the 2nd installation on a laptop, expecting to use the program in other	4 Intense way to learn	# 09 21, 2010
34	Okay, I admit it. I'm one of the millions of Americans that massively struggles with any language other than English. I have recently returned to college to finish out my Bachelors. The university i'n	5 Helped me actually get an A in a fore	# 06 16, 2011
35	I bought this product a little over a year ago. Now that I am done with it, I wanted to sell it. I put it on ebay and Craigslist. They had their lawyers contact the sites and have them remove the listing	1 DO NOT BUY IF YOU EVER WANT TO	# 09 21, 2012

Figure A.1: Clothing, Shoes and Jewellery Dataset

Figure A.1 shows some reviews of Clothing, Shoes and Jewellery dataset that is publicly available benchmark corpus in json format.

A.1.2 Home and Kitchen Reviews Labelled Dataset



id	reviewText	summary	reviewTime
1	reviewText	0 summary	ur reviewTime
2	My daughter wanted this book and the price on Amazon was the best. She has already tried one recipe a day after receiving the book. She seems happy with it.	5 Best Price	# 10 19, 2013
3	I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.	5 Zoku	# 06 18, 2014
4	There is no shortage of pop recipes available for free on the web, but I purchased the "Zoku Quick Pops" book, because Zoku has some good recipes for fruit pops on its blog. I was ho	4 Excels at Sweet Dessert Pops, but Falls Short on the	# 05 5, 2013
5	This book is a must have if you get a Zoku (which I also highly recommend). There is a large variety of recipes from simple, to fancy. The flavor combos are creative, and it gives you id	5 Creative Combos	# 08 4, 2011
6	This cookbook is great. I have really enjoyed reviewing all of the recipes. Sure - some are more difficult than others, but there is a nice variety in this book. For me it was totally wort	4 A must own if you own the Zoku maker...	# 06 7, 2014
7	If you have a Zoku Quick Pop maker (or two...) I have the original one-slot and then I bought the double-slot here at Amazon, I recommend you get this creative little book, which will	5 Love it	# 06 7, 2014
8	This book is so beautifully illustrated and easy to follow. It helps you to understand the Zoku better. It might have been even better if they made the recipe units for different sized mo	5 Beautifully illustrated recipe and instruction book fo	# 07 2, 2012
9	I bought this when I bought the pop maker. I think the recipes are fine, but there really isn't much in the book I wouldn't have figured out myself. A couple of good creamy recipes that	2 Okay but you can figure it out yourself without the b	# 12 8, 2011
10	The Zoku accessories are pricey, but the recipes have the exact amount you need to make the pops. I have made it through 15 varieties so far, and I like them all.	5 If you have the Zoku maker, buy this!	# 09 22, 2011
11	High quality book. Clear and helpful information, and then some superb recipe suggestions clearly outlined. I can highly recommend this product.	5 Satisfaction opinion	# 06 1, 2014
12	The recipes in this books are tedious to make. I found that easier and simpler recipes can be found online. I had no problem with the recipes online working in the Zoku maker.	4 I liked this book, but.....	# 12 26, 2012
13	This beautifully illustrated book featuring ten colorful, glittery butterflies is a favorite at our house. The nana-babies like to count the butterflies as they flit off! The book is also good f	5 A sweet book for babies and toddlers	# 07 27, 2009
14	I got this as a gift for my niece and she loved it. It's a super cute book and it'll be getting it for any other nieces that may come along.	5 Super cute niece present.	# 06 7, 2012
15	This is the story of a group of butterflies looking for a place to sleep. Each one is a different color, and each pair of facing pages (color-coordinated, of course!) tells the story of wher	5 My little one absolutely loved this book	# 04 6, 2012
16	This book is absolutely beautiful. A brilliant bedtime story with vibrant colors and gorgeous illustrations. Each page teaches a different color and the illustrations boldly burst with that	5 My Daughter's New Favorite	# 02 21, 2003
17	My daughter 1 year old really likes this book. I bought it because of its similarity to the 10 little ladybugs book. It has the same sort of colorful illustrations and I like it better as the	5 My daughter loves it	# 04 22, 2004
18	For a novice cake decorator who got talked into making her son's wedding cake, Wilton makes lovely cake decorating accessories to help with every step. This book gave me so many	5 Wilton Rocks!	# 05 4, 2014
19	I should not have bought it because it comes free from my cake decorating class. Oh well, at least I gave the other copy to my sister.	5 I should not have bought it.	# 06 7, 2013
20	had no idea about decorating cakes, by this booklet, now I can say that following the instructions and have the will and patience plus practicing one day I won't be ashamed of my cakes	5 follow the instructions	# 05 27, 2012
21	excellent book, shipped on time, packaged nicely, helpful tips in the book, good price. Very good experience, very happy with the product and vendor	5 book	# 06 12, 2012
22	I loved it, don't use it much but when I had it's been useful. I recommend it to all who wants to learn how to decorate cakes	5 great!	# 12 26, 2012
23	So I actually bought this thinking it was the entire Lesson 2 kit. However, I've found through experience that having the book without having to take the class is a good thing. I've taken	3 Good for what it's worth	# 08 16, 2012
24	helpful information on decorating basics. Good book for beginners who need step by step instructions and guidance. Good for all age groups.	4 This is a good book for anyone interested in cake de	# 12 13, 2013
25	Got this for my girlfriend who runs a baking business and she said it was handy to have for a reference book. She said it is handy to have when thinking of decorating ideas.	5 Handy to Have	# 06 8, 2013
26	I'm hoping that I can learn how to make gum paste flowers to perfection for my daughters wedding, and maybe a few other cakes.	5 learning	# 05 21, 2014

Figure A.2: Some reviews of Home and Kitchen Reviews Dataset

Figure A.2 shows home and kitchen appliance reviews with the reviewText, summary and reviewTime columns.

A.1.3 Flipkart Product Reviews Labeled Dataset with Sentiment

Summary	Review Text	Sentiment
1	5 sound bar is very good and sound quality is awesome size is perfect buttons are excellent quality sound bar material is very nice battery backup is awesome i personally like this product	positive
2	5 amazing quality speakersound is very loud and clear and decent bassspeaker has a very premium build qualitybattery backup is also good must buy it	positive
3	5 using with my tv and build quality is solid looks are elegant feels like premium aux bluetooth and memory are working perfectly fine sound is on a decent side no too much bass good buy	positive
4	5 first impression is its smaller than expected with good build quality the sound is top notch with adequate bassare are using it since 2 weeks bluetooth connectivity is good sound quality very good with extra base feat	positive
5	5 as expected a powerful mini soundbar! have been using blaupunkt 16w soundbar for a while and on the very first day of sale i purchased this product!fabulous2 21ch works perfect3 the connectivity range is really fc	positive
6	5 well the speaker i think this is one of the best speaker at this price range amazing sound quality is very nice and clear and the bass is extraordinary very good bluetooth speaker i love this speaker	positive
7	5 the speaker is excellent in soundbuild quality and unbelievable features within the price range ordered another speaker as wellive used many speaker of this price range but this one has really good sound quality bed	positive
8	5 amazing sound bar mivi very good quality sound and bass very rich battery backup very good and build quality good	positive
9	5 it came out very good than the expectedbass is good battery backup is best we almost use it 335 hrs everyday still its battery comes down to 80 only loved the product	positive
10	5 the packing was decent but what comes inside was an absolute boom bass can easily beat any speaker in the budget category of rs 20000 bass superb mid and low could have been better overall great value for money	positive
11	4 good product and sound quality very clear and good i am so happy to purchase it overall it is a good product as compared to its pricemust buy	positive
12	5 a very good speaker totally satisfied for my usage the bluetooth connectivity is very good and experienced no lags the bass is ok the output is sufficient for a good size room	positive
13	5 best sound quality easy to use happy with the product book now and enjoy the music with mivi fort s16 best and budget friendly product i just love mivi brandthanks flipkart	positive
14	5 optimum performance in this price range clear sound quality and good bass i have been using it for my mobile and even for my tv serves my purpose i would recommend buying this for a value for money experience	positive
15	5 a beast in its category! was looking for a speaker under 2000 and i came across this absolute beast and got it in offer for 1299finding another 16 watt speaker this good is going to be hardmusic plays well and bass is	positive
16	5 indeed the mivi speaker is made in india proclaimedit justifies the product with quite good in terms of built quality good sound output and if not thumping bass a decent bass is enjoyableoverall conclusion good built	positive
17	5 love this soundbar sound quality is awesome and the bass is fabulous the fabric design gives amazing finishone of the best speaker in this price range	positive
18	5 gifted it to my husband he majorly uses it while playing games and he totally loved it the sound quality battery back up and the sick design is too good at this price totally worth it	positive
19	5 value for money i already have sound but its bass is very high and sound clarity is also good when i compared to other sound wich have high price! like and love this sound bar this sound built quality also good when i	positive
20	5 im using this product for my tv as well as a perfect alternative to my small bluetooth speaker its very handy and the sound quality is much better than other bluetooth small speakers	positive
21	5 i was not fan of mivi speakers but after i purchased this i am totally happy with product and sound quality in this price segment bass is also good also happy with battery backupsound is too good and the c port is plu	positive
22	5 it has a great sound with amazing bass quality sounds neat and clear battery backup is also unbeatable overall loved the product very much	positive
23	5 great buy best in this money segment love the sound quality and the bass overall must buy and also love the made in india tag over the packagingound quality 55battery 45bass 45	positive
24	5 superb mivi bluetooth speaker with very good sound qualitybass is very goodbattery backup is awesome ntonstop 56hrscan be carry easily not so heavy personally liked itmust buy	positive
25	5 blast by mivi and a must have! am using it for some time now and i have to say the soundbar has amazing sound quality and deep bassgreat battery backup with reliable bluetooth connectivitysound is very loud and	positive
26	5 soundbar is amazing connectivity is very fast it gives proper sound delivery and perfect bass performance getting addicted to listening songs on this beautiful product thank you mivi	positive
27	5 excellent product supports multiple channels aux bluetooth usb long lasting battery with great sound output and the best part it is made in india highly recommended best product under budget	positive
28	5 amazing sound and deep bass great build quality speaker build design is also very good battery backup and bluetooth connectivity is also very good best soundbar at this price range	positive
29	5 i had purchased mivi green speakerwhen i got this product and used it i find extremely worth it! has a good voice clarity and its quite light this speaker is very good if you have a modest usageoverall i would say that	positive
30	5 i had purchased mivi green speakerwhen i got this product and used it i find extremely worth it! has a good voice clarity and its quite light this speaker is very good if you have a modest usageoverall i would say that	positive
31	5 i had purchased mivi green speakerwhen i got this product and used it i find extremely worth it! has a good voice clarity and its quite light this speaker is very good if you have a modest usageoverall i would say that	positive

Figure A.3: Flipkart Multiple Product Reviews Labeled Dataset

Figure A.3 shows the reviews of multiple categories of products with the columns product Id, Product Name and review and summary column.

A.1.4 Amazon Food Reviews Summarization Dataset

Produ:Usi:Profil:Hel:Score:Tim:Summary	Text
1B001A3:delmi: 1 5 ## Good Quality Dog Food	I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells
1B008A1:dll:pa 0 1 ## Not as Advertised	Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "
1B000AB:Natal 1 4 ## "Delight" says it all	This is a confection that has been around a few centuries. It is a light, pillowy citrus gelatin with nuts - in this case Filberts. And it is cut into tiny squares and then liberally coated with
1B000A3:Karl: 3 2 ## Cough Medicine	If you are looking for the secret ingredient in Robitussin I believe I have found it. I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry so
1B006A1:Mich: 0 5 ## Great taffy	Great taffy at a great price. There was a wide assortment of yummy taffy. Delivery was very quick. If your a taffy lover, this is a deal.
1B006AD:Twoa: 0 4 ## Nice Taffy	I got a wild hair for taffy and ordered this five pound bag. The taffy was all very enjoyable with many flavors: watermelon, root beer, melon, peppermint, grape, etc. My only complai
1B006A3:Pame: 0 5 ## Wonderful, tasty taffy	This taffy is so good. It is very soft and chewy. The flavors are amazing. I would definitely recommend you buying it. Very satisfying!!
1B000A1:R.Jar: 1 5 ## Yay Barley	Right now I'm mostly just sprouting this so my cats can eat the grass. They love it. I rotate it around with Wheatgrass and Rye too
1B001A2:Carol: 0 5 ## Healthy Dog Food	This is a very healthy dog food. Good for their digestion. Also good for small puppies. My dog eats her required amount at every feeding.
1B000A3:LT: 1 1 ## My Cats Are Not Fans of the Ne	My cats have been happily eating Felidae Platinum for more than two years. I just got a new bag and the shape of the food is different. They tried the new food when I first put it in t
1B001A1:Brian: 4 5 ## Lots of twizzlers, just what you	My daughter loves twizzlers and this shipment of six pounds really hit the spot. It's exactly what you would expect...six packages of strawberry twizzlers.

Figure A.4: Multiple Product Reviews Labeled Dataset on Hugging Face

Figure A.4 presents the hugging face multiple product reviews labeled dataset in multilingual but in english and dutch language.

A.1.5 Some Summaries with Evaluation Parameter's Scores

Table A.1: Summaries with Evaluation Parameter Scores

Summary	Reference	Feature, Product, Senti- ment	Rouge1, Rouge2, RougeL	Coherence, Informative, Relevance, Concise, Keyword, Fluency, Bi- gram_diversity, Human_Score
i order rayon ma- terial but i received cotton fabric . it is not actually rayon it is a mix fabric . fab- ric quality not good and product please return	not same product different and quality is not good received a defective prod- uct fabric is not same as shown. i order rayon material but i received cotton fabric. fabric quality not good. qual- ity not good and product please return material quality not good fabric is not same as shown. fabric quality is not good. product fabric is not good. it is not actually rayon it is a mix fabric	fabric, adult kurti, negative	0.888888884, 0.676056333, 0.888888884,	0.108070687, 0.807692308, 0.666666667, 0.959183673, 0.8 0.807692308, 0.420289855, 0.793103448,

Continued on next page

Table A.1 – continued from previous page

[illegible]

Continued on next page

Table A.1 – continued from previous page

Summary	Reference	Feature, Product, Senti- ment	Rouge1, Rouge2, RougeL	Coherence, tive,Relevance, cise,Keyword,Fluency, gram_diversity,Human_Score	Informa- Con- Bi-
the kids dress has very nice quality. the children's dress has super quality. The dress has size very tight quality issue. the dress has good quality but size of the neck is fit. quality is good as per the price . The dress is very nice and super	the kids dress has super quality. the kids dress has very nice quality. the kids dress has superb quality. the kids dress has good quality very nice. the kids dress has super quality. the kids dress has super quality. the kids dress has size very tight quality issue. the kids dress has good quality but size of the neck is fit. quality is good. the kids dress has super quality. the kids dress has super good quality. the kids dress has nice and super quality. quality is good. quality super. best quality. the kids dress has good quality on this price. very good better quality nice. quality is good as per the price	quality kids dress positive	0.857142852, 0.709677414, 0.857142852	0.298323169, 0.814814815, 0.636363636, 0.87, 0.968764147, 0.8	0.814814815, 0.424778761,

Continued on next page

Table A.1 – continued from previous page

Summary	Reference	Feature, Product, Senti- ment	Rouge1, Rouge2, RougeL	Coherence, tive,Relevance, cise,Keyword,Fluency, gram_diversity,Human_Score	Informa- Con- Bi-
the kids dress has good quality for cloth. quality is very good. quality super. quality . quality is really good. Quality is good for cloth . Quality super. Quality super . Quality for cloth quality is good. the kids . quality	quality super. the kids dress has good quality for cloth. the kids dress has very good quality. the kids dress has good quality. the kids dress has good quality. the kids dress has good qual- ity. quality is very good. quality is very good. the kids dress has super quality. the kids dress has nice qual- ity. the kids dress has good quality su- per. the kids dress has good quality. cloth quality is good. the kids dress has good quality. the kids dress has su- per quality. quality is really good. the kids dress has super quality. the kids dress has good quality	quality kids dress positive	0.923076918, 0.692307687, 0.923076918, 0.846153841	0.285318545, 0.928571429, 0.41, 0.846153846, 0.658536585, 0.9525,, 7.4	0.928571429,

Continued on next page

Table A.1 – continued from previous page

Summary	Reference	Feature, Product, Senti- ment	Rouge1, Rouge2, RougeL	Coherence, tive,Relevance, cise,Keyword,Fluency, gram_diversity,Human_Score	Informa- Con- Bi-
Quality of the clothe can be more good. quality is good but not in correct size. pattern wrong and quality poor very bad product. the kids dress has too good quality is super. the quality is not good but it is very poor not aspect quality very	quality is good but not in correct size. the kids dress has same quality. the kids dress has received small size so i returned it other wise quality super. quality is very poor not aspect quality very bed. quality is good but it is lose. quality is good but size is not correct. the kids dress has too good quality is super. the quality of the clothe can be more good. the kids dress has good quality but not proper size. pattern wrong and quality poor very bad product. quality poor. quality is not good. the kids dress has average quality. the kids dress has nice quality but size not good	quality, kids dress, negative	0.7999999995, 0.672268903, 0.7999999995	0.102826825, 0.69047619, 0.516129032, 0.974480151, 0.7	0.69047619, 0.423423423, 0.872340426,

Continued on next page

Table A.1 – continued from previous page

Summary	Reference	Feature, Product, Sentiment	Rouge1, Rouge2, RougeL	Coherence, tive,Relevance, cise,Keyword,Fluency, gram_diversity,Human_Score	Informa- Con- Bi-
Dress looking too good and cloth also better. Dress . has very beautiful cloth. the kids dress has super cloth. worth of money . Dress is very good for this price go for it. cloth quality is good. lining cloth is missing if it is missing .	worth of money super cloth. cloth quality is superb. dress looking too good and cloth also better. nice dress cloth quality super. cloth quality is good. lining cloth is missing if it is added then it will be in good quality. cloth quality is good for this price go for it. the kids dress has good quality nice cloth. the kids dress has excellent cloth. the kids dress has very beautiful cloth. the kids dress has super cloth. the kids dress has nice cloth. cloth is very good. the kids dress has good cloth. the kids dress has nice cloth. the kids dress has super cloth. good quality for cloth. cloth quality is good	cloth kidsdress positive	0.843749995, 0.64220183, 0.843749995	0.092346229, 0.777777778, 0.666666667, 0.82, 0.8	0.777777778, 0.438596491, 0.977092878,

Continued on next page

Table A.1 – continued from previous page

Summary	Reference	Feature, Product, Senti- ment	Rouge1, Rouge2, RougeL	Coherence, tive,Relevance, cise,Keyword,Fluency, gram_diversity,Human_Score	Informa- Con- Bi-
the kids kurti has poor quality of fabric as rate. after one wash quality is very bad. quality of product is not good. printed is going after wash material quality not good. the kids kurti has poor quality. printed is going after wash material quality not good. the kids have low quality cloth. the	after one wash quality is very bad. the kids kurti has poor quality of fabric as rate. leneth issue quality issue. quality of product is not good. printed is going after wash material quality not good. the kids kurti has poor quality. quality not that good. the kids kurti has very thin quality. the kids kurti has low quality. the kids kurti has low quality cloth. quality very bad. quality is low. quality not good. quality issue. the kids kurti has poor quality material. quality low. the kids kurti has average quality. the kids kurti has low quality	quality, kids kurti, negative	0.888888884, 0.696629209, 0.888888884	0.202742988, 0.833333333, 0.408163265, 0.8, 0.971729126, , 0.8	0.83333333333, 0.64,

Continued on next page

Table A.1 – continued from previous page

Summary	Reference	Feature, Product, Senti- ment	Rouge1, Rouge2, RougeL	Coherence, tive,Relevance, cise,Keyword,Fluency, gram_diversity,Human_Score	Informa- Con- Bi-
The saree has net is very bad material quality not good. material quality is not good . color is not same as shown safety and hygiene is not safe and safety is not shown . the saree is a defective product fabric fabric fabric is not	material quality not good. material quality not good. material quality not good. material quality not good. the saree has net is very bad material quality not good. material quality not good. the saree has cos-tumer not happy material quality not good. the saree has not good material quality not good. material quality not good. material quality not good. material quality not good. material quality not good color is not same as shown safety and hygiene. material quality not good received a defective product fabric is not same as shown. not like customer material quality not good	material, saree, negative,	0.862745093, 0.657534242, 0.862745093,	0.309339824, 0.821428571, 0.442307692, 0.891304348, 0.964938272, 6.4	0.821428571, 0.65, 0.821428571,

Continued on next page

Table A.1 – continued from previous page

Summary	Reference	Feature, Product, Sentiment	Rouge1, Rouge2, RougeL	Coherence, tive,Relevance, cise,Keyword,Fluency, gram_diversity,Human_Score	Informa- Con- Bi-
the saree has material quality not good color is not same as shown. color material quality not good color is not same as shown. the saree has colour and fabric quality is not good color is not same as shown	the saree has material quality not good color is not same as shown. color material quality not good color is not same as shown. the saree has colour and fabric quality is not good color is not same as shown	color, saree, negative	0.8888888, 0.823529407, 0.888888884	0.268804976,0.8125, 0.8125, 0.375, 0.7,0.86666666,0.928571429, 7.8	
saree quality is very bad only plain not design it is a defective item . the saree has material quality not good design is not matching fabric is not same as shown.	design is not matching. design is not matching. design is not matching. design is not matching. design are not same design is not matching. the saree has material quality not good design is not matching fabric is not same as shown. design is not matching. the saree has material quality not good design is not matching fabric is not same as shown. saree quality is very bad only plain not design it is a defective item	design, saree, negative	0.977777,0.87499, 0.977777	0.9583333,0.958333,0.4342105, 0.866666, 0.878787, 0.966796875, 8.2	

List of Publications

Conference Proceedings

- [P1] Sonia Rani, and Tarandeep Singh Walia, “**An experimental study of text pre-processing techniques on user reviews**”, in proceedings of *Recent Advances in Computing Sciences*, 2022.
- [P2] Sonia Rani, and Tarandeep Singh Walia, “**An Exploratory Analysis of Feature Extraction Techniques using NLP and Text Embedding Methods**”, in proceedings of *International Conference On Networks, Intelligence and Computing*, 2023.

Journal Proceedings

- [P3] Sonia Rani, and Tarandeep Singh Walia, “**Automatic summarisation of product reviews using natural language processing and machine learning methods: a literature review**”, *International Journal of Data Mining and Bioinformatics*, 2022.
- [P4] Sonia Rani, and Tarandeep Singh Walia, “**An experimental study of topic modelling techniques on Hinglish code-mixed product reviews**”, *International Journal of Business Innovation and Research*, 2024.

Under Review

- [P5] Sonia Rani, and Tarandeep Singh Walia, “**Aspect-based Product Reviews Summarization Using Rule-based and Transfer Learning Methods: A Hybrid Approach**”
- [P6] Sonia Rani, and Tarandeep Singh Walia, Radhika Mamidi ”**An Analysis of Hinglish Code-mixed Product Reviews Using Labelled and Unlabelled Topic Modelling Techniques**”

Future work

- [P6] Sonia Rani, and Tarandeep Singh Walia, Radhika Mamidi “**Feature-based Extractive Summarization of Hinglish Code mixed Reviews**”

List of Conferences and Workshops

Conference Proceedings

- [C1] Paper presented at ICONIC conference organised by LPU on 26 Apr 2023 and got the Best Paper Presentation Award, 2022.**
- [C2] Presented Poster at ARCS held by ACM India at Bhopal on 9-11 Feb 2023.**
- [C3] Attended CODS-COMAD 2022 The ACM India Joint virtual conference on Data Science and Management organized by ACM from 7-10 Jan 2022.**
- [C4] Presented paper at RACS conference organised by LPU on 27 Nov 2023.**
- [C5] Attended CODS-COMAD 2023 Virtual 6th Joint International Conference on Data Science and Management of Data organized by IIT Bombay, on 4-7 January 2023.**

Workshops and Short Term Courses

- [W1] Participated in Short Term Course on “Scientific Writing Using Typesetting Software Latex”, organized by LPU from 17th Oct-1st Nov 2020.**
- [W2] Participated in Online Short Term Course on “National Workshop on Statistical Analysis using SPSS”, organized by LPU from 22-26 Jun,2020.**
- [W3] Participated in Online Short Term Course on “Pandemic Analytics using Machine Learning with Python” organized by LPU from 1st-14th July, 2020.**
- [W4] Participated in Short Term Course on “Scientific Writing and Publication” organized by LPU from 31st Jan-5th Feb.**
- [W5] Participated in Workshop on “Unlocking Innovation through Intellectual Property Rights: Drafting Filling and Protection organized by LPU from 27 Nov-4 Dec 2024.**
- [W6] Completed Skill Development Course on “AI Odyssey: Mastering Deep Learning and Generative AI” organized by LPU from 5 Jun-21 July 2024.**
- [W7] Participated in Workshop on “Researcher’s Toolkit: A Short Term Course on LaTeX, Overleaf, Mendeley and Seaborn” for Sustainable Research organized by LPU from 26 Feb- 2 Mar 2024.**

- [W8] Participated in One Week Workshop on Mapping of Scholarly Output and its Impact using Open Source Software organised by Central Library, Indian Institute of Technology Delhi from 22-26 May 2023.**
- [W9] Participated in Short Term Course on “Hands-on Machine Learning using Python” organised by LPU from 18-23 Dec 2023.**
- [W10] Participated in the Workshop on “Art of Writing a Research Paper” organised by Chitkara University from May 25-29, 2020**
- [W11] Participated in the “Artificial Intelligence Summer School” organised by CAI IIT Delhi on 17th-18th May 2024.**
- [W12] Attended the conference “COMPUTE 2020” organised by ACM India Council on 9-12 December 2020.**
- [W13] Attended “ACM India Winter School on Natural Language Processing” hosted by IRNLP Lab, DA-IICT from 4th-15th January 2021.**
- [W14] Participated in Workshop on “The AI Frontier: Balancing Security and Explainability”, from 18-19 March 2023 organized by Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab.**
- [W15] Attended “One Week Faculty Development Programme on ”Cloud Computing and its Applications: Opportunities and Challenges” from 16-20 December 2022 organized by Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab.**
- [W16] Attended the “One Week Online Faculty Development Programme on ”Big Data and Analytics” from 3-7 April 2023, organized by Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab.**
- [W17] Completed Internship under Prof. Radhika Mamidi at IIT Hyderabad on “Anveshan Setu Fellowship Program” from 22 Apr-20 May 2024.**