CLASSIFICATION AND IMPROVEMENT OF E-LEARNING SYSTEMS WITH SENTIMENTAL ANALYSIS OF USER'S REVIEW

Dissertation submitted in fulfilment of the requirements for the Degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

KAZI MOSTAFIZUR RAHMAN

11502737

Supervisor

Mr. ADITYA KHAMPARIA

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab (India)

April 2017

PAC FORM

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program : P172::M.Tech. (Computer Science and Engineering) [Full Time]

COURSE CODE : CSE546 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGD0278

Supervisor Name : Aditya Khamparia **UID :** 17862 **Designation :** Assistant Professor

Qualification : ________________________ **Research Experience :** ________________________

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Kazi Mostafizur Rahman	11502737	2015	K1518	8876409761

SPECIALIZATION AREA : Intelligent Systems **Supervisor Signature:** ___________________

PROPOSED TOPIC : Classification and Improvement of E-learning systems with sentimental analysis of user's review

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	**Parameter**	**Rating (out of 10)**
1	Project Novelty: Potential of the project to create new knowledge	7.00
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	6.75
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	7.50
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	7.25
5	Social Applicability: Project work intends to solve a practical problem.	6.75
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	7.50

PAC Committee Members		
PAC Member 1 Name: Prateek Agrawal	UID: 13714	Recommended (Y/N): NA
PAC Member 2 Name: Pushpendra Kumar Pateriya	UID: 14623	Recommended (Y/N): Yes
PAC Member 3 Name: Deepak Prashar	UID: 13897	Recommended (Y/N): Yes
PAC Member 4 Name: Kewal Krishan	UID: 11179	Recommended (Y/N): Yes
PAC Member 5 Name: Anupinder Singh	UID: 19385	Recommended (Y/N): NA
DAA Nominee Name: Kanwar Preet Singh	UID: 15367	Recommended (Y/N): Yes

Final Topic Approved by PAC: **Classification and Improvement of E-learning systems with sentimental analysis of user's review**

Overall Remarks: Approved

PAC CHAIRPERSON Name: 11024::Amandeep Nagpal **Approval Date:** 05 Mar 2017

ABSTRACT

In the world of internet, it has a large number of data in form of text, images, strikers etc. which is also be called as feedbacks/reviews created by users to share their expressions or knowledge. Users like to express their feelings as it is in free format users provided the information in an unstructured form (feedbacks/reviews). Users share their knowledge, point of views, comments, feedbacks, etc. in the form of text. All those data may be in different format like positive, negative or neutral, sometime it may be in a single word or a single sentence or in document form. For the users to get a better future from the past experience held by other users through their valuable feedbacks/reviews in the form of text in document. Here it is intended to gain a better scope in E-learning and planned to extract knowledge from E-Learning sites. There are few techniques which has be measured to provide better classifier like Classification-Support Vector Machine (SVM), Naïve Bayes (NB) and some other techniques like J48 in decision tree. Here focusing on the different techniques we are trying to conclude a better result (accuracy, specificity, sensitivity etc.) using “WEKA”. These techniques will be tested so that user can get better knowledge in E-learning which is better technique and measure.

Keywords: Sentiment Analysis, Opinion Mining, Supervised Learning, Unsupervised Learning, SVM.

CERTIFICATE

This is to certify that "**Kazi Mostafizur Rahman**" has completed Dissertation-I titled "**Classification and Improvement of E-learning Systems with Sentimental Analysis of user's review**" under my guidance and supervision. To the best of my knowledge, the present work is the result of his original investigation and study. No part of the dissertation has ever been submitted for any other degree or diploma. The dissertation is fit for the submission and partial fulfillment of the conditions for the award of M.Tech Computer Science and Engineering.

Date: ______________ **Signature of supervisor: ______________**

Name: Aditya Khamparia

ACKNOWLEDGEMENT

No research can be done in an isolated environment and this, certainly was not an exception. It was a concerted effort of all my friends, family and above all me. I would like to thank I would like to express the deepest appreciation to my mentor, **Mr. Aditya Khamparia**, who has the attitude and the substance of a genius. He always helped to clear all doubts generated during different parts of this literature review and formulation of statement for my research work. His guidance is also a motivation for me to do work on time. His guidance was crucial for formulation of problem statement and carry forward approach.

DECLARATION

I hereby declare that the research work reported in the dissertation entitled **"CLASSIFICATION AND IMPROVEMENT OF E-LEARNING SYSTEMS WITH SENTIMENTAL ANALYSIS OF USER'S REVIEW"** in partial fulfilment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor **Mr. Aditya Khamparia**, I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

Signature of Candidate

Kazi Mostafizur Rahman

R.No-11502737

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation entitled "**CLASSIFICATION AND IMPROVEMENT OF E-LEARNING SYSTEMS WITH SENTIMENTAL ANALYSIS OF USER'S REVIEW"**, submitted by **Kazi Mostafizur Rahman** at **Lovely Professional University, Phagwara, India** is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

Mr. Aditya Khamparia

Counter Signed by:

1) **Concerned HOD:**

HoD's Signature: ________________

HoD Name: ____________________

Date: ____________________

2) **Neutral Examiners:**

External Examiner

Signature: _______________

Name: ___________________

Affiliation: ______________

Date: ____________________

Internal Examiner

Signature: _______________

Name: ___________________

Date: ____________________

TABLE OF CONTENT

TABLE OF FIGURES

LIST OF TABLES

CHECKLIST FOR DISSERTATION SUPERVISOR

Name: ___________________________ UID: _________

Domain: **M.Tech (CSE)** Registration No: **11502737**

Name of Student: **Kazi Mostafizur Rahman**

Title of Dissertation: **Classification and Improvement of E-learning Systems with Sentimental Analysis of user's review**

- ☐ Front pages are as per the format.
- ☐ Topic on the PAC form and title page are same.
- ☐ Front page numbers are in roman and for report, it is like 1, 2, 3…….
- ☐ TOC, List of Figures, etc. are matching with the actual page numbers in the report.
- ☐ Font, Font Size, Margins, line Spacing, Alignment, etc. are as per the guidelines.
- ☐ Color prints are used for images and implementation snapshots.
- ☐ Captions and citations are provided for all the figures, tables etc. and are numbered and center aligned.
- ☐ All the equations used in the report are numbered.
- ☐ Citations are provided for all the references.
- ☐ Objectives are clearly defined.
- ☐ Minimum total number of pages of report is 50.
- ☐ Minimum references in report are 30.

Here by, I declare that I had verified the above mentioned points in the final dissertation report.

Signature of Supervisor with UID

CHAPTER 1
INTRODUCTION

1.1 Motivation

In the world of internet because of large number of data stored regularly due to various types of communication process. It has provided a media through which users can share or discuss or exchanges information among themselves from the source opinion/sentiment can be extracted that users are sharing their point of view. The Web has important, large and unstructured information about opinion. User's opinion can be important when it is about to make any decision or selection from different various products.[1]

It is a Natural language processing and information extraction task to identify in the way to get about positive, negative or neutral comments/statements in the form of text.[2] Classifying a natural language i.e., any kind of review or opinion from blog, social site, e-commerce etc. not as per its topic but how the opinion has expressed in it. It is catching an attention that sentiment classification is to find the positive and negative review. [3]

Few words that are used mostly in any review or comment which represent positive or negative sentiment. For example –(Figure 1) Positive Sentiment(+ve) like awesome, good, mind blowing, fantastic, outstanding etc. and for negative sentiment(-ve) like bad, worst, ugly, disgusting, pathetic.[4] But sometimes it also difficult to analysis when the amount of data it very large from the web, moreover the positive sentiment or negative sentiment is in the form of large context that is another difficult to analysis. And sometimes it is also in the form of document like in details document or paragraph form, for example a subtitle of a movie, conference, description of any products, detail summary of a book etc. which is large in document.

Figure 1:User's Opinion [48]

1.2 Sentiment Analysis

Mining of Data is a process of extraction knowledge from a huge amount of data from internet that is transformed into an easy to understandable form for the users.[5] This process of analyzing data from huge data is Sentiment Analysis, This is also known as “Opinion Mining”.[6]

The manner of computationally classifying and grouping sentiments expressed in a piece of writing, exclusively in order to define whether the user's point of view for a particular subject, topic, material, etc. is positive, negative, or neutral. Using this knowledge user can get future scope to lead a better change or fix goal for their decision that can also be called decision making support system. Various source from which sentiment can be gathered i.e social-media (FB, twitter, YouTube etc), blogs, educational site, e-commerce, news, movie(IMDB) etc.

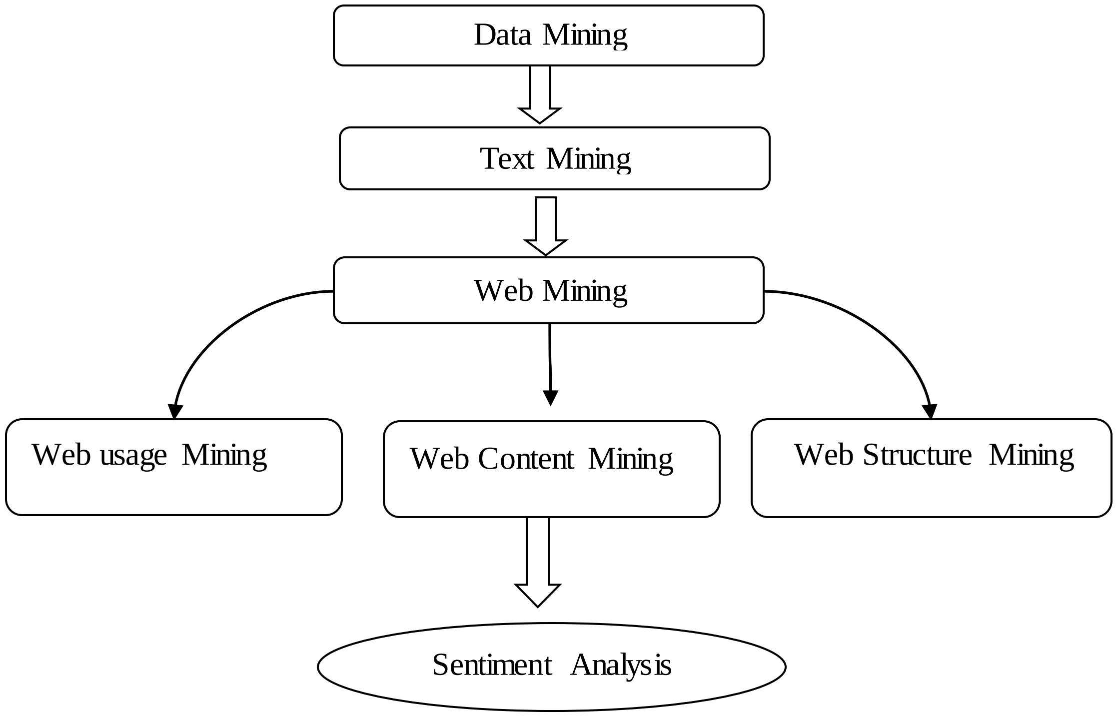


Figure 2: Data Mining Hierarchy

1.3 Types of Sentiment

Sentiment can be represent in different types[7][8]-

Direct – This sentiment can be expressed in direct view about any subject. e.g- The book is interesting and easy to understand

Indirect- This sentiment can be indirectly expressed the views about any subject.

e.g- I had a headache after reading the book which you gave me.

Comparative- This sentiment can be expressed in comparison about any subject with others.

e.g- Acer laptops are better than IBM laptop.

1.4 Different Kind of Sentiment

Sentiment Analysis can be classified into three ways Figure 3:

Positive Sentiment – If the popular number of positive sentiment is more than the number of negative sentiment word in review/comment in the form of document in web. For example the review like "The site looks good but the traveler service is not available. People like to visit that area is much more than any other site in that area.

Negative S*entiment* – If the popular number of negative sentiment is more than the number of positive sentiment word in review/comments presents in the form of document. For example the review like "In the movie acting was good but the story was not good. Many people didn't even feel watching it because graphics was not good".

Neutral Sentiment – If the popular number of positive sentiment is equal to the number of negative sentiment word in review in the form of document. For example the review like "I love the actor in the movie name "ABCXYZ" but i hate the performance of the actress and co-actress in that documentary".[9]

In this document the negative sentiment shows in italic bold and the positive sentiment shows in bold that makes easy to understand and identify the majority and perform calculation.

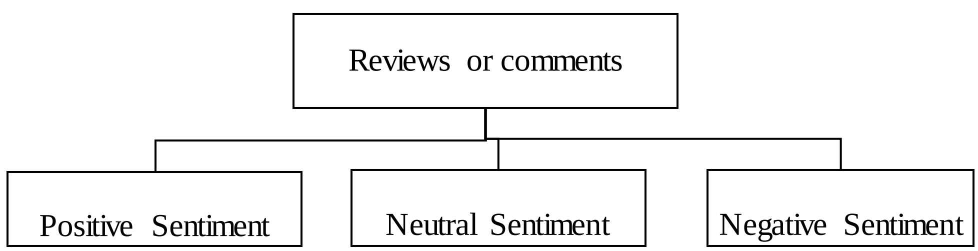


Figure 3: Types of Sentiment

1.5 Sentimental Analysis at Different Levels

Analysis of Sentiment that is used to classify input of users in the form document and which contains opinion of users. Opinion that can be expressed in positive, negative or neutral. Analysis of Sentiment can be performed by four levels Figure 4: [9][10]

- Word level: Classify reviews in word level as positive/negative or neutral which can be called as word-level sentiment classification
- Document level: Classify full statement as positive/negative or neutral which can be called as document-level sentiment classification.
- Sentence level: Classify opinions in the form of sentence as positive/negative or neutral which can be called as sentence-level sentiment classification.
- Aspect & Feature level: Classify both sentence & document as positive/negative and neutral according to the natural of opinion of that sentence & document which can be called as aspect-level sentiment classification.

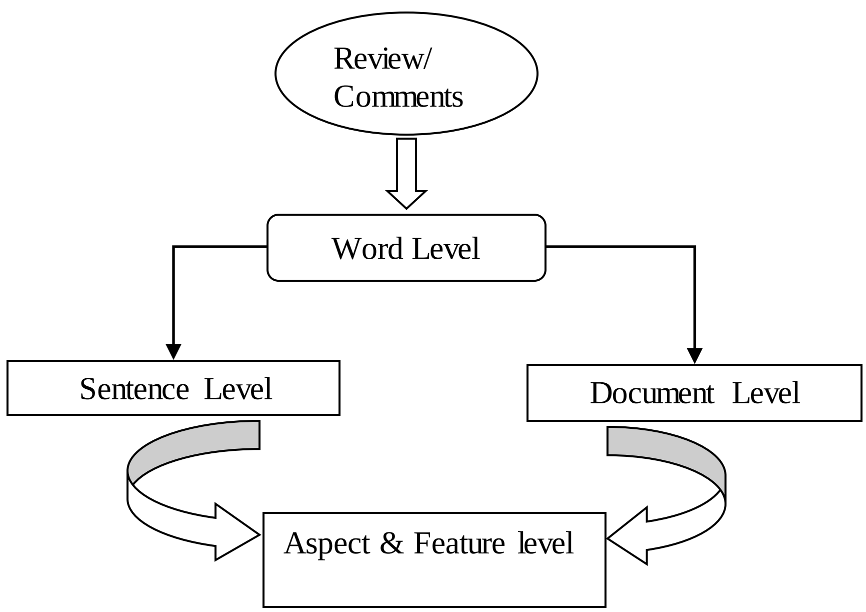


Figure 4: Sentimental Analysis at Different Levels

In Sentimental Analysis Document based classification in Figure 5 is done using unsupervised approach that find different opinions in it.[9] This concept extract the opinion from the document and these to users get an awareness to take decision depending on positive or negative result. Here we have an example of document-based Sentiment Analysis.

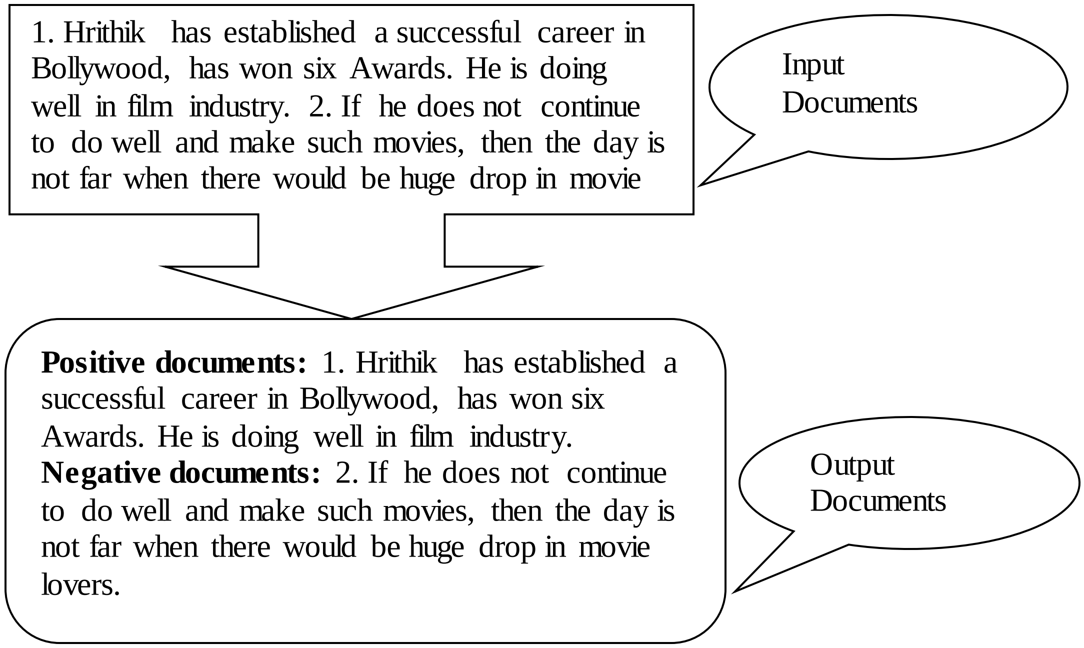


Figure 5: Example - Document based Sentiment Analysis

1.6 Shortcoming of Sentimental analysis

The shortcoming of sentiment analysis are as follows:[11][12][13]

a. Spelling Mistake
b. Special character/symbols/tokens like strickers
c. Regional language
d. Error in grammar
e. Abbreviation in short form like fine(fi9)
f. Fake reviews
g. Duplicate reviews
h. Diversity of contents
i. Spam

CHAPTER 2
REVIEW OF LITERATURE

Touhid Bhuiyan at al. [14] discussed in this paper about Sentiment Classification and Feature-based Opinion Mining where the classification can be done in different levels like sentence level or document level. In sentence level it is divided into two paper Corpus-based approaches (find co-occurrence patterns of words) & Dictionary-based approaches (synonyms and antonyms) both techniques are used to determine the sentiment of the word. Sentimental analysis has the probability to use from the singular level to hierarchical level, for example, organizations and government. Individuals and associations(organization) from a few spaces could be profited in different courses by utilizing the Opinion Mining procedures from online customer's opinion. In this paper it was evaluated the momentum explore work in the range of Opinion Mining. We have examined a few methodologies taken by the analysts to concentrate general sentiment from the unstructured content communicated as one's point of view, characterized and fundamentally assessed the current work. We firmly trust that this review will help to new analysts to uncover bleeding edge territory of enthusiasm for Opinion Mining.

George Stylios at al. [15] discussed in the paper that uses techniques like KNN, SVM, Naïve Bayes in the dataset of governmental decision . TPR (true positive rate) is more in SVM (67.74%) & less in Naïve Bayes(12.07%), TNR (true negative rate) is more in Naïve Bayes(99.37%) & less in KNN(96.06%), FPR(false positive rate) is more in KNN (3.905%) & less in Naïve Bayes(0.62%) ,FNR (false negative rate) is more in Naïve Bayes (87.96%) and less in SVM(32.25%).

The author concluded with a standout amongst the most essential issues for making e-Government compelling is to empower residents take an interest in the basic leadership handle. By means of our proposed approach we guarantee that national feelings and remarks are legitimately gotten by open bodies and that they are represented in consequent legislative activities and additionally we give both residents and governments with the way to viably connect with each other and effectively partake into com-mon activities from which both would profit. In spite of the fact that the work introduced in this part is still in early stages and just gives a general idea as for how sentiment mining systems can be successfully investigated over the span of e-Government and e-consideration approaches we trust that it will clear the ground for more activities in this regard. Actually, we are as of now chipping away at the fuse of extra societal angles in the feeling mining process and in addition towards the work of extra measurements that would assess the dependability of the subject remarks and suppositions on legislative choices. Another part of future work is depending on the mined and extremity clarified client assessments with a specific end goal to construct and prepare compelling expectation models that would have the capacity to rough the potential effect of arranged legislative choices on citizens‟ position. At long last, it is fascinating to apply our conclusion mining strategy towards a wide assortment of client feelings on administrative choices and distinguish the directions that intrigue subjects the most and along these lines offer them the foundation to associate with governmental bodies[15]

Ahmed Abbasi [3] mentioned in this paper that classifying a natural language i.e any kind of review or opinion from blog, social site, e-commerce etc., is not as per its topic but how the opinion has expressed in it. It is catching an attention that sentiment classification is to find the positive, negative and neutral review. He proposed an approach i.e. an intelligent "Feature Subsumption Hierarchy" that can also be called as FSH which integrates syntactic and semantic data. That de-scribe how unlikely, mixed feature set coupled with appropriate feature selection mechanisms can enhance and improve the performance of classification. Later feature selection like ("word n-gram"," information gain", "IFS", "Semantic IFS"," Syntactic IFS" etc.) are tested with best accuracy to "Digital Camera", "Automobiles" and "Movies".

Khairullah Khan et al. [16] in this paper the authors mentioned about users feeling or sentiment mining is an intriguing region of research due to its applications in different fields. Gathering sentiments of individuals about items and about social and political occasions and problems through the Web is winding up noticeably progressively well known each day. The suppositions of clients are useful for people in general and for partners when settling on specific choices. Supposition mining is an approach to recover data through web crawlers, web journals and informal communities. In light of the immense number of audits as unstructured content, it is difficult to outline the data physically. In like manner, proficient computational techniques are required for mining and compressing the audits from web documents. This review exhibits an orderly writing overview with respect to the computational systems, models and calculations for mining assessment segments from unstructured audits.

This review also misuses interpersonal organizations and web blogs, the most prevalent utilized sources for opinion retrieval, to look at conclusion portrayal, sentiment mining models, opinion segments, and related issues. Various computational models and semantic elements identified with sentiment mining, part investigation and one's point of view in target identification are systematically discussed.

Zhongwu Zhai et al. [17] discussed in this paper which it described that grouping feature expressions manual is maximum time consuming few. So here this author used few techniques and methods i.e Unsupervised [SHC, CHC, LDA, mLSA, k-means), Semi-supervised [LDA(L,H), DF-LDA(L,H), K-means(LC,H), EM-(LC,H)],Hard-constrained[Rand(LC,H), LDA(LC,H), DF-LDA(LC,H), K-means(LC,H),EM(LC,H)] ,Soft-constrained [LDA(LC,S), k-means(LC,S),SC-EM]used to check the result (Accuracy, Purity, Entropy)

Deepali Virmani et al. [18] discussed in this paper about an algorithm to find the positive ,negative and neutral documentation result, Frist of all it will check the sentiment word, then it will provide ranking based on positive or negative. If it is less than 5 then it will have ranked as negative or if it is more than 5 then it will be ranked as positive. If value is lesser than 2 then it is low and if value is greater than 8 is very high, ranking is mentioned from 0 to 10. Later the authors concluded as various case analyses are considered wherein instructors give comment about an understudy and a normal feeling worth is ascertained. The calculation contrasts each word and opinion and invalidation in the database. The calculation is actualized on the premise of score doled out to every estimation word in the database. The worked together feeling is assessed by analyzing instructor's comments word by word and afterward actualizing the calculation proposed. The assessed conclusion esteem for an understudy can be used while offering imprints to the under study. Recommendation might be given to an understudy as per the collaborated opinion value

Wei Wei [19] discussed in this paper about a research proposal of Ph.D on the research problem on opinion mining. The proposed work in the paper are Entity Related Opinion

Detection, Sentimental Analysis with Sentimental Ontology Tree, Multi-layer Neural Network Kernel for Sentiment Analysis with Sentiment Ontology Tree. The proposed research about issues will be separately tended to and exhibited in scholastic productions. In every production, we will show our proposed arrangements and comparatively consider and talk about them with regards to the current best in class approaches. Through this methodical research prepare, we go for profoundly understanding the studied issues and create critical arrangements individually and subsequently make great commitments to the exploration group of mining of opinions.

Richa Sharma et al. [9] discussed in this paper about the classification in different levels i.e in words, sentence, document and feature level in opinion mining. The authors used technique is unsupervised dictionary (WordNet) to find the sentiment word and synonym & antonym. Sentiment Mining assumes an important part in settling on a choice about item or services. Sentiment Mining has expansive application territories like Education in which assessment can be utilized to assess scholastics in light of suppositions communicated by understudies. Shopping, where sites like amazon.com enable clients to express their suppositions on their sites. Excitement where the general populations can without much of a stretch see the surveys of their most loved motion pictures and every day cleansers on the web. Advertising, Companies can now make investment funds on promoting costs by asking for surveys on their sites. Presently there is no compelling reason to direct reviews as organizations can now have every one of the information they require on the web. In this overview a few machine learning systems have been talked about and the related work has been finished by utilizing these strategies. However, still there are a portion of the difficulties that still to be settled like element recognizable proof, invalidation taking care of, many-sided quality in taking care of sentence or record and so on. Specialists have been completed to beat these difficulties.

Edison Marrese Taylor et al. [20] describe about the web sentiment mining in the paper which collects the data from that extract the information, analyzing and aggregate it. From that a point of view is observed in the form of opinion in majority what people actually feel for that particular subject. The author worked on twitter- “tweet” and the method is clustering using fuzzy logic.

Periakaruppan Sudhakaran et al. [21] in this paper the authors focused on limitation and problems in sentimental analysis. It is mentioned that it would be of direct or comparative sentiment where people can write their point of view in positive or negative individually. But sometimes users write the reviews in positive or negative individually and also in details moreover web pages have free format, so users can write in free format as well. Sentimental analysis finds the main words, intensifier, positive/negative words or neutral words in the comment in the form of sentence. And those each words are assigned a score by that score and calculating overall scores from that given data(reviews) gets a result which share as the opinion.

Alexander Pak et al. [22] in this paper authors mentioned about the social media's sentimental analysis and how those sentiment is Twitter as a Corpus for Sentiment Analysis and Mining reviews from it. And how the smiley has been moved from the reviews during mining the data.

David Osimo et al. [23] in this paper authors described about the limitations of sentiment analysis along with some future opportunities. Also shared some upcoming study for extended and small term which is helpful for society. The combination of increment in the volume of information accessible and more perplexing ideas to examine, as of late there has been a decline in enthusiasm on semantic-based application, and a move towards more noteworthy utilization of measurements and representation. Similarly, as some other logical teach, likewise computerized content examination is turning into an information concentrated science.

Poobana S et al. [24] in this paper the author discussed about the dataset that contains emotion , so to detect those emotion and remove during preprocessing using stop words removal and stemming. The author proposed that Senti-word Lexicon which words applying SVM to analyze the sentiment of the users. The sentiment analysis for sentence level is performed by naïve Bayesian classifier and perspective level sentiment analysis is for support vector machine. The user survey is breaking down and rank for a specific item. The audits are pre-processed to wipe out noise like stop words and stemming words are evacuated. The removed words are arranged into positive and negative in unigram utilizing machine learning guileless Bayesian classifier. We propose Machine Learning Based Senti-word Lexicon which depends on the sack of words created from applying Support Vector Machine to take in the critical Senti-word-as a notion word vocabulary. Our approach utilizes bigram and SVM grouping to break down the feeling of the clients. The bigram with high weight is considered in preparing to incorporate compound expressions like „not bad" in the yield pack of words.

Mubarak Himm et al.[8] in this paper the author discussed about the various techniques used to analysis the sentiment of customers review but along with that it has many challenges and limitation. It discussed about different types of sentiment, various level of reviews and approaches, also the class label (positive, negative, neutral). The author observed techniques are good in some manners but could get more better result as it still has many difficulties.

Ethan Zhang et al. [25] in this paper which patch important document from the set of topic/discussion. SentiWordNet lexicon & bigrams and trigrams are used as a technique for analyzing the sentiment of the user. here it collects discussion in the form of document is classified to examine of probability of the discussion contains opinions expressions. And with the confusion matrix it provided a result of accuracy is 61%.

S. Vasantharaj et al. [26] in this paper the author discussed about the sentiment analysis, describe all its techniques like supervised, unsupervised ,machine learning and CBR. And also mentioned few table with model, techniques, dataset and its respective result under various techniques and compared their result with each other which has a better result. Reviews like movie, online shopping blogs, student data etc , techniques like SVM, K-means, KNN, bayes etc and results like accuracy , precision etc.

Humera Shaziya et al. [27] in this paper the author discussed about the review of movie,according the author the text can be described in three ways i.e., supervised ,semi-supervised and unsupervised where different types of techniques can be used like SVM, Naïve Bayes,KNN etc , for evaluation it used cross-validation(folds) methods in Weka tools to avoid noise in training/testing data. It measures the result in accuracy, precision, recall and f-score using SVM and naïve bayes and in naïve 1702 classified correctly from 2000 (85%), in SVM 1691 classified correctly from 2000 (84%). To improve the result, recommend to use tokenizers for future work. At the end the authors added naïve bayes turns out to be superior to svm as for precision of the classifier. The model proposed in this paper is only an underlying stride towards the change in the methods for supposition examination. It merits investigating the abilities of the model for the audits dataset and amplifying the exploration utilizing cross breed procedures for slant examination. There is considerable scope for development in the creation and compelling preprocessing and include choice. The work can likewise be stretched out to enhance the outcomes utilizing different tokenization methods. Future looks into can be done to create better and quick models for higher request n-grams.

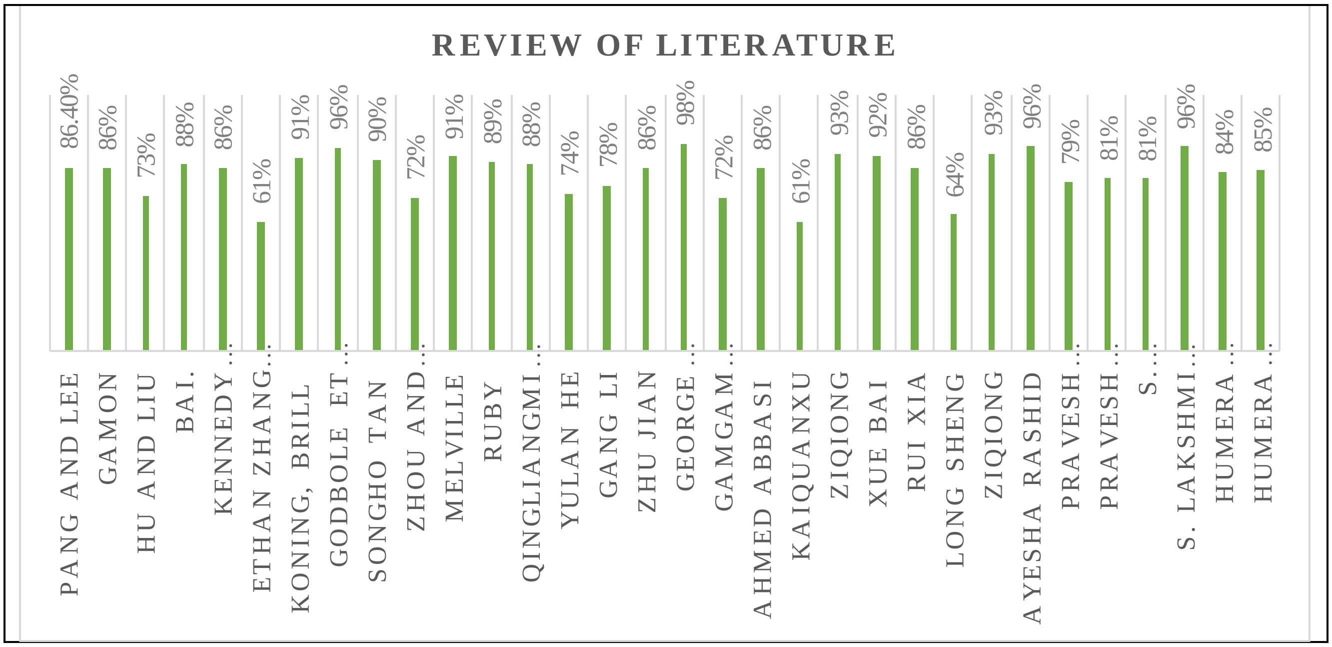
Gautami Tripati et al. [7] in this the author discussed about the huge growth of the data and to experience the knowledge how researcher is analyzing the data, extracting knowledge from that will help the society to make better decision like for business organization, government even for individuals it will be helpful. Collecting the data->processing the data ->features selection(POS). the author also mentioned about different application areas like government, business sector, movies, blogs etc., along with some challenges like spam, sarcastic reviews etc.,

G.Vinodhini et al. [28] the author has mentioned about the sentiment in the form of reviews by the users in large amount spread all over the network, discussed about SVM, machine learning it also provide few data source from which a better understanding level can be done of the product (review) like Blogs, review sites etc., Here the author has mentioned about other authors which dataset, techniques, performance they had obtained in their experiment from the year 2004-2011. Comparing all those results of different years with authors and check the better performance using various techniques and dataset.

Few more literature review from year (2004-2015) from different authors who have used different techniques like SVM, Naïve, CBR, Hybrid, NLP,J48, feature based selection etc., with different type of dataset like Movie review, car review, blogs, online site etc and it is observed with the performance. Also various tools are used like R, weka, matlab for the better performance and accurate result . After the study of many year it is found that among of the techniques SVM is the best techniques with 98% accuracy in 2010 by the author named George Stylios as in Table 1.

Table 1: Review of Literature 2004-2015 [29][30][28]

Yearwise	Authors	Datasets	Techniques	Results
2004	Pang and Lee	Movie Review	SVM	86.40%
2005	Gamon	Car Review	Naïve Bayes	86%
2005	Hu and Liu	Amazon Cnn.Net	Opinion Word extration	73%
2005	Bai.	Movie Review	Two-stage markov Blanket Classier	88%
2006	Kennedy and Inkpen	Online Site	SVM	86%
2006	Ethan Zhang et al.	Blog Opinion	SentiWordNet lexicon	61%
2006	Koning, Brill	Movie Review	Hybrid	91%
2007	Godbole et al.	Blog Posts	Lexical Approach	96%
2008	Songho tan	Chnsenticorp	SVM	90%
2008	Zhou and Chaovalit	Movie Review	ontology-supported polarity mining	72%
2009	Melville	Blogs	Bayesian Classification	91%
2009	Ruby	Movie Review	SVM	89%
2009	QingliangMiao	Amazon Review	Lexical Resource	88%
2010	Yulan He	Movie Review	Sentiment Lexicon	74%
2010	Gang li	Movie Review	K-means clustering	78%
2010	Zhu Jian	Movie Review	Back Propogation	86%
2010	George Stylios et al.	Governmental Decisions	SVM	98%
2010	Gamgam somprasti	Amazon Review	Mazimum Entropy	72%
2010	Ahmed Abbasi	Digital Cameras, Automobiles, Movies	N- Word	86%
2011	KaiquanXu	Amazon review	SVM	61%
2011	Ziqiong	Cantonese Review	SVM	93%
2011	Xue bai	Movie Review	Naïve Bayes	92%
2011	Rui Xia	Movie Review	SVM	86%
2011	Long Sheng	Movie Review	BPN	64%
2011	Ziqiong	Cantonese Review	SVM	93%
2013	Ayesha Rashid	Indian Hotel Review	NLP , Bayesian	96%
2014	Pravesh Kumar Singh	Product Review	SVM	79%
2014	Pravesh Kumar Singh	Movie Review	SVM	81%
2015	S. Vasantharaj et al	China Car Review	Word Kernal ,Path Kernal, N-gram kernal	81%
2015	S. Lakshmi Prabha	Maths	J48	96%
2015	Humera Shaziya	Movie Review	SVM	84%
2015	Humera Shaziya	Movie Review	Naïve Bayes	85%


REVIEW OF LITERATURE
86.40%
86%
73%
88%
86%
61%
91%
96%
90%
72%
91%
89%
88%
74%
78%
86%
98%
72%
86%
61%
93%
92%
86%
64%
93%
96%
79%
81%
81%
96%
84%
85%
PANG AND LEE
GAMON
HU AND LIU
BAI.
KENNEDY …
ETHAN ZHANG…
KONING, BRILL
GODBOLE ET …
SONGHO TAN
ZHOU AND…
MELVILLE
RUBY
QINGLIANGMI …
YULAN HE
GANG LI
ZHU JIAN
GEORGE …
GAMGAM…
AHMED ABBASI
KAIQUANXU
ZIQIONG
XUE BAI
RUI XIA
LONG SHENG
ZIQIONG
AYESHA RASHID
PRAVESH…
PRAVESH…
S.…
S. LAKSHMI…
HUMERA…
HUMERA…

CHAPTER 3

SCOPE OF STUDY

Sentiment analysis is implemented in many application domains where it is helps people (users) in every step in life. Using this analysis techniques people can come to a point where they can genuinely can decision for their better future and help others by giving suggestions. This analysis not only help to any step ahead but also to take back step if the data analysis (review/comments) doesn't seem to be positive, if negative then is not in favor to the 0society. Accurate and proper feedbacks always help for better understanding.

In this section we are discussing about few opinion mining application domain which really helping users to work on and take better decision.

a) Shopping (E-Commerce)

Online Shopping is important for people as it deliver package to their door steps because it is easy to get an overall idea about any product by the number of feedbacks mentioned by customers. In this kind of site users go through the feedback before they buy any product. [31] Shopping sites Flipkart, Amazon helps to compare products with all desired description and feedback of customers, information is displayed to the user in a "Graphical User Interface" for easy understanding about the quality/features and services of product. This is what really helps users to get their selection to product want to buy.

b) Entertainment

For Movie or show or TV programs can easily go through public feedback about any current release or any kind of popular program or movie. Internet movie database (IMDB) which helps the users online to view feedback for movies and other programs. It also helps users to understand the movie or program who are not sure it from this analysis. [31] [32]

c) Business

Companies are now asking for feedbacks on their portal which makes savings on marketing budget. Now companies can be able to analysis the data they from the given feedback online in their website. Recommendation to friends and family about products or services to each other which helps to gain a clear picture about the products or services before buy it. [31] [33]

d) R&D (Research and Development)

In any online shopping portal items feedback are used by industry to improve the quality of the products. Online sites can also ask the customers to provide their own design views to make modification or create new products for customers [31]. For example, "Play Store" in which it shows all the applications users has given all sort of feedback/point of view.

e) Politics

Political people can extract the views on public feedback. During Elections the participated candidates can have an overall knowledge about public opinion which helps to gain an idea about the progress (weakness and strength)[31].

f) Academics

In any online course, Student's opinion can be used to manage their academic. Student's feedback may help the institute to understand and analysis and make better improvement for student's future benefits[31]. Institution will be aware of their benefits and problems through their opinion in the form on feedback and comments.

g) Health

User's shares treatment which helps to take precaution while they suffer from headache, fever, high blood pressure, Diabetes and many more which is also called "Home Remedies". Users also share the exercise steps for body building, blood pressure, back pain , migraine etc which helps all to get rid of problems in their life. While doing so users also gives feedback and comment from which is it understandable that particular remedies is working or not to others.

h) Transportation

Transportation can be any movable items like animals, people, and goods from one place to another as per the requirement .There are many transport modes for example road, air, rail, water, cable, space and pipeline. Transport is necessary to trade between persons, which is essential for the development of civilizations. Transportation like tour and travel, logistic, cab service which is very helpful to people to choose best services mentioned in the feedback and comments from past experience shared by the users. For example - from Tour and Travel that is holiday plans it helpful to get an idea about the location and hotels and services provided to them.

i) Social Media

Social Media is another source from where sentimental analysis can be tested like Facebook, twitter and many more site from where reviews/comments can be extracted to find out the users' opinions on a particular subject. For example, a) I love that book b) I like the book c) it is awesome, from this we can extract that all the users like the book and it is positive sign[34][33]

CHAPTER 4

OBJECTIVE OF STUDY

In this chapter we study the following objectives are described as follows on which we have mainly focused on –

- Sentimental analysis on the E-learning reviews/comments with positive, negative and neutral data in Weka tool.
- Comparison of E-learning dataset with different tokenizer (word, ngram etc.,)
- Comparison of E-learning dataset with various classification algorithms in Weka tool like SVM, Naïve Bayes, KNN, J48, Decision Table.
- Comparing with different results (like Accuracy, Sensitivity, Specificity etc.) for better performance and enhancement in E-Learning System.

CHAPTER 5

RESEARCH METHODOLY

In this research of sentiment analysis, we have faced many limitations to extract the sentiment from the given valuable feedbacks/reviews/comments in the form of text in document. Those may be in different type single word or sentence form or in document containing positive, negative and neutral views of users for a particular topic. Due to those limitations like smiley, single-text or sentience or document level it has been used some methodology to filter and get the exact sentiment what a user is looking for during analysis of various feedbacks.

5.1 Solutions to the shortcoming Sentimental analysis

The Solutions to the shortcoming of sentiment analysis are as follows:

i. *Spelling Mistake*

If there is more number of mistakes in spelling like “thanq” is repeated many a time in the many comments, there is the is possibility to analysis the sentiment in weka tool. As weka has tokenizer is word, ngram , character ngram etc .

But there are also few case where there is full of mistakes like in Figure 6, those are difficult to analysis the sentiment of the user.

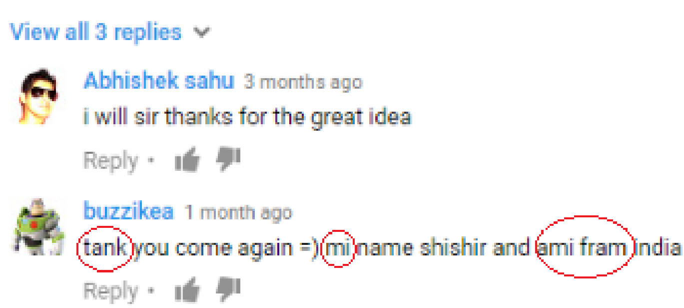


Figure 6: Spelling Mistake [49]

ii. Special character/symbols/tokens like stickers

Stickers are actually in the form of delimiters. Weka has that facility to remove delimiters from the user's comments/reviews during preprocessing. Few example of stickers used in comments/reviews like Figure 7

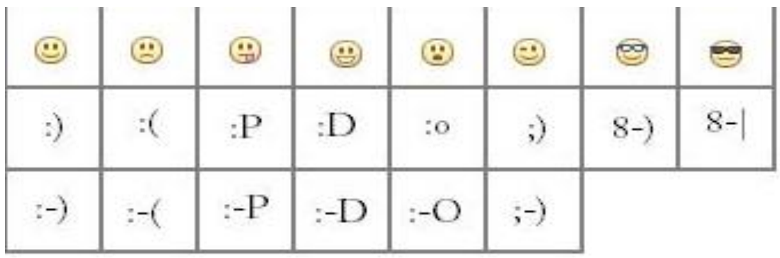


Figure 7: Special character/symbols/tokens like stickers [50]

iii. Regional language

Few regional languages can be translated manually at the time of preparing the dataset and labeling of data. This is the example which is used to translate a comment Figure 8

Figure 8: Translator [49]

iv. Error in grammar

As weka does has various tokenizer that splits the long sentences into words, this is in case of "word tokenizer", it also has some other tokenizer like ngram, character ngram and alphabetic. All of these has a different feature to deal with the error in grammar.

v. Abbreviation in short form like fine(fi9)

It works only when it has same abbreviation repeated in many comments/reviews. According to the labeling it will be able to analysis the sentiment of the users. If it has a single abbreviation in the dataset, then it doesn't use in analyzing the sentiment of the user.

vi. Fake reviews

Only the authorized user should be allowed to comment. For example: a) Facebook- if User-A is in the friend list of User-B, then both should be allowed to comment on each other's post, If User-A is not in the friend list of User-C then both should not be allowed to comment on each other's post. b) Online shopping site – if a user has bought a particular product then the user is allowed to give feedback/reviews/comments for that, but if a user has not bought a product then the user should not be allowed to give feedback/reviews/comments on any product or even if he/she is not a registered user, he/she must not be allowed to give any feedback/reviews/comments.

vii. Diversity of contents

Diversity of contents is a mixture of positive or negative or somethings it is both. For example, taking a label in class like positive and negative, neutral (Diversity of contents) can also be included as an another one in to for a better sentimental analysis.

Sentiment analysis can be implemented using a classifier which requires datasets of reviews/comments(test set and training set).this process of learning is called Machine Learning(ML).[7][35]

5.2 Sentimental classification (supervised learning)

Pang at al was the first paper to use classifier approach on a movie (reviews) which classified into two part i.e positive and negative. There are few most used features presents term presence and its frequency, parts of speech(POS), negations and positive/negative. [36] There are few algorithms used to classify the data to increase trend in the performance of opinion mining for all classifiers[37].They are as follows which will be used during mining –

The grouping of review text is directed by learning method. Using instance reviews the classifiers are taught. The trained classifier model is used to guess grouping of fresh text reviews. Support Vector Machines (SVM) and Naive Bayes (NB) are the most standard machine learning classifiers in sentiment analysis for text categorization from large data.[38]

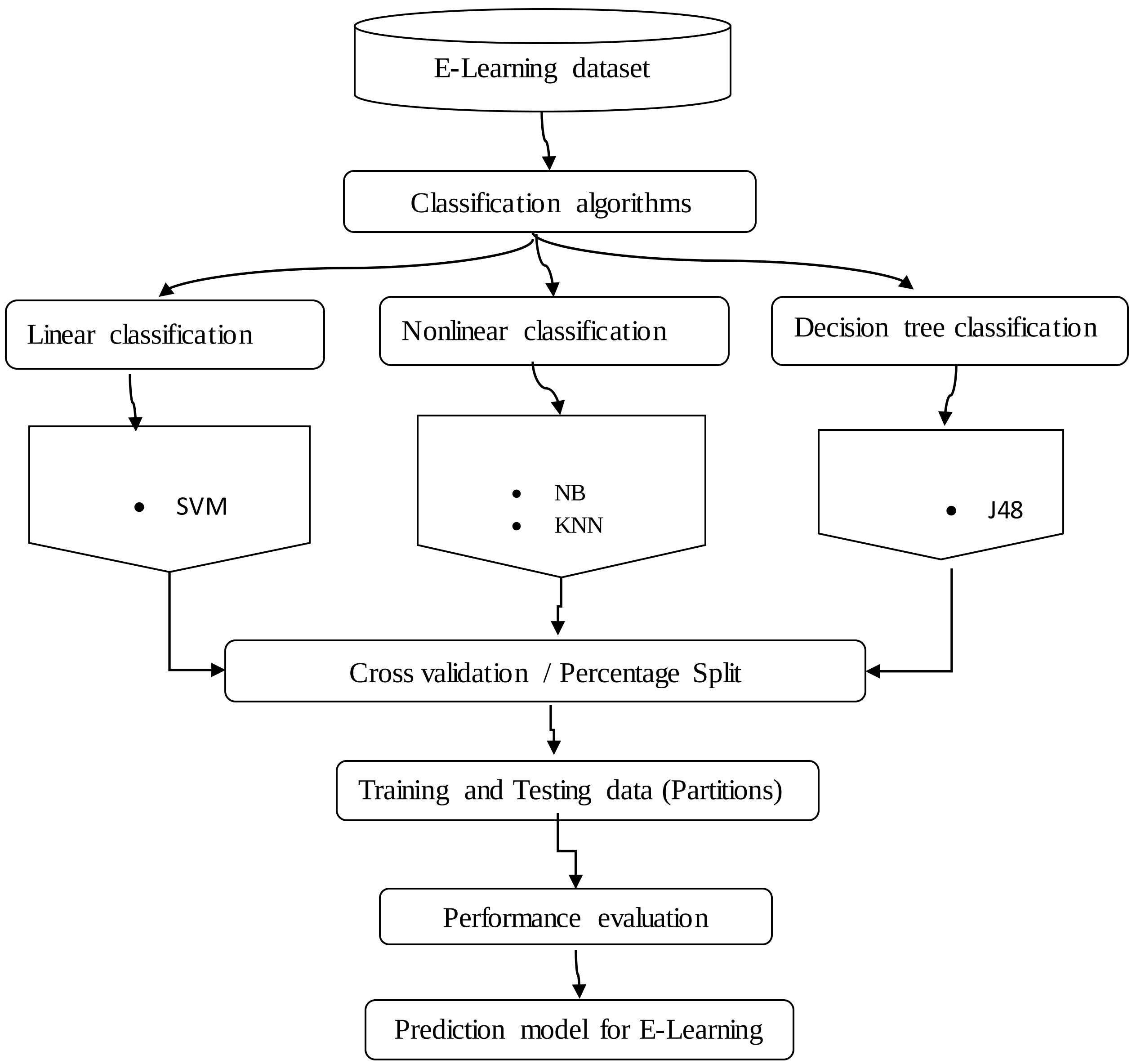


Figure 9: The Overall Structure

j) Support Vector Machines (SVM)

SVM have succeeded great victory in text labelling[35]. It is a linear learning machine that helps in categorizing the structures to their particular classes. Using this technique, from datasets i.e. text should be linearly separate. So that a boundary is made which is divided into different class (positive, negative, neutral)[39] [12]. If the judgement has a certain keyword in the attribute set, the equivalent value is 1 otherwise the value will be considered as 0. [40] The completed results have exposed that a classification can increase the categorization efficiency even in E-learning [41][42]

k) Naive Bayes (NB)

Naïve bayes is a non-linear learning machine, it is method that is broadly used algorithm for document classification and one of the standard techniques for text labelling. It is very prevalent algorithm as it is simple, effective and displays better presentation for genuine world problems. "Naive" accepts that features are entirely independent, this algorithm explains the problems fit to normal distribution. It assumes a probabilistic model and allows the capture of ambiguous aspects in the text, by calculating likelihoods of the results [10]. It has been presented to achieve exceptionally well in implementation by many scholars[27].

The additional most well- known machine learning approaches in the NLP area are like K-Nearest but from the previous research indicated that the SVM and Naive Bayes algorithms are the most accurate in analyzing the sentiment. So it has been panned to work on this in E-learning to classify the their reviews of users[43].

l) K-Nearest Neighbors (KNN)

KNN is a simple algorithm that stores all accessible cases and characterizes new cases in light of a similitude measure (e.g., remove capacities).[44] although it is used for estimate and prediction. KNN is an illustration of instance-based learning, in which the training data set is put in storage, so that a grouping for a new random data may be initiate simply by associating it to the most related data in the training set.

m) J48 classifier

Decision Tree is a verdict provision tool that uses a tree-structured graph or model that represents flow and their possible concerns, including unplanned event results, resource expenses, and utility.

J48 classifier predict the maximum number of class in training data. It predicts the mean for numeric values & mode for nominal classes[31]. The correctness is more and effective result in data mining. In the comprehensive correctness report by class specifies TP rate (True positive rate), TN rate (True negative rate), FP rate (False positive rate) and FN rate (False negative rate). J48 is the operation of algorithm ID3(interative Dichotomiser 3) technologically advanced by the project team of WEKA.

n) Decision Table (DT)

Decision Table is a good approach to manage with different combination inputs with their related outputs and furthermore called cause-impact table. Reason to call bring about impact table is a related consistent charting system called 'cause-impact diagramming that is fundamentally use to infer the choice table.

5.3 Sentimental classification (unsupervised learning)

In this classification the review is classified by differentiated against the word vocabulary. The sentiment value of the text vocabulary is analyzed according to the emotion. The sentiment vocabularies describe the collection of arguments and terms that are used to precise people's moods, visions and ideas. To know and get better understanding this we can primarily begin with a positive/negative word vocabulary. The text is examined and tested or the existence of positive/negative word vocabulary. If the text has more positive word vocabularies it is measured to be a positive text else if more number of negative word vocabularies are present, the text is measured negative. An important innovation in the unsupervised sentiment classification was done by Turney. He used the arguments "bad" and "good" as the seed words and considered the coordination of words based on them. He was effective in succeeding 66% accuracy for the movie reviews in the application domain that helps movie lover to analysis and take decision accordingly [45].

5.4 How it is working on Weka

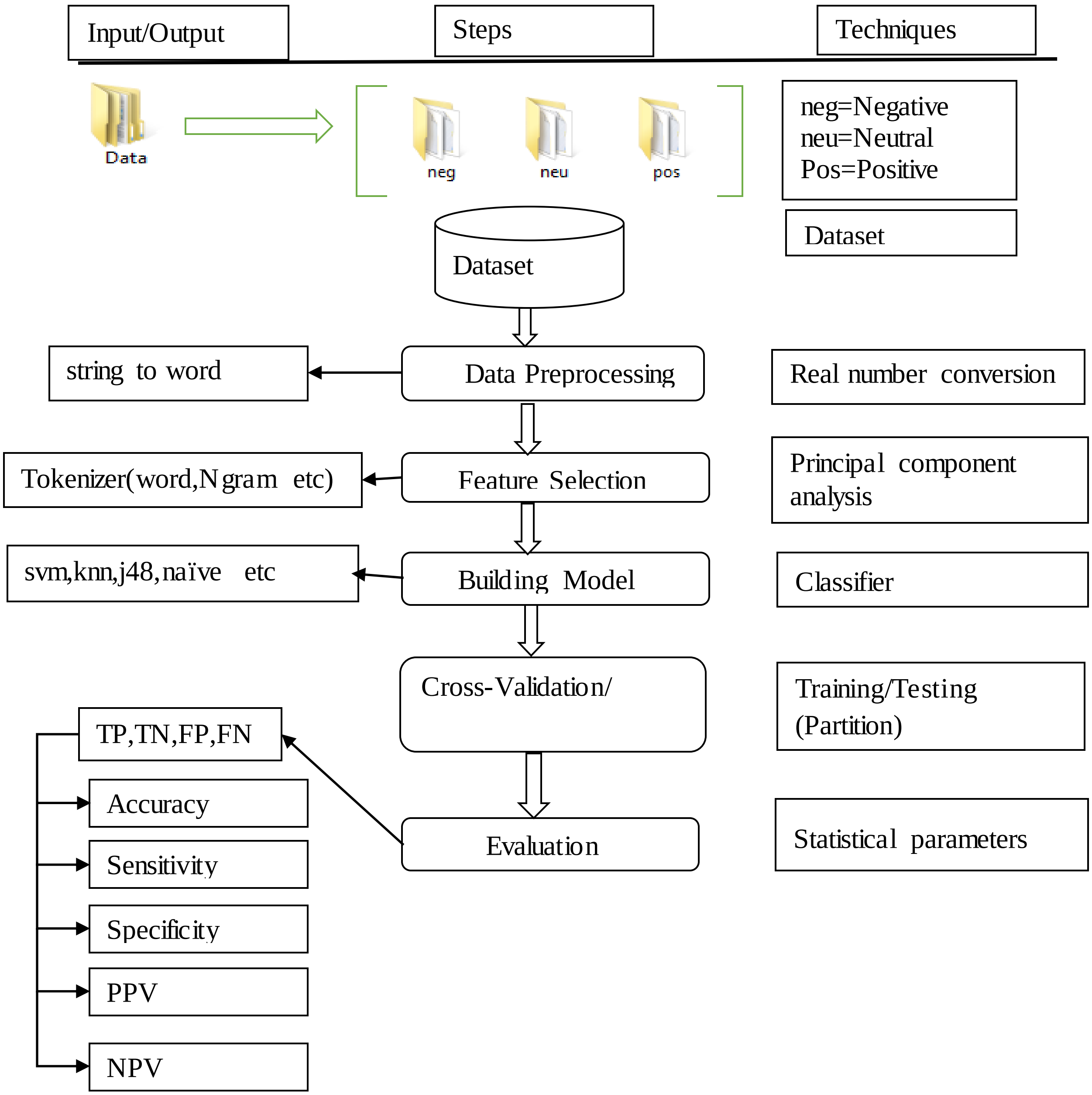


Figure 10: Block diagram of Weka- Classified integrated system [27][42]

Working in Weka

Using this tool and surveying various sites and links it is found that the use of dataset is done in the form of csv or arrf file, but here it has been implemented in a text file Figure 10.

Example of class labeling in Figure 11 i.e neg=negative , neu=neutral and pos=positive(150 file each).[46]

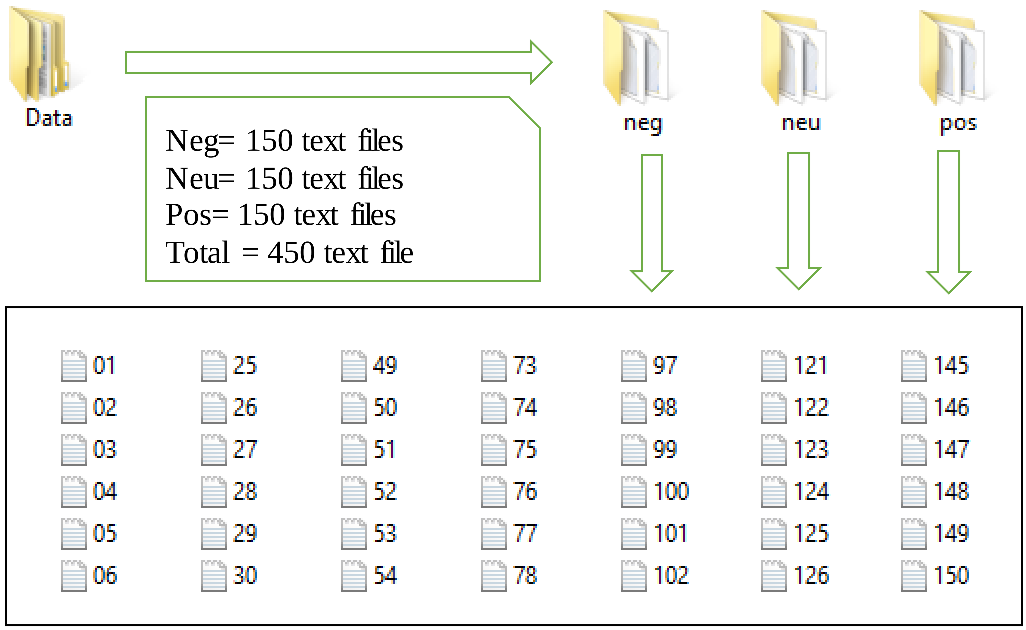


Figure 11: Dataset (class labeling)

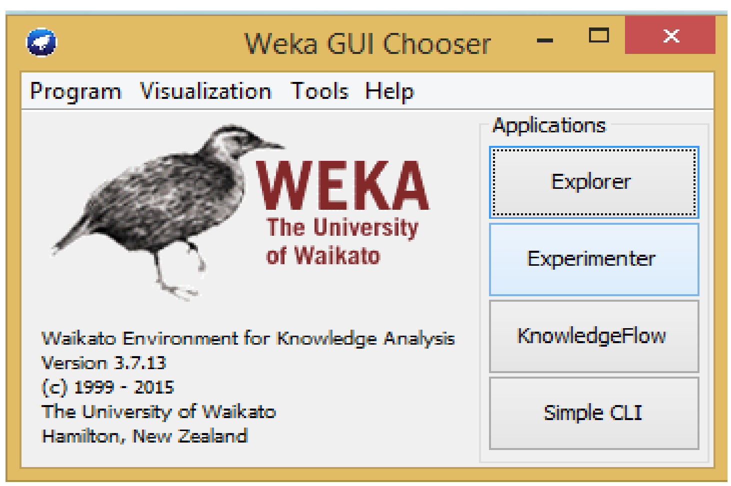


Figure 12: Weka Tool Version 3.7.13

a. Click on “Explorer” of Weka GUI Chooser Figure 12, a new popup window will open which will have many option on preprocess as open file, open url,open db etc as in Figure 13.

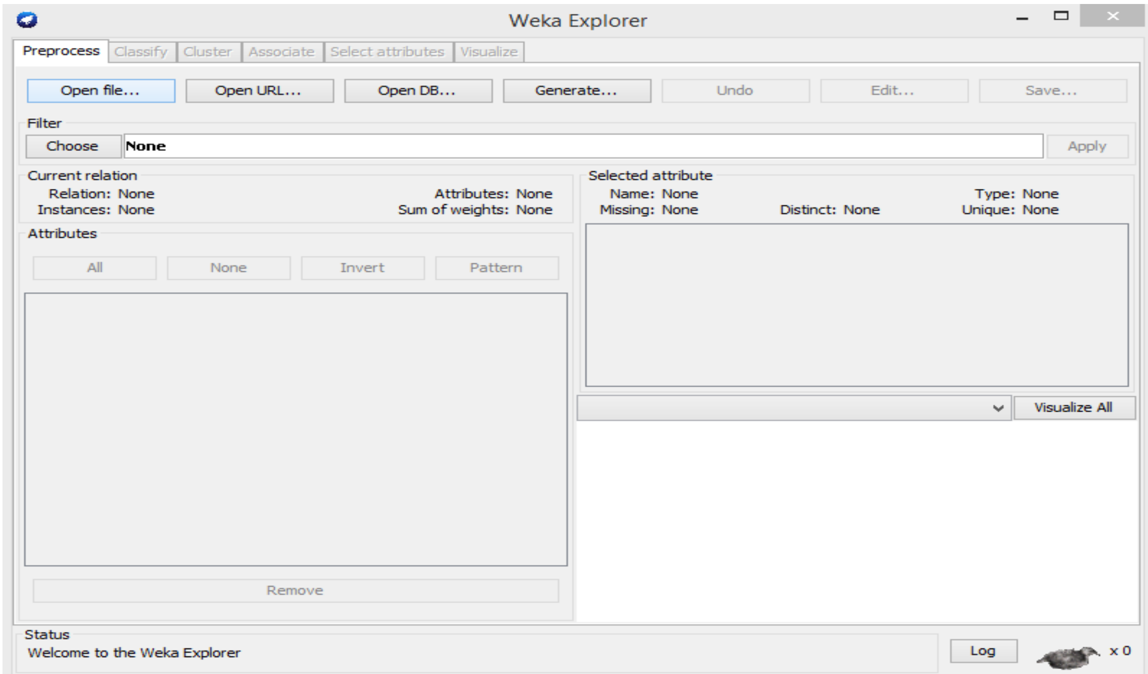
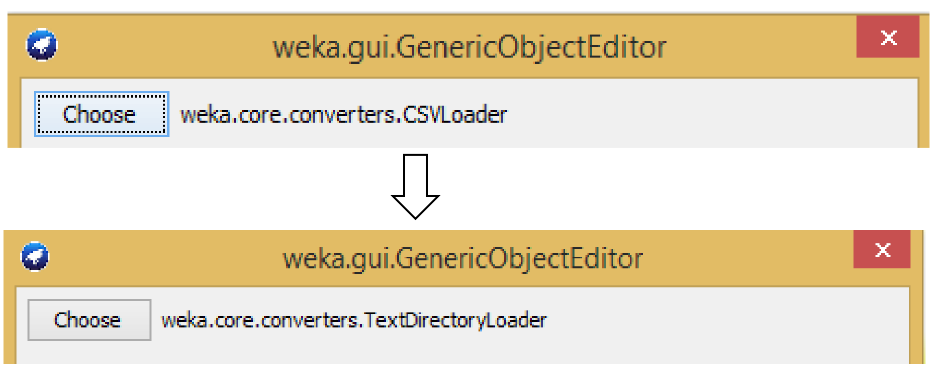


Figure 13: Weka Explorer

b. Click on “open file” and select the dataset “data” folder. It will generate an error “Cannot determine the file loader automatically, please choose one”.

c. Click on that error message “OK”, it generates an another popup window of file loader i.e weka.gui.GenericObjectEditor

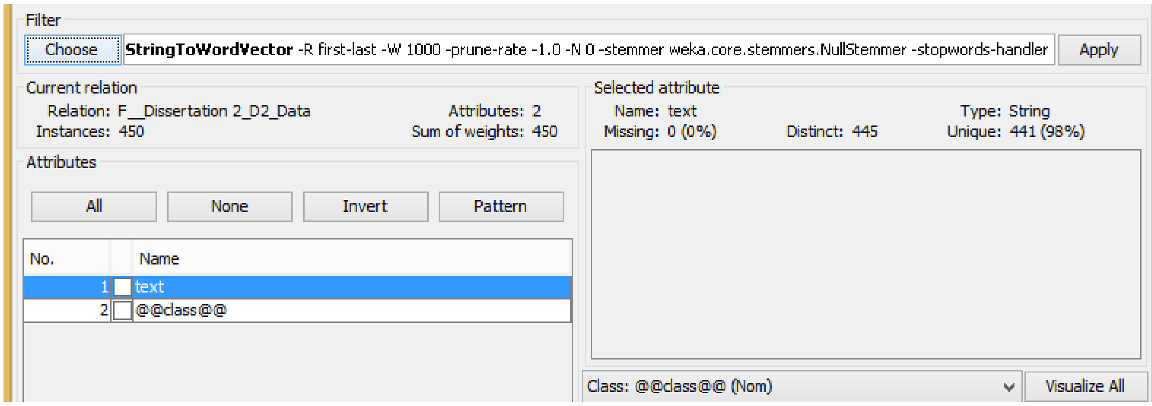
d. Choose in the option of weka.gui.GenericObjectEditor (popup window)and change “weka.core.converters.CSVLoader” to “weka.core.converters.TextDirectoryLoader” as in Figure 14 and click “ok”.



Textdirectoryloader- loads all the text files in a directory and uses the subdirectory as class labels

Figure 14: Text Directory Loader

e. Click on filter "Choose" Weka->filter->supervised->attributes->StringtoWordVector (filter can be used as per one's requirement).



StringTowordVector: Converts String attributes into a set of attributes representing word occurrence (depending on the tokenizer) information from the text contained in the strings.

Figure 15: Filter-String To Word Vector

f. Click on "StringToWordVector –R first-last" that will generate an another new popup window named "weka.gui.GenericObjectEditor" as mentioned in Figure 15.

g. According to the requirement choose "Tokenizer" as in Figure 16 from the list, even Stemmer selection is also made at this window (bydefault:NullStemmer) and also can add delimiters, min & max size and click on "Ok".[30]

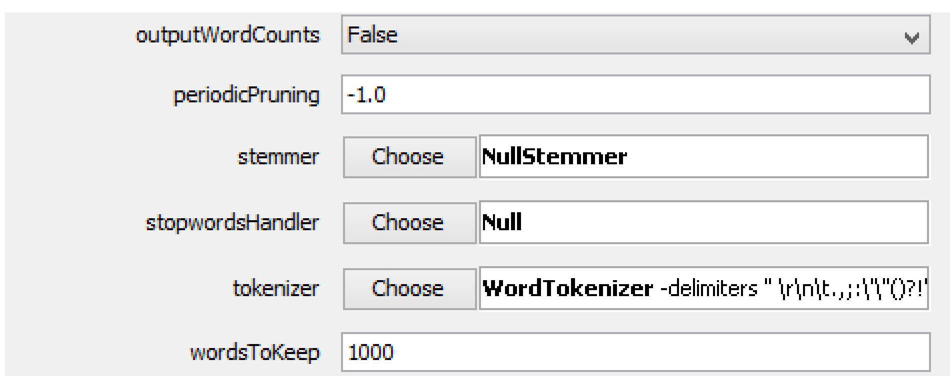


Figure 16: Tokenizer

h. Click on "Apply" to start the process, which will do the process as per the tokenizer selection.

Table 2: Tokenizer in details

Tokenizer	Description
Word	A simple tokenizer that is using the java.util.StringTokenizer class to tokenize the strings
	delimiters .,;:'"()?!
	Example- @twitter=twitter
N-gram	Splits a string into n-gram
	min and max grams
	delimiters .,;:'"()?!
	Example-this is good- this, this is, this is good
Character n-gram	Splits a string into all character n-gram it contains based on given min and max of n-grams.
	min and max grams
	even gap is counted
	No delimiters
	Example-thank you-tha nk you
Alphabetic	Alphabetic string tokenizer, tokens are to be formed only from contiguous alphabetic sequences
	No delimiters
	No min and max grams
	Example-welcome-welcome

i. Later the data will be processed and it will provide an output as in Figure 17.

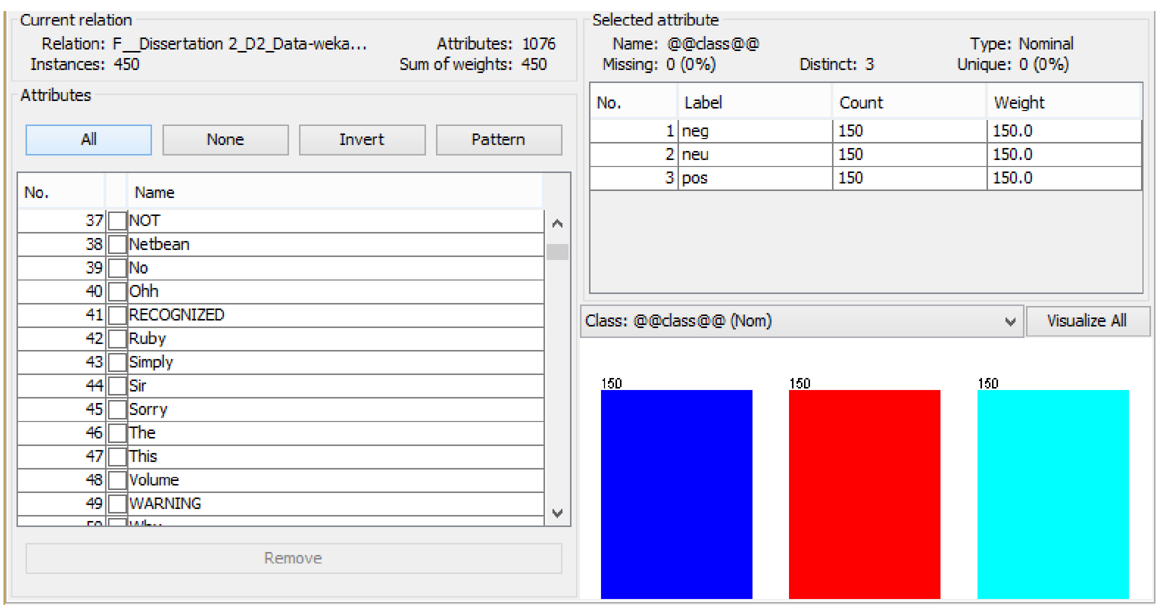


Figure 17: Processed data

j. Click on "Classify", Choose Classifier as per user requirement(choose->classifiers->bayes->Naivebayes) like in Figure 18.

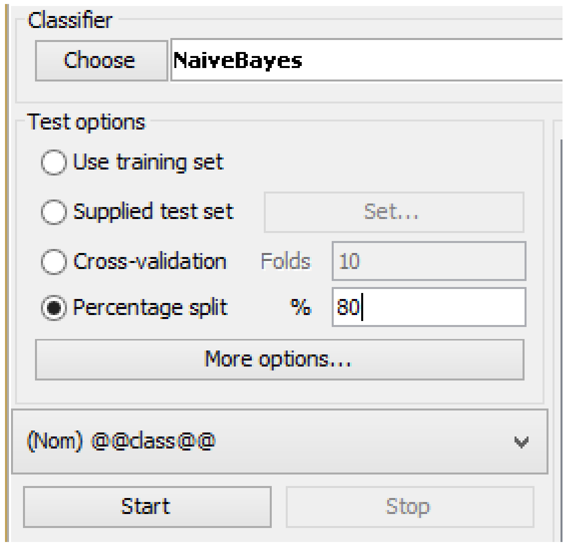


Figure 18: Classifier

k. After this process user has to select “Cross-Validation” or Percentage Split.

l. For example, if percentage split is selected using 80% then from the dataset it will take 80% as training data and 20% testing data but in case of cross-validation is it different if cross-validation is selected using 10 fold, it will make into 10 partition of data where 1 partition will be testing set and other 9 partition will be training set which will run for 10 iterations.

m. Click on “Start” and it will generate a confusion matrix as mentioned in Figure 19 .[24]

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
 19  7  2 |  a = neg
  6 27  0 |  b = neu
  2  4 23 |  c = pos
```

Figure 19: Confusion Matrix

n. Now find the TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative) values and find the result ACCURACY, SENSITIVITY, SPECIFICITY, PPV and NPV which are defined in Eq. (1), (2), (3), (4) and (5) respectively in .[39]

$$ACCURACY = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$SENSITIVITY = \frac{TP}{TP+FN} \tag{2}$$

$$SPECIFICITY = \frac{TN}{TN+FP} \tag{3}$$

$$PPV = \frac{TP}{TP+\mathrm{FP}} \tag{4}$$

$$NPV = \frac{TN}{TN+FN} \tag{5}$$

CHAPTER 6

RESULTS

6.1 Word Tokenizer

Using "Word Tokenizer" in the techniques like (SVM, Naïve, KNN, J48, DT) it is compared with different results (accuracy, sensitivity, specificity, ppv and npv). It is observed that SVM provided a better result in percentage split (80-20) % and average it is 85.98%, where Naïve (80.62%), KNN (77.66%), J48 (75.56%) and DT (63. 33%) as in Table 3.

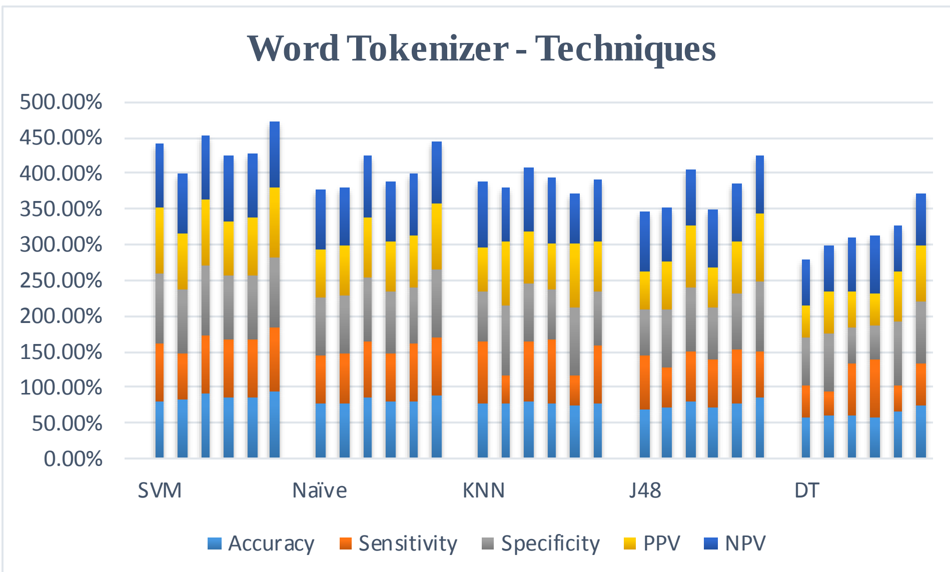


Figure 20: Word Tokenizer- Techniques

In case of measures(average) specificity (82.83%) where accuracy (77.18%), sensitivity (69.43%), ppv (72.89%) and npv (82.18%)

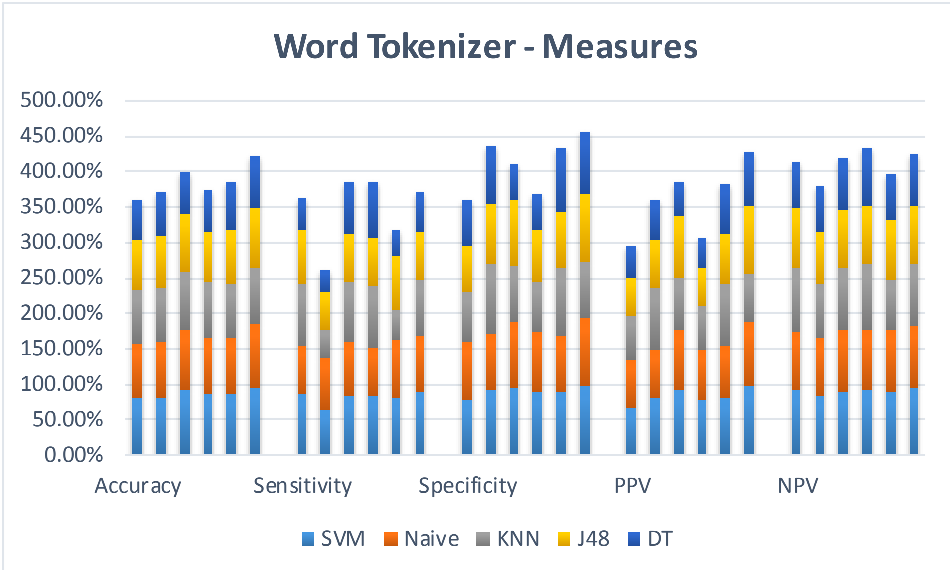


Figure 21: Word Tokenizer - Measures

Table 3: Word Tokenizer

Classification Methods			SVM	Naive	KNN	J48	DT
Accuracy	Cross Validation (10 fold)	a=neg	80.09%	76.94%	76.63%	68.93%	57.84%
		b=neu	81.78%	77.67%	76.81%	72.66%	61.98%
		c=pos	90.91%	85.83%	81.75%	81.72%	60.32%
	Percentage Split (80-20)%	a=neg	86.36%	80.23%	77.50%	71.59%	59.30%
		b=neu	86.36%	80.23%	74.70%	77.78%	66.23%
		c=pos	95.00%	89.61%	78.48%	85.14%	75.00%
Sensitivity	Cross Validation (10 fold)	a=neg	86.00%	68.67%	87.33%	75.33%	46.00%
		b=neu	64.67%	71.33%	40.67%	54.00%	31.33%
		c=pos	82.67%	78.00%	84.00%	67.33%	72.67%
	Percentage Split (80-20)%	a=neg	82.14%	67.86%	89.29%	67.86%	78.57%
		b=neu	81.82%	81.82%	42.42%	75.76%	36.36%
		c=pos	89.66%	79.31%	79.31%	65.52%	58.62%
Specificity	Cross Validation (10 fold)	a=neg	77.00%	81.45%	70.57%	65.47%	65.27%
		b=neu	91.01%	81.18%	97.35%	83.59%	83.57%
		c=pos	96.17%	90.91%	80.33%	91.94%	52.02%
	Percentage Split (80-20)%	a=neg	88.33%	86.21%	71.15%	73.33%	50.00%
		b=neu	89.09%	79.25%	96.00%	79.17%	88.64%
		c=pos	98.04%	95.83%	78.00%	97.78%	87.18%
PPV	Cross Validation (10 fold)	a=neg	66.15%	66.88%	62.68%	54.07%	45.39%
		b=neu	79.51%	67.72%	89.71%	65.85%	57.32%
		c=pos	93.23%	84.78%	72.83%	85.59%	50.46%
	Percentage Split (80-20)%	a=neg	76.67%	70.37%	62.50%	54.29%	43.14%
		b=neu	81.82%	71.05%	87.50%	71.43%	70.59%

		c=pos	96.30%	92.00%	67.65%	95.00%	77.27%
NPV	Cross Validation (10 fold)	a=neg	91.32%	82.66%	90.78%	83.11%	65.82%
		b=neu	82.68%	83.65%	74.28%	75.62%	63.35%
		c=pos	89.68%	86.42%	88.89%	79.84%	73.89%
	Percentage Split (80-20)%	a=neg	91.38%	84.75%	92.50%	83.02%	82.86%
		b=neu	89.09%	87.50%	71.64%	82.61%	65.00%
		c=pos	94.34%	88.46%	86.67%	81.48%	73.91%
Average (%)			85.98%	80.62%	77.66%	75.56%	63.33%

6.2 NGram Tokenizer

Using "NGram Tokenizer" in the techniques like (SVM, Naïve, KNN, J48, DT) it is compared with different results (accuracy, sensitivity, specificity, ppv and npv) as it has been done in case of word tokenizer but it has an additional feature (min and max size). Here it is observed that SVM provided a better result in percentage split (80-20) % and average it is 84.94%, where Naïve (80.99%), KNN (74.77%), J48 (76.33%) and DT (64. 31%) in Table 4.

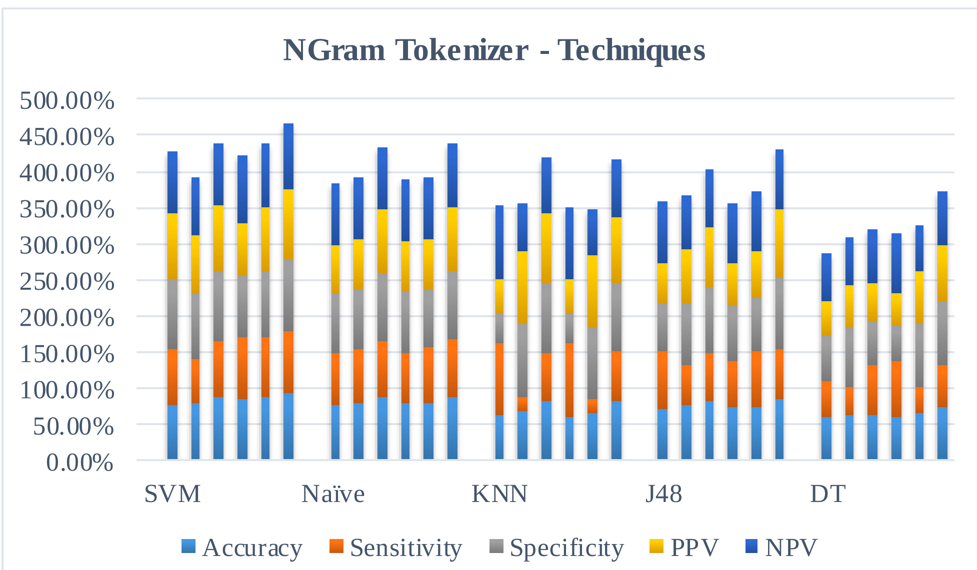


Figure 22: NGram Tokenizer - Techniques

In case of measures(average) specificity (82.69%) where accuracy (76.04%), sensitivity (67.71%), ppv (74.54%) and npv (81.68%).

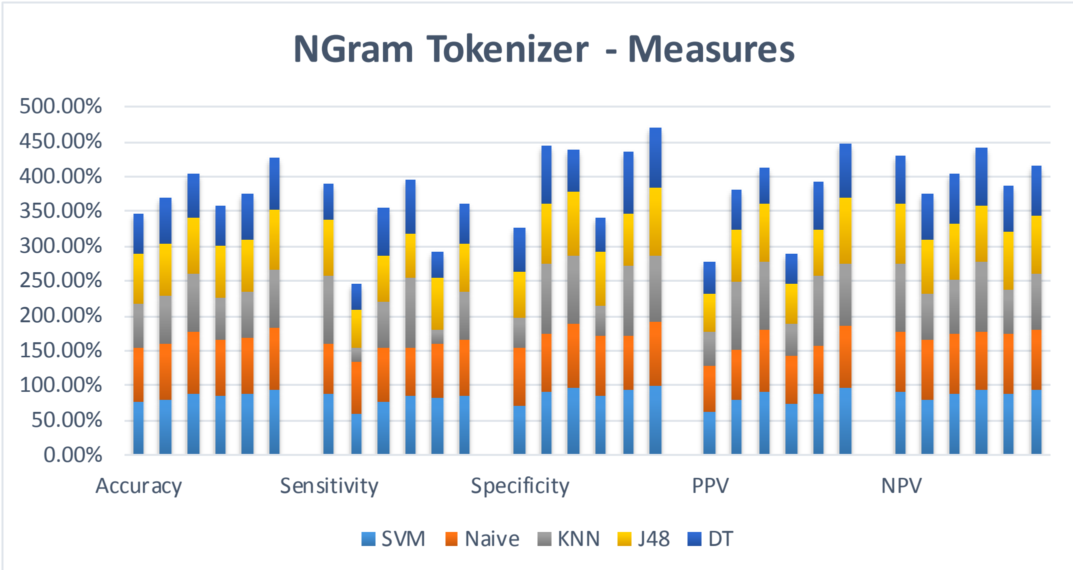


Figure 23: NGram Tokenizer - Measures

Table 4: NGram Tokenizer

Classification Methods			SVM	Naive	KNN	J48	DT
Accuracy	Cross Validation (10 fold)	a=neg	76.89%	78.14%	62.11%	71.03%	59.05%
		b=neu	80.19%	79.62%	69.60%	75.62%	64.03%
		c=pos	88.42%	87.50%	83.18%	81.28%	63.51%
	Percentage Split (80-20)%	a=neg	85.39%	80.00%	61.36%	73.26%	59.30%
		b=neu	88.37%	79.07%	66.67%	75.00%	66.23%
		c=pos	93.83%	88.31%	83.08%	86.30%	75.00%
Sensitivity	Cross Validation (10 fold)	a=neg	87.33%	71.33%	100.00%	80.00%	50.00%
		b=neu	59.33%	74.67%	19.33%	56.00%	38.00%
		c=pos	77.33%	78.00%	65.33%	66.67%	68.67%
	Percentage Split (80-20)%	a=neg	85.71%	67.86%	100.00%	64.29%	78.57%
		b=neu	81.82%	78.79%	18.18%	75.76%	36.36%
		c=pos	86.21%	79.31%	68.97%	68.97%	58.62%
Specificity	Cross Validation (10 fold)	a=neg	71.43%	81.79%	42.91%	66.19%	64.52%
		b=neu	91.82%	82.35%	100.00%	87.30%	82.03%
		c=pos	95.65%	93.59%	97.81%	91.07%	60.00%
	Percentage Split (80-20)%	a=neg	85.25%	85.96%	43.33%	77.59%	50.00%
		b=neu	92.45%	79.25%	100.00%	74.51%	88.64%
		c=pos	98.08%	93.75%	94.44%	97.73%	87.18%

PPV	Cross Validation (10 fold)	a=neg	61.50%	67.72%	47.02%	56.07%	46.01%
		b=neu	80.18%	70.00%	100.00%	72.41%	59.38%
		c=pos	92.06%	88.64%	96.08%	83.33%	53.93%
	Percentage Split (80-20)%	a=neg	72.73%	70.37%	45.16%	58.06%	43.14%
		b=neu	87.10%	70.27%	100.00%	65.79%	70.59%
		c=pos	96.15%	88.46%	90.91%	95.24%	77.27%
NPV	Cross Validation (10 fold)	a=neg	91.52%	84.19%	100.00%	85.98%	68.09%
		b=neu	80.19%	85.50%	67.21%	76.92%	65.68%
		c=pos	86.61%	86.90%	77.49%	80.31%	73.74%
	Percentage Split (80-20)%	a=neg	92.86%	84.48%	100.00%	81.82%	82.86%
		b=neu	89.09%	85.71%	64.00%	82.61%	65.00%
		c=pos	92.73%	88.24%	79.07%	82.69%	73.91%
Average (%)			84.94%	80.99%	74.77%	76.33%	64.31%

6.3 Character NGram Tokenizer

Using "Character NGram Tokenizer" in the techniques like (SVM, Naïve, KNN, J48, DT) it is compared with different results(accuracy, sensitivity, specificity, ppv and npv) as it has been done in case of word tokenizer and Ngram tokenizer it has an additional features (min and max size) but it doesn't have delimiters. Here it is observed that SVM provided a better result in Cross-validation (10 folds) and percentage split (80-20) % and average it is 84.52%, where Naïve (80.09%), KNN (77.34%), J48 (75.85%) and DT (71. 36%) in Table 5

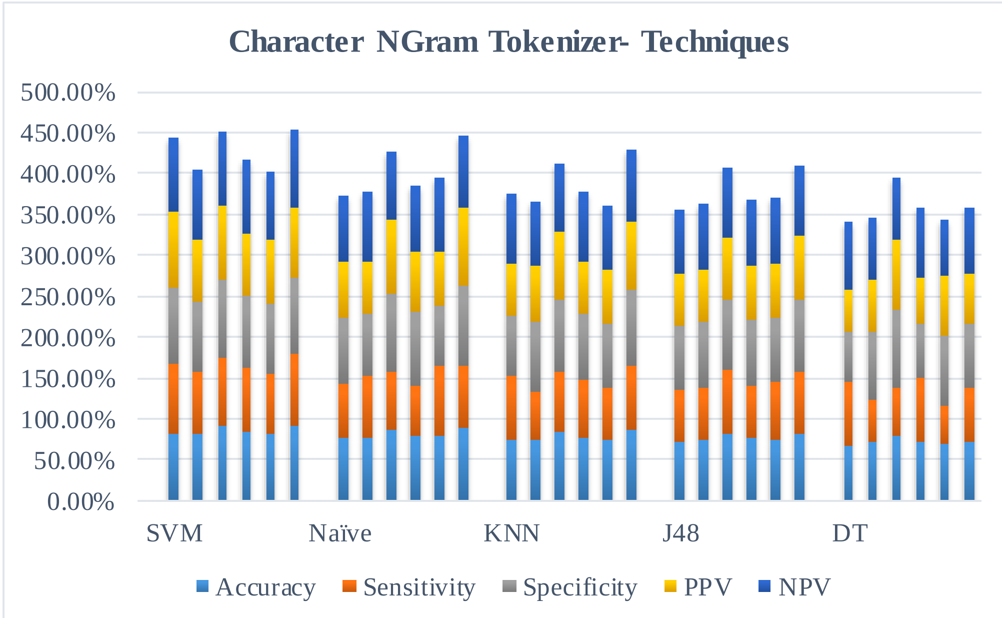


Figure 24: Character NGram Tokenizer- Techniques

In case of measures(average) specificity (83.77%) where accuracy (78.81%), sensitivity (71.42%), ppv (73.01%) and npv (83.42%)

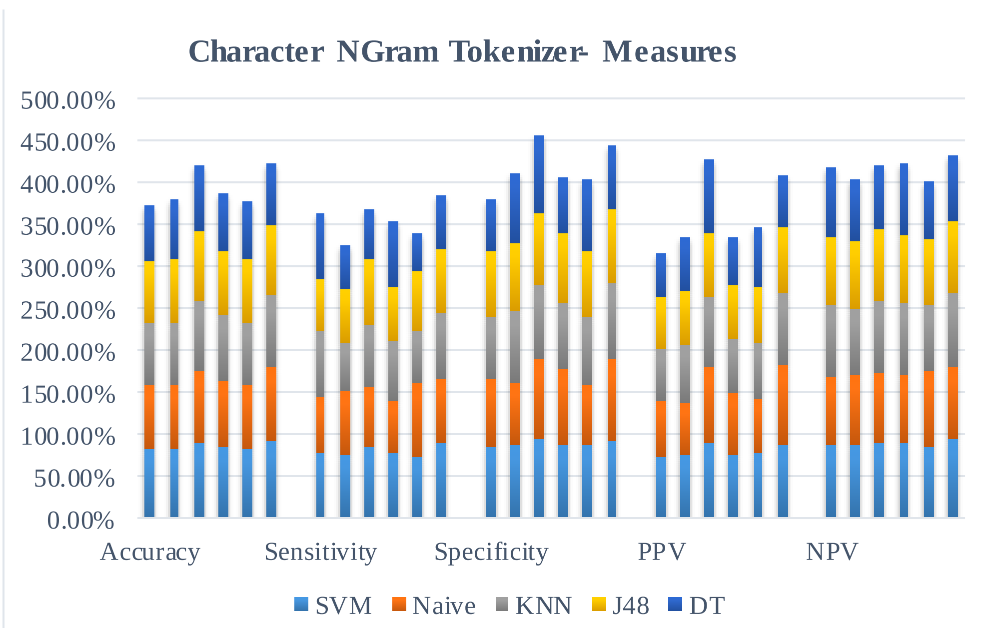


Figure 25: Character NGram Tokenizer- Measures

Table 5 : Character Ngram Tokenizer

Classification Methods			SVM	Naive	KNN	J48	DT
Accuracy	Cross Validation (10 fold)	a=neg	82.22%	76.30%	75.36%	73.10%	66.90%
		b=neu	82.60%	75.94%	75.36%	74.33%	71.25%
		c=pos	90.82%	85.64%	83.11%	82.09%	79.39%
	Percentage Split (80-20)%	a=neg	84.71%	80.00%	77.11%	75.90%	70.89%
		b=neu	81.82%	78.16%	73.56%	75.00%	70.00%
		c=pos	91.14%	89.47%	86.49%	82.89%	72.73%
Sensitivity	Cross Validation (10 fold)	a=neg	78.00%	66.67%	78.00%	62.67%	78.00%
		b=neu	74.67%	76.67%	58.00%	63.33%	52.67%
		c=pos	84.67%	71.33%	74.00%	78.67%	59.33%
	Percentage Split (80-20)%	a=neg	78.57%	60.71%	71.43%	64.29%	78.57%
		b=neu	72.73%	87.88%	63.64%	69.70%	45.45%
		c=pos	89.66%	75.86%	79.31%	75.86%	65.52%
Specificity	Cross Validation (10 fold)	a=neg	84.45%	81.62%	73.88%	78.89%	60.87%
		b=neu	86.83%	75.55%	85.07%	80.61%	82.40%
		c=pos	94.63%	95.13%	89.08%	84.38%	93.78%
	Percentage Split (80-20)%	a=neg	87.72%	89.47%	80.00%	81.82%	66.67%
		b=neu	87.27%	72.22%	79.63%	78.43%	87.23%
		c=pos	92.00%	97.87%	91.11%	87.23%	77.08%
PPV	Cross Validation (10 fold)	a=neg	72.67%	66.67%	62.57%	62.25%	52.00%
		b=neu	75.17%	63.19%	68.50%	65.07%	64.23%
		c=pos	90.71%	90.68%	81.62%	77.12%	87.25%
	Percentage Split (80-20)%	a=neg	75.86%	73.91%	64.52%	64.29%	56.41%
		b=neu	77.42%	65.91%	65.63%	67.65%	71.43%
		c=pos	86.67%	95.65%	85.19%	78.57%	63.33%
NPV	Cross Validation (10 fold)	a=neg	87.87%	81.62%	85.71%	79.18%	83.58%
		b=neu	86.52%	85.54%	78.35%	79.40%	74.37%
		c=pos	90.87%	83.33%	83.95%	85.52%	76.26%
	Percentage Split (80-20)%	a=neg	89.29%	82.26%	84.62%	81.82%	85.00%
		b=neu	84.21%	90.70%	78.18%	80.00%	69.49%
		c=pos	93.88%	86.79%	87.23%	85.42%	78.72%
Average (%)			84.52%	80.09%	77.34%	75.85%	71.36%

6.4 Alphabetic Tokenizer

Using "Alphabetic Tokenizer" in the techniques like (SVM, Naïve, KNN, J48, DT) it is compared with different results like (accuracy, sensitivity, specificity, ppv and npv) as it has been done in case of word ,ngram,character ngram tokenizer but it does not have any additional feature like (min and max size) or delimiters . Here it is observed that SVM provided a better result in percentage split (80-20) % and average it is 86.39%, where Naïve (80.92%), KNN (71.40%), J48 (75.80%) and DT (64. 39%) in Table 6.

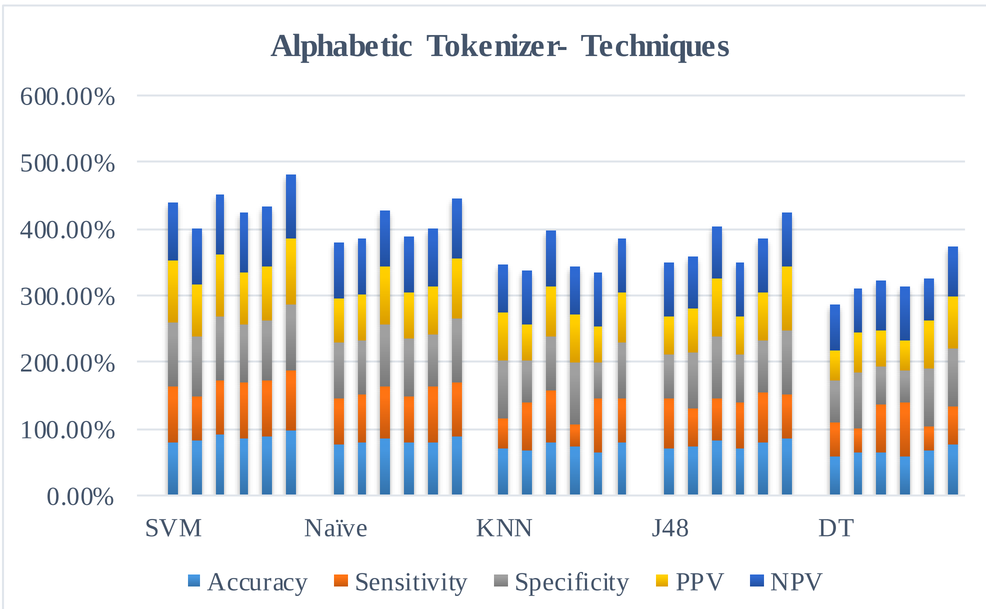


Figure 26: Alphabetic Tokenizer- Techniques

In case of measures(average) specificity (82.29%) where accuracy (76.50%), sensitivity (68.33%), ppv (71.80%) and npv (81.34%)

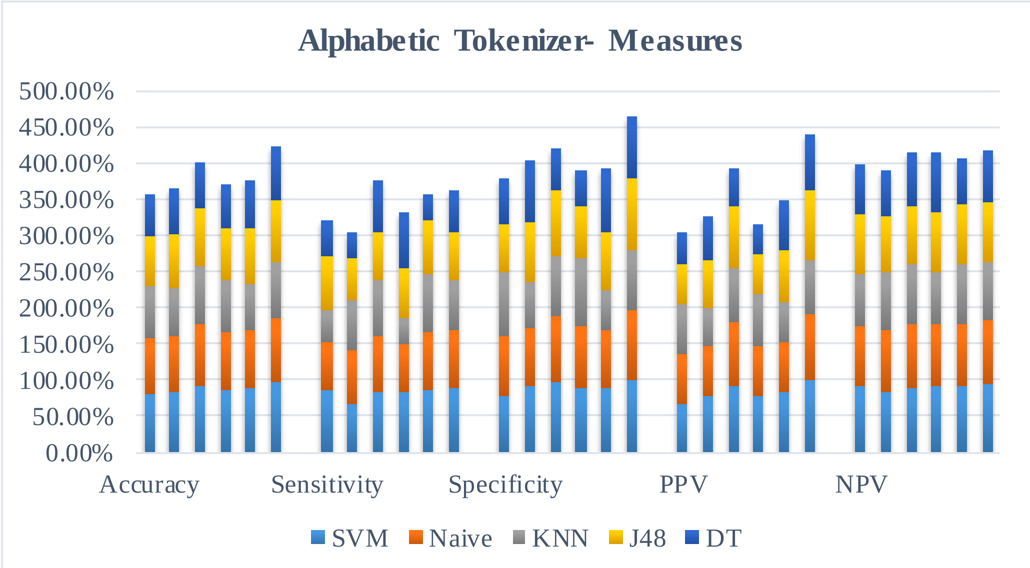


Figure 27: Alphabetic Tokenizer- Measures

Table 6: Alphabetic Tokenizer

Classification Methods			SVM	Naive	KNN	J48	DT
Accuracy	Cross Validation (10 fold)	a=neg	80.28%	77.28%	71.75%	69.79%	58.75%
		b=neu	81.78%	78.38%	67.06%	73.76%	64.21%
		c=pos	90.67%	86.39%	79.94%	81.20%	63.69%
	Percentage Split (80-20)%	a=neg	86.52%	80.23%	71.79%	71.59%	59.30%
		b=neu	87.50%	80.23%	65.12%	77.78%	66.23%
		c=pos	96.25%	89.61%	77.78%	85.14%	75.00%
Sensitivity	Cross Validation (10 fold)	a=neg	84.67%	68.67%	42.67%	75.33%	50.00%
		b=neu	66.00%	74.00%	71.33%	58.00%	34.67%
		c=pos	82.67%	77.33%	77.33%	65.33%	72.00%
	Percentage Split (80-20)%	a=neg	82.14%	67.86%	35.71%	67.86%	78.57%
		b=neu	84.85%	81.82%	78.79%	75.76%	36.36%
		c=pos	89.66%	79.31%	68.97%	65.52%	58.62%
Specificity	Cross Validation (10 fold)	a=neg	77.97%	81.95%	89.20%	66.79%	64.00%
		b=neu	90.29%	80.81%	64.75%	83.07%	84.72%
		c=pos	95.76%	92.24%	81.82%	92.17%	57.99%
	Percentage Split (80-20)%	a=neg	88.52%	86.21%	92.00%	73.33%	50.00%
		b=neu	89.09%	79.25%	56.60%	79.17%	88.64%
		c=pos	100.00%	95.83%	83.72%	97.78%	87.18%

PPV	Cross Validation (10 fold)	a=neg	66.84%	67.32%	70.33%	55.12%	45.45%
		b=neu	78.57%	68.10%	52.20%	66.92%	61.18%
		c=pos	92.54%	86.57%	75.32%	85.22%	54.00%
	Percentage Split (80-20)%	a=neg	76.67%	70.37%	71.43%	54.29%	43.14%
		b=neu	82.35%	71.05%	53.06%	71.43%	70.59%
		c=pos	100.00%	92.00%	74.07%	95.00%	77.27%
NPV	Cross Validation (10 fold)	a=neg	90.65%	82.85%	72.17%	83.33%	68.09%
		b=neu	83.11%	84.88%	80.72%	77.01%	65.12%
		c=pos	89.68%	86.29%	83.41%	79.37%	75.15%
	Percentage Split (80-20)%	a=neg	91.53%	84.75%	71.88%	83.02%	82.86%
		b=neu	90.74%	87.50%	81.08%	82.61%	65.00%
		c=pos	94.44%	88.46%	80.00%	81.48%	73.91%
Average (%)			86.39%	80.92%	71.40%	75.80%	64.39%

CHAPTER 7

CONCLUSION

Sentimental Analysis is an area of data extraction to get knowledge from huge amount of data from users in the form of text. Although there are many limitations exist in sentiment classification, many difficulties and gap like spelling mistake, grammar mismatch, writing technique, regional language, fake detection feedbacks etc. which using Weka tool it has reduced most of the limitations like google translator, various tokenizers, delimiters etc. In supervised learning (classification techniques) Support Vector Machine (SVM) and Naïve Bayes (NB) is more popular. Other techniques like KNN ,decision tree, J48 is also an important to classify data and here it is observed SVM is providing a better result with sensitivity in percentage split(80-20)% ,it has been observed that after using various tokenizer in weka tool(alphabetic tokenizer) has provided better result in measure i.e SVM (86.39%) in Table 6 comparing to techniques & specificity (83.77%) in average comparing to other measures.

For the future research areas of opinion mining scope is researched which is sentimental analysis is in trend, that is feedbacks of "twitter" comments or statements, status or comments on pictures, videos, status etc. and even on other online shopping sites product reviews. Sentimental analysis can be done on pictures and videos then researcher will use all these techniques to examine and that will provide a clear picture to future research which helps to get users recommend to make a better selection that can be also called "Recommendation System. [47]

.

CHAPTER 8
BIBLIOGRAPHY

[1] C. S. and S. C., "Opinion Mining and Sentiment Classification: a Survey," *ICTACT J. Soft Comput.*, vol. 3, no. 1, pp. 420–427, 2012.

[2] R. Sharma, S. Nigam, and R. Jain, "Supervised Opinion Mining Techniques : A Survey," vol. 3, no. 8, pp. 737–742, 2013.

[3] A. Abbasi, "AI and opinion mining, part 2," *IEEE Intell. Syst.*, vol. 25, no. 4, pp. 72–79, 2010.

[4] P. Tumsare, A. S. Sambare, and S. R. Jain, "Web Site : www.ijettcs.org Email : editor@ijettcs.org OPINION MINING IN NATURAL ANGUAGE PROCESSING USING SENTIWORDNET AND FUZZY," vol. 3, no. 3, 2014.

[5] M. S. Sagagi, "Investigating the Determinants of Wealth Creation in Nigeria: Does Oil Revenue Allocation Really Matter?," *J. Entrep. Innov. Emerg. Econ.* , vol. 1, no. 1, pp. 114–128, 2015.

[6] D. Zeng and R. Lusch, "Big data analytics: Perspective shifting from transactions to ecosystems," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 2–5, 2013.

[7] G. Tripathi and S. Naganna, "Opinion Mining: A Review," *Irphouse.Com*, vol. 4, no. 16, pp. 1625–1635, 2014.

[8] M. Himmat and N. Salim, "Obstacles and Challenges of Opinion Mining of Customers Reviews," *Int. J. Enhanc. Res. Sci. Technol. Eng.*, vol. 2, no. 2, pp. 1–4, 2013.

[9] B. Vidyapith, "Opinion Mining of Movie Reviews at Document Level," *Int. J. Inf. Theory*, vol. 3, no. 3, pp. 13–21, 2014.

[10] M. R. Jose, "Co-Extracting Opinions from Online Reviews," vol. 5, no. 2, pp. 95–98, 2016.

[11] T. S. Pawar, R. G. Sawant, P. S. Bothe, and Sh. A. Chopade, "A Survey on Login

Authentication System using Captcha as Graphical Password Technquies," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 3, no. 10, pp. 10131–10138, 2015.

[12] K. Patel, Vishakha, Prabhu, Gayatri, Bhowmick, "A Suvery of opinion mining and sentiment analysis," *Int. J. Comput. Appl.*, vol. 131, no. 1, pp. 24–27, 2015.

[13] B. P. and L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends® Inf. Retr.*, vol. 1, no. 2, pp. 91–231, 2006.

[14] T. Bhuiyan, Y. Xu, and A. Jøsang, "QUT Digital Repository : State-of-the-Art Review on Opinion Mining from Online Customers ' Feedback," *9th Asia-Pacific Complex Syst. Conf.*, no. November, pp. 4–7, 2009.

[15] G. Stylios *et al.*, "Public Opinion Mining for Governmental Decisions," *Electron. J. e-Government*, vol. 8, no. 2, pp. 203–214, 2010.

[16] K. Khan, B. Baharudin, A. Khan, and A. Ullah, "Mining opinion components from unstructured reviews: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 3, pp. 258–275, 2014.

[17] Z. Zhai, B. Liu, J. Wang, H. Xu, and P. Jia, "Product feature grouping for opinion mining," *IEEE Intell. Syst.*, vol. 27, no. 4, pp. 37–44, 2012.

[18] D. Virmani, V. Malhotra, and R. Tyagi, "Sentiment Analysis Using Collaborated Opinion Mining," *arXiv Prepr. arXiv1401.2618*, 2014.

[19] W. Wei, "Analyzing text data for opinion mining," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6716 LNCS, pp. 330–335, 2011.

[20] E. M. Taylor, C. Rodríguez O., J. D. Velásquez, G. Ghosh, and S. Banerjee, "Web opinion mining and sentimental analysis," *Stud. Comput. Intell.*, vol. 452, pp. 105–126, 2013.

[21] P. Sudhakaran *et al.*, "Research Directions, Challenges and Issues in Opinion Mining," *Int. J. Adv. Sci. Technol.*, vol. 60, pp. 1–8, 2013.

[22] S. V. Angiuoli, J. R. White, M. Matalka, O. White, and W. F. Fricke, "Resources and costs for microbial sequence analysis evaluated using virtual machines and cloud computing," *PLoS One*, vol. 6, no. 10, pp. 1320–1326, 2011.

[23] D. Osimo and F. Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis," *Using open data policy Model. Citiz. Empower. data Journal.*, 2012.

[24] S. Poobana and S. Rekha, "Opinion Mining From Text Reviews Using Machine Learning Algorithm," pp. 1567–1570, 2015.

[25] E. Zhang and Y. Zhang, "UCSC on TREC 2006 blog opinion mining," *Fifteenth Text Retr. Conf. Proc. (TREC 2006)', Gaithersburg, Maryl.*, pp. 1–3, 2006.

[26] S. Vasantharaj, A. M. Associate, and J. Madhusudnan, "A Survey on Sentiment Analysis Applied in Opinion Mining," *J. Netw. Commun. Emerg. Technol.*, vol. 1, no. 1, pp. 15–21, 2015.

[27] H. Shaziya, G. Kavitha, and R. Zaheer, "Text Categorization of Movie Reviews for Sentiment Analysis," pp. 11255–11262, 2015.

[28] B. Liu, "Sentiment Analysis and Opinion Mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[29] K. M. Rahman and A. Khamparia, "Techniques , Applications and Challenges of Opinion Mining," vol. 9, no. 41, pp. 455–461, 2016.

[30] P. K. Singh and M. Shahid Husain, "Methodological Study Of Opinion Mining And Sentiment Analysis Techniques," *Int. J. Soft Comput.*, vol. 5, no. 1, pp. 11–21, 2014.

[31] S. L. Prabha and A. R. M. Shanavas, "Ijcsit2015060562," vol. 6, no. 5, pp. 4440–4443, 2015.

[32] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, and K. Filippova, "Multi-lingual opinion mining on YouTube," *Inf. Process. Manag.*, vol. 52, no. 1, pp. 46–60, 2016.

[33] F. Ren and Y. Wu, "Predicting user-topic opinions in twitter with social and topical context," *IEEE Trans. Affect. Comput.*, vol. 4, no. 4, pp. 412–424, 2013.

[34] M. Ravichandran, G. Kulanthaivel, and T. Chellatamilan, “Intelligent topical sentiment analysis for the classification of e-learners and their topics of interest,” *Sci. World J.*, vol. 2015, no. 5, 2015.

[35] B. Liu, “Sentiment Analysis and Opinion Mining,” *Morgan Claypool Publ.*, no. May, 2012.

[36] N. Mishra and C. K. Jha, “Classification of Opinion Mining Techniques,” *Int. J. Comput. Appl.*, vol. 56, no. 13, pp. 975–8887, 2012.

[37] N. Samsudin, M. Puteh, A. R. Hamdan, and M. Z. A. Nazri, “Immune based feature selection for opinion mining,” *Lect. Notes Eng. Comput. Sci.*, vol. 3 LNECS, pp. 1520–1525, 2013.

[38] E. R. Kaur, E. T. Kaur, and E. H. Kaur, “International Journal of Advanced Research in Computer Science and Software Engineering,” *Vol. 3, Issue 2, Febr. 2013 ISSN 2277 128X Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 2, pp. 444–448, 2013.

[39] R. K. Maurya, M. R. Saxena, and N. Akhil, “Intelligent Systems Technologies and Applications,” *Adv. Intell. Syst. Comput.*, vol. 384, pp. 247–257, 2016.

[40] B. Liu, “Opinion Mining and Search,” *Present. slides,* 2007.

[41] Z. Kechaou and M. Ammar, “Improving e-learning with sentiment analysis of users opinions. Global Engineering Education Conference (EDUCON)--Learning Environments and,” pp. 1032–1038, 2011.

[42] Weilbach, “Spelling of Names and Affiliations, Accuracy of Chapter Titles and Subtitles, Figure/Table Accuracy, Minor Spelling Errors/Typos, Equation Display,” *Int. J. Cloud Appl. Comput.*, vol. 4, no. 1, pp. 33–44, 2014.

[43] H. V. Thakkar, “Twitter Sa Using Hybrid Naive Bayes.”

[44] S. Versus and U. Methods, “k-NEAREST NEIGHBOR ALGORITHM,” pp. 90–106.

[45] J. Prager, “Open-Domain Question–Answering,” *Found. Trends® Inf. Retr.*, vol. 1, no. 2, pp. 91–231, 2006.

[46] P. Gowtamreddy, "Opinion Mining of Online Customer Reviews," pp. 1–44, 2014.

[47] A. Khamparia and B. Pandey, "Knowledge and intelligent computing methods in e-learning," *Int. J. Technol. Enhanc. Learn.*, vol. 7, no. 3, pp. 221–242, 2015.

[1] C. S. and S. C., "Opinion Mining and Sentiment Classification: a Survey," *ICTACT J. Soft Comput.*, vol. 3, no. 1, pp. 420–427, 2012.

[2] R. Sharma, S. Nigam, and R. Jain, "Supervised Opinion Mining Techniques : A Survey," vol. 3, no. 8, pp. 737–742, 2013.

[3] A. Abbasi, "AI and opinion mining, part 2," *IEEE Intell. Syst.*, vol. 25, no. 4, pp. 72–79, 2010.

[4] P. Tumsare, A. S. Sambare, and S. R. Jain, "Web Site : www.ijettcs.org Email : editor@ijettcs.org OPINION MINING IN NATURAL ANGUAGE PROCESSING USING SENTIWORDNET AND FUZZY," vol. 3, no. 3, 2014.

[5] M. S. Sagagi, "Investigating the Determinants of Wealth Creation in Nigeria: Does Oil Revenue Allocation Really Matter?," *J. Entrep. Innov. Emerg. Econ.* , vol. 1, no. 1, pp. 114–128, 2015.

[6] D. Zeng and R. Lusch, "Big data analytics: Perspective shifting from transactions to ecosystems," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 2–5, 2013.

[7] G. Tripathi and S. Naganna, "Opinion Mining: A Review," *Irphouse.Com*, vol. 4, no. 16, pp. 1625–1635, 2014.

[8] M. Himmat and N. Salim, "Obstacles and Challenges of Opinion Mining of Customers Reviews," *Int. J. Enhanc. Res. Sci. Technol. Eng.*, vol. 2, no. 2, pp. 1–4, 2013.

[9] B. Vidyapith, "Opinion Mining of Movie Reviews at Document Level," *Int. J. Inf. Theory*, vol. 3, no. 3, pp. 13–21, 2014.

[10] M. R. Jose, "Co-Extracting Opinions from Online Reviews," vol. 5, no. 2, pp. 95–98, 2016.

[11] T. S. Pawar, R. G. Sawant, P. S. Bothe, and Sh. A. Chopade, “A Survey on Login Authentication System using Captcha as Graphical Password Technquies,” *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 3, no. 10, pp. 10131–10138, 2015.

[12] K. Patel, Vishakha, Prabhu, Gayatri, Bhowmick, “A Suvery of opinion mining and sentiment analysis,” *Int. J. Comput. Appl.*, vol. 131, no. 1, pp. 24–27, 2015.

[13] B. P. and L. Lee, “Opinion Mining and Sentiment Analysis,” *Found. Trends® Inf. Retr.*, vol. 1, no. 2, pp. 91–231, 2006.

[14] T. Bhuiyan, Y. Xu, and A. Jøsang, “QUT Digital Repository : State-of-the-Art Review on Opinion Mining from Online Customers ’ Feedback,” *9th Asia-Pacific Complex Syst. Conf.*, no. November, pp. 4–7, 2009.

[15] G. Stylios *et al.*, “Public Opinion Mining for Governmental Decisions,” *Electron. J. e-Government*, vol. 8, no. 2, pp. 203–214, 2010.

[16] K. Khan, B. Baharudin, A. Khan, and A. Ullah, “Mining opinion components from unstructured reviews: A review,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 3, pp. 258–275, 2014.

[17] Z. Zhai, B. Liu, J. Wang, H. Xu, and P. Jia, “Product feature grouping for opinion mining,” *IEEE Intell. Syst.*, vol. 27, no. 4, pp. 37–44, 2012.

[18] D. Virmani, V. Malhotra, and R. Tyagi, “Sentiment Analysis Using Collaborated Opinion Mining,” *arXiv Prepr. arXiv1401.2618*, 2014.

[19] W. Wei, “Analyzing text data for opinion mining,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6716 LNCS, pp. 330–335, 2011.

[20] E. M. Taylor, C. Rodríguez O., J. D. Velásquez, G. Ghosh, and S. Banerjee, “Web opinion mining and sentimental analysis,” *Stud. Comput. Intell.*, vol. 452, pp. 105–126, 2013.

[21] P. Sudhakaran *et al.*, “Research Directions, Challenges and Issues in Opinion Mining,”

Int. J. Adv. Sci. Technol., vol. 60, pp. 1–8, 2013.

[22] S. V. Angiuoli, J. R. White, M. Matalka, O. White, and W. F. Fricke, "Resources and costs for microbial sequence analysis evaluated using virtual machines and cloud computing," *PLoS One*, vol. 6, no. 10, pp. 1320–1326, 2011.

[23] D. Osimo and F. Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis," *Using open data policy Model. Citiz. Empower. data Journal.*, 2012.

[24] S. Poobana and S. Rekha, "Opinion Mining From Text Reviews Using Machine Learning Algorithm," pp. 1567–1570, 2015.

[25] E. Zhang and Y. Zhang, "UCSC on TREC 2006 blog opinion mining," *Fifteenth Text Retr. Conf. Proc. (TREC 2006)', Gaithersburg, Maryl.*, pp. 1–3, 2006.

[26] S. Vasantharaj, A. M. Associate, and J. Madhusudnan, "A Survey on Sentiment Analysis Applied in Opinion Mining," *J. Netw. Commun. Emerg. Technol.*, vol. 1, no. 1, pp. 15–21, 2015.

[27] H. Shaziya, G. Kavitha, and R. Zaheer, "Text Categorization of Movie Reviews for Sentiment Analysis," pp. 11255–11262, 2015.

[28] B. Liu, "Sentiment Analysis and Opinion Mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[29] K. M. Rahman and A. Khamparia, "Techniques , Applications and Challenges of Opinion Mining," vol. 9, no. 41, pp. 455–461, 2016.

[30] P. K. Singh and M. Shahid Husain, "Methodological Study Of Opinion Mining And Sentiment Analysis Techniques," *Int. J. Soft Comput.*, vol. 5, no. 1, pp. 11–21, 2014.

[31] S. L. Prabha and A. R. M. Shanavas, "Ijcsit2015060562," vol. 6, no. 5, pp. 4440–4443, 2015.

[32] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, and K. Filippova, "Multi-lingual opinion mining on YouTube," *Inf. Process. Manag.*, vol. 52, no. 1, pp. 46–60, 2016.

[33] F. Ren and Y. Wu, "Predicting user-topic opinions in twitter with social and topical context," *IEEE Trans. Affect. Comput.*, vol. 4, no. 4, pp. 412–424, 2013.

[34] M. Ravichandran, G. Kulanthaivel, and T. Chellatamilan, "Intelligent topical sentiment analysis for the classification of e-learners and their topics of interest," *Sci. World J.*, vol. 2015, no. 5, 2015.

[35] B. Liu, "Sentiment Analysis and Opinion Mining," *Morgan Claypool Publ.*, no. May, 2012.

[36] N. Mishra and C. K. Jha, "Classification of Opinion Mining Techniques," *Int. J. Comput. Appl.*, vol. 56, no. 13, pp. 975–8887, 2012.

[37] N. Samsudin, M. Puteh, A. R. Hamdan, and M. Z. A. Nazri, "Immune based feature selection for opinion mining," *Lect. Notes Eng. Comput. Sci.*, vol. 3 LNECS, pp. 1520–1525, 2013.

[38] E. R. Kaur, E. T. Kaur, and E. H. Kaur, "International Journal of Advanced Research in Computer Science and Software Engineering," *Vol. 3, Issue 2, Febr. 2013 ISSN 2277 128X Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 2, pp. 444–448, 2013.

[39] R. K. Maurya, M. R. Saxena, and N. Akhil, "Intelligent Systems Technologies and Applications," *Adv. Intell. Syst. Comput.*, vol. 384, pp. 247–257, 2016.

[40] B. Liu, "Opinion Mining and Search," *Present. slides,* 2007.

[41] Z. Kechaou and M. Ammar, "Improving e-learning with sentiment analysis of users opinions. Global Engineering Education Conference (EDUCON)--Learning Environments and," pp. 1032–1038, 2011.

[42] Weilbach, "Spelling of Names and Affiliations, Accuracy of Chapter Titles and Subtitles, Figure/Table Accuracy, Minor Spelling Errors/Typos, Equation Display," *Int. J. Cloud Appl. Comput.*, vol. 4, no. 1, pp. 33–44, 2014.

[43] H. V. Thakkar, "Twitter Sa Using Hybrid Naive Bayes."

[44] S. Versus and U. Methods, "k-NEAREST NEIGHBOR ALGORITHM," pp. 90–106.

[45] J. Prager, "Open-Domain Question–Answering," *Found. Trends® Inf. Retr.*, vol. 1, no. 2, pp. 91–231, 2006.

[46] P. Gowtamreddy, "Opinion Mining of Online Customer Reviews," pp. 1–44, 2014.

[47] A. Khamparia and B. Pandey, "Knowledge and intelligent computing methods in e-learning," *Int. J. Technol. Enhanc. Learn.*, vol. 7, no. 3, pp. 221–242, 2015.

[48] https://www.promptcloud.com/scraping-webpage-for-sentiment-analysis dt.21-10-2016.

[49] https://www.youtube.com/watch?v=Dji9ALCgfpM dt.22-03-2017.

[50] http://www.keywordhut.com/c2hhbWUgZW1vdGljb24/ dt.05-03-2017