Sentimental Analysis On Social Network Feeds

Dissertation - I

Report Submitted

By

Neha Gigi

To

Department of Computer Science and Engineering

In Partial Fulfillment of the Requirement of the Award of the degree of

Master of Technology in Computer Science and Engineering.

Under the Guidance of

Mr. Anupinder Singh

(Asst. Professor)

(April 2018)

LOVELY PROFESSIONAL UNIVERSITY

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

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

COURSE CODE : CSE545 **REGULAR/BACKLOG :** Backlog **GROUP NUMBER :** CSEBGD0405

Supervisor Name : Anupinder Singh **UID :** 19385 **Designation :** Assistant Professor

Qualification : _____ **Research Experience :** _____

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Neha Gigl	11508659	2015	KN708	9400707390

SPECIALIZATION AREA : System Architecture and Design **Supervisor Signature:** _____

PROPOSED TOPIC : Sentiment analysis of social media feeds

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	8.00
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	7.00
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.	8.00
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	9.00
5	Social Applicability: Project work intends to solve a practical problem.	8.00
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	8.00

PAC Committee Members		
PAC Member 1 Name: Er.Dalwinder Singh	UID: 11265	Recommended (Y/N): NA
PAC Member 2 Name: Harwant Singh Arri	UID: 12975	Recommended (Y/N): NA
PAC Member 3 Name: Harleen Kaur	UID: 14508	Recommended (Y/N): Yes
PAC Member 4 Name: Jangreet Singh	UID: 11266	Recommended (Y/N): NA
PAC Member 5 Name: Tejinder Thind	UID: 15312	Recommended (Y/N): NA
DAA Nominee Name: Kuldeep Kumar Kushwaha	UID: 17118	Recommended (Y/N): NA

Final Topic Approved by PAC: Sentiment analysis of social media feeds

Overall Remarks: Approved

PAC CHAIRPERSON Name: 11024::Amandeep Nagpal **Approval Date:** 30 Apr 2018

4/30/2018 1:30:16 PM

Acknowledgement

I have taken efforts in this Dissertation-1. However, it would not have been possible without the kind support and help of my Mentor **Asst. Pro. Anupinder Singh**. I would like to extend my sincere thanks to them.

I am highly indebted to my Mentor **Asst. Pro. Anupinder Singh** for his guidance and constant supervision as well as for providing necessary information regarding the Dissertation and also for his support in completing the Dissertation-1.

I would like to express my gratitude towards my parents and members of Lovely Professional University for their kind co-operation and encouragement to help me in completion of this Dissertation.
My thanks and appreciations also go to my colleague in doing the Dissertation and people who have willingly helped me out with their abilities.

CERTIFICATE

This is certified that Neha Gigi has completed the M.Tech Dissertation-1 report on **Sentimental Analysis On Social Network Feeds** 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 submission and partial fulfillment of the condition for award of M.Tech Computer Science and Engineering.

Date:-28/04/2018

Signature of Advisor

Mr. **Anupinder Singh**

(Asst. Professor)

DECLARATION

I hereby declare that the dissertation-1 proposed entitled, Sentimental Analysis on Social Network Feeds submitted for M.Tech. Degree is entirely my original work and all idea and references have been dully acknowledged. It does not contain any work for the award of any degree or diploma.

Date: 28/04/2018

Neha Gigi

Reg.No. 11508659

List Of Contents

List of Figures

CHAPTER 1

Introduction

Data mining is a task which is used to find out the hidden pattern / information to analysis any subject. Now a day's a lot of research is going on web mining i.e.to mine the web content for analysis. Web mining can be further classified into following categories i.e.,

- Static web mining
- Dynamic web mining, dynamic web mining is also known as Data Stream (DS)[1].

Stream mining is an area that gaining lots of practical significance and finding various application areas related to medicine, computer science, stock market prediction, online data generation etc. Since in web technology (stream data) has a challenging task because they are real time data which changes rapidly over the time.[1] In stream mining huge amount of online data is generated from several things like sensors, internet relay chat, twitter, Facebook, online transactions etc.

Stream mining over twitter data is a area where lots of research is going on because twitter is miniaturized scale blogging administration that checks with a huge number of clients from everywhere throughout the world. Sentimental analysis has improved in the last few years as well as its applications [6].This is used for product marketing for recognition of anti-social behavior. The advances in Facebook, twitter, YouTube and other smaller scale blogging and long range informal communication destinations have contributed change to the social locales as well as have in a general sense changed the way we utilize these locales and how we share our emotions, our perspectives with the more extensive gathering of people[2]. A huge number of messages are streaming day by day in prominent sites that provide access to smaller scale blogging. Users write about their life and share opinions on variety of topics and discuss current issues happened in the world. In light of a free arrangement of messages and a simple availability of miniaturized scale blogging stages, Internet clients tend to move from customary specialized instruments, (for example, conventional web journals or mailing records) to small scale blogging

administrations. The same number of clients post about items and administrations they utilize, or express their political and religious perspectives, smaller scale blogging sites get to be distinctly profitable wellsprings of individuals' assessments and suppositions.[4] Such information can be productively utilized for versatile UI. Data we get from these sources can be used in opinion mining and sentiment analysis tasks. For example:

• What do people think about these persons who post their status (comments)?
• How positive (or negative) are people about anything?

Previous work incorporates by the author Turney and Pang who connected the diverse techniques to identifying the extremity of item surveys of the product and film audits individually. There work is at the record level. One can likewise arrange an archive's extremity on a multi-way scale, which was endeavored by Pang and Snyder among others: Pang and Lee extended an essential assignment of grouping a motion picture survey either positive or negative to foresee star evaluations on either a 3 or a 4 star scale, while Snyder played out a top to bottom investigation of eatery audits, anticipating appraisals for different parts of the given eatery, for example, the sustenance and air (on a five-star scale).[3] Despite the fact that in most factual arrangement techniques, the impartial class is overlooked under the presumption that unbiased writings that lie to close the limit of the paired classifier in which a few analysts recommend that, in each extremity issue the three most important classifications must be recognized. Additionally, it can be demonstrated on particular classifiers, for example, the Max Entropy and the SVM can profit by the presentation of a nonpartisan class and enhance the general exactness of the characterization. An alternate strategy for deciding slant is the utilization by a scaling framework where the words ordinarily connected with a negative, nonpartisan or positive notion with them were given a related number as a - 10 to +10 scale (most negative to best This makes it conceivable to changed slant by a given term with respect with its surroundings (for a most part on that level of the sentence). At the point when a bit of no structured content is examined utilizing common dialect preparing, every idea in the predetermined environment is given the score to view the way of feeling words that identify with an idea and its related score.[4] This permits development to a more modern comprehension of opinion, since it is currently conceivable to conform the assumption estimation of an idea with respect to changes that may encompass it. Words, for instance, that escalate, unwind or refute the estimation communicated by the idea can influence its score.[5]

CHAPTER 2

Literature Review

Shoiab Ahmed and Ajit Danti, "Novel Approach for Sentimental Analysis and Opinion Mining based on Sentimental WordNet using Web Data", 2016, [12] in this paper the author novel approach is proposed in view of Sentiment WordNet, which produces tally of score words into seven classifications, for example, solid positive, positive, feeble positive, impartial, frail negative, negative and solid negative words for the supposition mining undertaking and assessed utilizing machine learning calculations like Naïve Bayes, SVM and Multilayer Perception (MLP). The proposed approach is probed motion picture and item web areas and got higher achievement rate regarding precision measured by different devices like Kappa insights with an exactness of 77.7% and has brought down mistake rates. Weighted normal of various exactness measures like Precision, Recall, TP Rate, F-Measure rate portrays higher proficiency rate and lower FP Rate for Naïve Bayes and MLP models. The creator first gathers the information from the web source then evacuate the stop word i.e. pre-handle the information and apply the pre-characterized calculation to get the outcome.

C. Castillo et al. [13] proposed the methodology to check or analysis the credibility of news propagated through Twitter. The main target of this is to check the credibility of data spread through social media network. They first collect the data (tweets) from the twitter using twitter API then after they twitter monitor to detect the tweets of about a 2-months period. Twitter Monitor is the on-line monitoring software system which detects the increases ("bursts") in the frequency of keywords found in messages. They collected every tweet which matches the query under a 2-day window centered on the peak of every burst. Newsworthy theme assessment was their first marking round was expected to isolate the subjects who spread data about a news occasion, from the cases which relate to individual suppositions and visit. Over different word, they separated the messages that need aid for possibility premium to an expansive situated for people, starting with discussions that are of minimal significance outside for this task we used Mechanical Turk8.

"Data Mining- Past, Present and Future" by M.S.B. Phridvi Raj [20] is generally a review paper in which the author try to catch attention to the importance of the data mining specially the stream mining. This paper provides detail information that how the concept of data mining is use in past and present to fetch the information from the bulk of storage or data as well as try to provide the future of research work which can be done. The paper tells about the data stream which can be classified into two types first one is Offline data stream and the second one is online data stream and the different types of algorithm which is used in the data mining as well as their disadvantages such as Frequent pattern, Naive Bayes, Decision Tree etc. The author try to provide the information that the algorithm which is used for Offline data mining is not suitable for the Online data mining because in the online world the data change rapidly according to time so the Online stream data mining is a challenging task in now a days. The paper also discuss some important challenge occur in the stream data mining such as handling stream clustering , classification and topic detection which is the major research topic for the future work in field.

"Mapping Of Real State Price Using Data Mining Techniques" by Eduard Hromada [16] in this paper author provide that how, Data Mining techniques is used in the mapping of real estate price i.e. the paper describe the software which is used for actual state assessment, mapping and analyzing of actual estate ad distributed on the web before developing system the author make a proper Literature Survey to find how the price and advertisement is differ from actual fact .The basic function of the software can be explain below

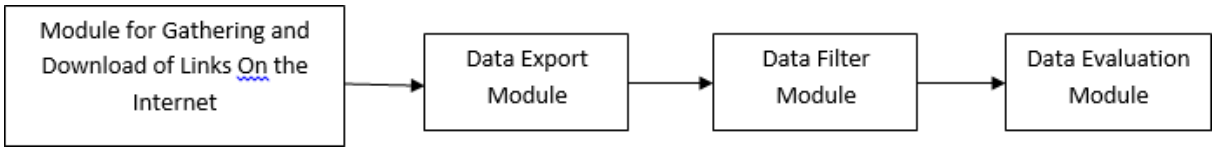

Fig:2.1 The basic structure of software

The information obtaining is scent in the accompanying Categories:-

- Apartment For Deal.
- Apartment For Rent.
- Apartment For Deal.
- Allotment For Deal.
- Corporate plots For Deal .

So by the help of output which is generated by the software contain the above module is used as ground for allocate investment or holder decision for both familiar person as well as companies. The paper also provides the challenges which was faced by the developer and still some problem which can be used for the next research work for the same.

"A Quick Review Of Data Stream Mining Algorithms" by Dr. Rajeev Tripathi & Dr. Santosh [22] in this the author proposed a review about the limitation of traditional data in supporting streaming application have been understood and why we developed the new system for manage the stream data. The paper provided the information that how data mining affects the data streaming, the author review the whole process for data stream for data, proper algorithm and model are analyzed. This paper provided the basics algorithm used in the data streams which can be listed as Segmentation, Clustering, Sampling, Sketching, Aggregation Sliding Window, Damped Window Algorithm Output Granularity. The paper also provide the challenges such as visualization of data, Efficient querying Mechanism, real data are irregular and unpredictable hence an algorithm should be able to manage the traffic, the technique should be intelligence to differentiate between noise and concept etc. The above challenge given by the author in this paper can be our future work.

"Visualization Of Streaming Data Using Social Media" by V.Punna Rao & Sagar Galandein [17] this paper the author proposed a new system/method to discuss the existing analysis of twitter and facebook dataset with text mining approach such as Natural Language Processing Algorithm and R- Programming . They give a new approach that automatically classified the sentiments of tweets, posts, comments taken from the twitter as well as facebook dataset .These tweets and comments are classifies into positive, negative and neutral with respect to a query term of people using graph. The tool is divided into four parts first one data which is

taken from twitter/ facebook, second one is Natural Language Processing which is used to derived meaning by the computer such as done by human. Third one is R programming which used for statistical computing and graphics as well as analysis the data and the last one is Lucene which is simple a java based search library which can added in any tool to make the capability of search. The main proposed methodology given by the author in this paper was explain by the help of flow chart.

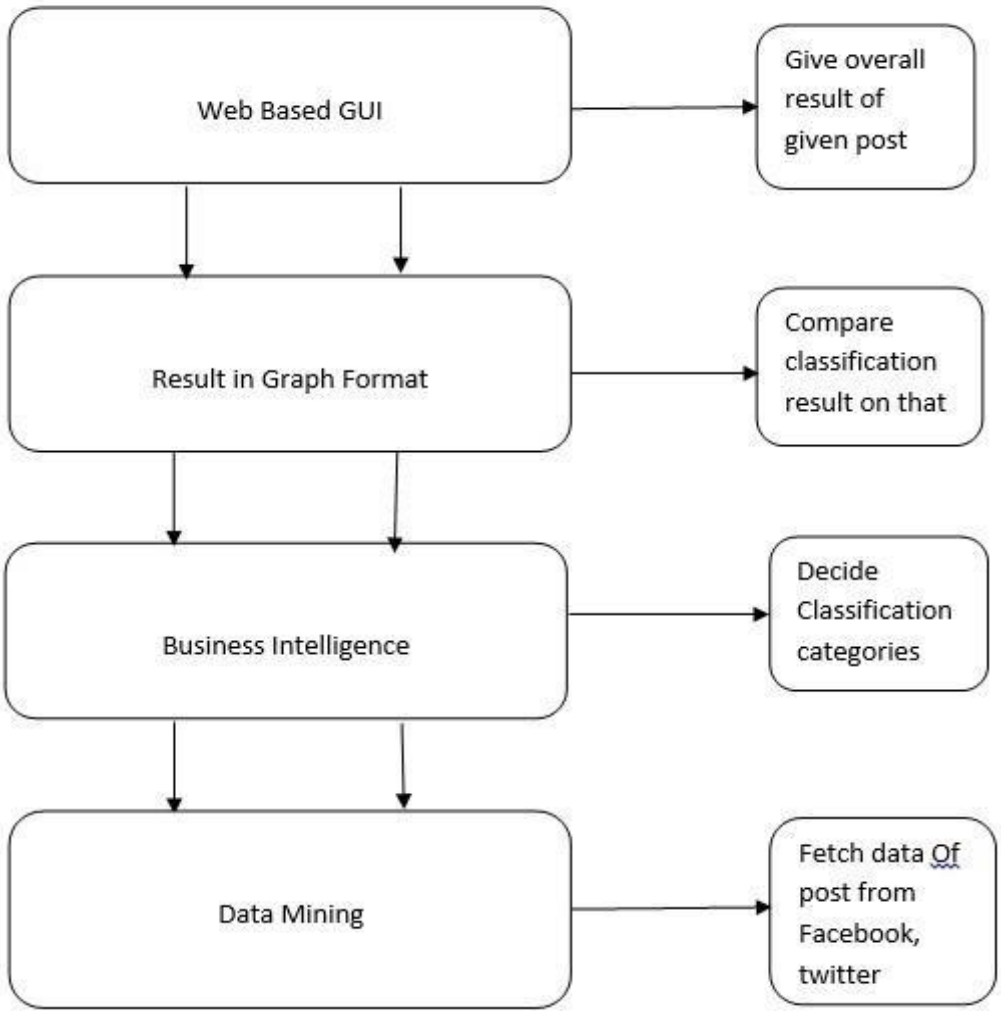

Figure 2.2: System Architecture

"Data Mining Application In E-Government Information System" by Tongwei Yuan & Peng Chen [19] in this paper the author proposed a present more develop Association Analysis Model after it gives an outline of information mining technique and behaviors a formal depiction. This paper told the significance of the security of the information as in now days the

entire world relies on upon web framework so security of information is the real worry to keep the framework from the infection assault, the spillage of privileged insights, framework disappointment. This paper also provided the security survey done by the FBI and the total loss occurs by the threats in the security of the system. The paper also provides the overview of the different method of data mining such as K-mean algorithms, Decision tree, artificial neural Network etc.

"Mining and Visualizing Social Data to Inform Marketing Decisions" by Jerome Treboux, Fabian Cretton, Florian Evequoz, Anne [21] in this paper the creator indicates how the mining and representation of online networking information can be useful really taking shape of the market choice, they introduce a strategy that plans to comprehend the advertising need of an organization and build up the framework/ apparatus that can bolster it, the philosophy was partitioned into two utilize case with Swiss organization, the entire technique utilized and characterized as a part of this paper depended on CRISP-DM(Cross industry standard procedure of information mining). The entire procedure contain seven stage and the work begin with the distinguishing proof of customer needs through meeting the rest of the means are recognizable proof of objective, characterize of precondition, characterize of post condition, depict of principle stream and the last stride is portrayal of exemption and the assistance of this they made a device which gives the investigation result based of the client criticisms and remark which settle on choice/showcase wanting to enhance the business procedure

"Web Usage Data Clustering using Dbscan algorithm and Set Similarities" by K.Santhisree and Dr. A. Damodaramin [25] this paper the creator attempt to exhibit another harsh set Dbscan grouping calculation which recognize the conduct of client page guest and request of event of visit. They exhibit the exploratory outcomes on MSNBC web route dataset as demonstrate that the roughly set of Dbscan grouping calculation is good productivity for execution bunching on web mining. This paper built up another Rough arrangement of Dbscan Clustering calculation and introduced an exploratory outcomes on msnbc.com which was helpful for finding the client to get designs and the request of visits of the hyperlinks of the every client and the bury bunch likeness among the groups. The roughly set Dbscan clustering algorithm is efficient when compared to the rough set agglomerative clustering As in rough set clustering the elements can be present in more than one cluster (soft clustering) , whereas in our proposed

algorithm the elements will not occur in other clusters.

"Mapping of real estate prices using data mining techniques" by Eduard Hromadain [14] which the author portray the imaginative programming that is utilized for land assessment and mapping and investigating of land promotion. The author of this paper offers an objective for fair-minded assessment for value improvement on land advertisement. The author presents data in light of broad research and a lot of factual information which has been gathered persistently from year 2007 until today. The principle idea of this exploration paper was the measurements correlations which can be made by the product empower experts and looks into in the field of land to pick up knowledge on genuine change in the market costs of land in the Czech Republic. This yield might be utilized a reason that suitable ventures and lodging choices for both regular people and organizations. We have seen a relentless long haul decline of land market costs since on second quarter of 2008. The negative pattern does not appear to an altogether evolving. Albeit land media endeavor to present positive data, there is no sign that costs ought to begin ascending in all districts of the Czech Republic (the special case is just the locale Prague and area Middle Bohemia).

"Condition Based Maintenance in Railway Transportation Systems Based on Big Data Streaming Analysis"by Emanuele Fumeo, Luca Oneto, and Davide Anguitain [18] this paper the creator say in regards to Streaming Data Analysis (SDA) of Big Data Streams (BDS) for Condition Based Maintenance (CBM) with regards to Rail Transportation Systems. SDA of BDS is the issue of breaking down, displaying and extricating data from gigantic measures of information that ceaselessly originate from a few sources continuously through computational mindful arrangements. Among others, CBM for Rail Transportation is a standout amongst the most difficult SDA issues, comprising of the usage of a prescient upkeep framework for assessing the future status of the observed resources keeping in mind the end goal to decrease dangers identified with disappointments and to maintain a strategic distance from administration disturbances. The author utilizes the Online Support Vector Regression computational mindful Models for heuristic approach. For this reason we propose to abuse the Online Support Vector Regression for overhauling the model when new information gets to be distinctly accessible. Our Proposal likewise comprises of a model choice system that can improve the exchange off between Accuracy of the last model and assets required with a specific end goal to play out the

model choice stage itself, an extra alluring component for the most part ignored because of its computational necessities.

"Novel Approach for Analyzing the Social Network"by Saranya Balaguru , Rachel Nallathamby, C.R.Rene Robin [20] in this paper the author propose a calculation to prepare an extensive symmetric framework of billion scale chart keeping in mind the end goal to concentrate information from diagram dataset. These fascinating examples are found by calculation of a few Eigen qualities and Eigen vectors. The principle challenge in breaking down the chart information are streamlining the diagram, tallying the triangles, discovering trusses. These difficulties are tended to in the proposed calculation by utilizing orthogonalization, parallelization and blocking methods. The proposed calculation can keep running on exceptionally versatile Map Reduce environment. We utilize an interpersonal organization dataset (Facebook around 2 to 7 TB of information) to assess the calculation. The primary point is to outline an Eigen solver that finds the top-k Eigen estimations of vast symmetric lattice framed from Billion scale chart in a parallel situation. The effective parallel environment for preparing the web scale chart is HADOOP. The plan of Eigen solver needs watchful consideration in picking the calculation. We pick the successive strategy and plan it in a manner that it will keep running in parallel. The principle commitment is to enhance the versatility, adaptability and proficiency contrasted with other Eigen solver.

"Smart Phone Based Data Mining For Human Activity Recognition" by Girija Chetty, Matthew White, Farnaz Aktherin [15] this paper the author say in regards to the Automatic action acknowledgment frameworks to catch the condition of the client and its surroundings by abusing heterogeneous sensors, and allow constant observing of various physiological signs, where these sensors are appended to the subject's body In this paper they introduce novel information scientific plan for shrewd Human Activity Recognition (AR) utilizing PDA inertial sensors in view of data hypothesis based component positioning calculation and classifiers in light of arbitrary backwoods, gathering learning and apathetic learning. We inspected a few learning methodologies and discovered sluggish learning, irregular timberlands and outfit learning based ways to deal with be promising as far as movement characterization exactness, show building time for programmed order, and Confusion network, with trial approval on openly accessible action acknowledgment dataset.

"Educational data mining: A review" by Siti Khadijah Mohamad, ZaidatunTasirin [27] this paper the author says in regards to the Data Mining is exceptionally helpful in the field of training particularly when looking at understudy's learning conduct in web based learning environment. This is because of the capability of information mining in examining and revealing the concealed data of the information itself which is hard and exceptionally tedious if to be done physically. The motivation behind this survey is to investigate how the information mining was handled by past researchers and the most recent patterns on information mining in instructive research. With respect of future research, maybe we can move our concentration from the e-learning, towards the utilization of long range interpersonal communication devices like Blog and Facebook since these applications as of now increased high prominence among understudies and appropriate to be utilized to connect with the understudies with synergistic learning [2021]. We might, of course, encounter some problems, like difficulties in gathering the log data since these applications are not able to provide us with the logs of learner activity as compared to other e-learning applications, but then again, this can be encountered by integrating the Google Analytics tool into the blog environment and the log data can be exported later for further analysis using the data mining techniques.

"Sentimental space based analysis of user personalized sentiments" by Xinzhi Wang and Xiangfeng Luo,[24] in this paper the author focus on the establishment of user sentimental space obtain from online documents to analysis the user behavior. The author take the three parameter which is used in their work, these parameter are Affection, sentiments behavior, and attributes of user. In this paper the author used the support vector machine to train the system for the analysis of the result. The author also used some basics law in support of their work to analysis the sentiments. These basics laws are as follows:-

- Law of Sentimental inertia.
- Law of origin asymptotically stable
- The law of sentimental confliction
- The law of sentimental diffusion

In short this paper used some basic law and attributes to analysis the user behavior and makes the classification of user based on their sentimental model alone.

“Novel Approach for Sentimental Analysis and Opinion Mining based on Sentimental Word Net using Web Data” [28]by Shoiab Ahmed and Ajit Danti, in this paper the author novel approach is proposed in view of Sentiment Word Net, which produces tally of score words into seven classifications, for example, solid positive, positive, feeble positive, impartial, frail negative, negative and solid negative words for the supposition mining undertaking and assessed utilizing machine learning calculations like Naïve Bayes, SVM and Multilayer Perception (MLP). The proposed approach is probed motion picture and item web areas and got higher achievement rate regarding precision measured by different devices like Kappa insights with an exactness of 77.7% and has bring down mistake rates. Weighted normal of various exactness measures like Precision, Recall, TP Rate, F-Measure rate portrays higher proficiency rate and lower FP Rate for Naïve Bayes and MLP models. The creator first gather the information from the web source then evacuate the stop word i.e. pre-handle the information and apply the pre-characterized calculation to get the outcome.

Chapter 3

SCOPE OF THE STUDY

The Sentimental examination on Social Media is an important field of study now a days and will be in future. It is a procedure of discovering/deciding a concealed passionate tone behind the arrangement of words which is utilized to pick up a comprehension of the conclusion, feelings, conduct of the general population communicated in the online mode. The utilization of Sentimental examination is extremely expensive now a days it is utilized wherever, for example, as a part of business, science, governmental issues social and so forth. Sentimental Analysis has been more than just a social sensible instrument. It is a most challenging field of research because the opinion of the people can randomly change with respect of time. In any case, it is a field that is up 'til now being examined, despite the way that not at magnificent lengths as a result of the multifaceted plan of this analysis. This field has limits that are exorbitantly jumbled for machines, making it difficult to get it. The ability to appreciate joke, misrepresentation, positive feelings, or negative suppositions has been troublesome, for machines that need feelings. The present system cannot have more than 70% exactness the opinions portrayed by people. So the examination of online networking information to foresee the future or any condition is an awesome territory of research in now a days.

Chapter 4

OBJECTIVES OF THE STUDY

The objective of the study is mentioned below:

- To analysis the sentiments of the user who updated their views/thoughts on social media.
- The events which occur/happen in the world, has its distinct effect on different regions of the world. The focus will be to analyze the recent news and what is people opinion about it whether people are happy or sad about it, whether people agree or disagree upon it and majorly which part of the globe is happy or sad, agree or disagree.
- The work will help to understand the effect of the event and reasons behind people opinion.

Chapter 5

EXPECTED OUTCOMES

After using the Classification based algorithm for Sentimental Analysis or Opinion mining, the expected outcomes are:

- Reducing completion time.
- Minimizing the error rate.
- Finding the behavior of the users.
- Increase in performance of Analysis.
- Self-trained algorithm for predicting future.

Consequently, overall efficiency of Analysis is expected to improve by minimizing speculative executing.

Chapter 6
REFERENCES

[1] S. Balaguru, R. Nallathamby, and C. R. R. Robin, "A NOVEL APPROACH FOR ANALYZING THE SOCIAL NETWORK," *Procedia - Procedia Comput. Sci.*, vol. 48, no. Iccc, pp. 686–691, 2015.

[2] P. Chen, "Procedia Engineering Data Mining Applications in E-Government Information Security," vol. 29, pp. 235–240, 2012.

[3] G. Chetty, M. White, and F. Akther, "Smart Phone Based Data Mining For Human Activity Recognition," *Procedia - Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 1181–1187, 2015.

[4] E. Fumeo, L. Oneto, and D. Anguita, "Condition Based Maintenance in Railway Transportation Systems Based on Big Data Streaming Analysis," *Procedia - Procedia Comput. Sci.*, vol. 53, pp. 437–446, 2015.

[5] M. M. Gaber, "Granularity-Based Approach," vol. 6, pp. 47–66.

[6] E. Hromada, "Mapping of real estate prices using data mining techniques," *Procedia Eng.*, vol. 123, pp. 233–240, 2015.

[7] S. Khadijah and Z. Tasir, "Educational data mining : A review," *Procedia - Soc. Behav. Sci.*, vol. 97, pp. 320–324, 2013.

[8] R. Mythily, A. Banu, and S. Raghunathan, "Clustering Models for Data Stream Mining," *Procedia - Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 619–626, 2015.

[9] S. Nasreen, M. Awais, K. Shehzad, and U. Naeem, "Frequent Pattern Mining Algorithms for Finding Associated Frequent Patterns for Data Streams : A Survey," *Procedia - Procedia Comput. Sci.*, vol. 37, pp. 109–116, 2014.

[10] K. K. Parameswari and A. S. Thanamani, "Frequent Item Mining Using Damped Window Model," vol. 3, no. 9, pp. 7902–7905, 2014.

[11] M. S. B. Phridviraj and C. V Gururao, "Data mining – past , present and future – a typical survey on data streams," *Procedia Technol.*, vol. 12, pp. 255–263, 2014.

[12] V. P. Rao, S. Galande, A. Nalla, S. Devghare, and S. Kadam, "Visualization of Streaming Data Using Social Media," no. 6, pp. 602–605, 2016.

[13] R. Tripathi and S. K. Dwivedi, "A Quick Review of Data Stream Mining Algorithms," no. 7, 2016.

[14] S. Ahmed and A. Danti, "A novel approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using web data," *Int. Conf. Trends Autom. Commun. Comput. Technol. I-TACT 2015*, pp. 0–4, 2016.

[15] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on twitter," *Collab. Electron. Message. anti-abuse spam Conf.*, vol. 6, p. 12, 2010.

[16] V. Martin, "Predicting the french stock market using social media analysis," *8th Int. Work. Semant. Soc. Media Adapt. Pers. SMAP 2013*, pp. 3–7, 2013.

[17] P. Tripathi, S. K. Vishwakarma, and A. Lala, "Sentiment Analysis of English Tweets Using RapidMiner," 2015.

[18] B. Wen, S. Duan, B. Rao, and W. Dai, "Research on Word Sentimental Classification Based on Transductive Learning," *2015 8th Int. Symp. Comput. Intell. Des.*, pp. 153–156, 2015.

[19] L. Wikarsa and S. N. Thahir, "A text mining application of emotion classifications of Twitter's users using Na??ve Bayes method," *Proceeding 2015 1st Int. Conf. Wirel. Telemat. ICWT 2015*, 2016.

[20] R. Batool, A. M. Khattak, J. Maqbool, and S. Lee, "Precise Tweet Classification and Sentiment Analysis," pp. 1–6, 2013.

[21] K. Chai, H. T. Hn, and H. L. Cheiu, "Naive-Bayes Classification Algorithm," *Bayesian Online Classif. Text Classic. Filter.*, pp. 97–104, 2002.

[22] R. Jin and G. Agrawal, "An algorithm for in-core frequent itemset mining on streaming data," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 210–217, 2005.

[23] M. Mathioudakis and N. Koudas, "Twitter monitor: trend detection over the twitter stream," *SIGMOD '10 Proc. 2010 ACM SIGMOD Int. Conf. Manag. data*, pp. 1155–1158, 2010.

[24] T. R. Patil, "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification," *Int. J. Comput. Sci. Appl. ISSN 0974-1011*, vol. 6, no. 2, pp. 256–261, 2013.

[25] J. D. M. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," *Proc. Twent. Int. Conf. Mach. Learn.*, vol. 20, no. 1973, pp. 616–623, 2003.

[26] P. K. Singh, A. Sachdeva, D. Mahajan, N. Pande, and A. Sharma, "An approach towards feature

specific opinion mining and sentimental analysis across e-commerce websites," *Proc. 5th Int. Conf. Conflu. 2014 Next Gener. Inf. Technol. Summit*, no. Statement 3, pp. 329–335, 2014.

[27] X. Wang and X. Luo, "Sentimental space based analysis of user personalized sentiments," *Proc. - 2013 9th Int. Conf. Semant. Knowl. Grids, SKG 2013*, pp. 151–156, 2013.

[28] H. Zhang, "The Optimality of Naive Bayes Naive Bayes and Augmented Naive Bayes," *Am. Assoc. Artif. Intell.*, 2004.

[29] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," *Proc. 20th Int. Conf. World wide web - WWW '11*, p. 675, 2011.

[30] P. Kumar, "An approach towards feature specific opinion mining and sentimental analysis across e-commerce websites," no. Statement 3, pp. 329–335, 2014.

APPENDIX A

ABBREVIATIONS USED

---

viz.: As follows

etal.: And others

NV: Naïve Bayes

i.e.: That is to say

FP: Frequent pattern

D- Tree: Decision Tree

ACL: Access Control List

D-Window: Damped window

SVM: Support Vector Machine

CFO: Central Force Optimization

SNA: Social Networking Analysis

CRISP-DM: Cross industry standard procedure of information mining