A Tour Towards Sentiments Analysis using Data Mining

Dr. Poonam Tanwar,
Associate Professor,
CSE, Lingayas University,
Faridabad

Nazar Parwez,
MCA, Lingaya's University
Faridabad,

Santosh Singh,
Seema Kushwaha Department-MCA,
Lingaya's University
Faridabad,

Seema Kushwaha
Department-MCA,
Lingaya's University
Faridabad,

Abstract - **In the current era where majority of people generally use social media for their agreement/disagreement about anything; Face book and Twitter are one of them. Twitter is the best example of micro-blogging site which uses maximum of 140 characters to show peoples view on any topic. Frankly speaking, sentiment analysis aims to determine the attitude of the person with respect to overall contextual polarity of the comment and the attitude might be his/her judgment or evaluation. Sentiment analysis is broadly applied to reviews and social media for a variety of the applications. Sentiment analysis or Opinion mining is one of the trending topics these days. The aim is to explore various approaches of opinion analysis and find out the best suited approach. We have discussed various algorithms and their effect based on previous research papers.**

Keywords - Opinion Mining, Sentiment Analysis, Text Mining, Comment Analysis

1. INTRODUCTION

Sentiment Analysis is a process of finding people's opinion towards a topic. The sentiment can be categorized in positive, negative or neutral. Sentiment analysis uses natural language processing and other modern technologies to identify and extract information on any subject. Sentiment analysis is also termed as Opinion Mining. Where the data is extracted from the data warehouse or other sources and perform various operations on the extracted data for the purpose of analysis. Sentiment analysis is one of the booming topics now days to work upon because each and every organization from government to corporate wants to know the views of majority of people on any topic, product, situation, movie etc. Soft computing and Artificial Intelligence have been the core revolution for this kind of analysis. It would have never been so easy and flexible to have such analysis. The continuous evolvement in the field of web technologies has also been one of the key for flexibility in real time analysis. Web 2.0 is plays crucial role for real time updates and analysis on any ERP related websites. There are many E-commerce websites like Amazon, Flipkart, Ebay, Myntra etc. which give the real time ratings and results that how people think about their product.

Sentiment analysis can be done using structured data and non-structured data. Structured data is easy to understand and manage by database systems. Structured data have rows and columns defined and are constrained by many clauses. This assures that the data present in the database are already crystal and understandable. However, unstructured and semi-structured data is difficult to manage and do analysis on it. Unstructured data can be in any format like text, csv etc. It is very difficult to find the conclusion out of the flat text based comments written on any social website. Hence, this has been very interesting topic for research purpose. People work on unstructured data to simplify the rules and make it easy to understand and use. Hundreds of the projects are going on these days in this topic. We are one of them to work on a project based on unstructured data. That's why the name of the project is "A novel approach on Sentiment Analysis". This projects aims to looking at various algorithms and tries to simplify up to some extent. The basic task in sentiment analysis is to classify the polarity of a given text at the comment. The objective is to get the outcome in terms of positive, negative or neutral. The advancement can be extended in emotional state as angry, sad or happy instead. These outcomes are termed as polarity. For example if the outcome expresses negative sentiment then it would be categorized as polarity="negative".

The sentence gets divided into the tokens and each word corresponds to the sequence of token numbers. Then the tokens are used to recognize subjectivity/objective or features.

There are various works which have been evolved as polarity by many of researchers. The early work area would use either positive or negative. This was easily understandable to categorize majority of people. People would either agree or disagree with the sentiment of the product. Later movie review did come into picture where movies would get one to five star ratings and this made the polarity to scale further into more groups like the one who gets two stars would be polarized as "negative", similarly three stars would get "neutral" and four or five stars would get "positive" polarity. These days even restaurants have their ratings divided into the five stars and the analysis performed on the basis of the ratings by the customers. In many of the approaches the neutral class is generally ignored as it is assumed that the neutral class will always fall near the boundary of the binary classifier however,

several researchers suggest that there must be three categories. This is proven Max Entropy [1] and SVM [2] can benefit from the neutral class and improve the overall accuracy of the classification rule. The algorithm works in two ways during the sentiment classification. The first way is to find out the neutral language, filter it then asses rest of the negative and positive sentiments. The second approach is to classify the sentiment into three way classifications in one step and this way involves Naive Baye's probabilistic distribution theory [3].

The use of neutral class depends upon the nature of data. If the data is clustered into neutral, negative and positive language then it's better to filter the neutral language out then focus into the polarity between positive and negative classes. However, if the data mostly neutral and there is a small deviation towards negative and positive then it is better to use three classes neutral, negative and positive.

There is a different method for considering the sentiment by using the scaling system between -10 and +10. Where -10 is the most negative, 0 is the neutral and +10 is the most positive sentiment. This increases the possibility to match the environment very close to its sentiment pole. When unstructured text gets analyzed by using NLP then each word in the environment is given a score based on the corresponding sentiment word present in the dictionary. This allows the movement of words towards more sophisticated understanding in terms of sentiment because it has now become possible to adjust the sentiment value of words relative to its modifications. For example the words that expressed by the concept, effect the sentiment can also effect its score. Alternatively, words can be given a sentiment strength score as negative and positive if the goal is to determine the sentiment in a text rather than the overall polarity of the text. Word Cloud is one of the examples of it. There are other applications which give the sentiment values as per their project.

1.1 Subjectivity/objectivity identification
The identification of a given sentence can be done into one of two ways, subject or object [4]. For example, Samsung provides best resolution camera. This problem sometimes becomes more challenging than polarity classification. The subjectivity of words or phrases might depend on their context and objective document may also contain subjective sentences [5]. However, Pang [6] has proven that removing objective sentences from a document before classifying its polarity improves performance.

1.2 Feature/aspect-based
Feature of an entity helps in determining the opinions or sentiment expressed in the sentence. A feature is an attribute of an entity, for example, screen of a cell phone, picture quality of a camera or processing speed of a laptop. The advantage of feature based sentiment analysis is the possibility to capture different shades of about the objects interest. Different features can generate different sentiments, for example a restaurant can have many Indian delicious cuisines however, in a remote location. The

problems may have several other problems e.g. identifying relevant entities and extracting their features, determining whether the opinion is positive, negative or neutral. There are automatic model methods to identify the features. More detailed discussion about sentiment analysis can be found in Liu's paper [7].

1.3 Methods:
There are many approaches to sentiment analysis; however, they are grouped into three main categories:

1. Knowledge Based Techniques: In this technique classification is done on the basis of definite words such as sad, happy, and bored [8]. There are other knowledge bases which are not only list obvious affect words but also similar probable to a particular emotion.
2. Statistical methods: These methods leverage on elements from ML (Machine Learning) such as Support Vector Machines (SVM), Latent Semantic Analysis and Semantic Orientation [9]. Grammatical dependency between the texts is obtained by deep parsing on them.
3. Hybrid approaches: These approaches leverage on both knowledge base representation and machine learning such as ontology's semantic networks to detect semantics. For example, the analysis concepts that do not convey the relevant information however, they are implicitly linked to other concepts that do so. There are many software tolls that deploy machine learning, natural language processing and statics to automate sentiment analysis on large datasets made up of web pages, online discussions, reviews, blogs etc. Knowledge-based systems on the other hand use these resources and do analysis with the help of natural language processing concepts [10]. Sentiment analysis can also be done on images and videos. The first approach in this area was SentiBank [11] which utilized adjective noun pair representation of visual contents.

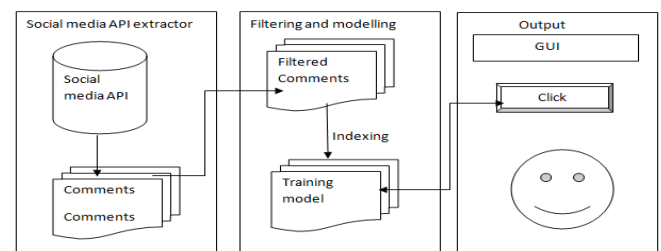**Architecture of Social media comment classification:**

Figure 1. Architecture of Social media comment classification:

1.4 Evaluation
The accuracy of a sentiment analysis system is usually measured by precision and recall. It relates that how well the analysis agrees with the human judgement. However, research shows human analyzer systems typically agree 79% of the time [12]
Thus, 70% accuracy by the program or humans is doing nearly good even though such accuracy might not sound

impressive. There are more sophisticated measures can be applied however, evaluation of sentiment analysis system remains a complex matter. For sentiment analysis correlation is better measure than precision because it detects how close the predicated value to the target value.

1.5 Web 2.0

The rise of Social Media has fuelled interest in sentiment analysis. The explosion of reviews, ratings, comments, recommends and online opinions have provided more business opportunities. E-commerce is one of the platforms, which is growing day by day and the level of trust is also increasing in e-business. Customers interact with the ERP of the organization and find desired product using some keywords or simply by navigating the mouse click. Online opinion has become a key to trust on particular product before buying. People who buy any product from the website leave a comment as per their overall experience towards that product or even that organization. It is also important from the organizations point of view. Organizations try to maintain the quality and its services by identifying the feedback from the customers. A business looks to automate the process of filtering noise, understanding conversation, identifying relevant content and auctioning appropriately on it. All these are done automatically on the web itself and this is possible due to the web 2.0. It is much flexible to give real time result on sentiment analysis and many more.

2. LITERATURE SURVEY

2.1 Study of Various existing Techniques

Shailendra Kumar et al have written their observations on sentiment analysis approaches using different datasets. As social networking sites generating comments, reviews, likes and dislikes every day. People are frequently using Facebook, Twitter etc and give their agreement and disagreements. This paper has focused on sentiment analysis using those social websites. Sentiment analysis is extraction of people opinions by using NLP (Natural Language Processing), Linguistic Computation and Text Mining. Authors have studied the datasets between 1997 and 2012 for the sentiment analysis. These days sentiment analysis has become very essential in every aspect like government, corporate, education and research etc.[13] Here are some of the existing works have been shown in this paper (Table 3)

1. Susan P. Imberman has concluded that the KDD(Knowledge Discovery in Databases) is a step by step process to find the knowledgeable data from the raw, unstructured data and non-volatile data.
    The author has explained data mining algorithms such as decision tree, neural networks and association rules. This paper identifies parts of the processes that are currently being used and some are not but may be helpful in analyzing performance data. Data mining is one step of the KDD process. The author further elaborates each steps in KDD and performs above algorithms and tried to find out the CPU usage and load performance. [14]
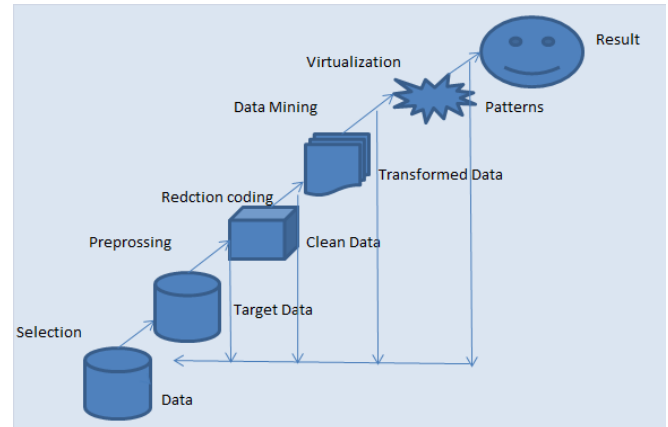

Figure 2. Knowledge Discovery in Databases Process

2. Parvesh Kumar Singh et al have explained about the decision making and it's criticality in finding the knowledge as well as its importance in the current era. Authors have taken various datasets from the social websites like Twitter, Facebook, blogs, multimedia websites and used data mining techniques to extract the knowledge. Authors have used important algorithms like Naïve Bayes Classifier, Support Vector Machine, Multilayer Perception and Clustering to find out the sentiments of people. The resources for the analysis on the web are likes, reviews, forum discussions, blogs etc. [15]

3. Aproov Agarval et al have performed sentiment analysis on Twitter data which is a micro-blogging site. Twitter uses 140 characters to comment on any topic, hashtag etc. Authors have introduced POS-specific prior polarity features and also explored the use of tree kernel. As per their observation the new features and the tree kernel perform approximately at the same level, both outperforming the state-of-the-art baseline Microblogging websites have evolved to become a source of varied kind of information.
    In this paper authors have taken data from Twitter and built models for classifying tweets into positive, negative and neutral sentiments. The tree kernel is the hierarchy system of words same as the file structure of kernel which starts from ROOT.[16]

4. Haseena Rahmat P has covered opinion mining and sentiment analysis as well as shown various application and its challenges. As several websites encourage users to express and exchange their views, suggestions and opinions related to product, services, polices, etc. The increased popularity of these sites resulted huge collection of people opinion on the web in much unstructured manner. Extracting the useful content from these opinion sources became a tedious task. Therefore, the situation now has created a new area of research called opinion mining and sentiment analysis. Opinion mining and sentiment analysis extract then classify the people's opinion automatically from the internet. This paper has discussed all the possible aspects and challenges related to Opinion Mining and Sentiment Analysis.

As opinion mining and sentiment analysis are the extension of data mining which uses NLP to extract comments from the web and try to find the sentiment of people using various algorithms and modern tools. Web 2.0 has simplified various issues from large volume of data extraction to dynamic data manipulation and at the same time the analysis result also performed automatically. The challenges of sentiment analysis include detection of spam and fake reviews which is very common now days. These are done intentionally to create rumors on the web and try to hit the sentiment of the people. Therefore, there are various agencies which are looking upon these challenges on real time basis. However, there are various limitation to this field still exist like ambiguity, linguistics etc, which needs to be look upon. [17]

5. Vijay B. Raut et al worked on Openion Mining/Sentiment Analysis using various techniques. Large amount of data generated by users has been used like web as blogs, reviews, tweets, comments etc. This data consist user opinion, views, attitude, sentiment towards particular product, topic, event, news etc. As opinion mining (sentiment analysis) is a process of finding users' opinion from user-generated content and Opinion summarization is useful in feedback analysis, business decision making and recommendation systems. In recent years opinion mining is one of the popular topics in Text mining and Natural Language Processing. This paper has presented the methods for opinion extraction, classification, and summarization. Authors presented different approaches, methods and techniques used in process of opinion mining and summarization, and comparative study of these different methods. [18]
Due to web and social network, large amount of data are generated on Internet every day. This web data can be mined and useful knowledge information can be fetched through opinion mining process. This paper has discussed different opinion classification and summarization approaches, and their outcomes. As the study shows that machine learning approach works well for sentiment analysis of data in particular domain such as movie, product, hotel etc., while lexicon based approach is suitable for short text in micro-blogs, tweets, and comments data on web. Authors have also shown some observations on Table (3).

6. Svetlana Kiritchenko et al have described their observation on sentiment analysis. According to the authors, state-of-the-art sentiment analysis system that detects (a) the sentiment of short informal textual messages such as tweets and SMS (message-level task) and (b) the sentiment of a word or a phrase within a message (term-level task). The system is based on a supervised statistical text classification approach leveraging a variety of surface form, semantic, and the sentiment features. The sentiment features are primarily derived from novel high-coverage sentiment lexicons of tweets. These lexicons are automatically generated from tweets with words, hashtags and emoticons. To adequately capture the sentiment of words in negated contexts, a separate sentiment lexicon is generated for negated words. Authors have created a supervised statistical sentiment analysis system that detects the sentiment of short informal textual messages such as tweets and SMS as well as the sentiment of a term (a word or a phrase) within a message. Authors have also implemented a variety of features based on surface form and lexical categories. Features have been derived from several sentiment lexicons for example, existing, manually created, general-purpose lexicons and high-coverage, tweet specific lexicons. There are negative processing of sentiment analysis have also been done to validate the result and it actually showed that the polarity reversing method may not always be appropriate. [19]

7. Bo Pang and Lillian Lee used blogs, social networking sites etc to find out sentiment of people on the basis of their comments, likes, dislikes, emoticons etc. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first class object.

Authors' survey covers techniques and approaches that promise to enable opinion-oriented information using systems. The main focus is on the methods that need to be addressed new challenges raised by sentiment aware applications, as compared to those which are already present in more traditional fact-based analysis. Thus, material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise are included in this research paper. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided. [20]

8. Renata Maria Abrantes Baracho et al have worked on sentiment analysis using social websites. This paper presents partial results of research project that aims to create process of sentiment analysis based on ontology's in the automobile domain and then to develop a new prototype. The process aims at making a social media analysis, identifying feelings and opinions about brands and vehicle parts. The method that guided the development process involves the construction of ontology's and a dictionary of terms that reflect the structure of the vocabulary domain. The proposed process is capable of generating information that answers questions such as: "In the opinion of the customer, which car is better: Audi or BMW? Which one is more beautiful? Which engine is stronger?" To answer these questions by comparison, one can show a general view reflected on different social networks,

indicating, for example, that for a given vehicle, a certain percentage of responses are considered positive, while for others, the percentage is considered negative. The results can be used for various purposes such as guiding decisions to improve the products or directing specific marketing strategies. The process is generalized and applied to other areas in which organizations are interested in monitoring views expressed about their products and services.

This paper presented a methodology to process, analyze and summarize sentiments as well as opinions from sentences extracted from the Web. This methodology was applied to the automotive field for a specific analysis of the FIAT brand and resulting in a prototype that is still incomplete, however, it an easily be applied to any other domain. Although the practical results were very simple, especially considering the potential for the methodology, it was successful to serve as a proof of concept that the methodology works and can provide interesting insights about the data. [21]

9.   Rudy Prabowo et al wrote this paper on a current research area called "Sentiment analysis". This paper combines rule-based classification, supervised learning and machine learning into a new combined method. This method is tested on movie reviews, product reviews and MySpace comments. The results show that a hybrid classification can improve the classification effectiveness in terms of micro and macro averaged function. Function measure takes both the precision and recall of classifiers effectiveness into consideration. In addition, authors have also proposed a semi-automatic, complementary approach in which each classifier can contribute to other classifiers to achieve a good level of effectiveness.

Authors concluded that the use of multiple classifiers in a hybrid manner can result in a better effectiveness in terms of micro and macro-averaged functions than any individual classifier. Authors have applied semi-automatic, complementary approach, i.e., each classifier contributes to other classifiers to achieve a good level of effectiveness by using a Sentiment Analysis Tool (SAT). Moreover, a high level of reduction in terms of the number of induced rules can result in a low level of effectiveness in terms of micro and macro averaged function. The induction algorithm can generate a set of induced antecedents that are too sparse for a deeper analysis. Therefore, in a real-world scenario, it is desirable to have two rule sets, one is the original set, and another one is the induced rule set. [22]

10.   The research paper uses the corpus obtained from Spanish twitter for opinion mining from users of iPhone. Machine learning techniques were used to classify sentiments (positive, negative, neutral and informative) in both English and Spanish language. The classifier methodology employed in this research

was SVM (supported vector machines), Naive Bayes and Decision trees. Paper also discusses how large size of corpus of at least 3000 tweets (data set extracted from twitter) improves accuracy of findings and unigram (among other variables of n-gram features) is best feature size because it provides most accuracy. Also, balancing the corpus, that is using proportional number of all categories of classified data gives slightly worse results and it also found out that SVM classifier gave results with best precision. [23]

11.   Most sentiment analysis is done through textual content, but in this paper multimedia content such as images and videos (that is visual content) from social media is studied. This paper uses half a million images from Flickr for visual sentiment analysis through use of Convolutional neural network (CNN). Problems in visual sentiment analysis are typically the weakly labeled image data, abstraction and subjective nature of images. For this supervised training of data set (images) through deep learning framework (CNN) was optimized for this research. It was followed by further fine-tuning for more generalizability and better performance of neural network. [24]

12.   Opinion analysis from online posts using word-based data set tool. Paper shows problem with tools like tweet feel as it can only monitor positive and negative connotations of a statement and it cannot know the proper context in which the words are used (example – bangalore royal challengers won the match and bangalore is a nice city are both categorized together instead of the fact that first post is about a team and second is about tourism). The architecture designed by research paper author follows steps like data collection, identifying word type (noun, adjective, adverb etc.), comparison of words with positive and negative keyword set, processing of that opinion value and aggregating all collected results [25].

13.   This paper discusses various sentiment classification models used in opinion analysis. The process which follows is extraction of study material from data source (blogs, review sites and micro blogging sites like twitter), sentiment classification through machine learning techniques like SVM, Naive Bayes, Centroid classification, K-nearest neighbour, Winnow and rule based classification. It also states about semantic orientation (how much positive or negative a word is) and the role of negation in sentences [26].

14.   This paper discusses the role of opinion mining in creating market intelligence for use by Corporates. Opinion mining is important for business owners to extract information about consumer attitudes and to determine what they require. This helps them frame market penetration strategies and measure the feedback from market. Real time feedback from social media is extracted through opinion mining without hindrances. Research paper talks about opinion orientation (negative, positive, neutral), opinion

strength, opinion model (constituent parts of an opinion like feature, target object, sentiment value, opinion holder and time of opinion) and its process of evaluation. Paper broadly categorizes opinion mining techniques into supervised or machine learning and unsupervised learning or natural language processing. Later, implications of opinion mining on marketplace are discussed in detail [27].

3. PROPOSED WORK

We have started working with Twitter for our experiments. Twitter is a micro-blogging social networking site where people post their opinions, comments etc. The contents of the messages vary from personal thoughts to public statements. The messages on the Twitter are known as tweets. Tweets have a limited size of 140 characters that is enough to convey the message as per Twitter. Therefore our work is limited to the sentence level.

We chose to use global polarity rating as the messages are short in Twitter. We have not considered processing the cases where tweets have more than one sentiment orientations.

We have compiled a corpus based on data extracted from Twitter. The corpus was built to process predefined entities. We have collected 1,000 tweets for a particular hash tag. We have shown the word cloud and the size of the word on that cloud would tell us its overall impact.
We would also show the sentiment based on the polarity: negative, positive or neutral.
Each class is described as follows:

- Positive: if it has a positive sentiment in general, like 'iPhone7 is the best model'.
- Negative: if it has a negative sentiment in general, like 'I hate using iPhone7 as it big in size'.
- Neutral: if it has no sentiment, like 'I am tweeting from my iPhone'.

3.1 Pre-processing

$$\frac{(\text{No. of word occurrence in class} + 1)}{\text{No. of words belonging to a class} + \text{Total No. of words}}$$

Tweets contain slangs, misspellings, words from other languages, short form of words etc. Tweets are basically normalized to deal with these kinds of problems and noises in the text. Tweets are normalized before training the classifiers using these procedures.

- Error Correction
- Special Tags
- POS-Tagging
- Negation Processing

Common Errors in Tweet
Error Correction deals with correcting the words which are similar to the words in English dictionary, for example

'goood' can be replaced with good. Special tags would deal with the tagging facility like user name is replaced with USER_TAG and ☺ is replaced by WINK_TAG. POS-Tagging done after normalization and is used to distinguish between noun, adjective, verb etc. Negation process is used to detect the negative sentiment/word like 'no'. These are explained thoroughly in [23].

There are various algorithms used in opinion mining and sentiment analysis. Some of the important algorithms are explained below.

- Naïve Bays Classification
- Support Vector Machine
- Multiplayer Perceptron

3.2 Classification Techniques
A. Naive Bays Technique:
Thomas Bayes proposed a probabilistic and supervised classification technique which is known as Naïve Bayes technique. As per the algorithm, if there are two events say, E1 and E2 then the conditional probability of occurrence of event E1 when event E2 has already been occurred is given by:

$$P(E1|E2) = \frac{P(E1|E2)P(E1)}{P(E2)}$$

Above algorithm is used to calculate the probability of data to be negative, positive or neutral.

Table 1. Different types of error and their example.

Error	Example
Slang	I forgot my cell on the table
Misspelling	I really lyk my cell phone.
Mixed languages	While I war travelling na, I lost my phone

The conditional probability of word is:

$$P(\text{Word} \mid \text{Sentiment}) =$$

Advantages

- Easy to interpret the model
- Efficient in computation

Disadvantage

- Assumptions may or may not valid.

B. Support Vector Machine (SVM):
Support Vector Machine is a supervised learning model. SVM is used to analyze the data and identify its pattern for classification. SVM is based on **decision plane** that defines decision boundaries. Decision plane is basically separates instances from one class to other.

In the binary categorization of text, the decision plane which classifies document α_j as C_j {1, −1} can be represented by weight vector $\vec{w}$ [15].

$$\vec{w} := \sum \alpha_j c_j \vec{d_j}, \ \alpha_j \geq 0, \quad (1)$$

Where α_j is a multiplier and for $\vec{dj}$ that α_j are greater than zero are support vectors. Test instance is classified by determining which side of $\vec{w}$'s hyper-plane they fall on.

Advantages of Support Vector Machine Method

- It gives very good performance.
- Dependency on dataset dimension is low.

Disadvantages of Support Vector Machine Method

- Dataset needs to be pre-processed in case of missing values.
- Difficulty in interpreting the result.
-

C.   Multi-Layer Perceptron (MLP):
It is actually a feed forward neural network, which has input layer, hidden layer and output layer. The flow starts from input layer and gets processed in hidden layer and ends with output layer as follows.

- When the prediction is binary then output layer will have one neuro
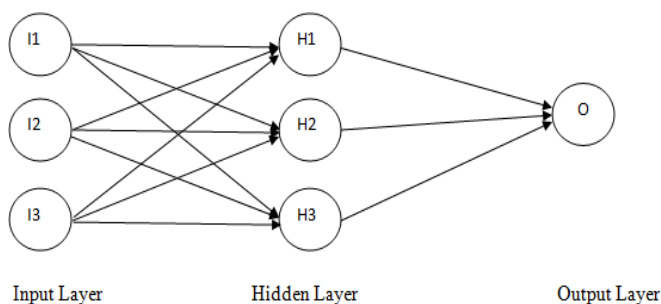- When the prediction is non-binary then output layer will have N neurons

Figure 3 Multilayer ANN Archtecture.

Multi Layer Perceptron is a back propagation and it works on two phases
- In the first phase the activation network propagate from input to output layer.
- In the second phase the activation network propagate back from output to input layer. It is basically to check if any error occurred.

MLP is one of the popular techniques and it acts as a universal function approximator. MLP is flexible tool because of "back propagation". In back propagation the network has at least one hidden layer and many non-linear entities. It does not enforce any constraint and does not require specific assumption to start as compared to traditional modelling.

Phase I: It is the forward phase where activation are propagated from the input layer to output layer.

Phase II: In this phase to change the weight and bias value errors among practical & real values and the requested nominal value in the output layer is propagate in the backward direction.

Accuracy
On the health care data Ludmila I. Kuncheva, (IEEE Member) calculate accuracy of MLP as 84.25%-89.50% [15].

Advantages of MLP
- It is flexible hence, acts like universal function approximate.
- It is unsupervised learning as it can learn every relationship among input and output variables itself.

Disadvantages of MLP
- It needs more time to execute due to its flexibility.
- It is complex to implement.

Authors have shown below accuracy figure based on their survey. As the table shows that the result of SVM has the highest accuracy, it is good to use SVM to get the best result. [15]

Table2 : accuracy table of different methods

	Movie Reviews			Product Reviews		
	NB	MLP	SVM	NB	MLP	SVM
N-gram Feature	75.50	81.05	81.15	62.50	79.27	79.40

REFERENCES

1. Vryniotis, Vasilis (2013). The importance of Neutral Class in Sentiment Analysis.
2. Koppel, Moshe; Schler, Jonathan (2006). "The Importance of Neutral Examples for Learning Sentiment". Computational Intelligence 22. pp. 100–109. CiteSeerX: 10.1.1.84.9735.
3. Ribeiro, Filipe Nunes; Araujo, Matheus (2010). "A Benchmark Comparison of State-of-the-Practice Sentiment Analysis Methods". Transactions on Embedded Computing Systems. 9 (4).
4. Pang, Bo; Lee, Lillian (2008). "4.1.2 Subjectivity Detection and Opinion Identification". Opinion Mining and Sentiment Analysis. Now Publishers Inc.
5. Mihalcea, Rada; Banea, Carmen; Wiebe, Janyce (2007). "Learning Multilingual Subjective Language via Cross-Lingual Projections" (PDF). Proceedings of the Association for Computational Linguistics (ACL). pp. 976–983.
6. Pang, Bo; Lee, Lillian (2004). "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based

on Minimum Cuts". Proceedings of the Association for Computational Linguistics (ACL). pp. 271–278.

7. Titov, Ivan; McDonald, Ryan (2008-01-01). "Modeling Online Reviews with Multi-grain Topic Models". Proceedings of the 17th International Conference on World Wide Web. WWW '08. New York, NY, USA: ACM: 111–120. doi:10.1145/1367497.1367513. ISBN 978-1-60558-085-2.

8. Cambria, E; Schuller, B; Xia, Y; Havasi, C (2013). "New avenues in opinion mining and sentiment analysis". IEEE Intelligent Systems. 28 (2): 15–21. doi:10.1109/MIS.2013.30.

9. Turney, Peter (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics. pp. 417–424. arXiv:cs.LG/0212032free to read.

10. Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86.

11. Borth, Damian; Ji, Rongrong; Chen, Tao; Breuel, Thomas; Chang, Shih-Fu (2013). "Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs". Proceedings of ACM Int. Conference on Multimedia. pp. 223–232.

12. Ogneva, M. "How Companies Can Use Sentiment Analysis to Improve Their Business". Mashable. Retrieved 2012-12-13.

13. Sentiment Analysis Approaches on Different Data set Domain: Survey , Shailendra Kumar Singh1, Sanchita Paul2 and Dhananjay Kumar1, International Journal of Database Theory and Application Vol.7, No.5 (2014), pp.39-50, http://dx.doi.org/10.14257/ijdta.2014.7.5.04

14. EFFECTIVE USE OF THE KDD PROCESS AND DATA MINING FOR COMPUTER PERFORMANCE PROFESSIONALS, Susan P. Imberman Ph.D. College of Staten Island, City University of New York, Imberman@postbox.csi.cuny.edu

15. METHODOLOGICAL STUDY OF OPINION MINING AND SENTIMENT ANALYSIS TECHNIQUES, Pravesh Kumar Singh1, Mohd Shahid Husain2, International Journal on Soft Computing (IJSC) Vol. 5, No. 1, February 2014

16. Sentiment Analysis of Twitter Data, Apoorv Agarwal BoyiXie IliaVovsha OwenRambow Rebecca Passonneau Department of Computer Science Columbia University New York, NY 10027 USA {apoorv@cs, xie@cs, iv2121@, rambow@ccls, becky@cs}.columbia.edu

17. Opinion Mining and Sentiment Analysis - Challenges and Applications , Haseena Rahmath P , Dept. of Computer Science and Engineering, Al-Falah School of Engineering, Dhauj, Haryana, India, ISSN 2319 – 4847

18. Survey on Opinion Mining and Summarization of User Reviews on Web, ISSN:0975-9646.

19. Sentiment Analysis of Short Informal Texts, Journal of Artificial Intelligence Research 50 (2014) 723–762 Submitted 12/13; published 08/14

20. Opinion mining and sentiment analysis, Bo Pang1 and Lillian Lee2, Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008) 1–135 c 2008 Bo Pang and Lillian Lee. This is a pre-publication version; there are formatting and potentially small wording differences from the final version.

21. Sentiment analysis in social networks: astudyonvehicles Renata Maria Abrantes Baracho, Gabriel Caires Silva, Luiz Gustavo Fonseca Ferreira , 1Progama de P´os-Graduac¸˜ao em Ciˆencia da Informac¸˜ao (PPGCI) Universidade Federal de Minas Gerais (UFMG) P.O. 486 31270-901 – Belo Horizonte – MG – Brazil

22. Sentiment Analysis: A Combined Approach, Rudy Prabowo1, Mike Thelwall, School of Computing and Information Technology University of Wolverhampton Wulfruna Street WV1 1SB Wolverhampton, UK Email:rudy.prabowo@wlv.ac.uk, m.thelwall@wlv.ac.uk

23. Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets , Grigori Sidorov1, Sabino Miranda-Jiménez1, Francisco Viveros-Jiménez1, Alexander Gelbukh1, Noé Castro-Sánchez1, Francisco Velásquez1, Ismael Díaz-Rangel1, Sergio Suárez-Guerra1, Alejandro Treviño2, and Juan Gordon2 , 1 Center for Computing Research, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz, s/n, esq. Mendizabal, Col. Nueva Industrial Vallejo, 07738, Mexico City, Mexico 2 Intellego SC, Mexico City, Mexico www.cic.ipn.mx/~sidorov, sabino_m@hotmail.com.

24. RobustImageSentimentAnalysisUsingProgressivelyTrainedandDomain,TransferredDeepNetworks, QuanzengYou and JieboLuo Department of Computer Science University of Rochester Rochester, NY 14623 {qyou, jluo}@cs.rochester.edu, Department of Computer Science University of Rochester Rochester, NY 14623 {qyou, jluo}@cs.rochester.edu, HailinJin and JianchaoYang Adobe Research 345 Park Avenue San Jose, CA 95110 {hljin, jiayang}@adobe.com

25. Opinion Mining And Sentiment Analysis On Comments , Volume - 5 | Issue - 1 | Jan Special Issue - 2015 | ISSN - 2249-555X

26. Sentiment Analysis and Opinion Mining: A Survey G.Vinodhini*, Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar-608002, RM.Chandrasekaran Professor, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar-608002. India, Volume 2, Issue 6, June 2012 ,ISSN: 2277 128X

27. Opinion Mining: A Tool for Market Intelligence Shweta T. Joglekar, Dr. Sachin A. Kadam Bharati Vidyapeeth Deemed University Institute of Management and Entrepreneurship Development Pune, Maharashtra, India , Volume 5, Issue 7, July 2015,ISSN: 2277 128X

28. J. Yi, T. Nasukawa, R. B and W. Niblack. "Sentiment Analyzer: Extracting sentiments about a given topic using Natural language Processing Techniques", Proceedings of the 3rd IEEE International Conference on Data Mining, (2003).

29. M. Karamibekr and Ali A. Ghorbani, "Sentiment Analysis of Social Issues", International Conference on Social Informatics (IEEE), (2012).

30. Aditya Joshi, Balamurali A. R., Pushpak Bhattacharyya "A Fallback Strategy for Sentiment Analysis in Hindi a Case Study" Proceedings of ICON 2010: 8th International Conference on Natural Language Processing, Macmillan Publishers, India.

31. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou "Movie Rating and Review Summarization in Mobile Environment", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol. 42, No. 3, May 2012, pp.397-406.

32. Jingjing Liu, Stephanie Seneff, and Victor Zue, "Harvesting and Summarizing User-Generated Content for Advanced Speech-Based HCI", IEEE Journal of Selected Topics in Signal Processing, Vol. 6, No. 8, Dec 2012, pp.982-992.

33. Alexandra Balahur, Mijail Kabadjov, Josef Steinberger, Ralf Steinberger, Andrés Montoyo, "Challenges and solutions in

the opinion summarization", Journal of Intelligent Information Systems Springer 2012, p.375-398.

34. Elena Lloret, Alexandra Balahur, José M. Gómez, Andrés Montoyo, Manuel Palomar, "Towards a unified framework for opinion retrieval, mining and summarization" Journal of Intelligent Information Systems Springer 2012, pp.711-747.

35. Alvaro Ortigosa, José M. Martín, Rosa M. Carro, "Sentiment analysis in Facebook and its application to e-learning", Computers in Human Behavior Journal Elsevier 2013.

36. Alexandra Trilla, Francesc Alias "Sentence-Based Sentiment Analysis for Expressive Text-to-Speech", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, No. 2, February 2013, pp.223-23

Table 3 Comparative study of various methods of Sentiments analysis

Author	Methods	Dataset	Result (Accuracy %)	References
Aditya Joshi, Balamurali A.R., Pushpak Bhattacharyya (2010)	Machine Learning, Machine Translation (MT), Hindi SentiWordNet Dictionary	Travel Reviews	In – language sentiment analysis- 78.14%, MT – based sentiment analysis - 65.96%, Resource based - sentiment analysis- 60.31%	[30]
Chien-Liang Liu, Wen-Hoar Hsaio, Chia- Hoang Lee, Gen-Chi Lu, Emery Jou (2012)	SVM classifier, LSA method for feature based summarization	Movie Reviews	SVM classifier - 85.4%	[31]
Jingjing Liu, Stephanie Seneff, Victor Zue (2012)	Word phrase Extraction, word phrase sentiment scoring and classification for sentiment analysis	Hotel Reviews	Decision Tree classifier for word phrase classification - 77.9%	[32]
Alexandra Balahur, Mijail Kabadjov, Josef Steinberger, Ralf Steinberger, Andrés Montoyo (2012)	Sentiment dictionary Resources, LSA based opinion summarization	Blogs	Sentiment analysis And Summarization ROUGE R1 negative class - 0.268, positive class-0.275	[33]
Elena Lloret, Alexandra Balahur, José M. Gómez, Andrés Montoyo, Manuel Palomar (2012)	Sentiment Resource, Machine learning, Term Frequency based Summarization	Product Reviews	Summary- ROUGE-1 (10% compression) Precision-30.16, Recall-20.54,	[34]
José M. Martín, Rosa M. Carro (2013)	Machine Learning, Word Lexicons	Facebook Messages	Sentiment analysis-83.27%	[35]
Alexandra Trilla, Francesc Alias(2013)	Machine Learning	Semeval Dataset, Twitter Dataset	SVM-58.12% - Semeval Dataset, SVM-72.76% - Twitter Dataset	[36]
Jeonghee Yi et al., 2003	NLP	Feature Extraction	91-93 85.6	[28]
M.Karamibe kr et al.,2012	POS, Pattern based	Extraction of Verb & opinion term	65	[29]