CHAPTER 1

INTRODUCTION

This chapter emphasizes the significance and characterization of Sentimental Analysis.

1.1 SENTIMENTAL ANALYSIS

Sentiment analysis is a Place of Natural Language Processing refers to an opinion or attitude expressed by a person towards their goal. SA is process of collecting and analyzing information based upon the individual opinion, views and thoughts. SA is done with the help of Natural Language Processing (NLP), Statistical models and different machine learning algorithms for extracting features from large data. Attitudes --"relatively enduring, affectively colored beliefs, tastes, and predispositions towards things or individuals (enjoying, loving, hating) are distinct from emotions --"short episodes of synchronized responses (mad, sad, joyous, fearful, humiliated, joyful)" as a response to outside influences. This distinguishes belief analysis from various other problems like emotion evaluation where the overall psychological state (affected by several external factors) is of interest, rather than the mindset towards a particular target.

The human perception getting change at different time stamp. The mentality and the mood of any person has great influence in different activities performed by any person. For example, consider a medical practitioner who provides treatment for different patients, but the performance of treating the patient or curing the disease is depending on the selection of medication which is highly based on the mood of the person. When the medical practitioner is in

good or normal mood, then he would be choosing effective medication for the disease than the worst mood. Similarly, the sentiment or mood of any person affects the performance of the work being done.

The sentiment analysis is the process of identifying the mood of any person. By identifying the sentiment of any person, necessary action can be taken. It has been performed in several ways according to different features. However, the text mining plays vital role in identifying the sentiment of any person. It can be performed by using any document or text and using set of bio signals. Using all these, the problem of sentiment analysis can be handled effectively. The easiest and most frequent polarity strategy supposes three groups, positive, negative and neutral.

However, there are number of techniques and measures available which are discussed in detail in this section.

1.1.1 Twitter Sentimental Analysis

The micro blogging website like twitter signifies a significant Platform for users to share their opinions and discuss their thoughts. In twitter the subjects related all domain names are discussed in which it contributes to supplying variety of information. Twitter has come to be the silent important illustration of “Big Data" due to the access to huge volume and variety of information made accessible by consumer contributions. In twitter the same information might be shared amongst the public, within their subgroups. This produces the amount of the accessible data much more. As a consequence of its substantial, diverse and growing user base, Twitter has surfaced as a significant resource for internet opinions and opinion indicators.

1.2 TYPES OF SENTIMENTS

In today's study, lots of distinct perspectives on automatic opinion analysis exist, which contributes to different task. The most obvious distinction between them is that the granularity of investigation. Sentiment analysis is done on different linguistic levels. In document level, the task would be to classify if or not a whole opinionated record has a positive, neutral or negative sentiment. In the sentence level, the task would be to classify whether a single paragraph includes a positive, neutral or negative sentiment. In the aspect level (the entity level), the task would be to classify the opinion of individual phrases or sentences planned towards particular entities or facets.

Types of Sentiments

The sentiments can be exposed by human can be classified as follows:

Happy	:	This sentiment is exposed by human when the person is happy or glad.
Sad	:	This sentiment is exposed by human due to feeling or showing sorrow; and unhappy.
Exclamation	:	This is exposed when the person is surprised due to specific action or event.
Anger	:	This sentiment is exposed due to a strong feeling of being upset because of something wrong or bad annoyed because of something wrong or bad.
Silent	:	This is a neutral emotion expressed by human without any emotion.
Fear	:	This emotion or sentiment is expressed by human due to unpleasant emotion caused by the threat of danger, pain, or harm.

Disgust	:	This emotion or sentiment is expressed by feeling about something went wrong.
Shame	:	This sentiment or emotion is expressed when a person feeling distress caused by the consciousness of wrong or foolish behaviour.
Kindness	:	This is exposed when a person feeling good and considerate to other people.
Pity	:	This sentiment is expressed when the person feels sorry to others.
Indignation	:	Feeling bad or angry because of some activity which is not fair to them.
Envy	:	Exposed when the people have something which is not fair to them.
Love	:	Expressed intense feeling of deep affection with opposite gender and same gender.
Surprise	:	The emotion is expressed due to some unexpected event or sudden happening.
Trust	:	The emotion is expressed when someone is being honesty or integrity or truthful.

1.2.1 Principle of Sentiment Analysis

The sentiment analysis approach first reads the text present in any corpus or tweets. The method extracts the text from the input document or corpus or tweet. Further, the method performs preprocessing to eliminate the noisy words or terms. Using the noise removed feature, the method identifies the root words by applying stemming and tagging. From the noise removed feature set, the method estimates similarity measure towards different class of sentiment or emotion. Using the similarity value, the method performs sentiment detection. The general working principle of sentiment analysis

process has been presented in Figure 1.1. The sentiment analysis model receives the Text and performs preprocessing to eliminate the noisy features. Further, feature extraction is performed to measure the similarity between different documents of the data set to produce result. Each process has been discussed in detail in this section.

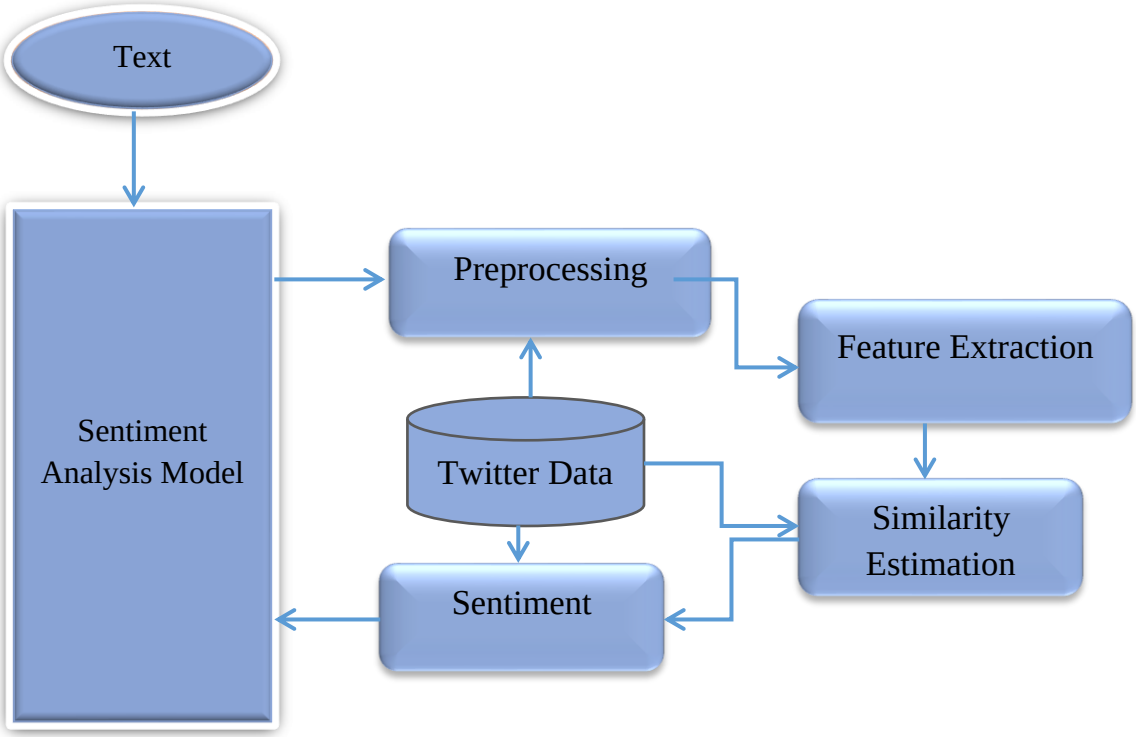

Figure 1.1 General Block Diagram of Sentiment Analysis Model

1.2.2 Preprocessing

The preprocessing is the process of preparing the input text and document towards sentiment analysis. In the preprocessing stage, the query submitted has been read and the text feature has been extracted. From the terms identified from the text, the method identifies the stop words which have no meaning. Consider the statement “I really happy to announce to join the meeting”, which has stop words like “and to” which has no meaning. Such stop words are eliminated from the text term set. Such term set identified has been used in further process of sentiment analysis.

Pseudo Code of Preprocessing:

Given: Text or Document

Obtain: Collection of Terms

Start

Read the input text

Text = Extract the text feature from input text

Terms = split the text into terms by removing punctuation

Read the stop words set

for each term from the terms set

Verify whether it is a stop word

if stop word then

Remove the term from the term set

End if

End for

Stop

The above pseudo code represents how the preprocessing is performed to remove the noisy features from the set. According to this, the method would remove the stop words and generates a term set to perform sentiment analysis.

1.3 FEATURE EXTRACTION

Feature extraction is a procedure in data mining which involves the measures for reducing the number of information available to describe a large set of information. The feature extraction is the process of extracting the required features from the text given. In this case, the text mining process

considers only the text feature which tends to identify only the pure nouns or mono gram. It has been performed by performing stemming operations on each term. While doing the Sentiment Analysis (SA) evaluation of complicated texts one of the more important issues originates from the number of factors involved. Analyzing enormous and intricate texts normally needs a lot of memory and computation power. Blessy Selvam *et al.* (2013) have discussed that in application involving substantial number of features, attribute extraction is comparable to the practice of dimensionality reduction. Each term has been tagged with the part of speech tagger provided by Stanford University as natural language processing tool. According to the result of tagger, the method identifies the pure noun to be used in text mining.

Pseudo Code of Feature Extraction:

```
Given: Term Set Ts
Obtain: Pure Terms
Start
        Read term set
        for each term
            Perform stemming on the term
            Perform tagging on the term
        if the term is noun then
            add to the result
        end if
        end for
Stop
```

The above pseudo code represents how the feature extraction is performed in selecting the required features from the text given. With the features identified and selected, the method performs similarity measurement towards sentiment analysis.

1.4 SIMILARITY ESTIMATION

The result of sentiment analysis is depending on the similarity measure being estimated between different terms of sentiment class and the terms obtained from the input text. According to that the presence of terms in both the set has been used in measuring the similarity between them.

The similarity among the text features are measured in several ways. Numbers of similarity measures are available and described earlier to measure the similarity between any documents or text and listed as follows:

1.4.1 Cosine Similarity Measure

The term vectors extracted from the documents are used to compute the similarity between them. The cosine similarity is the cosine angle present between the vectors and it's a most popular similarity measure used for text clustering and other information retrieval applications. For example given two documents Ta and Tb the cosine similarity between the term sets are computed as follows.

$$C_{Sim} = \frac{Ta \times Tb}{|Ta| \times |Tb|} \quad (1.1)$$

In Equation (1.1), Where Ta and Tb represent the vectors and each dimension shows the term weight in the document which is a non-negative value.

1.4.2 Euclidean Distance

Euclidean distance is a standard metric for geometrical problems. It is the ordinary distance between two points and can be easily measured with a ruler in two- or three-dimensional space. Euclidean distance is widely used in clustering problems, including clustering text. It satisfies all the above four

conditions and therefore is a true metric. It is also the default distance measure used with the K-Means algorithm. Measuring distance between text documents, given two documents da and db represented by their term vectors ta and tb respectively, the Euclidean distance of the two documents is defined as:

$$E_{Sim} = (\sum_{t=1}^{m} |Wt.a - Wt.b|2)1/2 \quad (1.2)$$

1.4.3 Jaccard Coefficient

The Jaccard coefficient, which is sometimes referred to as the Tanimoto coefficient, measures similarity as the intersection divided by the union of the objects. For text document, the Jaccard coefficient compares the sum weight of shared terms to the sum weight of terms that are present in either of the two documents but are not the shared terms. The formal definition is:

$$J_{Sim} = \frac{Ta \times Tb}{|Ta|2 \times |Tb|2 - (Ta.Tb)} \quad (1.3)$$

1.5 APPLICATIONS OF SENTIMENT ANALYSIS

As there's enormous growth in e-commerce websites, blogs, micro blogging websites and social media people used to share and post their views through these websites. This produces the research community to extract data's from these websites to mine the opinion within their reviews and discussion. Walaa Medhat *et al.* (2014) have explained the different applications of sentiment analysis.

- Purchasing of Service or product
- Quality Improvement in Service or product
- Marketing study
- Recommendation Systems

- Detecting of odd words use
- Detection of Opinion Spam
- Decision and Policy making

Some of Major applications of sentiment analysis are listed as follows. The sentiment analysis has been applied in several domains towards the development of different sectors. Such applications are detailed in this section.

1.5.1 Web Search with Sentiment Analysis

The web search is the activity performed by the most human in the daily life. In the activity of web search, the user submits the query to the search engine and view list of web pages. Apart from that the user would search number of web pages and perform different activities. By monitoring the web activity of different web users, the sentiment of any person can be identified. The method would maintain number of sentiment classes and each of them would have number of terms related. According to the terms from the web page or query and the terms of class available, the method would estimate different similarity measure to identify the sentiment of the person. By identifying the sentiment of any person, the web results related to the sentiment identified can be populated. Such mood-based sentiment analysis is recently integrated with the popular Google Search Engine. This would improve the performance of web search and the user would get the feeling that the web search engine is working for him.

1.5.2 Product Marketing

The product manufacturers produce various products and market them in any part of the world. In recent times, the product marketing is performed through the web and social environment. The emotion or sentiment of any person surfing in web environment or social environment can be identified by monitoring the messages and tweets. From their messages, the sentiment of the user can be identified. By identifying the user sentiment, the products related to their mood can be populated to them. This would improve the performance of product marketing.

1.5.3 Social Network

The social network is the most dominant environment being used by different users of the world. The web users or the social network users share many things between them. By identifying the mood or sentiment of users, the information related to their mood can be exchanged and shared.

1.5.4 Election Promotion

The countries facing parliamentary elections would use the candidate and party promotions through the analysis of sentiment and tweets analysis. From the tweets of different persons, you can identify the interested party and could identify the similar interested users in their friends group. According to them, the method would perform promoting the candidate and party to improve the performance.

1.5.5 Entertainment Promotion

The sentiment of any person can be identified by monitoring the activities of them. By identifying the activities and sentiment, the entertainment channels or environment can play back the songs according to their mood.

1.5.6 Feature Selection

It refers to the procedure of finding the Meaningful inputs in the extracted information for making simple processing and evaluation of the data that is available. Contribution of feature selection technique towards creating a fantastic model is essential for many reasons. one major reason for using attribute selection is that it indicates a degree of cardinality decrease, to impose a cutoff on the number of characteristics which may be taken into consideration when building a classifier model. Data offered for constructing a classifier model typically contains more information than it's required to construct the model.

1.6 AIM OF SA INVESTIGATION IN MICRO BLOGGING WEBSITES

The aim of SA is to analyze the positive, negative and neutral opinions regarding service or product posted by agency, people, society or group. Given a group of opinionated data’s, the purpose is to find all comment from the specified tuples by doing these tasks.

Twitter data analysis has come to be widely popular. A Twitter Sentiment classification performance remains elusive due to many problems: significant class imbalance at a multi-class issue, representational abundance issues for opinion cues, and using diverse colloquial linguistic patterns. These issues are problematic since many kinds of social networking analytics rely on true underlying Twitter sentiments. For instance, in Micro-blogging websites like Twitter data is represented as a short message referred to as "Tweet".

1.6.1 Challenges of SA investigation in Micro Blogging Websites

Despite many different applications of opinion Mining, it has its challenges. (Walaa Medhat *et al.* 2014). Sentiment could be expressed as a record, sentence or quality that's labelled negative or positive. However, monitoring and finding view web sites and distilling info inside them are a formidable task because of the proliferation of diverse websites. (Ayesha Rashid *et al.* 2013) have suggested an average individual reader has difficulty in identifying applicable websites, extracting data and summarizing opinions in them. The following challenges in opinion mining to perform sentimental classification are listed as follows,

- Finding out the sense of a word in a Specific sentence
- Locating immediate changes to negative polarity from positive
- Contradictions in managing the words polarity
- Placing right word in given sentence

1.7 FACTORS OF SENTIMENT ANALYSIS

The sentiment analysis is the process of mining the sentiment of any person according to the tweets obtained from the twitter data set. However, there are number of factors to be considered in sentiment analysis.

1.7.1 Relevancy

The relevancy is the factor which represents the performance of any sentiment detection algorithm in producing relevant result to the text submitted. For an efficient approach, the relevancy should be higher.

1.7.2 Irrelevancy

Any sentiment analysis algorithm would generate number of irrelevant result and it should be less for a better algorithm.

1.7.3 False Ratio in Classification

The classification of text is performed in identifying the sentiment class. For a better and efficient sentiment analysis algorithm, the false classification ratio should be less.

1.7.4 Time Complexity

The time complexity represents the total time being taken by the algorithm in producing result for a submitted text. The time complexity should be less for any efficient algorithm.

1.7.5 Sentiment Detection Performance

The sentiment detection performance represents the performance of the algorithm in all the above-mentioned factors. So for a sentiment detection algorithm the above mentioned factors should be considered.

1.8 SENTIMENT ANALYSIS WITH TWITTER DATA

The twitter organization maintains the data set from the year of 2001, which contains set of collections of tweets produced by the conversation of different uses of the environment. In general, the users of the environment perform chatting and shares various information. For example, when a user purchases a KIA-Seltos vehicle, they share the features of the car to their friends or the users present in their group. Also, even they share their mood and they

share their day to events with their friends. Such tweets has been used in different problems..

For example, if the automobile sector looks to get the implicit feedback from social network users, the tweets generated by different users towards their friends can be used. By collecting the tweets, the rating and the opinion of the car can be obtained. Similarly, the sentiment of the user can be obtained from the tweets generated by different users to solve variety of problems. The twitter data set contains various information and they are being maintained for several years. They can be grouped under various class of sentiments in different time stamp. Such grouped tweets can be used in identifying the sentiment of different users and the users of same sentiment can be grouped to perform various activities in a social environment. The analysis of sentiment with the twitter data set has been pictorially represented in Figure 1.2. The method receives the twitter data set and preprocesses each tweet to eliminate the noise. Further the noise removed data set has been extracted for its feature as text and the tweets are grouped under several categories. The clustered tweets are used to perform sentiment analysis. The identification of sentiment class is performed by measuring the similarity among the tweets and used to perform several activities like playing a song according to the user sentiment and propagating a product according to the user interest and so on.

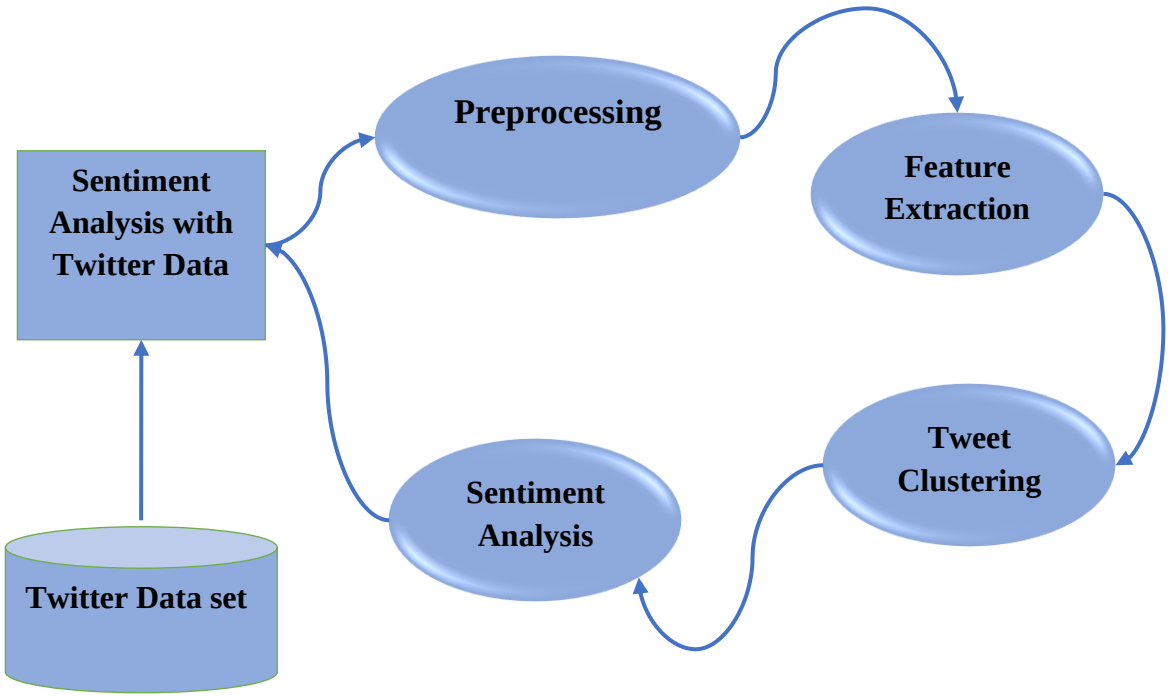

Figure 1.2 Block Diagram of Sentiment Analysis with Twitter Data

1.9 NEED FOR CLUSTERING IN SENTIMENT ANALYSIS

Sentiment analysis is the process of identifying the sentiment from the tweets available in the twitter data set. The twitter data set would contain number of tweets which are huge in size and it would belong to number of sentiments. Identify a tweets class is a challenging task when the size of data set increases. To identify the class of text towards sentiment classes, grouping them is most important. The performance of identifying the sentiment class is depending on the performance of clustering.

For example, consider the popular search engine “Google” which receives a input query and produces set of web pages related to the query. It searches over the meta data of the pages stored in the data set to produce result. But the result is highly irrelevant when you look at the pages given after few links. This really affect the performance of text mining and questions the relevancy of the content present in the web document. So, by clustering the documents using efficient algorithms, the retrieval also will be improved.

Clustering is the task of identifying groups in a data set based upon some criteria of similarity (Junaid Rashid *et al.*). Clustering aims to discover sensible organization of objects in a given dataset by identifying and quantifying similarities or dissimilarities between the objects [Donghyeon Kim]. Clustering process divides a heterogeneous structured population into more homogeneous subgroups. Process of clustering records is made according to the similarity with each others.

Clustering is applied in various fields, including data mining, statistical data analysis, compression and vector quantization. In data mining, clustering is used especially as preprocess to another data mining application. A variety of clustering formulation exists. Mostly used clustering approach is k-means and is one of the best for implementing the clustering process. The k-means algorithm implementing for cluster analysis as data mining approach has been discussed on several researches (Shu Tian *et al.*). Clustering is the process of grouping related information's under single logical name. For example there are information available about different animals in many files. Then the files can be separated and grouped based on the names of animal or the family of animals and so on. Similarly any information can be grouped under a common name so that the retrieval of the data becomes easier. Also the clustering helps indexing similar document easily.

On the other side, in the retrieval or mining phase, the categorization is performed. Categorization is the process of classifying the data or information towards N number of category. For example there exist a number of categories about the animals and at each category there exists a number of files available then the process of categorization is to identify the category to which the input file or data belongs to. This can be also called classification which is performed with the help of already trained data sets.

1.10 MACHINE LEARNING IN SENTIMENT ANALYSIS

Machine learning is the subfield of artificial intelligence that is concerned with the design and development of algorithms that allow computers to improve their performance over time based on data, such as from sensor data or databases. A major focus of machine learning research is to automatically produce (induce) models, such as rules and patterns, from data. Hence, machine learning is closely related to fields such as data mining, statistics, inductive reasoning, pattern recognition, and theoretical computer science. Similarly, machine learning can be used in several applications where the dimensionality is higher and the there are many missing values.

1.10.1 Supervised Classification Strategy

Supervised learning is an process of data mining in which a attribute is inferred from a purpose that's formed from a labeled training data. The training data includes set of training examples. Data classification in supervised technique is categorized as a two step procedure namely training stage and classification stage. Figure 1.3 describes the functioning of supervised classification strategy.

1.10.2 Supervised Classification Methods

In supervised classification algorithms, a training data set is used categorize the classification. Classification algorithm to categorize the newest data in line with the observations of this training data collection. Fundamental concepts of distinct supervised classification approaches are explained as follows.

1.10.3 Tree based Classification

Tree based target is to make a model that's capable of getting various input factors and predict that the value of a target by making the proper decisions.

1.10.4 Bayesian Classification

Classification technique that makes the classification of the available data collection by assessing the value of data tuples. Within this process a classification method is framed by assigning the class label to problem cases and represented as vectors of feature values.

1.10.5 Neural Network based Classification

It is an method of biological Neural Networks for operating in a similar manner concerning passing the data, processing the available data and making conclusions. Neural Network aren't equal to traditional calculating methods rather it needs to be regarded as complementary since the very successful neural solutions are those that function together with existing, conventional practices.

1.10.6 SVM based Classification

To discover a hyper-plane dividing the various types of the training cases with the most error margin. In this process an effort is made to form distinct hyper plane for creating the classification as well as also the best separating hyper plane is calculated by measuring the distance task.

1.11 APPROACHES IN SENTIMENT MINING

The sentiment mining process can be classified into number of cases as follows:

1.11.1 Frequency Based Sentiment Mining

The frequency-based sentiment mining scheme consider the frequency of terms in the tweet. For any tweet given, the method identifies the list of pure nouns by preprocessing the text features. Further, a set of terms are selected from feature extraction process. According to the features extracted, the method estimates the frequency of terms of input term set over the terms of different sentiment class. Using the number of occurrences, the method would compute the Term Frequency (TF) and Inverse Document Frequency (IDF) values. Using these two, the method would compute a weight measure for different category of the sentiment class and based on that the method selects a sentiment class.

Consider, there are P number of classes available, where the frequency value should be measured for the query which has Q number of terms. Then, for the class C with R number of terms, then the value of term frequency TF is measured as below:

$$TF = \frac{Number\ of\ terms\ \ present\ in\ the\ tweet\ or\ term\ set\ of\ the\ class}{total\ number\ of\ terms\ of\ the\ query}$$

The Term frequency value represent the similarity of the query with the specific class of tweet. Similarly, the similarity with other class of sentiment is measured by computing IDF measure as follows:

$$IDF = \frac{Number\ of\ Terms\ of\ tweet\ present\ in\ other\ class\ sentiment}{Total\ number\ of\ tweets}$$

Now using these two, the weight measure for the query towards the class C can be measured as follows:

Weight = TF×IDF.

According to the value of TF and IDF, the method would select a class of sentiment as result and retrieve set of tweets from the class identified.

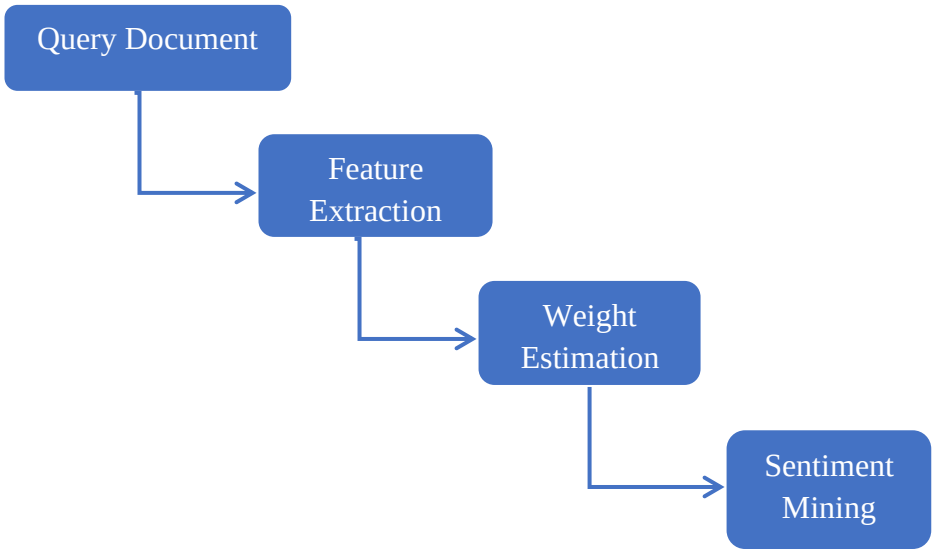

Figure 1.3 Process of Frequency Based Sentiment Mining

The process of sentiment mining according to the frequency measure is presented in Figure 1.4, which extract the text features from the tweet and measure frequency measure towards various sentiment class. According to that the method identifies a result sentiment class as result.

1.11.2 Probability based Sentiment Mining

The probabilistic approaches are more important and efficient solution in many scientific problems, where there is no exact solution identified. In this case, the same set of terms would be shared in different sentiment class. To find the sentiment class, the method first identifies set of

terms by preprocessing them and extracting the features from the tweet. Further, towards each class of sentiment the method would estimates the probability measure. According to the probability measure, the sentiment mining is performed. Consider the input tweet contains K number of terms and a class C which has P number of terms. Then the probability of the query towards the class C is measured as follows:

$$\text{Class Probability} = \frac{Number\ of\ terms\ from\ K\ appear\ in\ P\ number\ of\ target\ class\ terms}{number\ of\ terms\ present\ in\ the\ target\ class\ as\ P.}$$

The above equation represents how the class probability for the input tweet is measured. The probability value is measured towards each class of tweets and based on the value of probability, the target class is identified. From the class identified, a set of tweets are populated as result to the user.

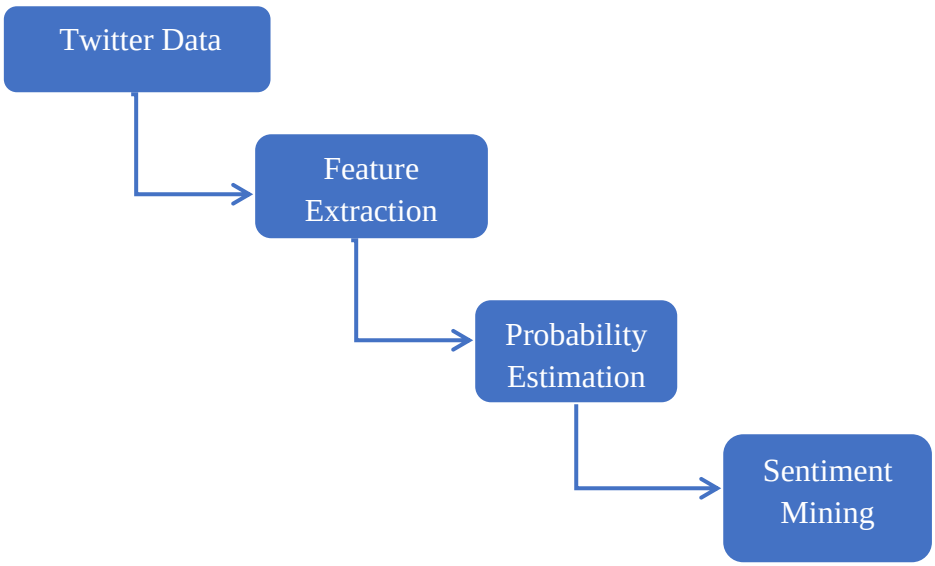

Figure 1.4 Process of Probability Based Sentiment Mining

The working of probability-based sentiment mining algorithm is presented in Figure 1.5. The method extracts the features from the query and estimates the class probability towards various classes. According to the value of probability, a single class has been identified and the tweet class has been used to perform different activities.

1.11.3 Relation based Sentiment Mining

The performance of sentiment mining is highly depending on the number of relations covered. In this approach, the method first extracts the text features from the tweet submitted, and for each of them the method performs stop word removal which has no meaning and performs stemming in each term. According to the result of stemming and part of speech tagging, the method identifies the pure nouns from the tweets. The terms of different tweets represent the relationship between them. By measuring the relationship contained in the tweet submitted, the class of tweet can be identified.

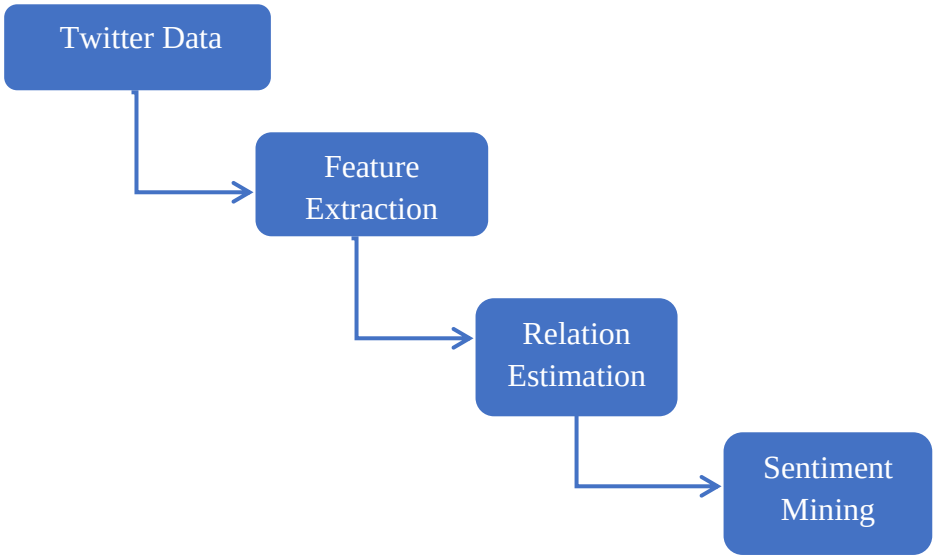

Figure 1.5 Process of Relation Based Sentiment Mining

The principle of mining sentiment according to the relationship covered by different tweet class has been presented in Figure 1.5. The method extracts the features from the tweet submitted and using the features, the method computes the relation measures and based on that the sentiment analysis is performed.

1.11.4 Rule based Sentiment Mining

Any sentiment class has set of features in terms of text and relations. This has been represented as rule. As like this, each class of sentiment has certain rule which says that in order to become a member of any sentiment class, the tweet should follow the rule mentioned. The rule based approach is method of sentiment mining which works based on certain rules. The method maintains number of rules for each class of sentiment. So, from the input tweet, the method would identify the set of terms and perform the preprocessing operations to get the pure nouns from them. Using the terms identified, the method would match with the rule available for each class. According to the rule match obtained, the query has been classified. Identified class and the tweets of the class has been selected to produce result.

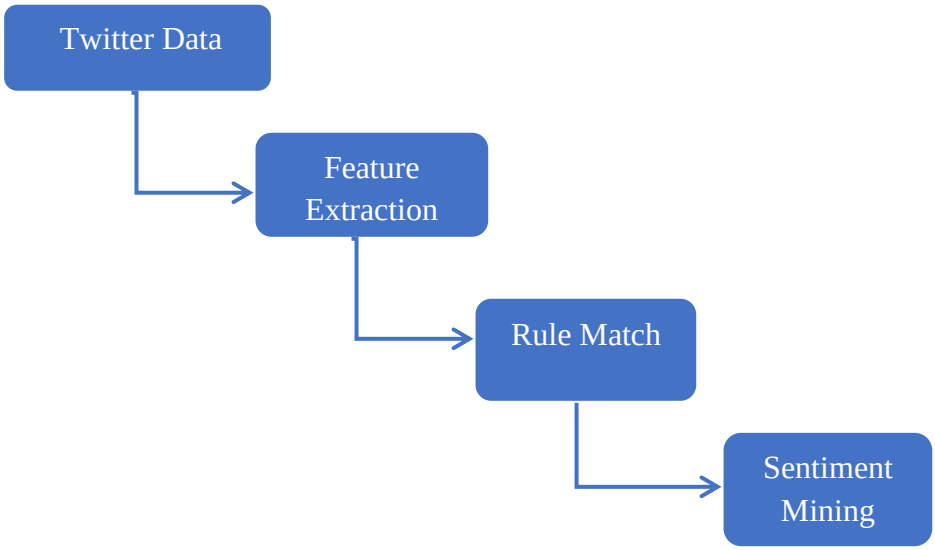

Figure 1.6 Process of Rule based sentiment mining

The general working principle of rule-based sentiment mining is presented in Figure 1.6, where the method receives the tweet and extracts the text features. Using the text features, the method measures the rule match measure to identify the sentiment class of given tweet.

1.11.5 Pattern based Sentiment Analysis

The pattern-based sentiment analysis approach maintain et of pattern of tweet where each pattern has set of terms related to the sentiment. By receiving a tweet, the method estimates the pattern similarity with each class of sentiment. According to the value of sentiment pattern similarity, the method selects a class with higher similarity as resultant sentiment class.

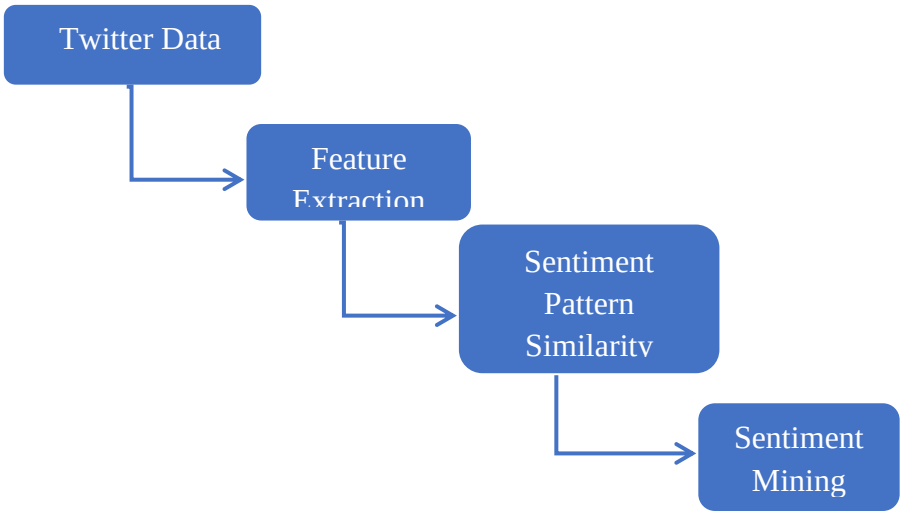

Figure 1.7 Process of pattern based sentiment mining

The working principle of sentiment mining according to pattern similarity is measured for different class of sentiment. Finally, a single class with higher similarity is selected as result.

1.12 SEMANTICS IN SENTIMENT ANALYSIS

The semantics is the most novel way of presenting the information but unlike traditional methods, it covers various relational corners of any data. The semantic is a relational meaning of any data, which represents different features and relations of data with other data. Any data has a relation with other for example; the customer data is named as relational data, because it combines personal, professional and private information.

Similarly, the semantics can be used in analyzing the semantics of the tweet. For any sentiment class, the terms and semantic meanings are combined in a class. By analyzing the tweets submitted according to the semantic features, the classification of tweets can be performed in most efficient way.

Sentiment	Types
<Happy>	<Glad, Smile, Joy>
<Sad>	<Unhappy, worry>
<Anger>	<Tensed, shout>
<Silent>	<No sound, normal>
< Fear>	<Worried, fearful>
<Disgust>	<Wrong, confused>
<Shame>	<Bad, worst feeling>
<Kindness>	<good, peace>
<Pity>	<Sorry, forgive>
<Indignation>	<Bad, angry>
<Envy>	<Not good, not fair>
<Love>	<good feeling, love>
<Surprise>	<Exclamation, Waugh>
<Trust>	<Sure, Correct>

The above text shows the semantic details of different sentiments which can be used towards various methods of classifying the tweets towards different sentiment classes.

1.13 NEURAL NETWORKS IN SENTIMENT ANALYSIS

The neural network has been used as efficient environment for any problem and it has been used for several problems. The problem of missing features and dimensionality reduction has been approached with the neural network. In the same way, the neural network can be used as the tool or key or environment for the problem of sentiment analysis. The neural network with K number of neurons on P number of layers can be designed for the problem. The tweets of each sentiment have been organized as separate neuron where for each sentiment class, a separate layer has been dedicated. Each neuron can be assigned with the job to measure the weight according to the similarity of input tweet and the tweets of sentiment class. This would return m number of weights where for each sentiment class there will be a weight assigned. According to the weight measure, the sentiment with higher weight has been selected as result.

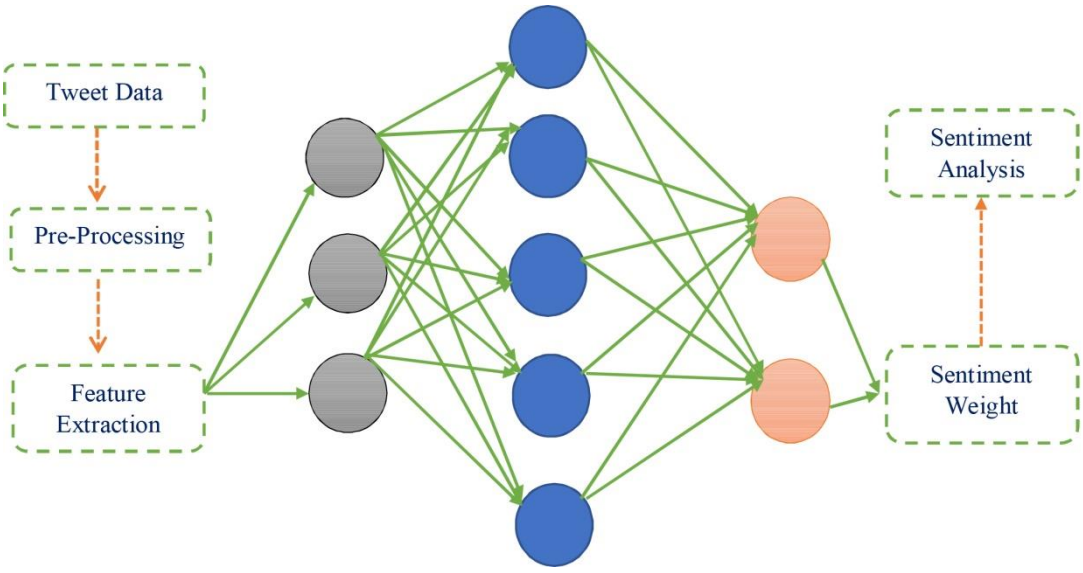

Figure 1.8 Architecture of NN Based Sentiment Analysis

The general working block diagram of NN based sentiment analysis approach is presented in the above figure 1.9 which receives the tweets and preprocesses them to extract the features. Extracted features are trained with set

of neurons at each sentiment layer. Finally, a set of weight values are obtained towards each class of sentiment and based on that a single sentiment is gets selected as result to the user.

1.14 ORGANIZATION OF THE THESIS

The following are the different sections to intent our research to produce the best performance of Sentiment analysis feature selection and classification approach.

Chapter 2: Discuss the reviews and survey of sentiment analysis information, measure with different methods on Sentiment classification and implementation of algorithm in differential stages.

Chapter 3: Elucidates the detailed view and performance of Deep Learning Modified Neural Network-based proficient Sentiment Analysis on twitter. Also, it describes the different process involved in DLMNN.

Chapter 4: Provides details and performance of Gradient Boosted Decision Tree-based Sentiment Classification of twitter data.

Chapter 5: This section presents the result and discussion to test with various parameters in various stages to prove the proposed implementation performance in the form of chart and tabulated contents.

Chapter 6: This section concludes with a summary of the achievements and limitations of the proposed methods and suggests Future research directions.

1.15 SUMMARY

In this chapter, brief analysis and procedure involved with Sentimental analysis is clarified also presented a detailed introduction on the problem of sentiment mining. It has covered the factors to be considered and the applications of sentiment mining. The methods of sentiment mining are well briefed and classified. Significance of micro-blogging websites like twitter and participation of different techniques(algorithms) in Sentimental Analysis is discussed. The organization of the entire thesis is presented in detail. The next chapter provides the literature survey around the problem of sentiment mining.