

# **MINI PROJECT-I (2020-21)**

## **SENTIMENTAL ANALYSIS**

### **REPORT**



**Institute of Engineering & Technology**

**Submitted by-**

Deepanshi Garg

(181500206)

Divya Baghel

(181500218)

Ayushi Maheshwari

(181500183)

AakhyaSwamiwal

(181500001)

***Supervised By:***

**Mr. Mandeep Singh**

*Technical Trainer*



Department of Computer Engineering & Applications

**Department of computer Engineering and Applications**

**GLA University, Mathura**

**17 km. Stone NH#2, Mathura-Delhi Road, P.O. – Chaumuhan,**

**Mathura – 281406**

## **CERTIFICATE**

I hereby declare that the work presented in this report entitled “ **Sentimental Analysis using Machine Learning**” in fulfilment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology, GLA University Mathura is an authentic record of my own work carried out over a period

under the supervision of **Mandeep Sir.**

Ayushi Maheshwari 181500183

Deepanshi Garg 181500183

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Mr. Mandeep sir

## ACKNOWLEDGEMENT

It is with profound gratitude that I express our deep indebtedness to our supervisor, **Mr Mandeep sir** without whose support and guidance it would not have been possible for us to successfully implement our project. His readiness for consultation at all the times, his educative comments, his concern and assistance even with practical things have been invaluable to us.

We are also highly grateful to all other staff members of the Department, Computer Science of Engineering for providing us the necessary opportunities for the completion of our project and owe our debt to them for their invaluable help and guidance.

Ayushi Maheshwari 181500183

Deepanshi Garg 181500206

## Table of contents

### CHAPTER 1

Introduction .....	1
1.1 What is Sentimental Analysis? .....	1
1.2 Problem statement .....	6
1.3 Objective.....	7

1.4 Sentimental Analysis Uses .....	8
1.5 Phases. ....	8
1.6 Framework.....	9
1.7 Technologies .....	10
a) Knowledge Based .....	10
b) Statistical Based .....	10
c) Hybrid.....	10

## CHAPTER 2

Literature Survey.....	11
------------------------	----

## CHAPTER 3

System Design .....	19
3.1 Concept .....	19
3.2 Language.....	19
3.3 Methodology.....	20
3.4 Tools .....	22
3.5 Data set .....	24
a) Data input.....	24
b) Part of Speech Tagging.....	25
c) Features and Opinion Words Extraction.....	25
d) Identify Sentence Polarity.....	26
e) Identify Review Polarity.....	26
f) Storing word in an array.....	26
g) Point to Point relationship.....	26
h) Conversion of Unstructured to structured words.....	27
i) The Sentiment Directory.....	27
j) Testing Dataset.....	27
k) Trained Dataset.....	28
l) Classification of Review.....	28
m) Evaluate.....	29
3.6 Extracting Features.....	30
3.7 Deep learning.....	31
3.8 Support Vector Machine.....	32

3.9 Natural Language Processing.....	33
CHAPTER 4	
Algorithms .....	25
4.1 Data set .....	35
4.2 Storing word in an array.....	36
4.3 Binary Classifying approach.....	37
4.4 Bag of word.....	38
4.5 Words of vector.....	39
4.6 Word to reviews.....	39
4.7 Libraries Used.....	40
4.8 Applications.....	41
CHAPTER 5	
Test Plan.....	46
5.1 Dataset .....	46
5.2 Files .....	46
CHAPTER 6	
Results .....	47
CHAPTER 7	
Conclusions .....	48
References .....	49

## LIST OF FIGURES

S.NO.	Title	Page No.
1	Fig 1.1 - Sentimental Analysis	1

2	Fig 1.2 - Process of determining polarities I	2
3	Fig 1.3 - Process of determining polarities I	3
4	Fig 1.4 – Polarity of Reviews	4
5	Fig 1.5 – Determining Polarities	5
6	Fig1.6 – Classification models	7
7	Fig 1.7- Division of Machine Learning	9
8	Fig 1.8 – Various Techniques	10
9	Fig 2.1- Minimum cut in shorter and cleaner way	15
10	Fig 2.2 - Framework associated with research paper	16
11	Fig 3.1 - Flowchart showing the concept used for doing the sentiment analysis.	19
12	Fig 3.2 - Basic Framework	20
13	Fig 3.3 - Expanded Framework	21
14	Fig 3.4 - Flowchart of proposed System	22
15	Fig 3.5 - Classification of neural network	23

16	Fig 3.6 - Part of Speech Tagging	25
17	Fig 3.7 - Sentence Polarity	26
18	Fig 3.8 - Determining sentimental model	29
19	Fig 3.9 - Deep Learning	31
20	Fig 3.10 - Deep learning mechanisms	32
22	Fig 3.11 - Support Vector Machine	33
23	Fig 3.12 - Natural Language Processing	34
24	Fig : 4.1 - Various steps Involved	39
25	Fig 4.2 - Framework of Train/Test Dataset classification	40
26	Fig 4.3 –Expanded Framework of Train/Test Dataset classification	40
27	Fig 5.1- Flowchart for given process.	45

## LIST OF TABLES

S.NO.	Title	Page No.
1	Table 2.1 Comparison of different polarities	12

2	Table 2.2.1 Comparison of favourability and unfavourability(Human)	13
3	Table 2.2.2 Comparison of favourability and unfavourability (System)	13
4	Table 2.3 Distribution of various categories	16
5	Table 2.4 Table Literature survey summary	17

## **ABSTRACT**

Sentimental Analysis is a new variant in the research area. It is basically refers to as opinions or views of the different data that is being collected using surveys , comments and reviews



over the web. Through this we can classify the review as positive or negative, large amount of data is being generated daily. This will determine the polarity of reviews and these reviews will be classified as two types positive and negative.

## **CHAPTER-1**

### **INTRODUCTION**

#### **1.1 What is Sentimental Analysis?**

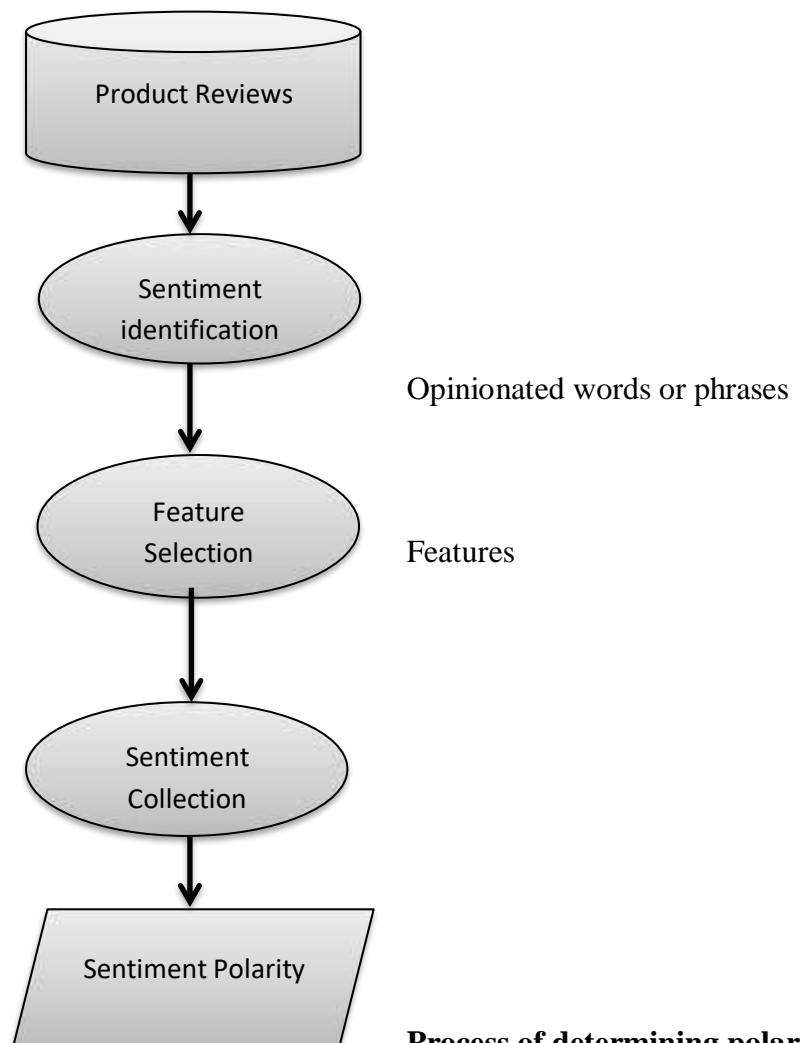
Sentiment Analysis deals with interaction between machine such as computers and natural languages used by human beings in short we can say training machine in accordance to the problem statement We can derive high quality information by using simple text entered by the user through different patterns and trends leading towards the output we expect by the various evaluations and interpretations, thus we categorize and cluster our text with in

respect to the problem statement we are dealing with. After that we compute our output by using the data set we have using different models, algorithms, mathematical computations with comes under the category of computational linguistics. This field is mainly applied when we have to take reviews or survey from our customers on products or services. Sentimental Analysis is a new variant in the research area. It is basically refers to as opinions or views of the different data that is being collected using surveys , comments and reviews over the web.



**Fig 1.1 Sentimental Analysis**

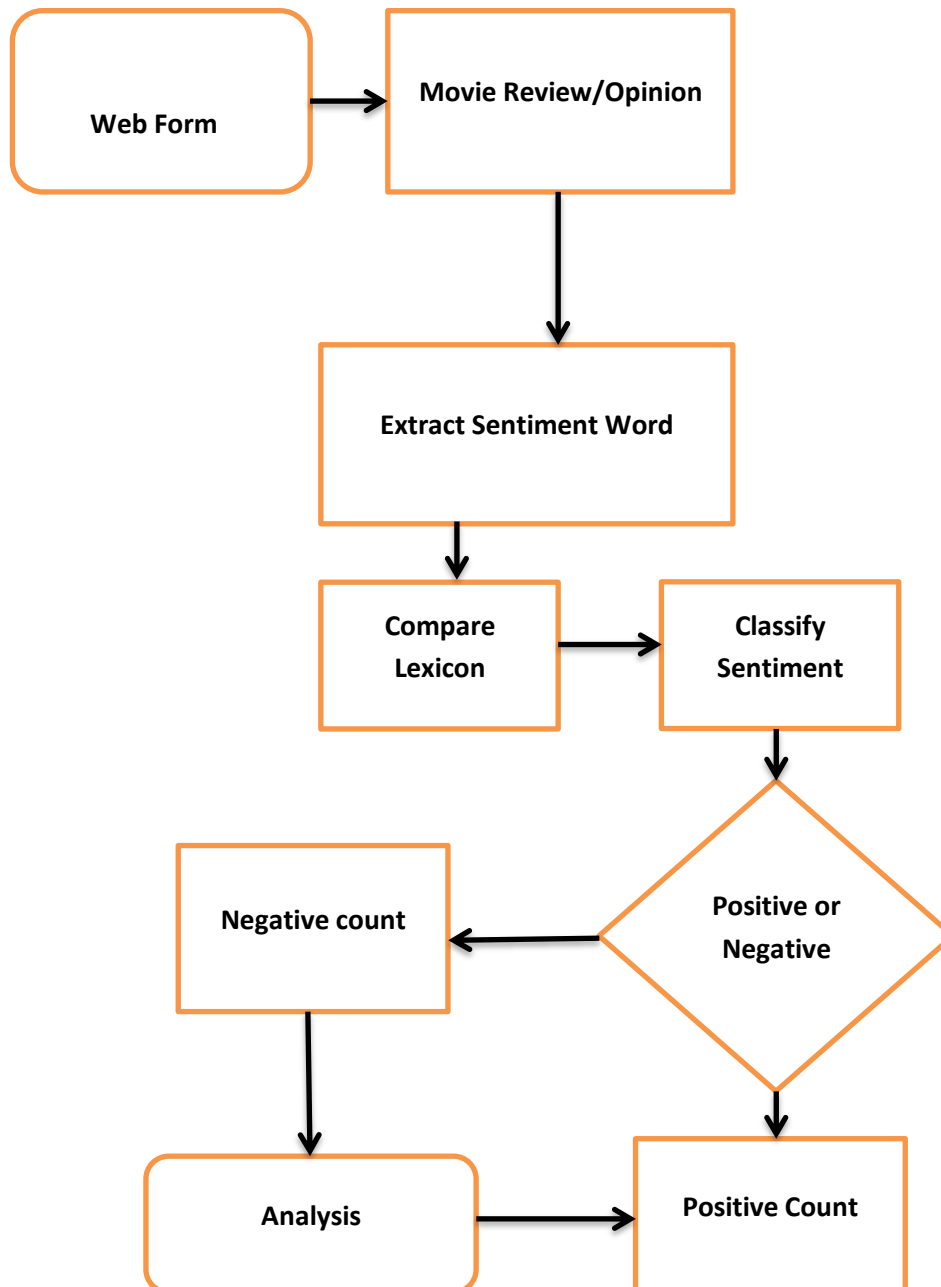
Large amount of data is being generated daily which is processed by using natural language processing, text analysis and computational linguistics. Opinion mining and sentiment analysis has been rapidly growing and exploring the views on different platforms through machine learning algorithms , techniques and tools. This research paper will determine the polarity of reviews and these reviews will be classified as two types positive and negative. Reviews of IMDB is used as source data and different algorithms like Naïve Bayes, K nearest neighbour, SVM(support Vector machines) and Ensemble classifier on different data streams has been provided. The reviews can be polar or neutral based on the how the customer felt by consuming that product or service.



**Fig 1.2**

**Process of determining polarities I**

It also determines the attitude of a person on some topic or his\her emotional reaction about some incident, documents or events. Sentiment Analysis works on the principle of Machine learning, so we took the different concepts and algorithms of machine learning, tried joining them so that we can reach the destination of our project. With the increase in technology, various social media platforms such as twitter, facebook, instagram, linked in and many other These platforms contain huge amount of data being generated daily in the form of tweets, blogs, posts, status etc. Sentimental analysis predicts the mood of these texts, tweets, reviews or posts which are available online on the platform by determining the polarity of emotions like happiness, affection, grief, anger and hatred.



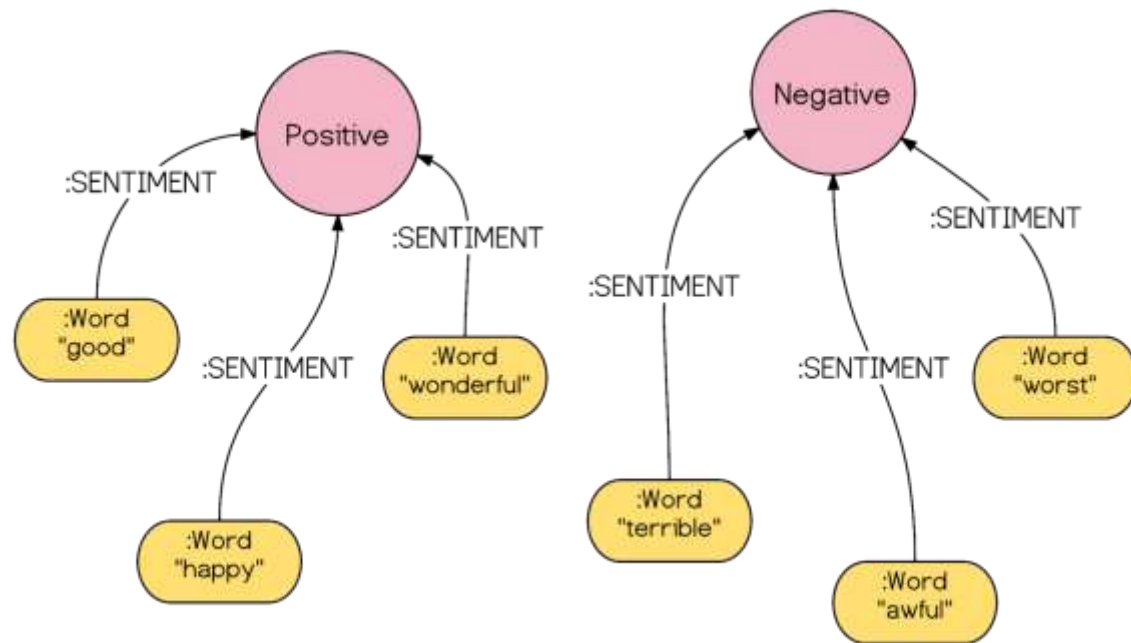
**Fig 1.3 Process of determining polarities II**

However, this task is not easy as people do not always express in the same way. Comments and reviews differ from person to person in terms of their regional languages, internet slangs, emoticons. Sentimental analysis is mainly concerned with identification and classification of opinions. It is broadly classified into two types first is knowledge based approach and other one is using machine learning techniques. Using first approach, it requires large database of predefined.



**Fig 1.4 Polarity of Reviews**

Emotions and an efficient knowledge representation for recognising opinions. But using machine learning techniques, we have train data and test data which will be used as a data set to develop a classifier. It is quite simpler also. Another task in sentiment analysis is subjectivity/objectivity identification where it focuses on classifying a given text (usually a sentence) into one of the two classes (objective or subjective). As the subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (a news article quoting people's opinions), this problem can sometimes be more difficult than polarity classification.



**Fig 1.5 Determining Polarities**

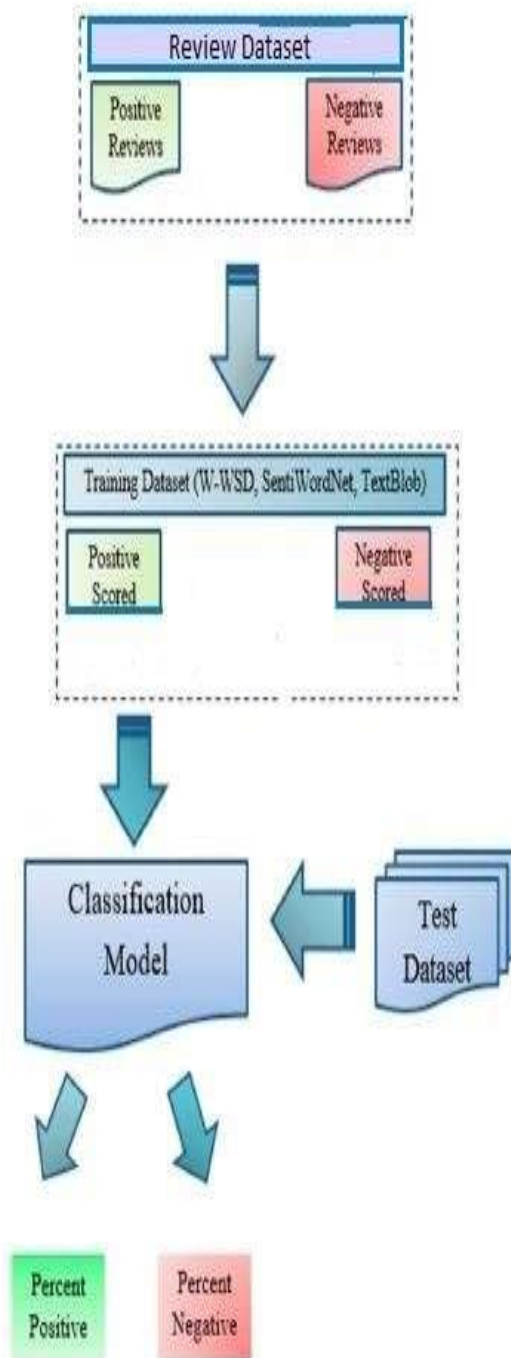
Existing approaches to sentiment analysis can be grouped into four main categories. They are keyword spotting, lexical affinity, statistical methods, and concept-level techniques. Keyword spotting classifies text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Lexical affinity improves keyword based approach by considering not only obvious affect words. It also assigns arbitrary words a probable “affinity” to particular emotions. Statistical methods influence on elements from machine learning such as latent semantic analysis, support vector machines, bag of words and Semantic Orientation. Unlike above mentioned purely syntactical techniques, concept-level approaches leverage on elements from knowledge representation such as ontologies and semantic networks so that they also able to detect semantics that are expressed in a subtle manner i.e. through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so. University of Stanford has proposed a novel approach of sentiment analysis. Most of the conventional sentiment prediction systems work just by looking at words in isolation, giving positive points for positive words and negative points for negative words and then summing up these points. In that approach, the order of words is ignored and important information is lost.

## 1.2 Problem Statement:

By using the concept of machine learning we want to get the reviews of various users about the various activities on social media and then categorize these reviews in accordance with the polarity, we want to know whether the attitude of the user for a particular issue is positive, negative and see whether the machine is able to detect that polarity and evaluate the correct output.

```
labels.txt      :      reviews.txt

NEGATIVE       :      this movie is terrible but it has some good effects . ...
POSITIVE       :      adrian pasdar is excellent is this film . he makes a fascinating woman . ...
NEGATIVE       :      comment this movie is impossible . is terrible very improbable bad interpretat...
POSITIVE       :      excellent episode movie ala pulp fiction . days suicides . it doesnt get more...
NEGATIVE       :      if you haven t seen this it s terrible . it is pure trash . i saw this about ...
POSITIVE       :      this schiffer guy is a real genius the movie is of excellent quality and both e...
```



**Fig 1.6 Classification model**

### **1.3 Objectives:**

To determine the thinking of people on a particular issue: this will tell about the extent to which people can go, thinking about a topic given to them and how they put forward their opinions, views and ideas plus how our machine will deal with this thought processing and



how correctly it will categorize the same into positive, negative To determine how positive or negative are people about a particular issue: once we categorize the reviews and get the correct output we can calculate the percentage of people who are having positive reviews about the topic, same in case of negative and people who have mixed emotions on a issue, this is very useful in rating a product or service such as if we take an example of movie rating on imdB ratings: totally depends on the percentage of people who had positive and negative reviews about the movie. So the objective of our topic is same as mentioned above and tries to maintain higher accuracy rates in judging the opinions.

#### **1.4 Sentimental Analysis various Uses:**

##### ***a)Feature identification***

Feature identification is one of the major application of sentimental analysis. For ex: the sentence “this movie has amazing plot and excellent characters”. Selected features are “Plot” and “characters. They will be used to identify the features of a review.

##### ***b)Opinion identification***

Determining the polarity i.e. positive and negative is also a very important task. For ex: the sentence “this movie has amazing plot and excellent characters” is of positive polarity because both opinion words are of positive polarity. Many words in natural language have similar meaning. We will combine them or group them as synonyms as a group of similar words together.

##### ***c) Synonyms grouping***

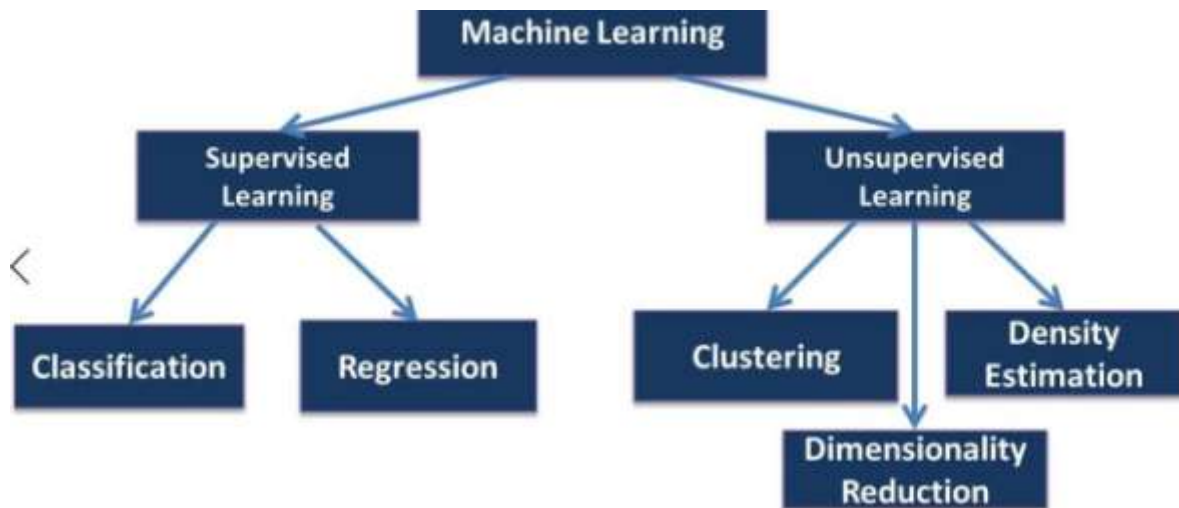
Many words in natural language have similar meaning. We will combine them or group them as synonyms as a group of similar words together.

#### **1.5 Phases:**

***a)Pre-Processing phase:*** Data set is first cleaned.

***b)Feature extraction:*** Keywords are given token and this token is now put under certain analysis.

***c)Classification phase:*** Based on different algorithms, then these keywords are put under certain category.



**Fig 1.7 Division of machine learning**

## **1.6 Framework**

To commence our project we decided to follow the following steps:

- Project Title: Semantic Analysis
- Concept: Machine Learning
- Refers to: Programming of machine to process and analyse large amount of human language data.
- Text Analysis: Deriving of the high quality information from text.

## **1.7 Techniques:**

**a)Knowledge based method:** To look for the presence of unambiguous words such as happy, sad, nice, bored etc.

**b)Statistic based method:** This method provides machine with the ability to learn as per human requirements.

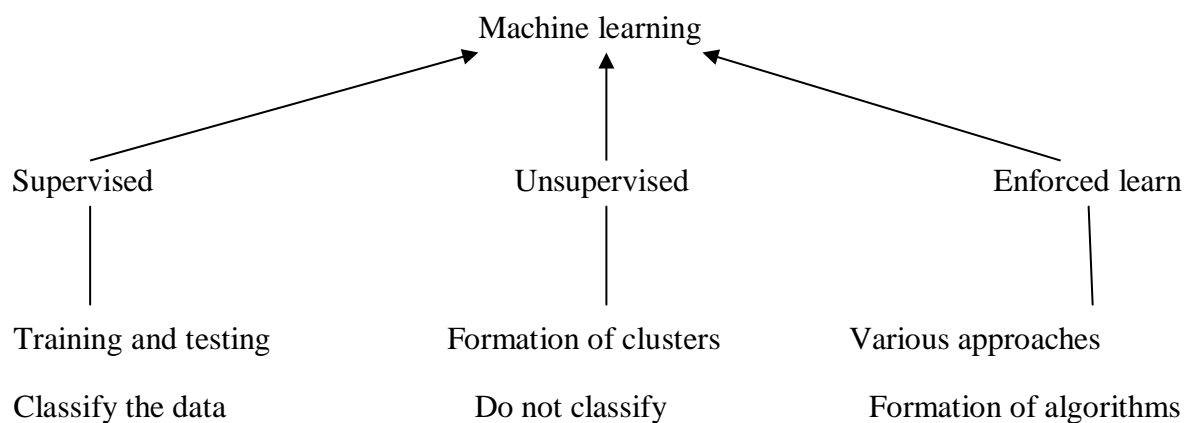
**c)Hybrid approach:** Mixture of both the above procedures that is Knowledge based and Statistical methods. After studying all the three methods in deep we decided statistic approach for the project.

## CHAPTER-2

### SYSTEM DESIGN

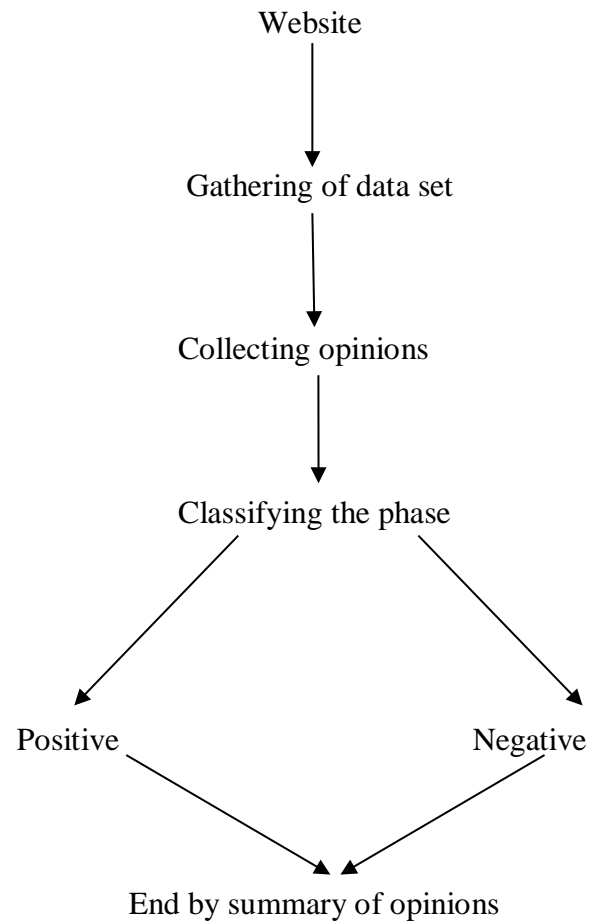
#### 3.1 Machine Learning:

As the name suggests in this concept is based on training and testing, first step is to train the machine to take different reviews or opinions from the user and give useful outputs and then comes the second step of testing if the machine gives correct outputs of similar opinions or reviews.



**Fig 3.1 Flowchart showing the concept used for doing the sentiment analysis.**

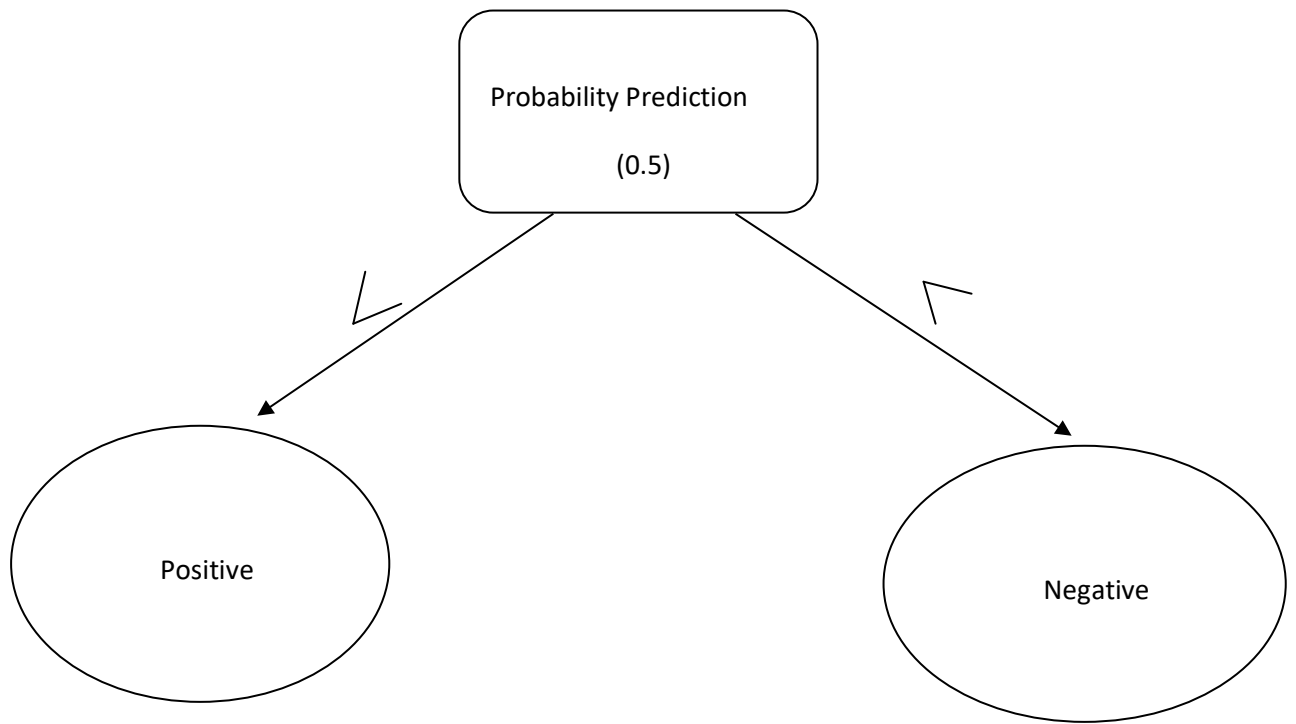
#### 3.2 Language : Python



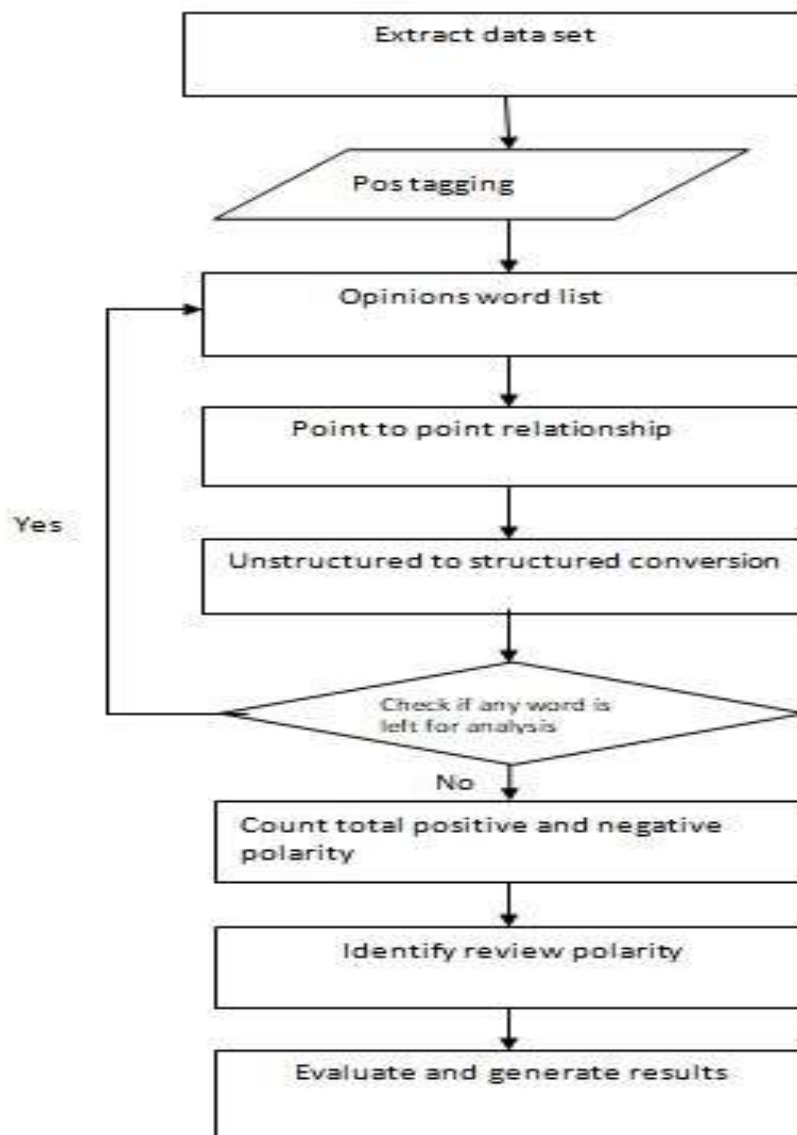
**Fig 3.2 Basic Framework**

### **3.3 Methodology:**

- a) Review line, paragraph or expression
- b) Store each word in the array
- c) Store the word count
- d) Pass this word count through a neural network
- e) Predict the Probability



**Fig 3.3: Expanded Framework**



**Fig 3.4: Flowchart of proposed System**

If the predicted probability after passing through the neural network is greater than 0.5 the sentiment is positive and if it is less than 0.5 then the sentiment is negative this will totally depend on which algorithm we will use in the project to classify the sentiments which we will come to see in the next chapter, before which we should learn about the different tools used in or system design.

### **3.4 Tools:**

### ***a) Google Colaboratory***

This is one of the tool in our project work which helps us to provide a free platform for python and it is an open source distribution used when we have to apply the concept of machine learning on large data set

## **3.5 Data set:**

Determining how the movie was and as per different views and opinions from the end users. The data set we have used in our project from imdb movie rating. Through this we can classify the review as positive or negative by using neural network that helped us formation of various ML algorithms and determine the viewpoint of different users that whether the movie is liked by them or not. The given task was not that easy like it seems but we have tried our best to reach higher precision.

### ***a) Data Input***

There are two ways to give input to the movie review sentiment analyzer. One by providing a list of reviews in JSON file format. Or by providing the TMDB ID of Movie Title.

In Case if TMDB ID a TMDB JSON API is used to fetch and store reviews in MySQL Database.

### ***c) Features and Opinion Words Extraction***

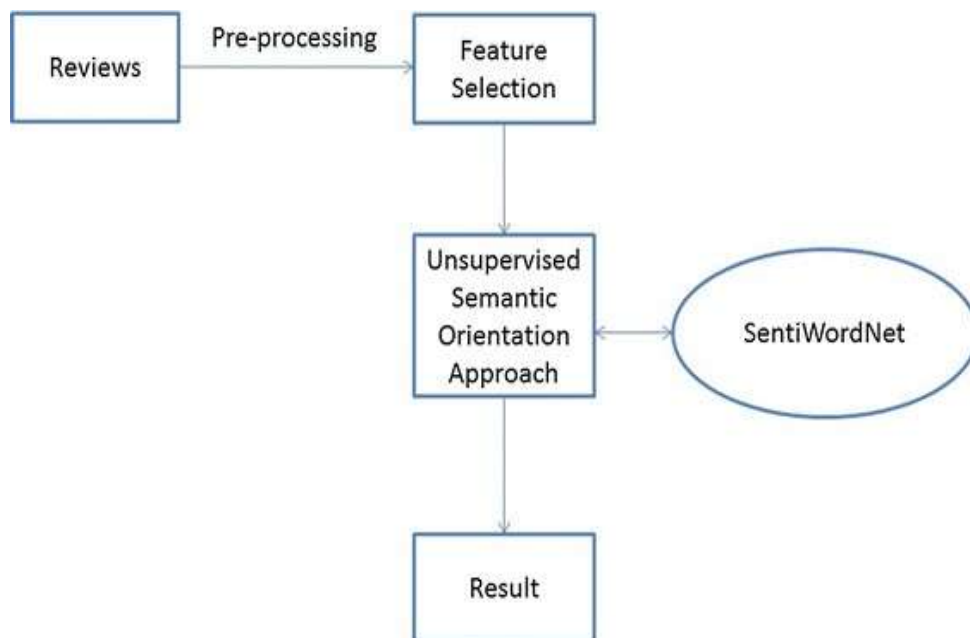
All opinion words are selected from the sentence. The system extracts all nouns, noun phrases, verbs and adjectives from the movie review and compares with the existing list of words. These words are classified on basis of their polarity. For Example “good” word is of positive polarity. On the other hand, features are selected on basis of number times occurrence of opinion words. If opinion word is an occurrence in review higher than the



threshold value then it is added features list. For this system API is trained only for movie reviews with keyword and phrases dictionary which includes “good acting”, “solid story” and “awesome action”.

#### **d) Identify Sentence Polarity**

After extracting all features and Opinion words, it is very easy to find the polarity of the sentence. Sentence polarity follows the same rules as arithmetic expressions. A negative sentiment contains all negative opinion words and positive sentiment contain all positive opinion words. A negative sentiment may contain a positive opinion word. For Example: “This movie Story is not good” sentence in a movie review. In this sentence, “good” opinion word is of positive polarity but “not” is a negative word. Therefore, the overall polarity of this sentence will be negative.



**Fig 3.7 Sentence Polarity**

#### **e) Identify Review Polarity**

Whole review polarity depends on a number of total positive or negative sentences found in a review. If the number of total positive sentences is greater than the number of total negative sentences then review polarity will be positive. Similarly, a review polarity will be negative if

the number of total negative sentences is greater than the number of total positive sentences.

***f) Storing word in an array***

We have studied the various method to store words in an array so all the subjects, objects, predicators be it adjective, verb, adverb, noun is stored in a continuous array so that classification of sentiments become more easy and then we have to store this word count in a variable called count and then keep this iteration on-going so that each word of the statement is tested in accordance with the neural network and we can easily predict the probability using the prediction algorithm that will be discussed in the coming chapters.

***g) Point to Point relationship***

Information tool helps to determine or helps in finding the relationship with one another, formulates every word association in the document to the predefined adjectives for determining various sentiments that is related to the words.

***h) Conversion of Unstructured to structured words***

Movie review are generally constructed into 'awsme' as 'awesome', 'happpy' is 'happy', 'btr', 'better' in real. The final transformation of unconstructed to constructed words is determine by data set that is converted from unstructured to structured followed by addition of vowels.

Unstructured Words to structured one

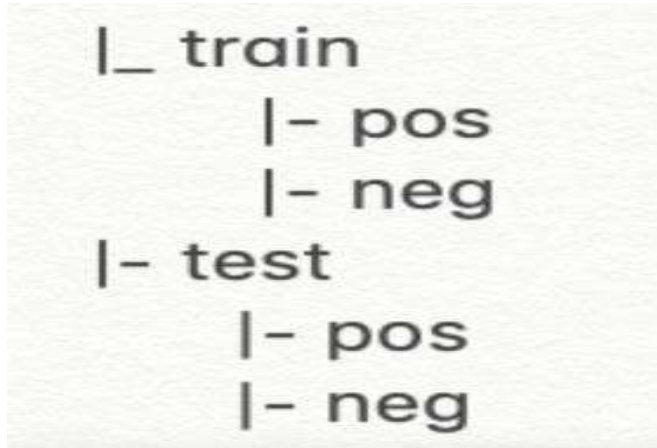
happyyy-happy  
awsm-awesome  
btr-better

***i) Testing Dataset***

The dataset we further used to test our algorithm .The test contains 2745 reviews. The goal is to find the classifier to determine and labels accurately which will further determine the polarity.

#### ***j) Trained Dataset***

The tool which helps the learning algorithm ,obtaining different features from the dataset and classifying the reviews as positive or negative.

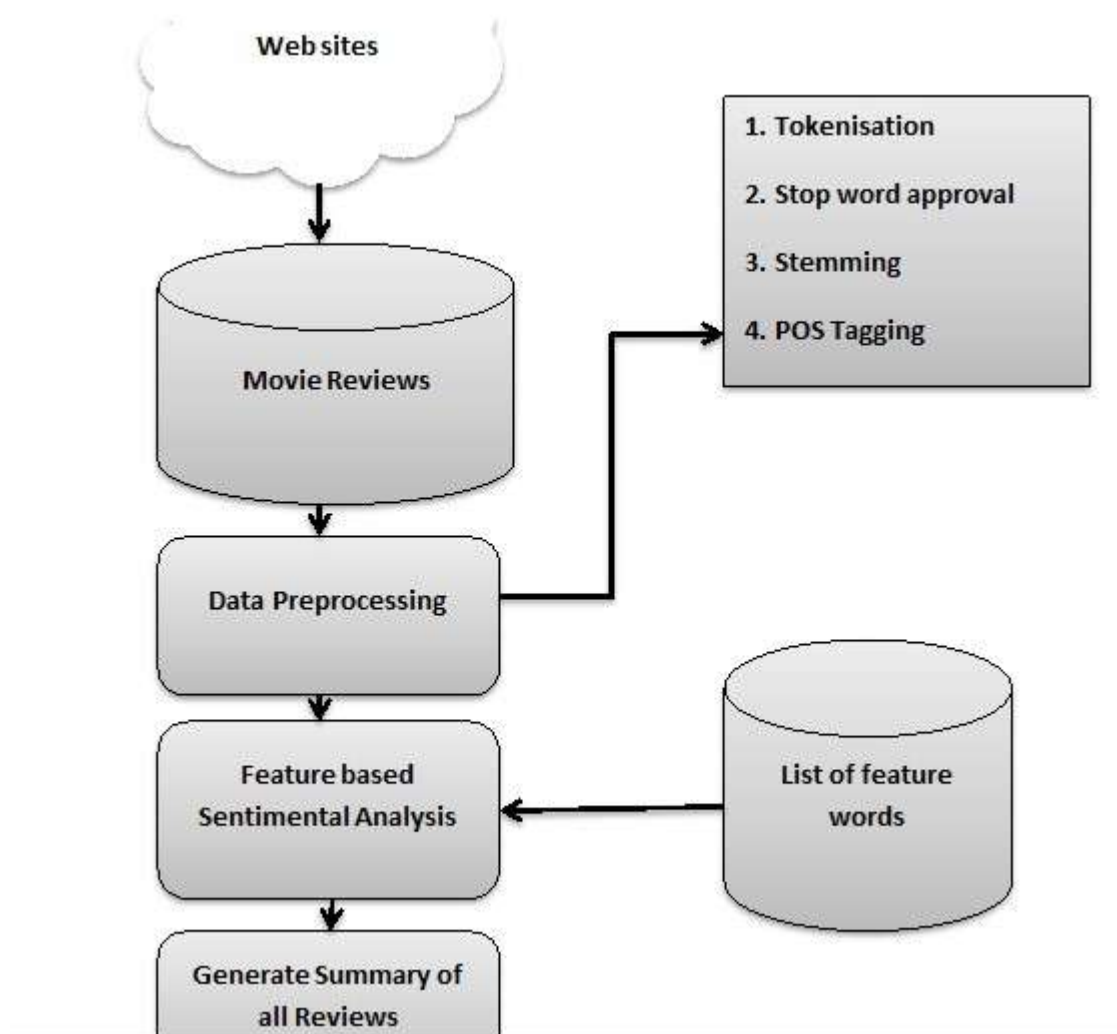


#### ***k) Classification of Review***

Once, review polarity is calculated. Review polarity percentage and polarity (Positive or Negative) classified [28] and saved for further analysis. With further analysis, box office collection can be predicted and overall performance of movie can also be predicted. ***m)***

#### ***Evaluate***

The need to evaluate our algorithm is there for understanding how it performs. For doing this we need to calculate the prediction accuracy which is nothing but the percentage of labels that were predicted correctly, higher will be the prediction accuracy better is the algorithm.



**Fig 3.8 Determining sentimental model**

Positive	Feature Words
Bajrangi Bhaijann The film is exceptionally positive .Celebrate Humanity. Doesn't take any religion or country's side.	'positive', 'Humanity', 'religion', 'country's', 'side'

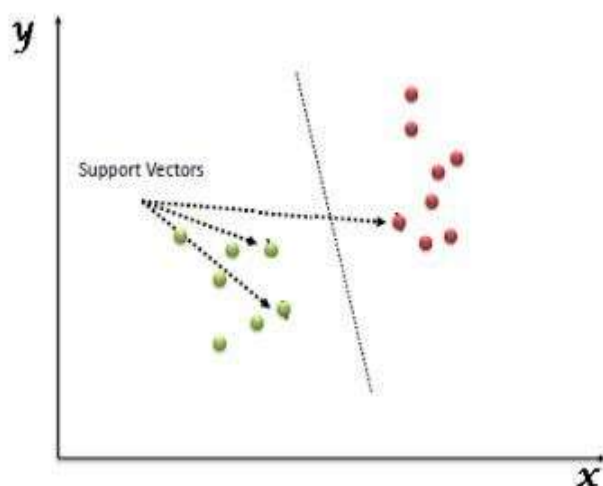
Negative	Feature Words
AT_USER disappointed. Watched a movie. It is a waste of time.	'disappointed', 'watched', 'movie', 'waste', 'time'
I miss my mom and dad. I hate this life.	'miss', 'hate'

### 3.6 Extracting Features:

The process which deals with the extraction of information containing and non redundant values present in our dataset which are then used to learn the procedure of machine learning algorithms to produce the classifier models. Once the cleaning up of 75,000 reviews was done from the training set then the creation of vocabulary using the word models was done which computes the frequency of occurrence of these words in the form of features which are then used for training the classifier. We used a kit to learn the feature extraction module for performing the actions. This module derives numeric features from the movie reviews which are in the format of text such as each string is converted into a 'tokens' which are provided with token IDs then the frequency of occurring of each token is calculated and then the tokens are organized based on how often they are occurring.

### 3.8 SVM(Support Vector Machine):

SVM stands for support vector machine. It is differentiating classifier which is formally defined by a unravelling hyper plane or for a labelled training data this algorithm results in an best hyper plane which classifies the other fresh examples. When in two dimensional spaces this hyper plane looks as a line which divides the plane into two parts where in every class lies in its each side. This algorithm moderately separates the two classes and if any of the point is in the left of line cascade in a dark circle class and the one on right falls into blue square class this is the action that the SVM performs and also looks for the hyper-plane.



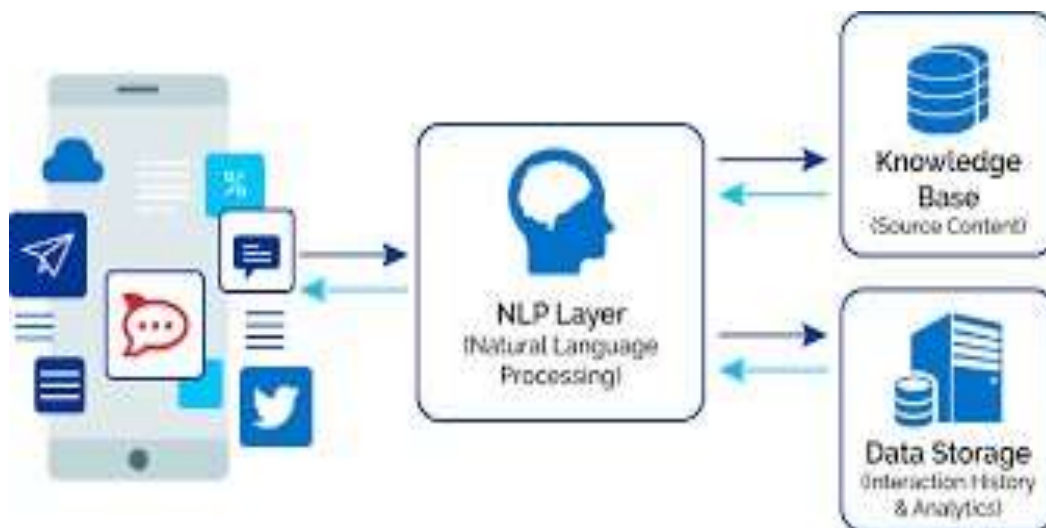
**Fig 3.11 Support Vector Machine**

### 3.9 Natural Language processing:

NLP works as follows:

- A human interacts with machine □ The machine takes the audio.
- Audio gets converted to text
- Processing of the text's data
- Data to audio conversion takes place.
- The machine responds to the human by playing the audio file.

Some of the applications where NLP is used are Language translation applications such as Google translate, Interactive Voice Response(IVR) applications, Assistant applications such as OK Google, Siri, Cortana and alexa. NLP is difficult to implement because of human nature. Even the single 's' in the language can be used for so many purpose in a word like 's' signifies plurality of items.



**Fig 3.12 Natural Language Processing**

## CHAPTER-3

### ALGORITHMS

In our project we have used the subset of concept of machine learning that is deep learning and various algorithms associated with the same. In this chapter we will define those algorithms only.

#### 4.1 Data set

The data set that we have used for our movies is from IMDB movie ratings as per different views and opinions from the end users about how the movie was. We will try to classify these reviews into positive and negative by using neural network that has helped us in the formation of various ML algorithms and prediction about the viewpoints of different users that whether they like a movie or not. Even though the task is not that easy like it seems but we have tried our best to reach higher precision.

---

So there is no way for me to plug it in here in the US unless I go by a converter. 0

Good case, Excellent value. 1

Great for the jawbone. 1

Tied to charger for conversations lasting more than 45 minutes.MAJOR PROBLEMS!! 0

The mic is great. 1

I have to jiggle the plug to get it to line up right to get decent volume. 0

If you have several dozen or several hundred contacts, then imagine the fun of sending each of them one by one. 0

If you are Razr owner...you must have this! 1

Needless to say, I wasted my money. 0

What a waste of money and time!. 0

And the sound quality is great. 1

He was very impressed when going from the original battery to the extended battery. 1

If the two were seperated by a mere 5+ ft I started to notice excessive static and garbled sound from the headset. 0

Very good quality though 1

The design is very odd, as the ear "clip" is not very comfortable at all. 0

Highly recommend for any one who has a blue tooth phone. 1

I advise EVERYONE DO NOT BE FOOLED! 0

So Far So Good!. 1

Works great!. 1

It clicks into place in a way that makes you wonder how long that mechanism would last. 0

I went on Motorola's website and followed all directions, but could not get it to pair again. 0

I bought this to use with my Kindle Fire and absolutely loved it! 1

The commercials are the most misleading. 0

I have yet to run this new battery below two bars and that's three days without charging. 1

.....

.....

-----



#### **4.2 Storing word in an array:**

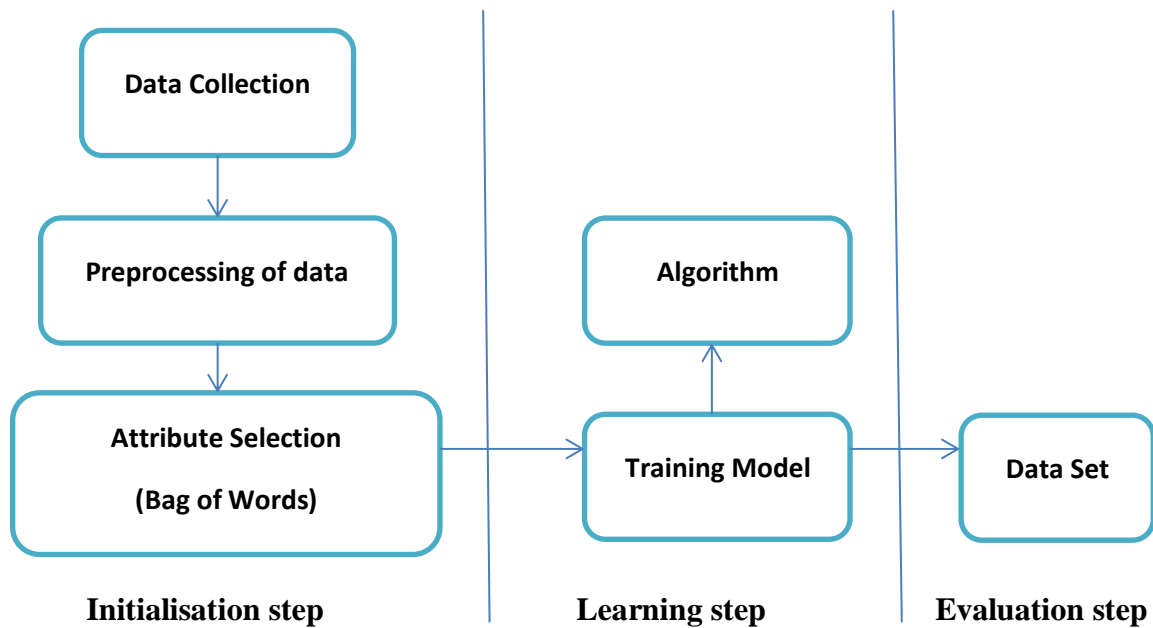
We have studied various method to store words in array to all the subjects, objects, predicators they may be verb, adjective, adverb ,noun are to be stored in a continuous array so that they can be easily classified and then we have to store this word count in a variable called count and keep the iteration going on, so that each word of the statement is tested through the neural network and determine it polarity accordingly.

#### **4.3 Binary classifying approach:**

As trying various basic sentimental approaches or methods as described above ,we have performed several pre-processing steps in order to clean up the data which includes removal of the HTML and CSS which is done using python. Then comes not so needed punctuations which were removed from regular expression which we have studied in Theory of Computation an then the conversion from lower case and the removal of stop words for conversion of a cleaned words numeric figure vectors.

#### **4.4 Bag of Word:**

This way to numerically represent the text was the simplest way. Let us take ext  $M$  we have to assign vector  $v(M)$  that belongs to our dataset, as  $V(M)$  is the frequency for the  $m$ th word in the text.  $a$  is the size of the dataset which is containing the word in the dataset and some are rare. After Learning about the bag of words, we will use different classifiers to classify our data.



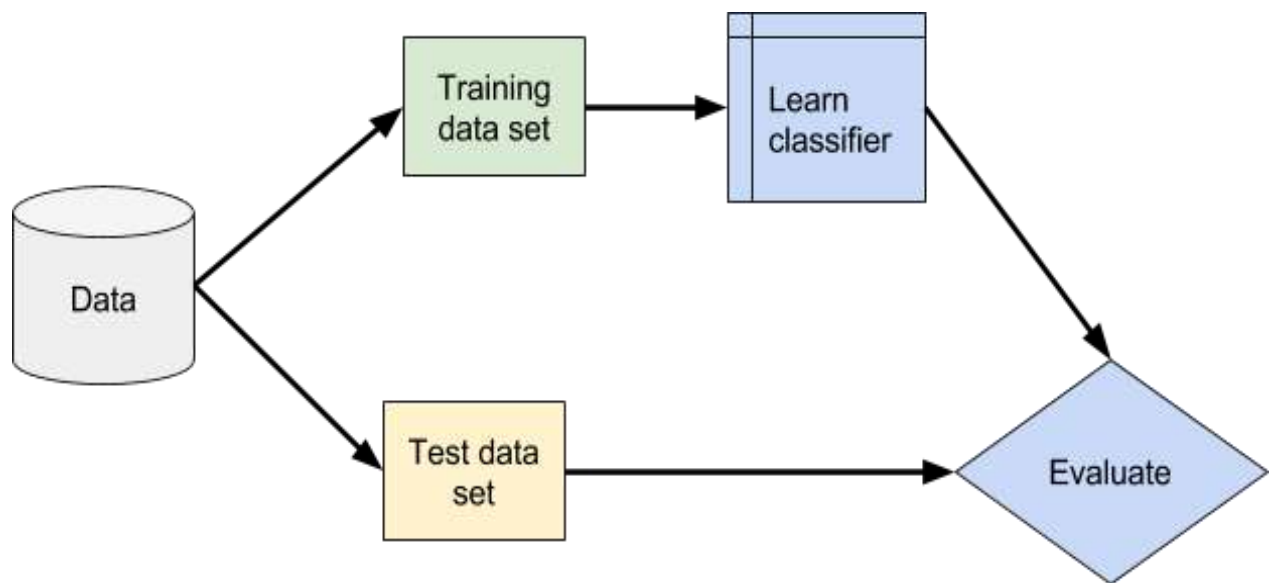
**Fig : 4.1 Various steps Involved**

#### **4.5 Words to Vector:**

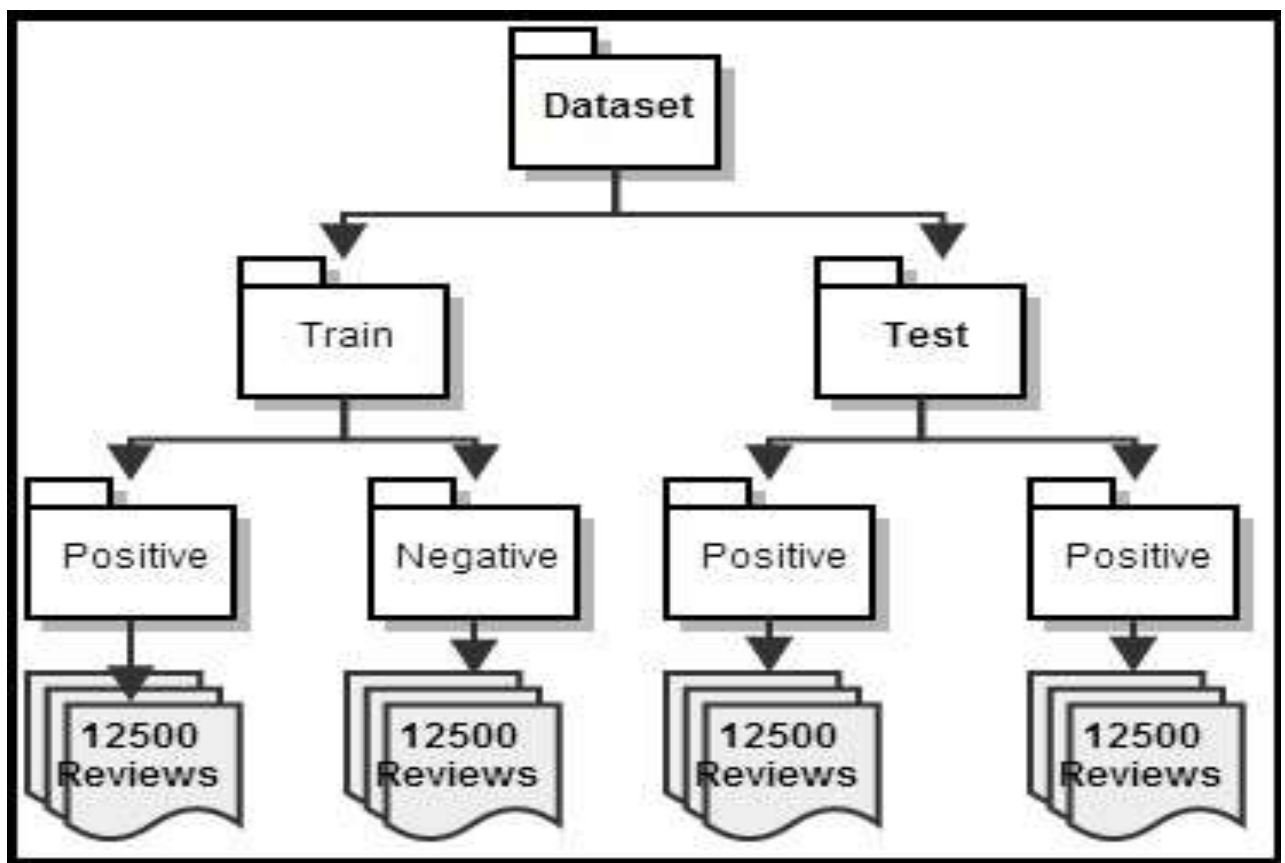
This is the another way to numerically represent the texts by converting every word present in the text to its respective vector. One of th important aspect of the word to vector concept is, it is not dependent on sentiment analysis and does not require to have a labelled dataset. Therefore we have taken 75,000 words out of which 35k are labelled and 35K are unlabelled one.

#### **4.6 Words to reviews:**

The simplest way to assign a feature vector to the words of the data set which is also known as the review two stop words are frequent words they are ‘the’ and ‘is’ they do not have any determined sentiments. The Natural language toolkit package identifies the stop words by statistic analysis on some large data corpus, the two word vectors of all the words present in our dataset. As the stop words do not change the polarity of the paragraph so we can ignore them as they will affect our precision at a negligible amount.



**Fig 4.2 Framework of Train/Test Dataset classification**



### Fig 4.3 Expanded Framework of Train/Test Dataset classification

#### 4.7 Libraries used:

a) **pandas** In computer programming, **pandas** is a software **library** written for the **Python** programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license

b) **NLP. Python libraries. Natural language processing (NLP)** is a field located at the intersection of data science and Artificial Intelligence (AI) that – when boiled down to the basics – is all about teaching machines how to understand human languages and extract meaning from text..

c) **spacy spaCy** is a free and open-source **library** for Natural Language Processing (NLP) in Python with a lot of in-built capabilities. It's becoming increasingly popular for processing and analyzing data in NLP.

d) **Scikit-learn** is a free machine **learning library** for **Python**. It features various algorithms like support vector machine, random forests, and k-neighbours, and it also supports **Python** numerical and scientific **libraries** like NumPy and SciPy

.

#### 4.8 Applications:

##### a)Feature identification:

Feature identification is one of the major application of sentimental analysis. For ex: the sentence “this movie has amazing plot and excellent characters”. Selected features are “Plot” and “characters. They will be used to identify the features of a review.

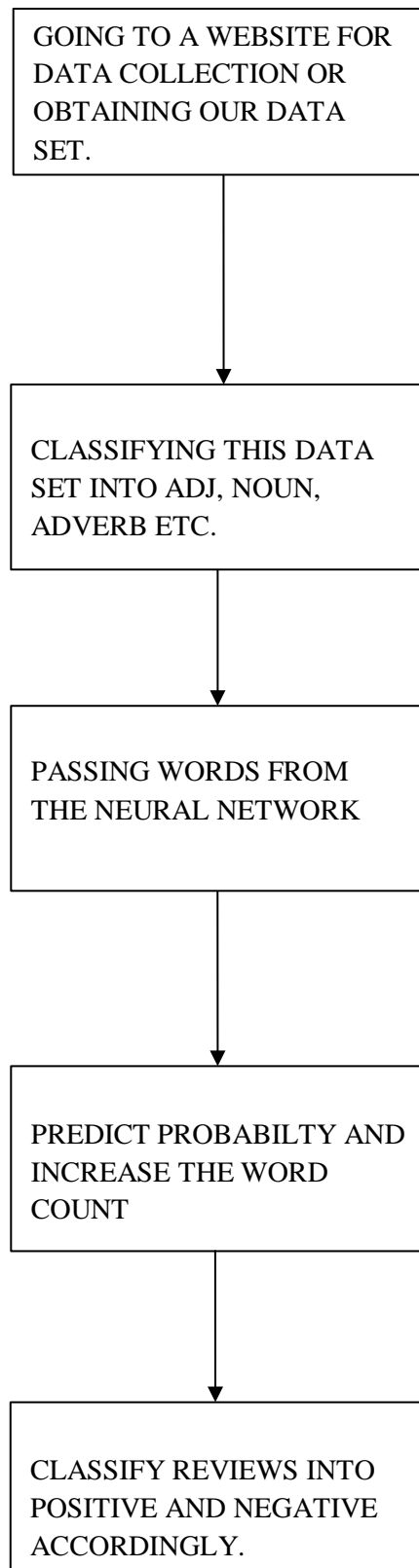
##### b)Opinion identification:

Determining the polarity i.e. positive and negative is also a very important task. For ex: the sentence “this movie has amazing plot and excellent characters” is of positive polarity

because both opinion words are of positive polarity. Many words in natural language have similar meaning. We will combine them or group them as synonyms as a group of similar words together.

**c) Synonyms grouping:**

Many words in natural language have similar meaning. We will combine them or group them as synonyms as a group of similar words together.



**Fig 5.1 Flowchart for given process.**

## **CHAPTER-6**

### **RESULTS**

The system is tested for 50 plus different movie titles each with max 10 reviews. It shows no of provided features, the accuracy of the system (Accurate Result Percentage), Error Percentage and False Negative Percentage and False Positive Percentage. False Negative means a positive polarity review considered as negative. Similarly, False Positive means a negative polarity review considered as positive. The average accuracy of this system for test review is 81%.

## **CHAPTER-7**

### **CONCLUSION**

In this paper, movie reviews are classified into positive or negative polarity. The system proposed by author in the paper can be used to classify a huge database of movie reviews. This will help movie producers to check the status of their movie. Future work, this API can be trained for other reviews like smartphones, laptops or clothes etc.



