

**SENTIMENTAL  
ANALYSIS AND  
DEPRESSION  
DETECTION IN TWEETS  
USING MACHINE  
LEARNING AND  
PYTHON**

# **SENTIMENTAL ANALYSIS AND DEPRESSION DETECTION IN TWEETS USING MACHINE LEARNING AND PYTHON**

Submitted in partial fulfilment of requirement for the award of the degree of

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE ENGINEERING**

**by**

**Aditi Tiwari (35127202718)**

**Soumya Bhardwaj (40827202718)**

**Hitanshu Vig(42127202718)**



**Department of Computer Science and Engineering**

**Session 2018-2022**

**GREATER NOIDA INSTITUTE OF TECHNOLOGY**

**PLOT NO. 6C, KNOWLEDGE PARK – II, GREATER NOIDA**

**(Affiliated by Guru Gobind Singh Indraprastha University)**

# ABSTRACT:

The advent of different social networking sites has enabled anyone to easily create, express, and share their ideas, thoughts, opinions, and feelings about anything with millions of other people around the world. With the advancement of technology, minicomputers and smartphones have come into human pockets and now it is very easy to share your idea about anything on social media platforms like Facebook, Twitter, Wikipedia, LinkedIn, Google+, Instagram, etc. Due to the tremendous growth in population and communication technologies during the last decade, the use of social networks is on the rise and they are being used for many different purposes. One such service for which their use may be explored is an analysis of users' posts to diagnose depression.

Depression is a common illness worldwide with potentially severe implications. Early identification of depressive symptoms is a crucial first step towards assessment, intervention, and relapse prevention. With an increase in data sets with relevance for depression and the advancement of machine learning, there is a potential to develop intelligent systems to detect symptoms of depression in written material.

# OUR APPROACH:

Our goal is to define whether a sentence has a sentiment or not and if it does, to determine whether the emotion is positive, negative, or neutral.

We have worked on modeling two models with different approaches in order to find the algorithm which provides maximum precision and accuracy. The two models are similar in working but one focuses on Detecting Depression or negativity in the input message from a user after training and testing the model using a dataset. This model uses both Naive Bayes and Logistic Regression. This model gives an accuracy of 95-96%. This model uses the dataset to train the model and then forms a wordcloud to depict the words or phrases that have the maximum frequency. More the space occupied in the wordcloud, more is the frequency of the word in the dataset. Once it is trained and tested, the accuracy and precision table is generated. Ultimately, the model has an option where a user can enter a message, paragraph, word, phrase etc and predict the sentiment of the user.

The other model mainly focuses on the sentimental analysis of texts/tweets by segregating them into categories of positive, negative, or neutral, ultimately giving us data that can be put to use to enhance further models already in use. This model clearly visualizes the data into categories using graphs and wordcloud. The precision of this model is 0.9456 and the accuracy of this model is 94%

# INTRODUCTION

Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

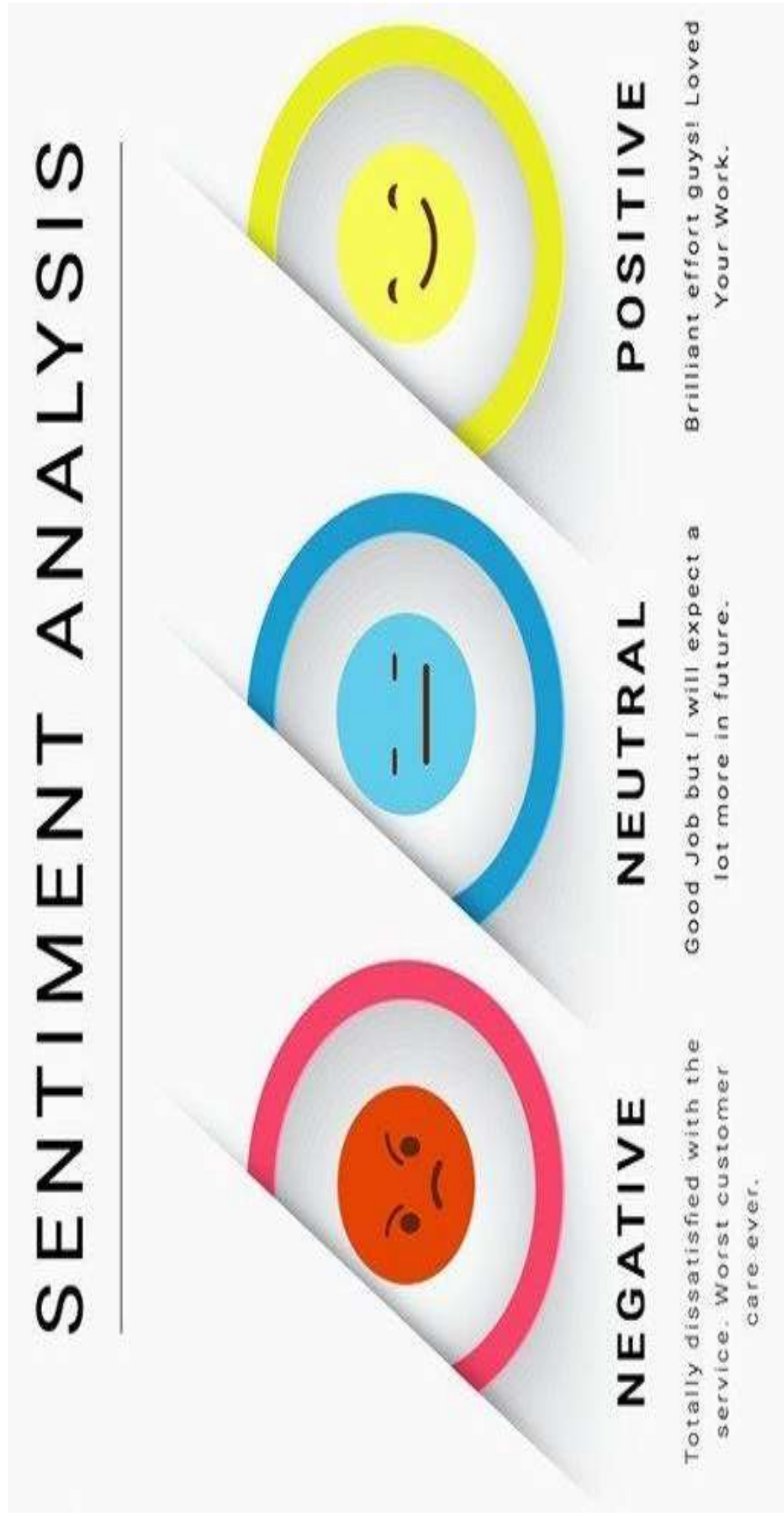
Sentiment analysis focuses on the polarity of a text (positive, negative, neutral) but it also goes beyond polarity to detect specific feelings and emotions (angry, happy, sad, etc), urgency (urgent, not urgent) and even intentions (interested v. not interested). Depending on how you want to interpret customer feedback and queries, you can define and tailor your categories to meet your sentiment analysis needs.

Since humans express their thoughts and feelings more openly than ever before, sentiment analysis is fast becoming an essential tool to monitor and understand sentiment in all types of data.

Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs.



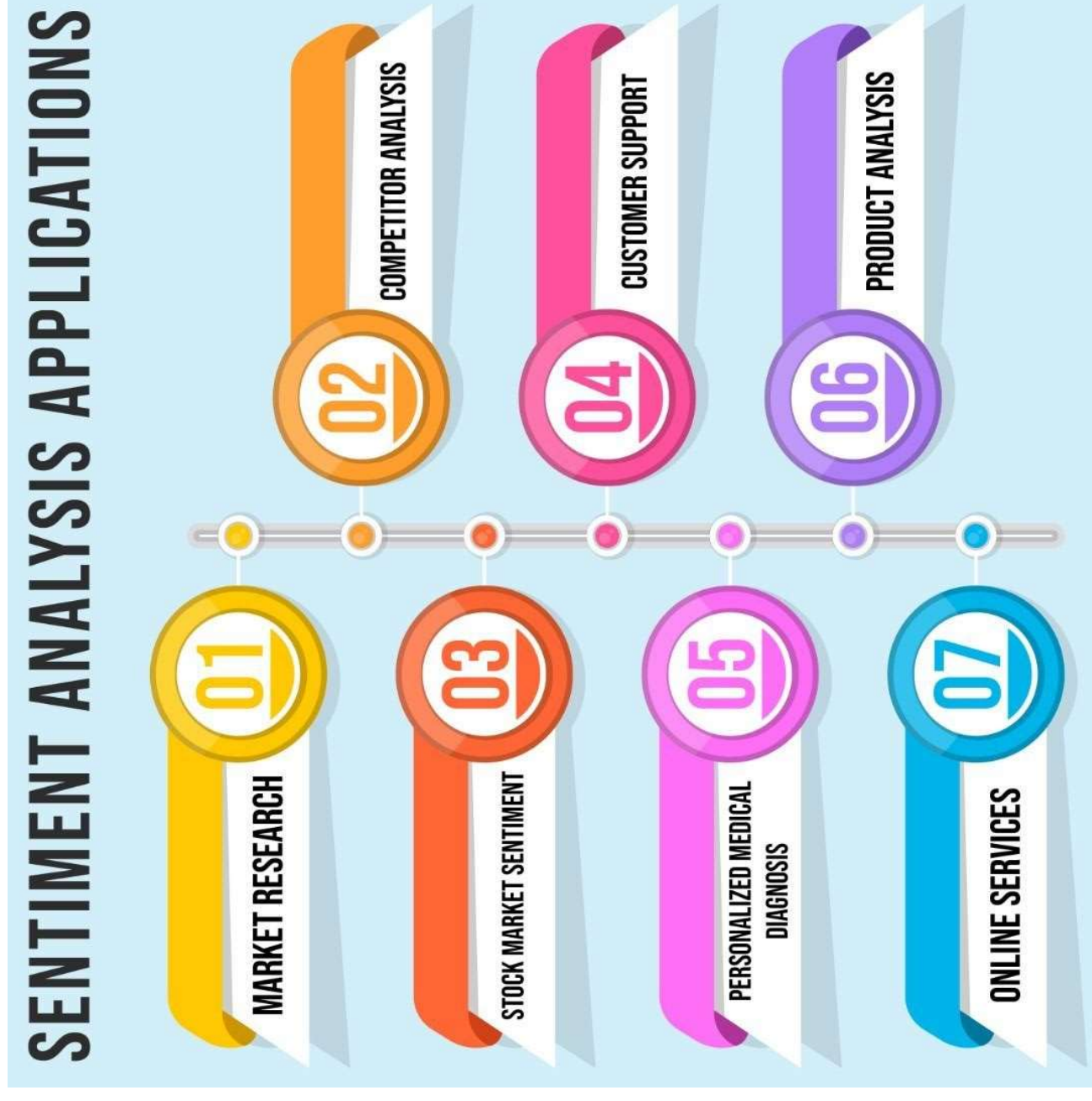
A common practice of almost all imaginable brands, companies, services, political campaigns, hospital services, etc. is that they require user feedback. They build a huge dataset with such feedbacks and ultimately hire a team of data scientists to work on the data and help them improve their value/prestige/service experience etc. The following is an example of one such survey.



# Applications of Sentimental Analysis:

The most popular applications of sentiment analysis:

- Social media monitoring
- Customer support ticket analysis
- Brand monitoring and reputation management
- Listen to voice of the customer (VoC)
- Listen to voice of the employee
- Product analysis
- Market research and competitive research
- Social Media Monitoring
- Analyzing customer feedback
- Customer Support Management
- Improve the brand performance
- Block the tweets/posts reflecting a particular sentiment



## An example of how sentiment of text is determined:

*"I am happy with this water bottle."*



Positive

*"This is a bad investment."*



Negative

*"I am going to walk today."*



Neutral



# MACHINE LEARNING IN SENTIMENTAL ANALYSIS AND DEPRESSION

## DETECTION:

Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.

Similar to how the human brain gains knowledge and understanding, machine learning relies on input, such as training data or knowledge graphs, to understand entities, domains and the connections between them. With entities defined, deep learning can begin.

The machine learning process begins with observations or data, such as examples, direct experience, or instruction. It looks for patterns in data so it can later make inferences based on the examples provided. The primary aim of ML is to allow computers to learn autonomously without human intervention or assistance and adjust actions accordingly.

# Methods of Sentimental Analysis:



- ❑ Data Collection - Consumers usually express their sentiments on public forums like the blogs, discussion boards, product reviews as well as on their private logs – Social network sites like Facebook and Twitter. Opinions and feelings are expressed in different way, with different vocabulary, context of writing, usage of short forms and slang, making the data huge and disorganized. Manual analysis of sentiment data is virtually impossible. Therefore, special programming languages like ‘R’ are used to process and analyze the data.
- ❑ Text Preparation - Text preparation is nothing but filtering the extracted data before analysis. It includes identifying and eliminating non-textual content and content that is irrelevant to the area of study from the data.

- ❑ Sentiment Detection - At this stage, each sentence of the review and opinion is examined for subjectivity. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded. Sentiment analysis is done at different levels using common computational techniques like Unigrams, lemmas, negation and so on.
- ❑ Sentiment Classification - Sentiments can be broadly classified into two groups, positive and negative. At this stage of sentiment analysis methodology, each subjective sentence detected is classified into groups-positive, negative, good, bad, like, dislike.
- ❑ Presentation of Output - The main idea of sentiment analysis is to convert unstructured text into meaningful information. After the completion of analysis, the text results are displayed on graphs like pie chart, bar chart and line graphs

# Dataset Used:

The Sentiment140 dataset available on Kaggle contains 1,600,000 tweets extracted using the Twitter API. The tweets have been annotated (0 = negative, 2 = neutral, 4 = positive) and they can be used to detect sentiment. It contains a numerous fields but the following we have used:

label/target: the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)

ids: The id of the tweet ( 2087)

date: the date of the tweet (Sat May 16 23:58:44 UTC 2009)

user: the user that tweeted (robotickilldozr)

message/text: the text of the tweet (Lyx is cool)

You can find the dataset here: <https://www.kaggle.com/kazanova/sentiment140>.

For our project, we just took a sample of 50,000 tweets, containing tweets of all polarities ranging from 0 to 4.

# Algorithms Used:

## 1. NAIVE BAYES

The Bayes theorem is used by the Naive Bayes classifier to forecast membership probabilities for each class, such as the likelihood that a given record or data point belongs to that class.

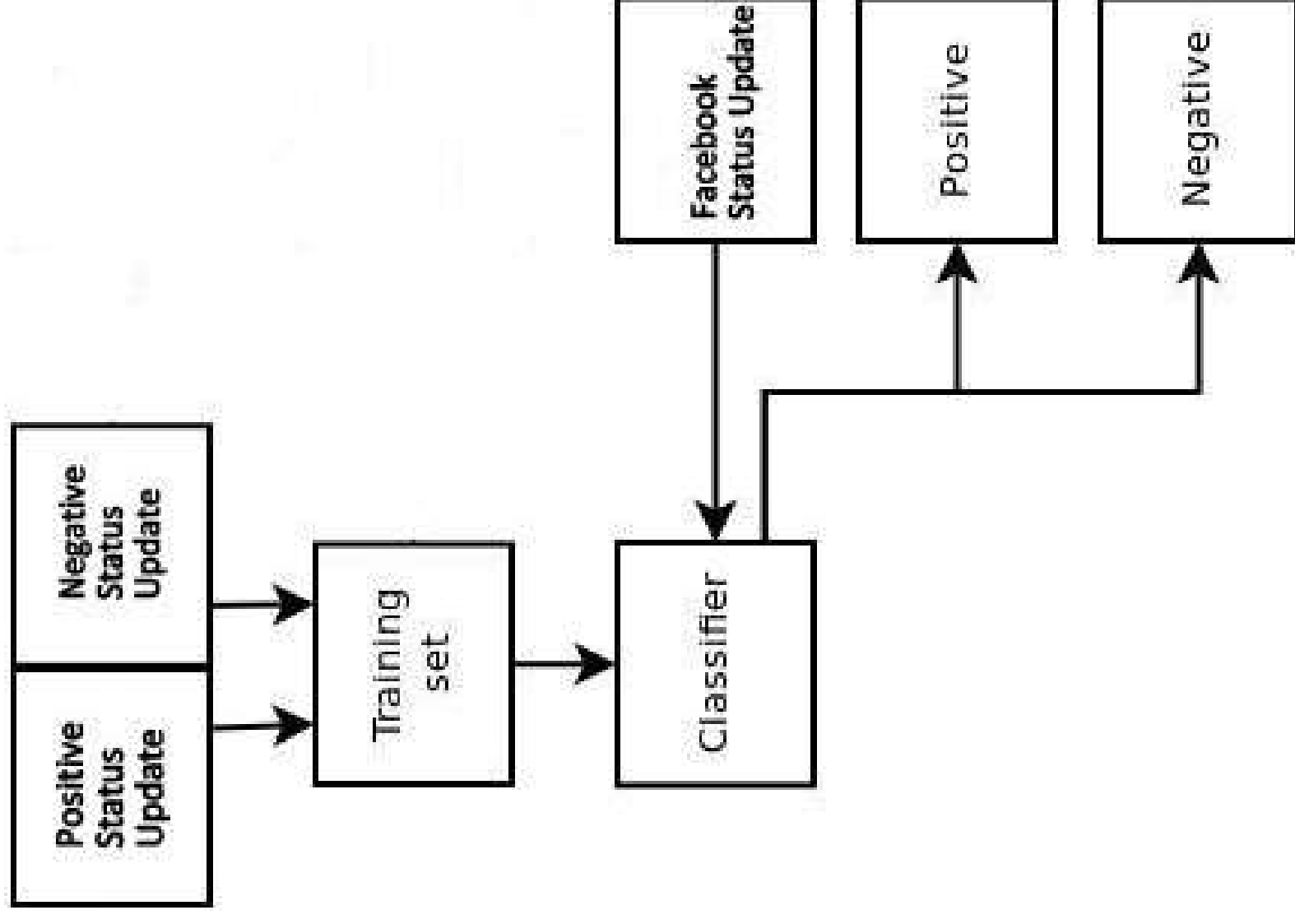
The Naive Bayes model uses features that are reasonably **easy to understand**. It supports **large-scale** Sentiment Analysis efforts since the computation required for training is fast.

It requires a small amount of training data to learn the parameters. It can be trained relatively fast compared to other models.

The most likely class is defined as the one having the highest probability.

Posterior = likelihood \* proposition / evidence

$$P(A|B) = P(B|A) * P(A) / P(B)$$



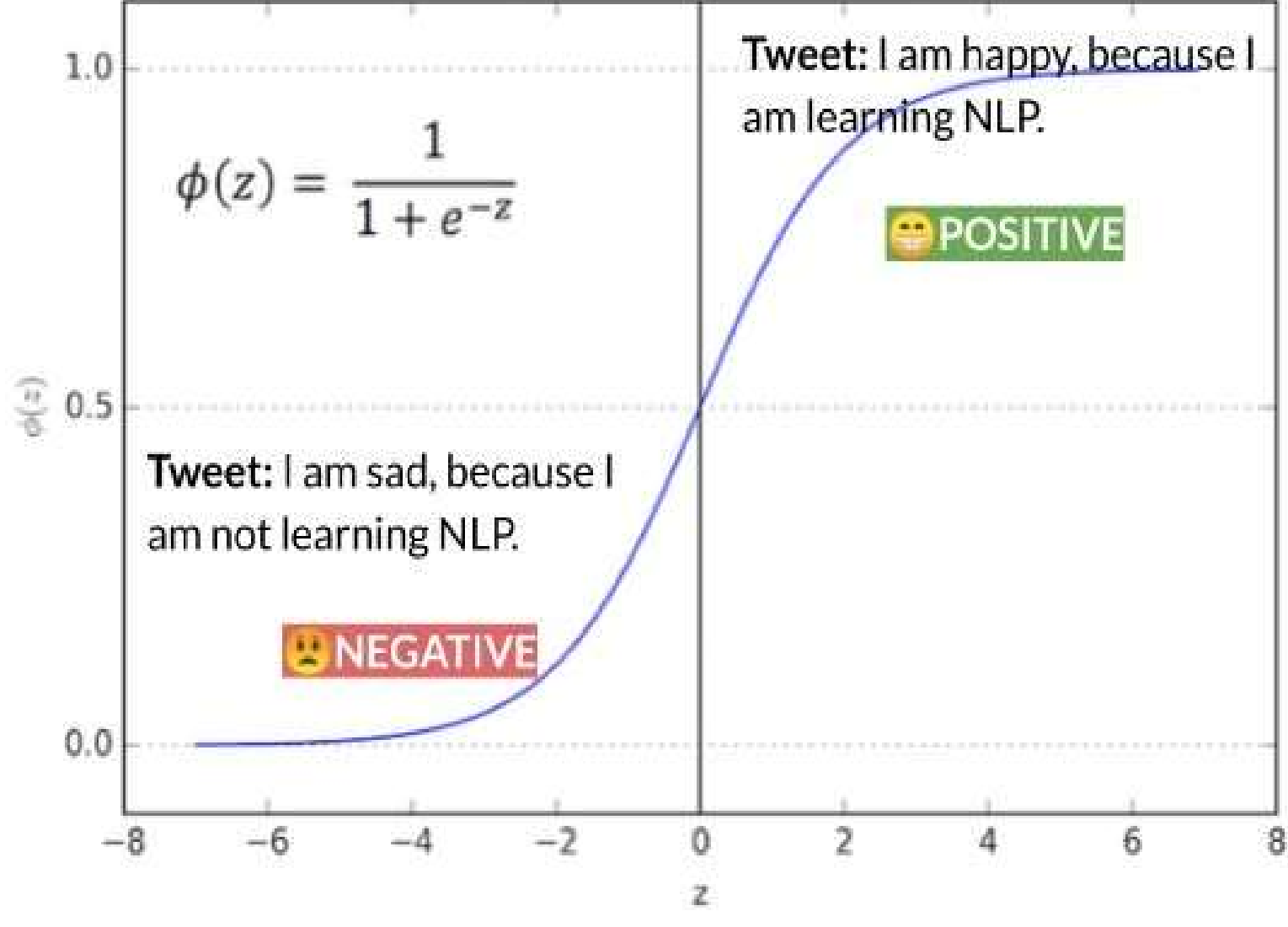


## 2. LOGISTIC REGRESSION:

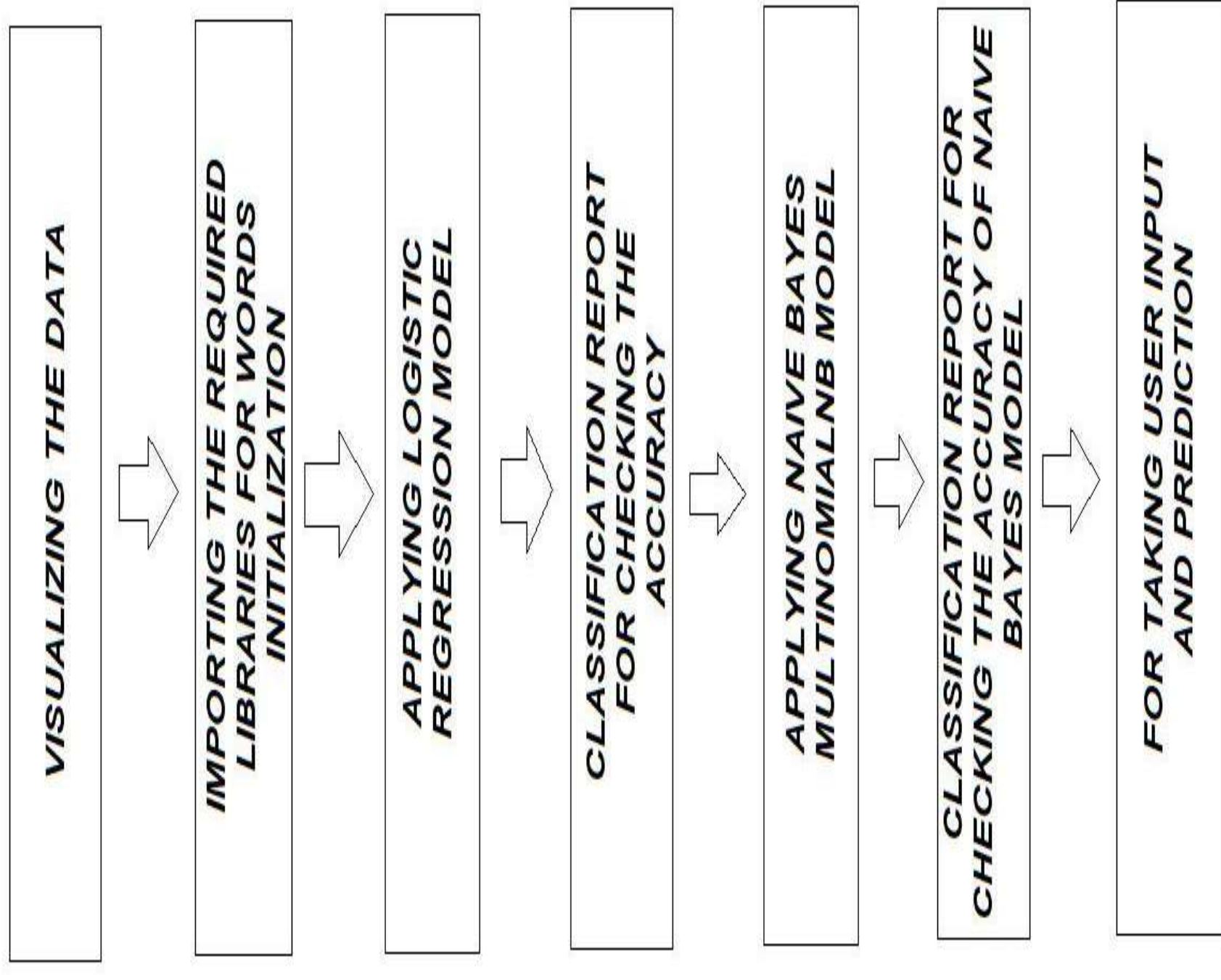
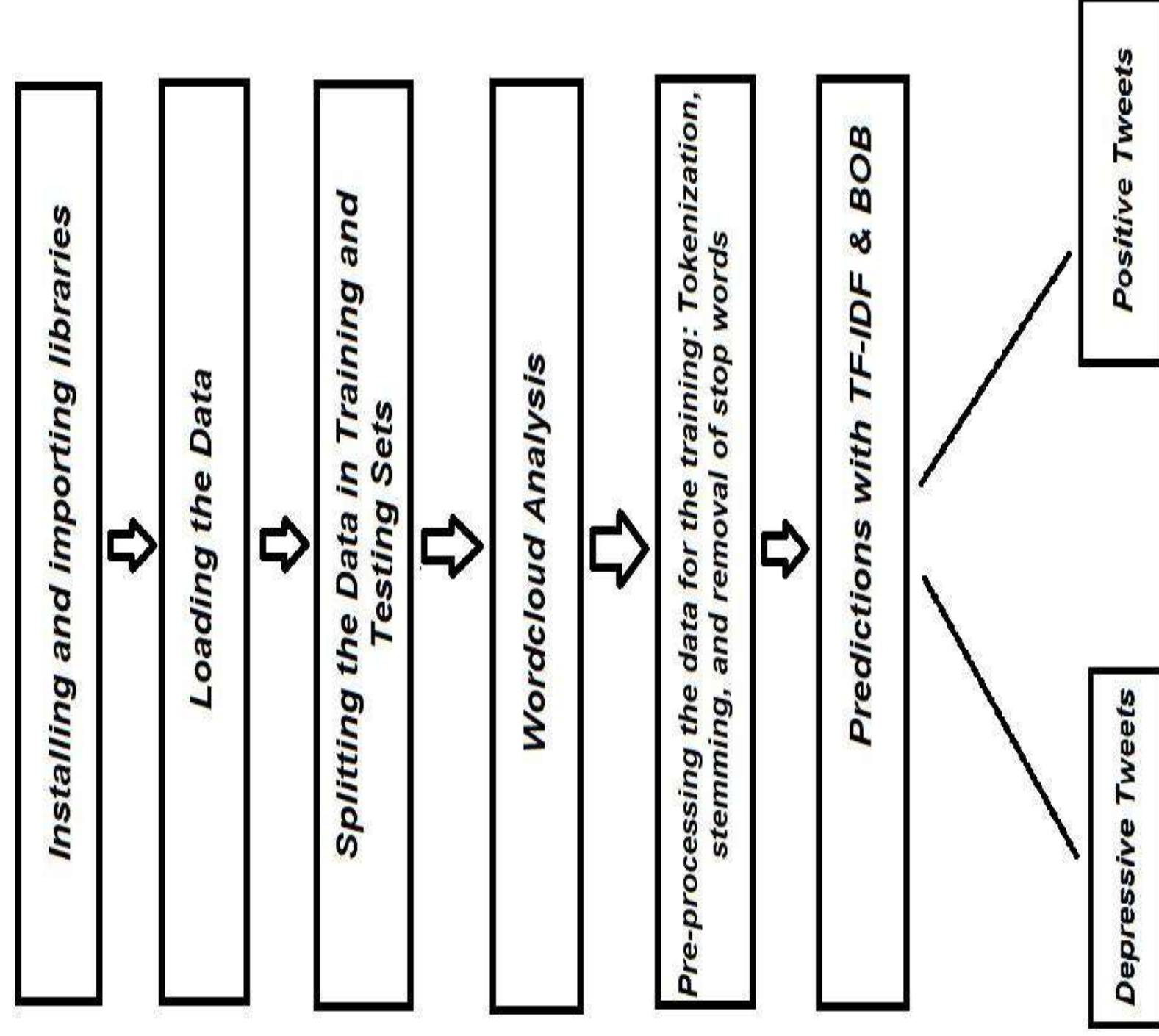
Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes.

In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

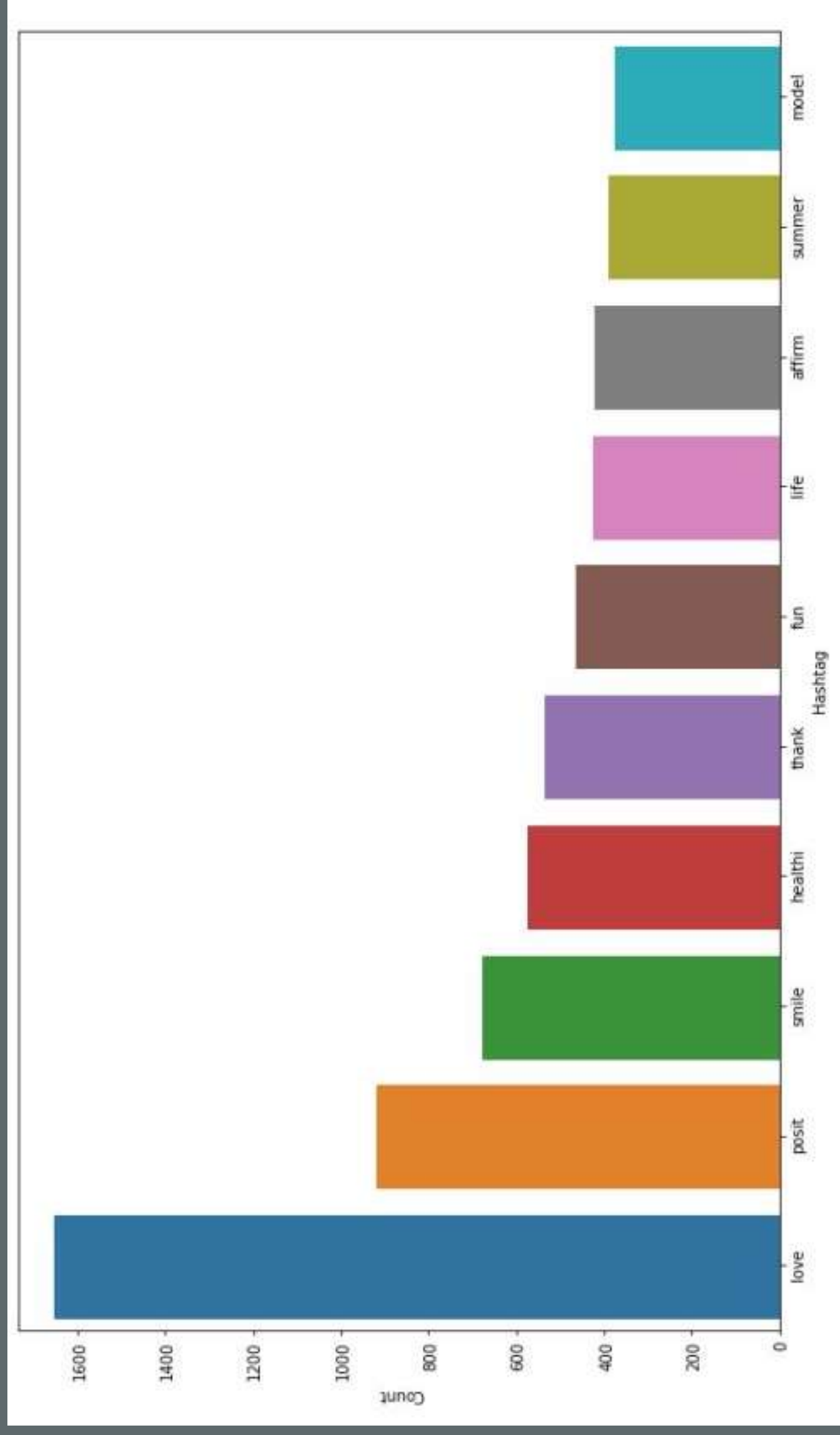
Mathematically, a logistic regression model predicts  $P(Y=1)$  as a function of  $X$ . It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.



# SYSTEM ARCHITECTURE



# Visualizing the positive sentiment data:













# APPLICATIONS OF SENTIMENT ANALYSIS:

- ❑ **Purchasing Product or Service:** While purchasing a product or service, taking right decision is no longer a difficult task. By this technique, people can easily evaluate other's opinion and experience about any product or service and also he can easily compare the competing brands. Now people don't want to rely on external consultant.
- ❑ **Quality Improvement in Product or service:** By Opinion mining and sentiment analysis the manufactures can collect the critic's opinion as well as the favorable opinion about their product or service and thereby they can improve the quality of their product or service. They can make use of online product reviews from websites such as Amazon and C|Net , RottenTomatoes.com and IMDb
- ❑ **Marketing research:** The result of sentiment analysis techniques can be utilized in marketing research. By sentiment analysis techniques, the recent trend of consumers about some product or services can be analyzed.
- ❑ **Keeping an eye on the brand's image:** Sentiment analysis is frequently used to investigate user perceptions of a product or topic. You can also use it to conduct a product analysis and provide all relevant data to the development teams.

- ❑ **To predict the outcome of an election:** To predict the outcome of an election, anyone can use sentiment analysis to compile and analyze large amounts of text data, such as news, social media, opinions, and suggestions. It takes into account how the general public feels about both candidates.
- ❑ **Data from customer feedback is being analyzed:** Data from customer feedback can be used to identify areas for improvement. Sentiment analysis can help you extract value and insights from customer feedback data, as well as develop effective customer satisfaction strategies.
- ❑ **Observing and analyzing conversations on social media:** Conversations on social media are a gold mine of information. Look at conversations about your brand on social media to see what your customers are saying with sentiment analysis; this can help any company plan its future strategies much more effectively.

# LIMITATIONS:

Sentiment analysis tools can identify and analyse many pieces of text automatically and quickly. But computer programs have problems recognizing things like sarcasm and irony, negations, jokes, and exaggerations - the sorts of things a person would have little trouble identifying. And failing to recognize these can skew the results.

The patterns a machine learning system trained on review data has learnt to recognize as evidence for predicting sentiment in one domain will generally not be useful for predicting sentiment in other domains.

Even more problematically, most online review data is in English. For global organizations, successful reputation management requires monitoring media sources in many languages. In order to use sentiment analysis systems trained on English data exclusively, special steps must be taken that either involve costly translation of all relevant news articles and social media posts, or complex, state-of-the-art methods that allow the trained system to transfer what it has learned from one language to another.

# FUTURE SCOPE:

The future of sentiment analysis is going to continue to dig deeper, far past the surface of the number of likes, comments and shares, and aim to reach, and truly understand, the significance of social media interactions and what they tell us about the consumers behind the screens.

As a result of deeper and better understanding of the feelings, emotions and sentiments of a brand or organization's key, high-value audiences, members of these audiences will increasingly receive experiences and messages that are personalized and directly related to their wants and needs.



# CONCLUSION:

Applying sentimental analysis to extract the sentiment became an important work for many organizations and even individuals. Sentiment analysis is an emerging field in decision making process and is developing fast. Our project goal is to analyze the sentiments on a topic which are extracted from the Twitter and determine its nature (positive/negative/neutral) of the defined topics. The development of techniques for the document-level sentiment analysis is one of the significant components of this area. Recently, people have started expressing their opinions on the Web that increased the need of analyzing the opinionated online content for various real-world applications. A lot of research is present in literature for detecting sentiment from the text. Still, there is a huge scope of improvement of these existing sentiment analysis models. Existing sentiment analysis models can be improved further with more semantic and commonsense knowledge.