

Sentiment Analyser using Natural Language Processing(NLP)

*A project report,
submitted in partial fulfillment of the requirements for B. Tech project*

by

**Vinay Pratap Indoriya
(2018IMT-105)**

Under the Supervision of

Dr. Vinal Patel



विश्वजीवनमृतं ज्ञानम्

**ABV INDIAN INSTITUTE OF INFORMATION
TECHNOLOGY AND MANAGEMENT
GWALIOR-474015**

2021

CANDIDATES DECLARATION

I hereby certify that the work, which is being presented in the report, entitled Sentiment Analyser using Natural Language Processing(NLP), in partial fulfillment of the requirement for the award of the Degree of Bachelor of Technology and submitted to the institution is an authentic record of my own work carried out during the period June 2021 to october 2021 under the supervision of **Dr. Vinal Patel**. I also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.



Date:14/08/21

Signatures of the Candidates

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date:14/08/21

Signatures of the Research Supervisors

Acknowledgements

I am highly indebted to **Dr. Vinal Patel**, and obliged for giving me the autonomy of functioning and experimenting with ideas. I would like to take this opportunity to express my profound gratitude to them not only for their academic guidance but also for their personal interest in my project and constant support coupled with confidence boosting and motivating sessions which proved very fruitful and were instrumental in infusing self-assurance and trust within me. The nurturing and blossoming of the present work is mainly due to their valuable guidance, suggestions, astute judgment, constructive criticism and an eye for perfection. My mentor always answered myriad of my doubts with smiling graciousness and prodigious patience, never letting me feel that I am novices by always lending an ear to my views, appreciating and improving them and by giving me a free hand in our project. It's only because of their overwhelming interest and helpful attitude, the present work has attained the stage it has. Finally, I am grateful to our Institution and colleagues whose constant encouragement served to renew my spirit, refocus our attention and energy and helped me in carrying out this work.

Contents

1	Abstract	4
2	Introduction	4
3	Motivation and Literature	5
4	Project Objectives	5
5	System Architecture	5
5.1	Natural Language Processing	6
5.2	Bag of Words Model	7
5.3	Convolution Neural Network	7
6	Progress Made so far	8
7	Test Result	8
8	Task to be Completed	8
9	Gantt Chart	9
10	References	10

List of Figures

1	System Architecture	6
2	CNN Architecture	7
3	Result Screenshot	8
4	Gantt Chart	9

1 Abstract

Natural language processing is used in sentiment analysis. Emotion extraction or opinion mining are other terms for the same thing. Text mining research in this area is quite popular. The primary idea is to determine the text's polarity and categorize it as negative, positive, or neutral. It aids decision-making in humans. Various tasks, such as aspect term extraction, feature extraction, subjectivity detection, sentiment classification, and so on, are required to accomplish sentiment analysis. Data analysis can be approached from a range of perspectives and ways, combining a number of strategies to improve accuracy. Data mining is one of the most widely used data analysis techniques, with a focus on modelling and knowledge acquisition for prediction rather than descriptive purposes. Text analytics uses structural, linguistic, and statistical techniques to gather information from text sources, whereas predictive analytic is primarily used for classification. The most of the text resources are in the form of unstructured data.

2 Introduction

The practice of determining emotions or views from a piece of text from certain topics known as sentiment analysis. It enables us to comprehend the text's attitudes, ideas, and feelings. It gathers information about a users dislikes and likes from web content. It entails evaluating or predicting the text's concealed content. This concealed data is extremely beneficial for gaining insight into a users dislikes and likes. The objective of sentiment analysis is to figure out what a writer or speaker thinks about a certain issues. Audio, photos, and videos can all be used for sentiment analysis.

The internet has now become an integral component of our daily lives. The majority of people utilize social networking or online blogging sites to voice their thoughts on various topics. They also use these websites to find out what people think. As a result, data mining and sentiment extraction have become key research areas. Data mining is one of the most widely used data analysis techniques, focusing on modelling and knowledge acquisition for prediction rather than descriptive purposes. Text analytics uses structural, linguistic, and statistical techniques to gather information from text sources, whereas predictive analytic is primarily used for classification. The most of the text resources are in the form of unstructured data.

To obtain a sentiment, Sentiment Analysis employs three terms. That is, there is an object and a feature, as well as an opinion, an opinion holder, and an orientation. Sentiment Analysis deals with a variety of technological issues, including object

recognition, categorization of opinion orientation, and feature extraction. Typically, unsupervised and supervised learning techniques like as naive Bayes, Neural Networks, and SVM are used to accomplish sentiment analysis. SVM is seen to be the most suitable of these three algorithms for sentiment analysis.

3 Motivation and Literature

Opinion mining or Sentiment analysis plays an important role of our day to day life decision-making process. These selections can vary from buying a product like a phone to reviewing a movie to making investments—all of them will have a significant impact on one's everyday life. People used to seek advice from their friends, neighbours, and other acquaintances before purchasing a product or service. However, in the age of the internet, it is simple to obtain opinions from people all over the world. People nowadays will look at review sites, e-Commerce sites such as e-Bay, Amazon, and social media sites such as facebook, twitter before purchasing any product or service to receive feedback about the specific product.

Sentiment analysis is a field that has seen a lot of investigation. Today, enterprises, corporations, and institutions employ uses sentiment analysis to understand better and measure their customers. People tend to evaluate and express their emotions on their accounts on social media platforms such as Twitter, which is invaluable for opinion mining because the outreach is high. Traditionally, sentiment analysis has focused on dividing emotions into negative and positive categories.

4 Project Objectives

The main objective of the project is to design a Deep Learning Model which is able to classify the text as positive or negative. This is essentially a binary classification problem. The main focus of this project will be to classify the review of different domains such as e-commerce sites, restaurant or twitter which is the most extensively utilised data source in the study of abusive language. We use all publicly available data-sets with tweets categorised as various sorts of abuse and written in English. Convolutional Neural Networks seems to be the best performing model here and we will try to implement that extending it to Hindi-English mix text if possible.

5 System Architecture

In this section we will go through some of the Machine Learning and Deep Learning techniques that are closely related to the project.

5.1 Natural Language Processing

Natural language processing (NLP) is an artificial intelligence field that aids computers in comprehending, interpreting, and manipulating human language. In order to bridge the gap between human communication and machine comprehension, NLP depends on a variety of fields, including computer science and computational linguistics. Natural language processing is divided into two stages: algorithm development and data pre-processing. Data pre-processing is the process of preparing and "cleaning" text data so that machines can examine it. Pre-processing transforms data into usable form and emphasises textual characteristics that an algorithm can use. The following are some applications of Natural Language Processing:

- Speech Recognition
- Sentiment Analysis
- Part of speech tagging

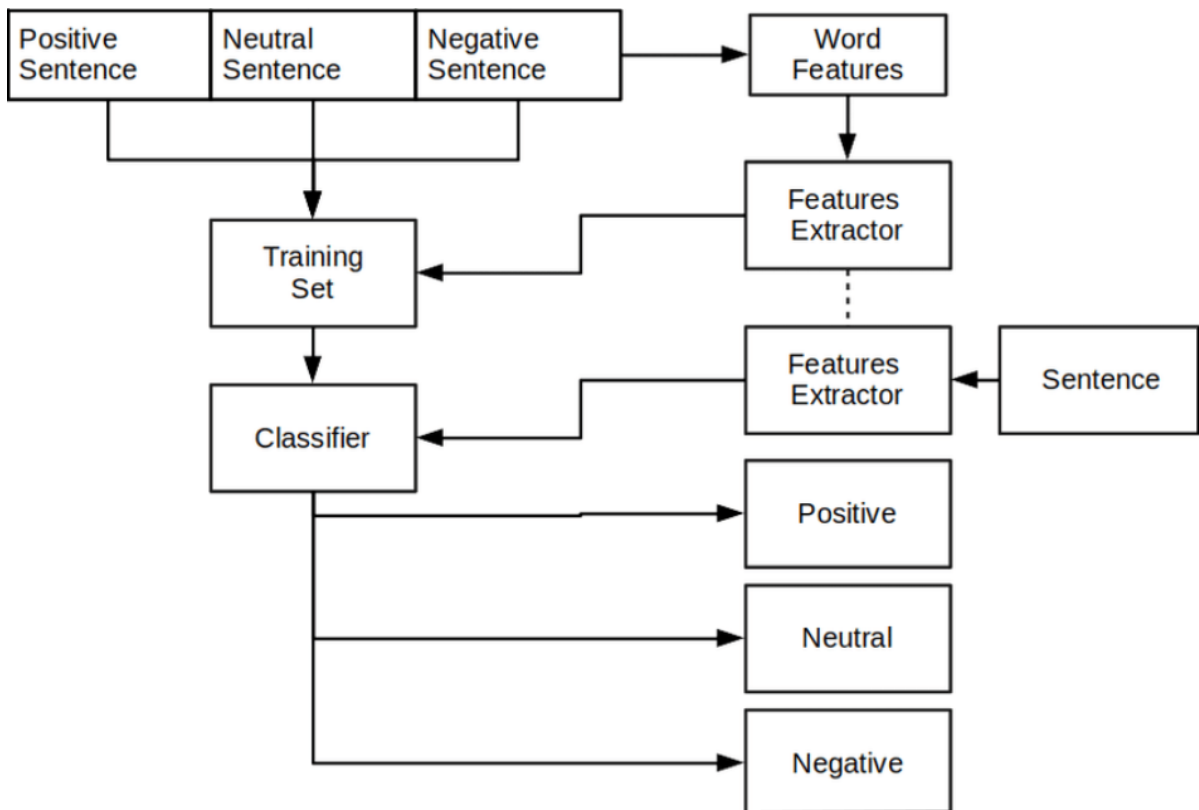


Figure 1: System Architecture

5.2 Bag of Words Model

Text modelling with a bag of words is a Natural Language Processing method. In technical words, we may call it a technique for extracting features from text data. This method of extracting characteristics from documents is easy and versatile.

A bag of words is a text representation that describes the frequency with which words appear in a document. We only keep track of word counts and don't pay attention to grammatical intricacies or word arrangement. Because all information about the sequence or structure of words in the text is lost, it is referred to as a "bag" of words. The model simply cares about whether or not recognised terms appear in the document, not where they appear. It entails two steps:

- A vocabulary of known words.
- A measure of the presence of known words.

5.3 Convolution Neural Network

CNN is a deep neural network that was created with image analysis in mind. It has been recently discovered that CNNs have exceptional capabilities in sequential data analysis, such as natural language processing. Convolution and pooling are two essential procedures that are always included in CNN. The convolution procedure, which employs many filters, is capable of extracting features from the data set while preserving their spatial information. Pooling, also known as sub sampling, is a technique for reducing the dimensionality of feature maps created by the convolution procedure.

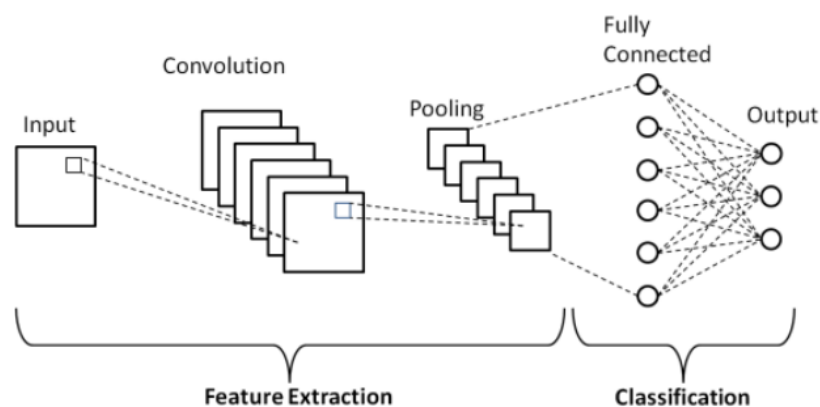


Figure 2: CNN Architecture

6 Progress Made so far

Step by step, we developed various sentiment detection models. This is a binary classification problem, as previously stated. Initially, logistic regression is tried, but it does not perform well. Then we use Natural Language Processing tools to improve our outcomes. Because our task is text sentiment analysis, a similar architecture and approach are used. The data-set consists of different sentences that have been acquired using the kaggle.com and processed as stated in the preceding sections. These annotated data sets distinguish between input features and desired output. Currently, the Bag of Words Model is in use, and the Naive Bayes Model is being trained on a training set.

7 Test Result

So far, I have build a classifier based on Bag of words algorithm and trained the data set on naive bays model whose accuracy is 73% which can be improved. So to improve the accuracy of classifier different models will be tried to improve accuracy.

Making the Confusion Matrix

```
In [8]: from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)

[[55 42]
 [12 91]]

Out[8]: 0.73
```

Figure 3: Result Screenshot

8 Task to be Completed

Further task is to transition of this model towards Convolution Neural Networks (CNN) Model, that seems to be a better choice with higher accuracy based on the research done previously [5](mentioned in Literature Survey). Further we aim to include Hindi-English text and extend the task to detecting sentiment of speech in English or Hindi text.

9 Gantt Chart

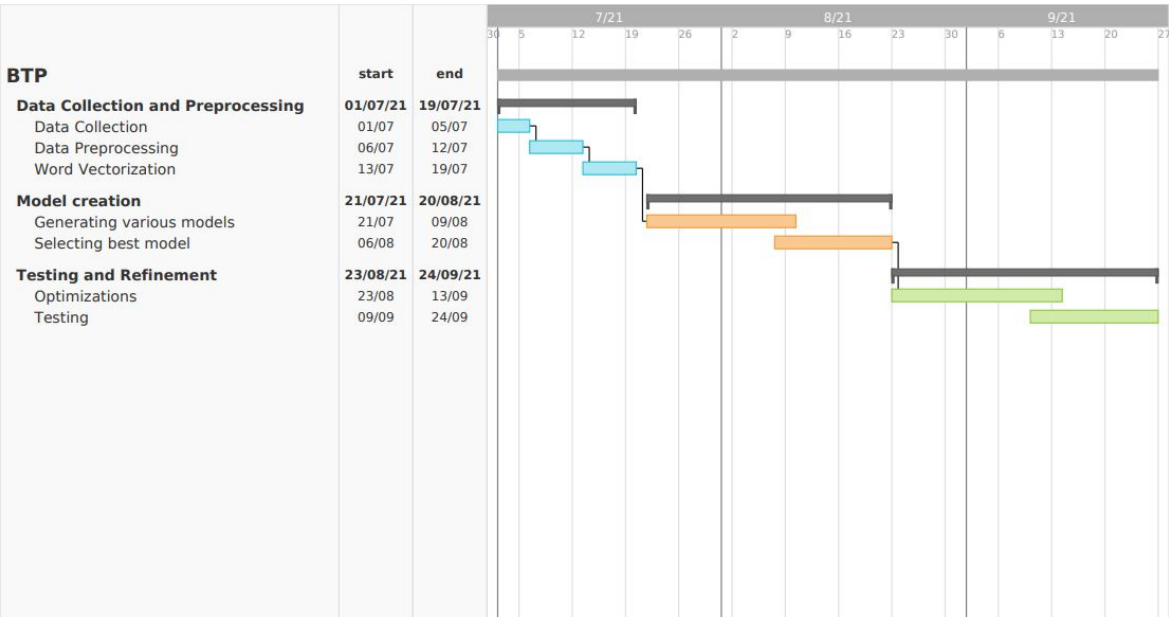


Figure 4: Gantt Chart

10 References

1. H. Kaur, V. Mangat and Nidhi, "A survey of sentiment analysis techniques," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 921-925, doi: 10.1109/I-SMAC.2017.8058315.
2. M. R. Hasan, M. Maliha and M. Arifuzzaman, "Sentiment Analysis with NLP on Twitter Data," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), 2019, pp. 1-4, doi: 10.1109/IC4ME247184.2019.9036670.
3. A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 628-632, doi: 10.1109/ICATCCT.2016.7912076.
4. A. M. Abirami and V. Gayathri, "A survey on sentiment analysis methods and approach," 2016 Eighth International Conference on Advanced Computing (ICoAC), 2017, pp. 72-76, doi: 10.1109/ICoAC.2017.7951748.
5. Y. Chen and Z. Zhang, "Research on text sentiment analysis based on CNNs and SVM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018, pp. 2731-2734, doi: 10.1109/ICIEA.2018.8398173.