

# **SMART ELECTION ANALYSER USING FREE AVAILABLE DATA**

## **PROJECT REPORT**

submitted by

**MAHESH M NAIR (PKD15IT035)**

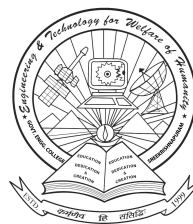
to

The APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the award of the degree

of

**Bachelor of Technology in  
Information Technology**



**DEPARTMENT OF INFORMATION TECHNOLOGY  
GOVERNMENT ENGINEERING COLLEGE SREEKRISHNAPURAM  
PALAKKAD**

**MAY 2019**

## DECLARATION

I **MAHESH M NAIR**, student of eighth semester B.Tech Information Technology, Government Engineering College Sreekrishnapuram, Palakkad, hereby declare that the project report entitled as **Smart Election Analyser Using Free Available Data** submitted to the APJ Abdul Kalam Technological University during the academic year 2018-2019 is a bonafide record of an original work done by me under the supervision of Assistant Professor Sneha C P (ad-hoc). This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. The work has not been submitted to any other University or Institute for the award of any degree.

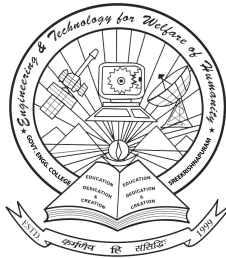
Place: Sreekrishnapuram

MAHESH M NAIR (PKD15IT035)

Date:

# CERTIFICATE

DEPARTMENT OF INFORMATION TECHNOLOGY  
GOVERNMENT ENGINEERING COLLEGE SREEKRISHNAPURAM  
PALAKKAD - 678 633



## IT492 PROJECT

*This is to certify that this project report entitled **Smart Election Analyser Using Free Available Data** submitted by **MAHESH M NAIR (PKD15IT035)** to the Department of Information Technology, Government Engineering College, Sreekrishnapuram, Palakkad - 678633, in partial fulfilment of the requirement for the award of B.Tech Degree in Information Technology is a bonafide record of the work carried out by him.*

Place: Sreekrishnapuram, Palakkad

Date:

Asst. Prof. Sneha CP    Asst. Prof. Sasinas Alias Haritha    Prof. Remesh Babu KR

**Guide**

**Project Co-ordinator**

**Head of the Department**

## Acknowledgement

First and foremost I wish to express my wholehearted indebtedness to God Almighty for the gracious constant care and blessings showered over me for the successful completion of the project.

I am extremely grateful to **Prof. Sneha C P (ad-hoc)**, our Guide, Assistant Professor, Department of Information Technology, Government Engineering College Sreekrishnapuram, for her sincere guidance, inspiration and support throughout the project.

I am thankful to **Prof. Sasinas Alias Haritha Z A**, **Prof. Ranjith K (ad-hoc)**, Project Co-ordinator, Assistant Professor, and **Prof. Remesh Babu K R**, Head of the Department, Department of Information Technology, Government Engineering College Sreekrishnapuram, for their support and inspiration throughout the project.

Gratitude is extended to all teaching and non teaching staffs of Department of Information Technology, Government Engineering College Sreekrishnapuram for the sincere directions imparted and the cooperation in connection with the project.

I am also thankful to my parents for the support given in connection with the project. Gratitude is extended to all well-wishers and my friends who supported me to complete the project in time.

# Table of Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Abbreviations</b>	<b>vii</b>
<b>Abstract</b>	<b>1</b>
<b>1 INTRODUCTION</b>	<b>2</b>
<b>2 LITERATURE REVIEW</b>	<b>4</b>
2.1 Deep Convolution Neural Networks for Twitter Sentiment Analysis. .	4
2.2 NLP Based Sentiment Analysis on Twitter Data Using Ensemble Clas- sifiers . . . . .	4
2.3 Sentiment Analysis in TripAdvisor . . . . .	5
2.4 Fine-grained Sentiment Analysis with 32 Dimensions . . . . .	6
2.5 Sentiment Analysis in Twitter using Machine Learning Techniques . .	6
2.6 Sentiment Analysis for Social Media: A Survey . . . . .	7
<b>3 METHODOLOGY</b>	<b>8</b>
3.1 Sentiment Analysis on Social media . . . . .	8
3.2 PredictA . . . . .	9
3.3 Naive Bayes Classifier . . . . .	11
3.4 Natural Language Toolkit (NLTK) . . . . .	13
3.5 Sentiment Analysis . . . . .	13
3.5.1 Data Collection . . . . .	14

3.5.2	Text Preparation . . . . .	14
3.5.3	Sentiment Detection . . . . .	14
3.5.4	Sentiment Classification . . . . .	15
3.5.5	Presentation of Output . . . . .	16
3.6	Challenges . . . . .	16
3.7	Advantages . . . . .	17
3.8	Gantt Chart . . . . .	18
<b>4</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>19</b>
<b>5</b>	<b>FUTURE SCOPE</b>	<b>21</b>
5.1	Future Scope Of Project . . . . .	21
<b>6</b>	<b>CONCLUSION</b>	<b>22</b>
	<b>Bibliography</b>	<b>23</b>

## List of Figures

3.1	Working overview of PredictA. . . . .	10
3.2	Methodology for sentiment analysing of data . . . . .	14
3.3	Pie chart representation of the result of analysis . . . . .	15
3.4	Histogram representing the continuous periodic analysis . . . . .	16
3.5	Design procedure schedule of PredictA. . . . .	18

# List of Abbreviations

ABSA	Aspect Based Sentiment Analysis
API	Application Programming Interface
NLP	Natural Language Processing
NLTK	Natuaral Language Tool Kit
SAMs	Sentiment Analysis Methods
SVM	Support Vector Machine
URL	Uniform Resource Locator



# Abstract

Today's generation of people is more open and expressive, a lot in social media and other platforms. Thus the amount of big data generated even in an hour is very huge. The freely available data from these social platforms (comments and tweets of users) are collected for useful analytical purposes. This data can be used for sentimental analysis. It can be determined if a tweet or review is positive, negative or neutral, using sentimental analysis method. This sentimental analysis technology can be used to analyse and predict the election results in political field. The data from specific users in social media is collected and their opinions and views on a particular government or political party or person, is analysed and provides a numerical percentage favoring or disfavoring the government. This can be used in election result predictions. Election campaigns can also be designed with this information from the data that is provided. Natural Language Processing techniques are used here, that uses real-time data from the social media platforms and more accuracy can be obtained for the predicted results or outputs.

*Keywords: Big Data, Sentimental Analysis, Machine Learning, Natural Language Processing*

# CHAPTER 1

## INTRODUCTION

Nowadays, the use of social networking sites has increased tremendously. Millions of people conveniently share their data and opinions on social media platforms. Collecting, processing and analysing of this freely available Big Data [7] may help to gather some general or individual information of users. The age of getting meaningful insights from social media data has now arrived with the advance in technology.

At early days common people expressed their political opinions at public places like tea shop, clubs, wedding party or at other public gatherings. But, the present generation mainly expresses their personal opinions in social media such as Facebook, Twitter, Instagram, etc. Comparing with the olden days, analysing the general political review at present, is easy due to the influence of social media among ordinary peoples. But the problem of this is, it consumes more time to read and analyse the entire documents about some political party or some politicians available in such social medias. Some of the reason for time consumption are,

- Reading each and every tweet is a tedious process.
- Analysing and calculation of these data.
- Visualization of these data.

Time is the most valuable thing in one's life. Effective utilization of time is very important for success of a person, especially if he is trying to achieve something in this competitive world. Processing and analysing of these freely available opinions within a short time can make highly effective result in competition like elections.

Here a proposal for a smart election analyser using this freely available data from social medias is made. Using sentiment analysing one can analyse the opinions in each text data within short interval of time. Sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

The interest in sentiment analysis has increased significantly over the last few years due to the large amount of stored text in Web 2.0 applications and the importance of online customer opinions. As a result, many research papers contain the term sentiment analysis and various start-ups have been set up to analyze sentiments in social media companies. The proposed system is a designed to provide a friendly and simple user interface. The system is efficient as it consumes lesser memory than other tools hence leading us to give more time and memory for other parallel processing.

# CHAPTER 2

## LITERATURE REVIEW

### **2.1 Deep Convolution Neural Networks for Twitter Sentiment Analysis.**

Twitter sentiment analysis technology provides the methods to survey public emotion about the events or products related to them [1]. Most of the current researches are focusing on obtaining sentiment features by analysing lexical and syntactic features. These features are expressed explicitly through sentiment words, emoticons, exclamation marks, and so on. In this paper, the author introduce a word embeddings method obtained by unsupervised learning based on large twitter corpora, this method using latent contextual semantic relationships and co-occurrence statistical characteristics between words in tweets.

These word embeddings are combined with n-grams features and word sentiment polarity score features to form a sentiment feature set of tweets. The feature set is integrated into a deep convolution neural network for training and predicting sentiment classification labels. Experimentally compare the performance of the proposed model with the baseline model that is a word n-grams model on five Twitter data sets, the results indicate that the model performs better on the accuracy and F1-measure for twitter sentiment classification.

### **2.2 NLP Based Sentiment Analysis on Twitter Data Using Ensemble Classifiers**

Most sentiment analysis systems use bag-of-words approach for mining sentiments from the online reviews and social media data [2]. Rather considering the whole sen-

tence or paragraph for analysis, the bag-of-words approach considers only individual words and their count as the feature vectors. This may mislead the classification algorithm especially when used for problems like sentiment classification.

Traditional machine learning algorithms like Naive Bayes, Maximum Entropy, SVM etc. are widely used to solve the classification problems. These machine learning algorithms often suffer from biasness towards a particular class. Monisha Kanakaraj and Ram Mohana Reddy Guddeti propose Natural Language (NLP) based approach to enhance the sentiment classification by adding semantics in feature vectors and thereby using ensemble methods for classification. Adding semantically similar words and context-sense identities to the feature vectors will increase the accuracy of prediction. Experiments conducted demonstrate that the semantics based feature vector with ensemble classifier outperforms the traditional bag-of-words approach with single machine learning classifier by 3-5%.

### **2.3 Sentiment Analysis in TripAdvisor**

Sentiment analysis is a natural language processing tool that is useful for monitoring Web 2.0 applications, as it can reveal public opinion about numerous issues without requiring satisfaction enquiries [3]. According to the Oxford dictionary, sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text to determine whether the writers attitude toward a particular topic, product, and so on is generally positive, negative, or neutral [1]. TripAdvisor emerged in 2004 as a Web 2.0 application for the tourism domain. This user-generated content website offers a plethora of reviews detailing travelers experiences with hotels, restaurants, and tourist spots. TripAdvisor has since been ranked as the most popular site for trip planning, with millions of tourists visiting the site when arranging their holidays. Clear and exact opinion of users on the various services provided could be easily identifiable and classifiable using the sentiment analysis methodology.

Multiple studies on TripAdvisor exist, but there is no complete analysis from the sentiment analysis viewpoint. The article proposes TripAdvisor as a source of data for sentiment analysis tasks. The author develop an analysis for studying the matching between user's sentiments and automatic sentiment-detection algorithms. Finally, discuss some of the challenges regarding sentiment analysis and TripAdvisor, and conclude with some final remarks.

#### **2.4 Fine-grained Sentiment Analysis with 32 Dimensions**

Xianchao Wu, Hang Tong, Momo Klyen proposed a fine-grained sentiment analysis system which classify emotions into 32 categories because of understanding humans complicated emotions remains a fundamental challenge [4]. For one direction, the paper cover more detailed emotions and for the other direction and further measure each emotion with strength, such as describing angry by annoyance, anger and range. Taking Japanese as a test language, description methods of building the training data, of constructing deep neural network classifiers, and of evaluating the models are made.

#### **2.5 Sentiment Analysis in Twitter using Machine Learning Techniques**

Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in source text [5]. Social media is generating a vast amount of sentiment rich data in the form of tweets, status updates, blog posts etc. Sentiment analysis of this user generated data is very useful in knowing the opinion of the crowd. Twitter sentiment analysis is difficult compared to general sentiment analysis due to the presence of slang words and misspellings.

The maximum limit of characters that are allowed in Twitter is 280. Knowledge base approach and Machine learning approach are the two strategies used for analysing sentiments from the text. An effort to try to analyse the twitter posts about elec-

tronic products like mobiles, laptops etc using Machine Learning approach is made. By doing sentiment analysis in a specific domain, it is possible to identify the effect of domain information in sentiment classification. The model present a new feature vector for classifying the tweets as positive, negative and extract people's opinion about products.

## **2.6 Sentiment Analysis for Social Media: A Survey**

Harshali P. Patil did a detailed survey on different techniques used in Sentiment Analysis, carried out to understand the level of work this fascinating problem is increasingly important in business and society [6]. It offers numerous research challenges but promises insight useful to anyone interested in opinion analysis and social media analysis. A comprehensive introduction to the topic from a primarily natural language processing point of view to help readers understand the underlying structure of the problem and the language constructs that are commonly used to express opinions and sentiments is given. It covers all core areas of sentiment analysis, includes many emerging themes, such as debate analysis, intention mining, and fake opinion detection, and presents computational methods to analyse and summarize opinions [9]. It is a valuable resource for researchers and practitioners in natural language processing, computer science, management sciences, and the social sciences.

# CHAPTER 3

## METHODOLOGY

### 3.1 Sentiment Analysis on Social media

Sentiment analysis can be defined as a process of analyzing user reviews, opinions, emotions, sentiments, attitude regarding various entities such as products, services, organization, key issues. The Web 2.0 is place where anybody can express their opinions or post reviews about various entities [3]. Social Media is actually influencing the users point of view or individual opinions. It has now become an essential part of digital marketing. Sentiment analysis is a process of identifying and aggregating opinions of various about any issues or products. Sentiment analysis is similar to opinion mining, with the help of which one can analyse the users view by detecting the polarity of the text. Polarity means, it can be used to analyse the text on the basis of their views posted online i.e. to determine whether the piece of writing is positive, negative or neutral [1].

Sentiment analysis can be used to give valuable insights into how people feel about your product brand or service. When applied to social media channels, sentiment analysis can be used to identify spikes in sentiment, thereby allowing you to identify potential product advocates or social media influencers. To buy a product the user checks the various social networking sites for reviews [9]. These reviews and ratings on the Internet matters a lot to evaluate the product. Monitoring social media activities is necessary to evaluate the customers loyalty and also to keep track on their sentiments regarding a brand or a product, impact of various marketing activities and many other aspects [4].



### 3.2 PredictA

PredictA is a software tool which helps to analyse and predict the election result based on freely available social media political opinions. This tool works on the base of sentimental analysis of textual documents. Here, NLP (Natural Language Processing) for sentimental analysis is used. Natural language processing (NLP) is a sub field of computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyse large amounts of natural language data.

In this system, analysis of peoples opinions about a political person or party which is nominated for election in a particular area is done. A graphical interface in the form of a pie chart and histogram depicting the periodic changes in the influence of the customer in the social media, is also developed. This is done for the customer so they can easily analyse peoples opinion about them and can make strategical changes in their election campaigning. Faster and efficient process can always make more fruitful results.

In PredictA, analysing the opinion of person who is nominated to election process is done. Tweets about that person written by people about them is extracted. Then analyses the data and info to client, about how much people like or dislike them, with in short period of time. Also the project provides a graphical analysis interface to the client for simple understanding. The client can improve opinion about them through election campaign according to our analysis and will continue this evaluation periodically until the last moment of election process. This help the client to improve Election winning possibility. Comparing with direct social media analysis this helps client to act immediately and they can consume their time and cost for the huge data analysis, because our system can analyse and predict with in short time in more accurate. There might be cases when users are not responding with any of their opinion

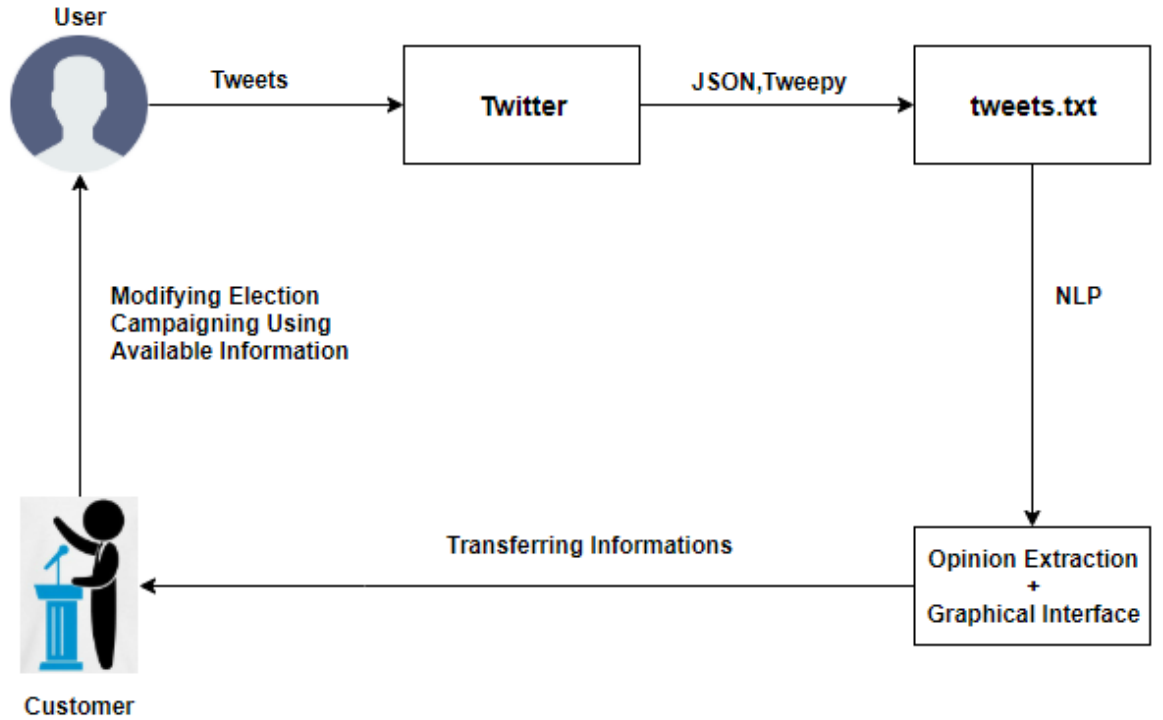


Figure 3.1: Working overview of PredictA.

or views, so in this system we consider the polarity of those unexpressed opinions as neutral polarity and this mostly deviates or favours to the majority polarity.

Here, extraction of tweets is done using the Twitter API. Twitter API is simply a set of URLs that take parameters. These URLs let you access many features of Twitter, such as posting a tweet or finding tweets that contain a word, etc. Tweepy is used for Twitter API [8]. Tweepy is a Python library for accessing the Twitter API. It is great for simple automation and creating twitter bots. Tweepy has many features. The extracted tweets are then applied to the Natural Language Processing Toolkits and classified with the Naive Bayes Classifier which functions on the basics of the Bayes Theorem [11]. The classification is a supervised one and properly done to obtain the positivity and negativity in the tweets expressed by a social media user.

The data storage technique of this project enforces a text (.txt) file initially for storing the extracted data from twitter collected via the Twitter API. File properties of the python language is used to do the necessary function executions of read, write, etc on this text file. Then this text file is preprocessed for noise removal and filtering of data. Sentiment is extracted from this file and it is then stored to a comma separated variable (.csv) file. This is used because when compared to the text file, comma separated variable file has a better structured arrangement. Inserting and retrieving of the data is efficient in comma separated variable file.

Using matplotlib libraries, a graphical representation of the classification is prepared. Matplotlib is nothing but a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits. Pyplot is a Matplotlib module which provides a MATLAB-like interface. Matplotlib is designed to be as usable as MATLAB, with the ability to use Python, and the advantage of being free and open-source. The project is also added with a text to speech module to say out the positive, negative tweets and positivity of the analysis. The module is imported from the package to be added as a light feature for the project.

### **3.3 Naive Bayes Classifier**

Naive Bayes is a simple but surprisingly powerful supervised learning algorithm for predictive modeling. Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems [11]. Naive Bayes classifiers are a collection of classification algorithms based on Bayes Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, that is, every pair of features being classified is independent of each other. In the Bayesian analysis, the final classification is produced by combining both sources of information, i.e., the prior and the likelihood, to form a posterior probability using the so-called Bayes' rule.

Bayes Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes theorem is stated mathematically as the following equation:

$$P(h | d) = \frac{P(d | h) P(h)}{P(d)}$$

where,  $P(h|d)$  is the probability of hypothesis  $h$  given the data  $d$ , this is called the posterior probability,  $P(d|h)$  is the probability of data  $d$  given that the hypothesis  $h$  was true,  $P(h)$  is the probability of hypothesis  $h$  being true (regardless of the data), this is called the prior probability of  $h$  and finally  $P(d)$  is the probability of the data (regardless of the hypothesis).

The representation for naive Bayes is probabilities. A list of probabilities are stored to file for a learned naive Bayes model. This includes the probabilities of each class in the training dataset, called as class probability and the conditional probabilities of each input value given each class value. Learning a naive Bayes model from your training data is fast. Training is fast because only the probability of each class and the probability of each class given different input values need to be calculated. No coefficients need to be fitted by optimization procedures. The class probabilities are simply the frequency of instances that belong to each class divided by the total number of instances. The conditional probabilities are the frequency of each attribute value for a given class value divided by the frequency of instances with that class value. Naive Bayes was introduced into the text retrieval community, and remains a popular baseline method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines. There is very little explicit training in Naive Bayes compared to other common classification methods. The only work that must be done before prediction

is finding the parameters for the features individual probability distributions, which can typically be done quickly and deterministically. This means that Naive Bayes classifiers can perform well even with high-dimensional data points and a large number of data points.

### **3.4 Natural Language Toolkit (NLTK)**

The Natural Language Toolkit, or more commonly NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces, along with a suite of text processing libraries for classification, tokenization, stemming, lemmatization, punctuation, character count, word count, tagging, parsing, and semantic reasoning. Here with the use of NLTK libraries for sentiment extraction of these tweets collected with the help of Twitter API, as the stage progresses. The toolkit is one of the most powerful NLP libraries which contains packages to make machines understand human language and reply to it with an appropriate response.

### **3.5 Sentiment Analysis**

Sentiment analysis is similar to opinion mining, with the help of which one can analyse the users view by detecting the polarity of the text. There are mainly 5 steps to analyse sentiment of data, and it is based on [10]

- Data Collection
- Text Preparation
- Sentiment Detection
- Sentiment Classification
- Presentation of Output



Figure 3.2: Methodology for sentiment analysing of data

### 3.5.1 Data Collection

Consumers usually express their sentiments on public forums like the blogs, discussion boards, product reviews as well as on their private blogs such as 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 python and R are used to process and analyse the data.

### 3.5.2 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. The data extracted from twitter using the Twitter API comes with a lot of noisy data and thus a proper cleansing of the data is done and it is made in the necessary format for the beginning of analyzing.

### 3.5.3 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.

### 3.5.4 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. This is done with the usage of Naive Bayes and a proper training data. The classifier will be prepared with respect to the information and training from the training dataset. Supervised learning for the classifier development is thus a best approach for obtaining better precision in the results that follow.

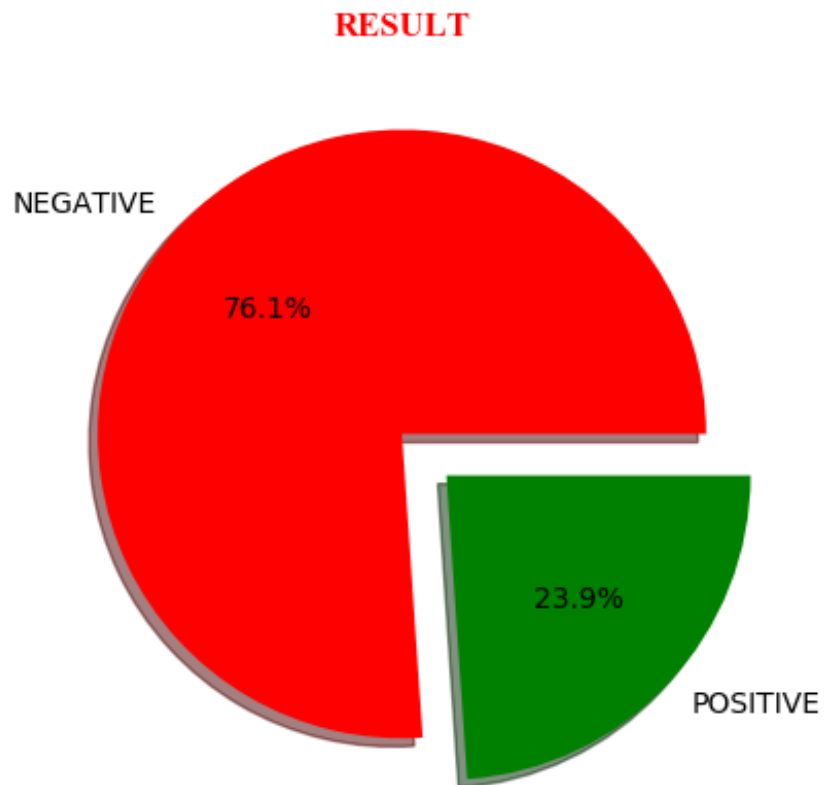


Figure 3.3: Pie chart representation of the result of analysis

### 3.5.5 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. Also a histogram depicting the periodic analysis is given such that a better presentation of the results can be obtained. The output is also given as a audio ouput using the python text to speech module. This is just added to enhance the feature in the overall project.

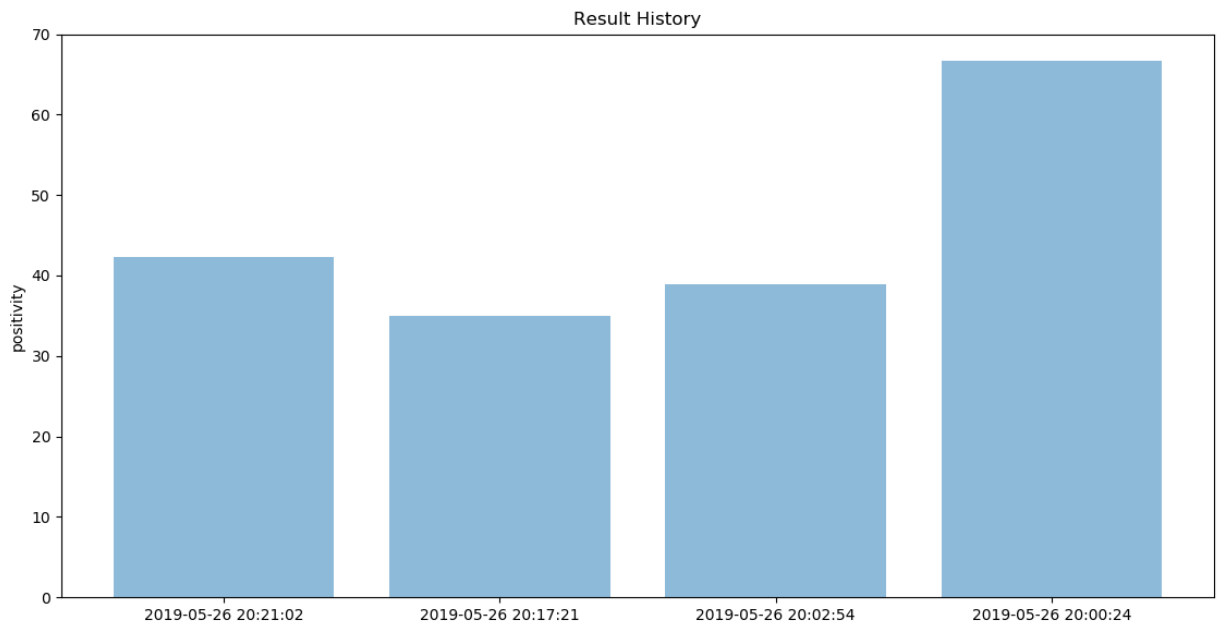


Figure 3.4: Histogram representing the continuous periodic analysis

### 3.6 Challenges

- The detection of irony and sarcasm is a complex sentiment analysis task.
- Communication range.
- Spam contents.



- Intentionally manipulated or fake contents.
- Processing of some native languages.
- Processing of video and audio data.

### **3.7 Advantages**

- Analyse opinions beyond user ratings.
- Can handle the negative opinions identified with SAMs.
- Discover relationships among common aspects to characterize the cause behind the negative comments.

### 3.8 Gantt Chart

The below gantt chart is the design procedure schedule of PredictA software tool.

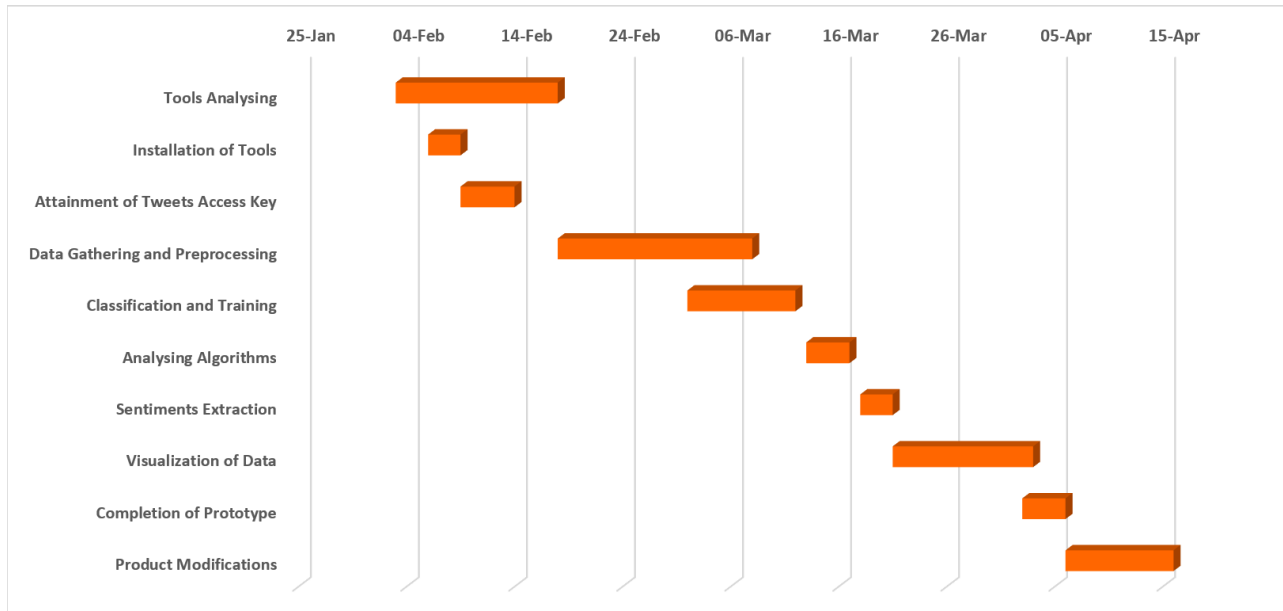


Figure 3.5: Design procedure schedule of PredictA.

## CHAPTER 4

### RESULTS AND DISCUSSIONS

Managed to successfully apply the project objectives. The application turned out to be user friendly and effective as expected. The project is user friendly in the sense that the options like analysing, preprocessing and data extraction are given in a good graphical user interface. Effective results can be obtained with just a click on the options that are available in the main screen. The project itself is a light weight tool because it consumes lesser amount of RAM usage. Currently there seems to be no common light weight election prediction tool available in the internet.

Here the data for the project is collected from the freely available sources over the vast internet. Extraction is mainly carried out from Twitter, so there occurs no data privacy issue here. There is even an option to change the training data at any time, this is done because the sentiment of some tweets collected may change with respect to place, time, person and so on and an effective adaptability provides better results. In this way, consistency of the project could be maintained. The application will generate result as per the training data give by user. There is a facility that allows to modify the data models for user and the accuracy of result completely depending on the data in data model is given by the analyser.

The output from the analysis of extracted data is given as both a visual result and an audio result. User can generate the result history at any time according to the needs. The time complexity of application is quite more than from what we expected. Because python is a interpreter language, this leads to the compilation time being greater than any other high level languages. Another reason behind this high time

complexity is the bigger size of the data set being used. There are two different data-sets employed here, training set and testing set, so the processing of these two data-sets will take more time and this results in a higher time complexity scenario.

# CHAPTER 5

## FUTURE SCOPE

### 5.1 Future Scope Of Project

- Range of influence of each tweet.
- Processing of video and audio data.
- Processing of some native languages.
- Extraction of fine grained information.
- Multiple parallel analysing is a complex task.
- The detection of irony and sarcasm contained tweets.
- Data extraction from more social platforms or resources.
- Removal of spam contents, intentionally manipulated or fake contents, etc.
- Increasing efficiency of classification strategy using neural network technology.
- Effective functioning of project with minimum number of file execution, thus reducing time complexity.

## CHAPTER 6

### CONCLUSION

PredictA can analyze the sentiment of the people in an area and thus all the strategies and estimation can be based on how the people of that area respond to various political mandates. The PredictA tool can be very useful to identify the grey areas for a political leader and thus help the respective party to ennoble the status of the leaders elections within a short interval of time. It helps to detect consistency and inconsistency between statements and actions at the government or individual level. It can also be a useful tool in predicting the results of an election with high accuracy and thus provide help to take the required action accordingly. All in all, with such a tool, to judge the preferences of the people in terms of politics could hold the key to a leader's win in an election and also could help improve the strategies that are to be taken up by the party so that they are in the best interest of the people being based on the people's own preference.

# Bibliography

- [1] Zhao Jiaqing, Gui Xiaolin and Zhang Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis" *DOI: 10.1109/ACCESS.2017.2776930, IEEE Access VOLUME 6, 2018 IEEE.*
- [2] Monisha Kanakaraj and Ram Mohana Reddy Guddeti, "NLP Based Sentiment Analysis on Twitter Data Using Ensemble Classifiers", *2015 Third International Conference on Signal Processing, Communication and Networking (IC-SCN), DOI: 10.1109/ICSCN.2015.7219856, 2015 IEEE.*
- [3] Ana Valdivia, M. Victoria Luzn and Francisco Herrera "Sentiment Analysis in TripAdvisor", *IEEE Intelligent Systems, Volume 32, Published by the IEEE Computer Society, DOI: 10.1109/MIS.2017.3121555, 2017 IEEE.*
- [4] Xianchao Wu, Hang Tong and Momo Klyen, "Fine-grained Sentiment Analysis with 32 Dimensions", *2017 International Conference on Asian Language Processing (IALP), 978-1-5386-1981-0/17/31.00, DOI: 10.1109/IALP.2017.8300624, 2017 IEEE.*
- [5] Neethu M S, R Rajasree , "Sentiment Analysis in Twitter using Machine Learning Techniques", *Fourth ICCCNT, DOI: 10.1109/ICCCNT.2013.6726818, 2013 IEEE - 31661.*
- [6] Harshali P. Patil and Dr. Mohammad Atique, "Sentiment Analysis for Social Media: A Survey", *2015 2nd International Conference on Information Science and Security (ICISS), 978-1-4673-8611-1/15/31.00, DOI: 10.1109/ICISSEC.2015.7371033, 2015 IEEE.*
- [7] Sas.com. (2018). "What is Big Data and why it matters." [online] Available at: [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html) [Ac-

cessed 18 Nov. 2018].

- [8] GeeksforGeeks. (2018). "Extraction of Tweets using Tweepy - GeeksforGeeks." [online] Available at: <https://www.geeksforgeeks.org/extraction-of-tweets-using-tweepy/> [Accessed 20 Nov. 2018].
- [9] Gage, J. (2018). "Introduction to Sentiment Analysis Algorithms." [online] Algorithmia Blog. Available at: <https://blog.algorithmia.com/introduction-sentiment-analysis/> [Accessed 25 Nov. 2018].
- [10] Edureka Blog(2018). Sentiment Analysis Methodology - Edureka Blog. [online] Available at: <https://www.edureka.co/blog/sentiment-analysis-methodology/> [Accessed 23 Jan. 2019].
- [11] Jason Brownlee (2016). Understand Machine Learning Algorithms [online] Available at: <https://machinelearningmastery.com/naive-bayes-for-machine-learning/> [Accessed 11 Feb. 2019].