A PROJECT REPORT ON

ELECTION RESULT PREDICTION USING TWITTER SENTIMENT ANALYSIS

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

VARUN KANADE	Exam No. **B150054330**
CHINMAY MOTORWAR	Exam No. **B150054396**
AJAY RAO	Exam No. **B150054436**

Under the guidance of
Prof. S.N. GIRME

DEPARTMENT OF COMPUTER ENGINEERING
PUNE INSTITUTE OF COMPUTER TECHNOLOGY
DHANKAWADI, PUNE – 43

SAVITRIBAI PHULE PUNE UNIVERSITY
2021-2022

CERTIFICATE

This is to certify that the project report entitles

ELECTION RESULT PREDICTION USING TWITTER SENTIMENT ANALYSIS

SUBMITTED BY

VARUN KANADE	Exam No. **B150054330**
CHINMAY MOTORWAR	Exam No. **B150054396**
AJAY RAO	Exam No. **B150054436**

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof. S.N. Girme** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

Prof. S.N. Girme
Internal Guide
Dept. of Computer Engineering

Dr. G.V. Kale
Head
Dept. of Computer Engineering

Dr. R. Sreemathy
Principal,
Pune Institute of Computer Technology

Place: Pune
Date:

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the project report on ‘Election Result Prediction Using Twitter Sentiment Analysis’.

We would like to take this opportunity to thank **Prof. S.N.Girme** for giving us all the help and guidance we needed.We are really grateful to her for her kind support. Her valuable suggestions were very helpful.

We are also grateful to **Dr. G.V. Kale**, Head of Computer Engineering Department, PICT for her indispensable support, suggestions.

Varun Kanade
Chinmay Motorwar
Ajay Rao

ABSTRACT

The rapid increase of social media in the recent past has provided end users a powerful platform to voice their opinions. Our approach is to gather a collection of Tweets of top political parties within the election, then compute the sentiment score. Collection of tweets is done using Tweepy API.Dataset contains mixture of both popular as well as recent tweets related to specific political party. Specific keywords are used to extract tweets for a party like 'BJP elections 2022', '#UPelections BJP', '#Punjabelections BJP', etc.

We utilized a combination of VADER Sentiment Analyzer and classic machine learning algorithms like Random Forest Classifier, SVM, etc. to build our classifier and classify the test data as positive, negative and neutral tweets. Therefore, this work analyzes tweets collected from twitter and predicts election results by performing sentimental analysis on them.

Keywords: Sentimental Analysis, Tweets, Twitter, Tweepy, Supervised learning, Natural Language Processing, Machine Learning

Contents

List of Abbreviations

Abbreviation	Full form
ML	Machine Learning
SVM	Support Vector Machine
DT	Decision Tree
LR	Logistic Regression
VADER	Valence Aware Dictionary for Sentiment Reasoning
SDLC	Software Development Life Cycle
API	Application Programming Interface
NLP	Natural Language Processing

Table 1: Abbrevations

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Introduction

Elections play an important role in a democratic country. Indian parliamentary system gives its people the right to decide who will govern them for the next five years. During the tenure of Feb 22 to March 22, five state elections are lined up, with the important one being at Uttar Pradesh, which sends the largest number of MPs to parliament. The major national political parties contesting in the elections are Bhartiya Janata Party(BJP), Indian National Congress (INC), Aam Aadmi Party(AAP) and Nationalist Congress Party(NCP).

Social Media has become a powerful tool to share one's views. It has provided strong mediums such as Facebook, Twitter and Google+ to share opinions, reviews, and ratings. All major political parties and their members all over the world have their official accounts on Twitter with millions of followers. They consider this platform as a medium to connect with young people who might vote them. With significant rise of Indian users on Twitter during the pandemic, people have been more vocal to criticize or appreciate a political decision.

Sentimental Analysis is a method to teach a machine to extract emotion from a given text. A text can be anything, a simple review, a social statement, tweets or messages. Twitter Sentiment Analysis of tweets regarding elections can be used by the general public as well as the political parties to understand the positive or negative views of people regarding a particular political party, thus, helping to predict the election results during that period.

1.2 Motivation

There was an 87% increase in the usage of social media in India during the lockdown. As of July 2021, India has 22.1 million active Twitter users and 18% of them look at this social media platform as a source of news. 2020-21 saw an increase in people taking a stand against problems everyone had traditionally stayed away from before. These topics include politics, social injustice, mental health, and so on. Majority of political parties in India see this platform as a way to understand the view of the general public towards issues currently happening in India. Thus, it becomes important to analyze the sentiment of people during election periods, so that authorities can get a better view of their decisions and accordingly improve their current status.

1.3 Problem Definition

Building a classification model to predict the popularity of Indian political parties on social media and infer their chances of winning the upcoming State Elections-2022, by utilizing the sentiment analysis of Twitter data.

CHAPTER 2

LITERATURE SURVEY

The most important factor while doing research is literature survey. Before we get started with any kind of project, we have to study few previous published papers and study the architecture of the model based on the papers and predict the drawbacks to avoid any complications while building the model.

Below, we have reviewed few papers that we studied related to Twitter Sentiment Analysis system related to elections.

Parul S and Teng-Sheng Moh [3] predicted the results of 2016 Indian general elections using tweets in Hindi language. They performed text mining on 42,235 tweets collected over a month. They applied three ML algorithms. The accuracy of the Naïve Bayes' algorithm was 62.1% and the accuracy of Support. Vector Machine was 78.4%. Final prediction was done by utilizing SVM, since the accuracy was higher.

Dr D. Rajeswara Rao and team [4] gathered a dataset of more than 500,000 tweets out of which 80% was used for training and rest for testing. They predicted which political party had more influence on social media. Proposed a system that trained the dataset for more than 2 days and a classifier was built. Experiments proved that SVM was the most accurate model built with an accuracy of 80%.

Ferdin Joe and John Joseph [5] used decision tree to predict 2019 Indian General Elections. The results obtained in the proposed methodology showed it had a promising future in predicting Indian election results.

Meng-Hsiu Tsai and his team [6] at Middle Georgia State University presented a machine learning strategy to analyse Twitter data for predicting the results of local elections in US. They categorized their results into 5 classes namely very positive, positive, neutral, negative and very negative. They used RNTN model to calculate weighted sentiment scores.

Payal Khurana Batra and her team [7] predicted election results of Lok-Sabha 2019. After pre-processing, they split the data into two parts containing BJP and Congress tweets in separate sets. They trained their model using five different ML algorithms. Decision tree and XGBoost gave higher accuracy above 80%.

CHAPTER 3

SOFTWARE REQUIREMENTS AND SPECIFICATION

3.1 Introduction

3.1.1 Project Scope

The proposed plan will focus on predicting the outcome of elections based on sentiments of user tweets about electoral teams. The system will predict that the tweet is at parties next to or against it by using NLP techniques and getting the feel of a particular tweet that you should check out if positive or negative. The final result of each tweet will predict the feeling you might have have been instrumental in training various machine learning models. Feature Extraction Techniques are used done. The accuracy of each model will be calculated and will tell us which model is best predicts the maximum number of votes cast in a particular party.

3.1.2 User Classes and Characteristics

This application will be used by Users where they will use our application to check which party has more positive votes and compare it with other parties in the election. They can visualize graphs for accuracy of votes for parties.

3.1.3 Assumptions and Dependencies

- Assumptions are as follows:
  - The dataset which contains tweets of users are in English so that Stopwords on English language can be applied while performing data preprocessing techniques.
  - The system the application is executing on will have the required resources available as necessary.
  - Another assumption is that the software and hardware components work in the same way as used while developing this project.
- Tweepy API of Twitter is the main dependency to fetch current tweets about election on twitter.

3.2 Functional Requirements

3.2.1 Input Format of Dataset

- All items in the database should not have NULL values.
- The system must able to extract current tweets from twitter to classifiy the sentiments of each tweet.
- The system must able to process dataset entries so that unwanted words are removed and it becomes easy for further preprocessing techniques.
- The system must able to transform given entry into cleaned one.

3.2.2 Transformed Dataset

- The system should identify the language of input text.
- The system shall output the polarity score of each and distinguish between positive and negative sentiment of tweets.
- The polarity score is huge factor in predicting sentiment of any text.
- The main functionality of the project is to generate proper sentiments of users tweets for given party.

3.3 External interface requirements

3.3.1 User Interfaces

- The interface can be accessed via Google Chrome, Mozilla Firefox and any browser.
- A browser with its latest update is preferred to use.

3.3.2 Hardware Interfaces

- The development is based on calling the prediction engine to generate the Required Results.
- A good RAM would be necessary to enable faster processing.
- Hardware Devices required are - Mouse, Keyboard, Monitor

3.3.3 Software Interfaces

- . Operating System: 64-bit Linux OS (Ubuntu 16.04+, Fedora 20+) or Windows 7 / 8 / 10 or Mac OS
- Web Browser

3.3.4 Communications Interfaces

- This project supports all types of web browsers.
- The project can be accessed or communicated from any device.

3.4 Non-functional requirements

3.4.1 Performance Requirements

- System performance should be fast and accurate. Where tweets are extracted must be converted into a refined tweet, it should be converted to the token form of each sentence so that any stop-words present in the tweet can be ignored.
- Response delay, if any, should be only due to lack of computation power and not due to inefficiency of the algorithm.
- The result should be very accurate as it will be used to calculate the accuracy of various machine learning models to predict the required results. various machine learning models to predict required results.

3.4.2 Safety Requirements

- System should not crash due to overabundance of queries.
- Abrupt failure of system should not happen.

3.4.3 Security Requirements

- It should preserve the privacy of the twitter user whose tweet is being fetched and should not disclose any kind of information.

3.4.4 Software Quality Attributes

The software quality attributes for the system under consideration are:

- Robust - The software should be able to process multiple entries at a time and train model .
- Portability - The proposed system should be compatible with all browsers and versions.
- Usability - The system should be easy to use so that everyone can easily use it without any problems and visualize the status of each group.
- Accuracy - The accuracy of the system can be high considering the method it has used. The prediction will be accurate based on the data provided.
- Reliability - System reliability can be high considering the above reasons. The main reason for high reliability is that, the right information will be stored and then predicted.

3.5 System requirements

3.5.1 Database Requirements

This model has no database requirement as we save the training weights in our machine learning model itself.

3.5.2 Software Requirements

- Python: Python is the software that we have used, for which the interface is Jupyter Notebook. It is the most common tool, in some basic statistics, and is one of the easiest tools to use. Some of the libraries or frameworks to be used are: -
- NLTK: Natural Language Tool Kit is the library the provides facility for the python programs to work on Human language data as it helps in performing sentence detection, tokenization, lemmatization, stemming, parsing, chunking, and POS tagging.
- os : This provides the interface for interacting with the operating system in python which in term helps in interacting with the file system as well.
- re: This library helps in writing the regular expressions for specifying rules for the possible set of strings using the different symbols in python.
- sklearn: This library helps us to fetch various machine learning models and train our dataset on these models to predict required results.

3.5.3 Hardware Requirements

- Laptop/Desktop with a minimum of 4GB RAM.
- Stable internet connection.
- Processor which can handle intensive loads and provide efficient throughput.

3.6 Analysis Models: SDLC Model to be applied

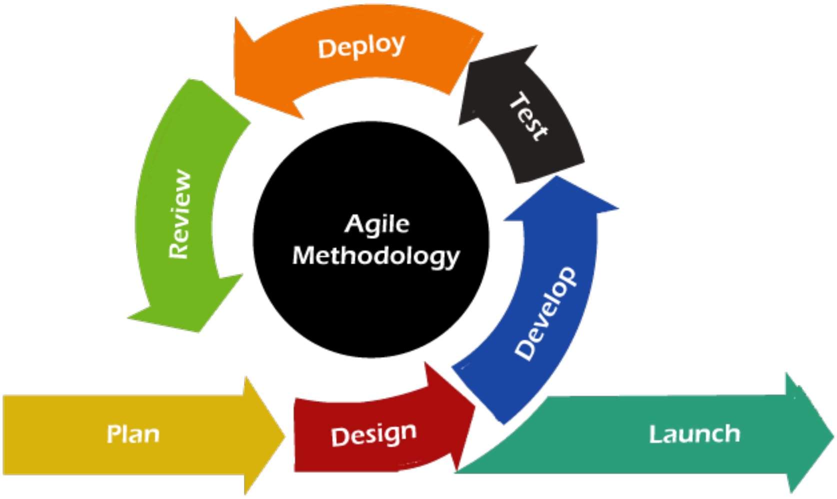


Figure 3.1: Software Development Life Cycle - Agile Model

This project makes use of the agile model. It involves following steps.

- SDLC is a process which is followed for any software project, within an organization.
- It contains a detailed program description of software development using the required techniques, how to save software after completion and replacing, altering or improving certain parts of the project.

1. **Planning**
   This section is responsible for analyzing needs and creating a map for future product development..
2. **Design**
   This is the stage at which system language is determined and other technical details of the project. We have improved the design of our project and we decided that different parts of our application can be integrated. We have created a learning design for our machine models in this category.

3. **Development**
   After the design stage, comes the development stage, where coding of the software is performed. Here various optimization techniques are applied to make the model more accurate and as useful as a predictive or segmentation tool.

4. **Testing**
   In this phase, the developed software is verified that it is built as per the specifications provided by the client and check if it is working properly or not.

5. **Deployment**
   In this phase we make sure that the environment is up and we deploy the application in the respective environment. We do overalls a request to verify that the request is not violated.

6. **Maintenance**
   The upgraded project is now maintained by performing standard updates bug detection and appropriate testing for the project to produce the appropriate output when delivered proper input.

3.7 System Implementation Plan

This section presents an outline of how the system will be used effectively in order to to satisfy all needs.
This includes the following activities:

1. An in-depth understanding of the tweepy twitter API and NLP methods for classifying tweets into categories available in the database and predicts the required output.

2. Separation of functions for performing various methods such as pre-processing of data, feature extraction, data cleaning, modeling training, etc.

3. Implementation of each part by team members with continuous meetings to track progress.

4. The test of each module uses a variety of test conditions.

CHAPTER 4

SYSTEM DESIGN

4.1 System Architecture

The high level system design shows how the different computational components, APIs and clients will be laid out.

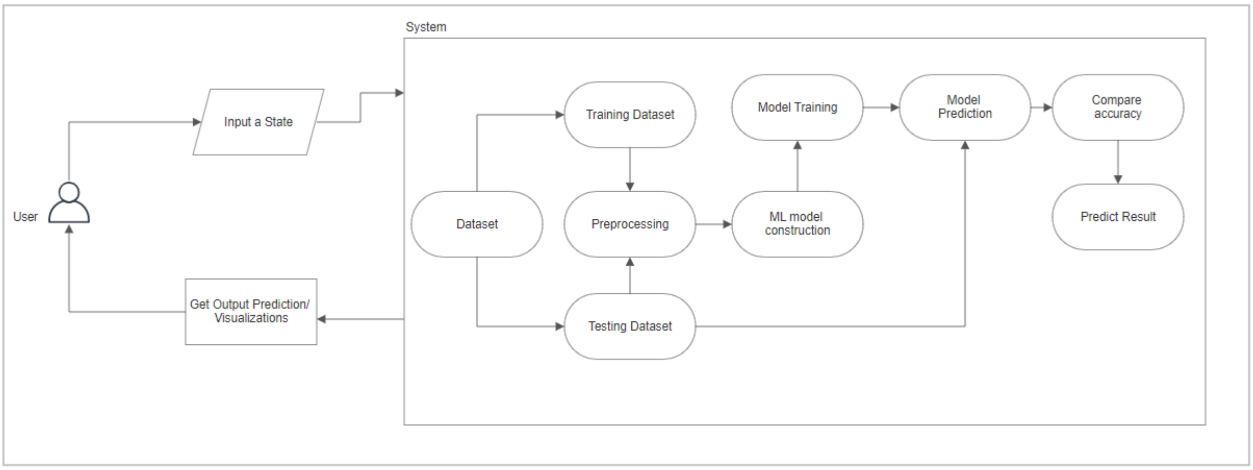


Figure 4.1: System Architecture Diagram

4.2 Data Flow Diagram

The data flow diagram illustrates the flow of data throughout the life time of the application, for either of the two main use cases

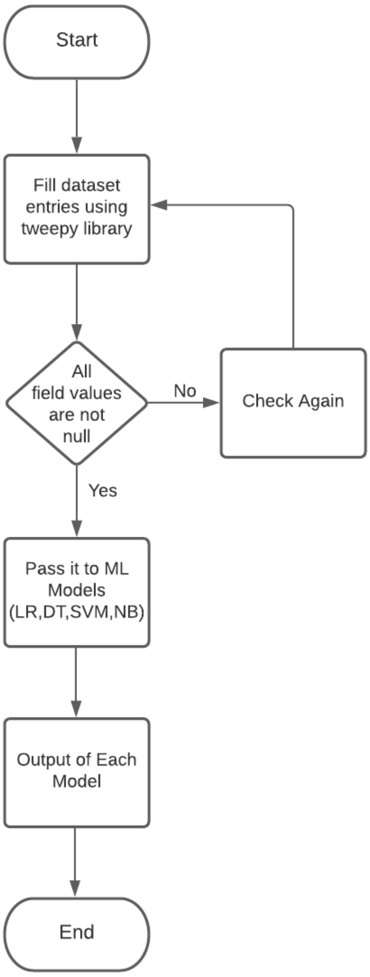

Start
Fill dataset entries using tweepy library
All field values are not null
No
Check Again
Yes
Pass it to ML Models (LR,DT,SVM,NB)
Output of Each Model
End

Figure 4.2: Data Flow Diagram

4.3 Usecase Diagram

The major use cases of the application are illustrated here, with users as the actors interacting with different modules

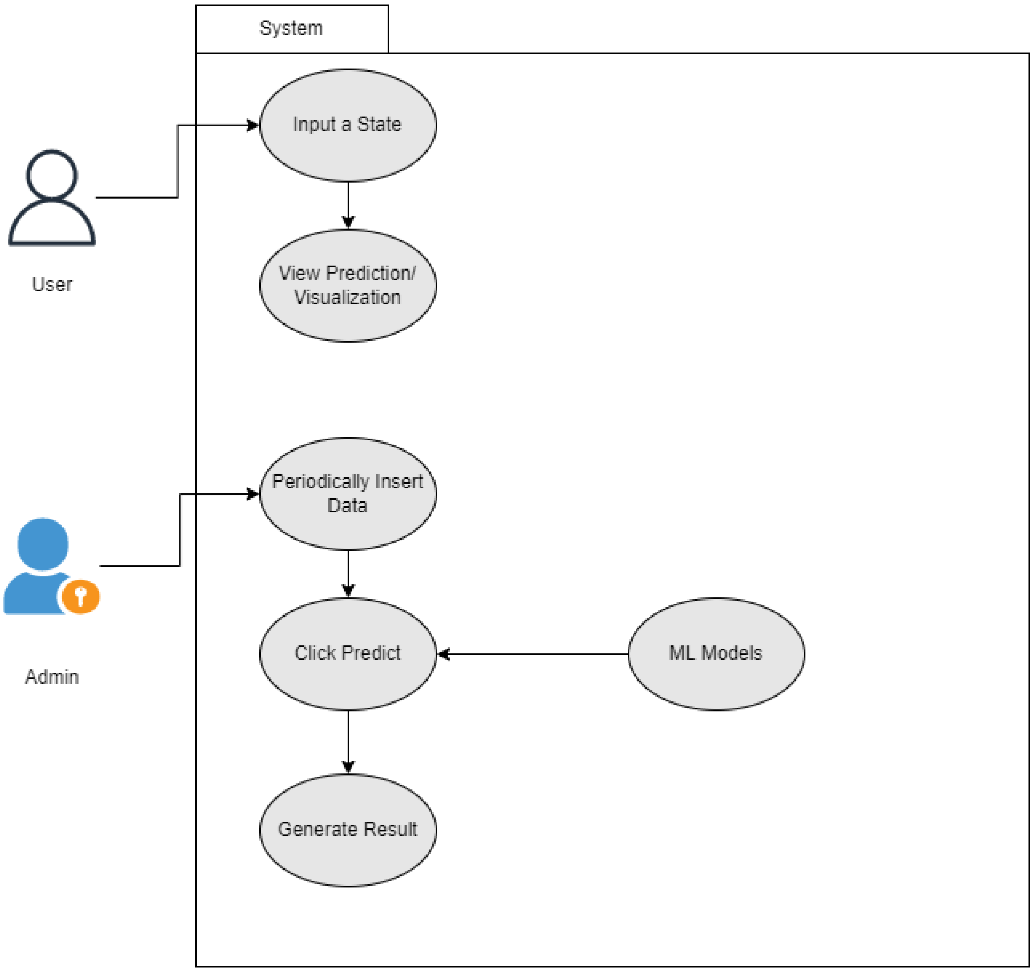


Figure 4.3: Usecase Diagram

CHAPTER 5

PROJECT PLAN

5.1 Project Estimates

A project estimate is a measure of the time and money invested to complete the project.Time required to complete the entire project was around 7-8 months.Any additional financial costs were not required as we used open source software.

5.1.1 Reconciled Estimates

- Cost Estimate: The Application is built using open-source libraries of python-3 for the back-end.The software used for the application will incur no cost as it is all released under free open source licenses.
- Time Estimate: The time estimate of the project, after follwing the software development life cycle turns out to be 150-180 working days, considering the development team of 3 members.

5.1.2 Project Resources

- Guidance :- Prof. S.N. Girme
- Group Members :-
  - Varun Kanade
  - Chinmay Motorwar
  - Ajay Rao
- Hardware Resources :-
  - 64-bit i5 core processor
  - 8GB RAM
  - 10GB Memory Space
- Software Resources:-
  - Windows or Linux OS
  - Visual Studio Code
  - Jupyter Notebook
  - Web Platform : Flask Framework
  - Libraries : NLTK,numpy,pickle,sklearn,tweepy
  - Programming Language : Python

5.2 Risk Management

Risk management is responsible for identifying potential risks project in advance, analyze and take safety measures to reduce risk.

5.2.1 Risk Identification & Analysis

The risk analysis is performed using the following guidelines:

Risk Probability

a.	High Probability	$75\% \leq x \leq 100\%$
b.	Medium High Probability	$50\% \leq x \leq 75\%$
c.	Medium-Low Probability	$25\% \leq x \leq 50\%$
d.	Low Probability	$0\% \leq x \leq 25\%$

Table 5.1: Risk Probability Levels

Risk Impact

a.	Very High	Catastrophic
b.	High	Critical
c.	Medium	Moderate
d.	Low	Marginal

Table 5.2: Risk Impact Severity

5.3 Project Schedule

5.3.1 Project Task Set

Task	Description
Task 1	Domain Study and Literature Survey
Task 2	Study different ML model for analysis and choose the best one for project.Study the NLP process for sentiment extraction from tweet dataset.
Task 3	Defining the problem domain and determine the data sets for case study
Task 4	Designing the system architecture and related diagrams
Task 5	Selecting feasible models and applying few shot learning techniques.
Task 6	Constructing the ML Models on given dataset
Task 7	Evaluation of model performance on selected datasets and metrics
Task 8	Project Review with Powerpoint Presentation

Table 5.3: Project Task Set

5.3.2 Timeline Chart

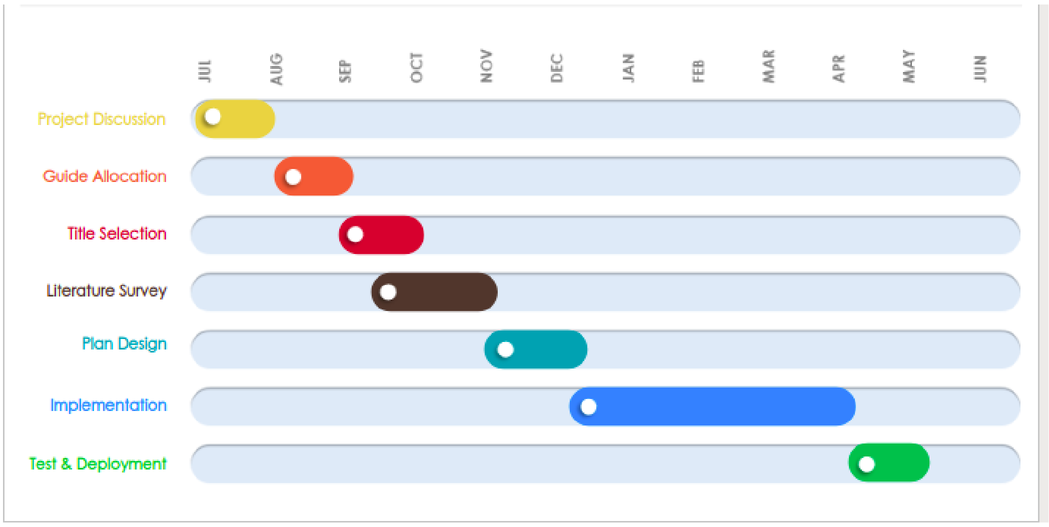


Figure 5.1: Timeline Chart

5.4 Team Organization

5.4.1 Team Structure

Role	Participants	Responsibilities
Internal Guide	Prof. S.N.Girme	1. Guidance for the project 2. Monitor project plan 3. Feedback and corrective action plan 4. Focus the team on project objectives
Team Members	1. Varun Kanade 2. Chinmay Motorwar 3. Ajay Rao	1. Implementation Of Front-end and Documentation 2. Integration of ML Models and Documentation 3. Back-end and ML Model Implementation and Documentation

Table 5.4: Team Structure

5.4.2 Management reporting and Communication

- Reporting to Management in form of reviews: 4
- Team members also collaborate in-person and work together to improve efficiency.
- Communication for review was through presentations with proof of work and project demos.
- Group discussions using face-to-face meetings and zoom meeting video call.

CHAPTER 6

PROJECT IMPLEMENTATION

6.1 Overview Of the Project Datasets

Tweets used for training the dataset was collected during the period of November-December, 2021. A total of nearly 12000 tweets were collected. Different hashtags and phrases like 'UPelection', 'Punjabelection', 'Yogi Adityanath', 'BJP elections 2022', 'INC Punjab elections', etc. were used. Thus, a domain-specific corpora was created and other features like 'like count', 'retweet count', 'user name' and date-time of tweet' were also included in the dataset.

Datasets of top political parties for every Indian state(contesting elections-2022) were collected every 5 days during Jan-Feb period for testing the model.

Two important libraries used were:

1. **Tweepy :** It is provided by Twitter. A collection of latest as well as popular tweets of a particular hashtag were collected and combined together.
2. **Snscrape :** As tweepy has a restriction on the amount of tweets to be extracted and tweets older than 7 days cannot be extracted, snscrape was used to overcome these limitations.

6.2 Tools & Technologies Used

6.2.1 Programming Languages

- The machine learning model and backend infrastructure were written using Python with the help of Flask and pickle.

- The client-side application is written in JavaScript.

6.2.2 Front End Tech Stack

- **HTML**
  HTML stands for Hyper Text Markup Language which gives markup for website. Markup is nothing but the entire structure the website you can say that it is a skeleton of the website. With HTML, all the necessary building blocks are created like forms, navigation, buttons, headings footer.

- **CSS**
  CSS stands for Cascading Style Sheets. On the other hand CSS is a stylesheet language which is used to style the html markup. Once the building blocks are ready we can begin with styling it.

- **Javascript**
  JavaScript is a programming language commonly used in web development. It is a client-side scripting language, which means the source code is processed by the client's web browser rather than on the web server.

- **Tableau**
  Tableau is a data visualization tool helpful in designing flexible front end and intuitive dashboards. Tableau dashboards can be embedded into a website by adding external data sources and visualizations on Tableau Server or Tableau Public.

6.2.3 Back End Tech Stack

- **Flask**
  Flask is a web framework. This means flask provides you with tools, libraries, and technologies that allow you to build a web application. This web application can be some web pages, a blog, a wiki, or go as big as a web-based calendar application or a commercial website.It uses jinja2 template for html rendering .Flask is part of the categories of the micro-framework. Micro-framework is normally a framework with little to no dependencies on external libraries. This has pros and cons. Pros would be that the framework is light, there is little dependency to update and watch for security bugs, cons is that sometimes you will have to do more work by yourself or increase yourself the list of dependencies by adding plugins. Flask constructor takes the name of current module as an argument. The route() function of the flask class is a decorator, which tells the application which URL should call the associated function

- **Pickle**
  Pickle is the standard way of serializing objects in Python. You can use the pickle operation to serialize your machine learning algorithms and save the serialized format to a file. Later you can load this file to deserialize your model and use it to make new predictions.Pickling is a way to convert a python object (list, dict, etc.) into a character stream. The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.

  To save the model,

  filename = 'model.csv'

  pickle.dump(model, open(filename, 'wb'))

  To load it again,

  loadedModel = pickle.load(open(filename, 'rb'))

6.3 Algorithm Details

6.3.1 Random Forest Classifier

Random Forest is a supervised learning algorithm that can be used for regression and classification problems. Random Forest is mainly used for classification problems. First, consider a completely random data set. In these cases, the entropy that defines randomness is very high. Splitting the root into sub-branches reduces the gini index.The Gini index determines the impurity when constructing a decision tree. Gini index is lower at leaf node of the decision tree.Decision trees also evaluate missing values in a data set. Model is mainly used for large data sets and provides accurate results. In case if we have overfitting, we can use concept of hyper-parameter.We have different parameters like min impurity split, ccp alpha which is stand for cost complex puring there are different parameters which can be used to reduce overfitting. Ccp alpha means impurity of nodes while building decision tree. As we grow towards leaf node purity is increase and impurity is decrease

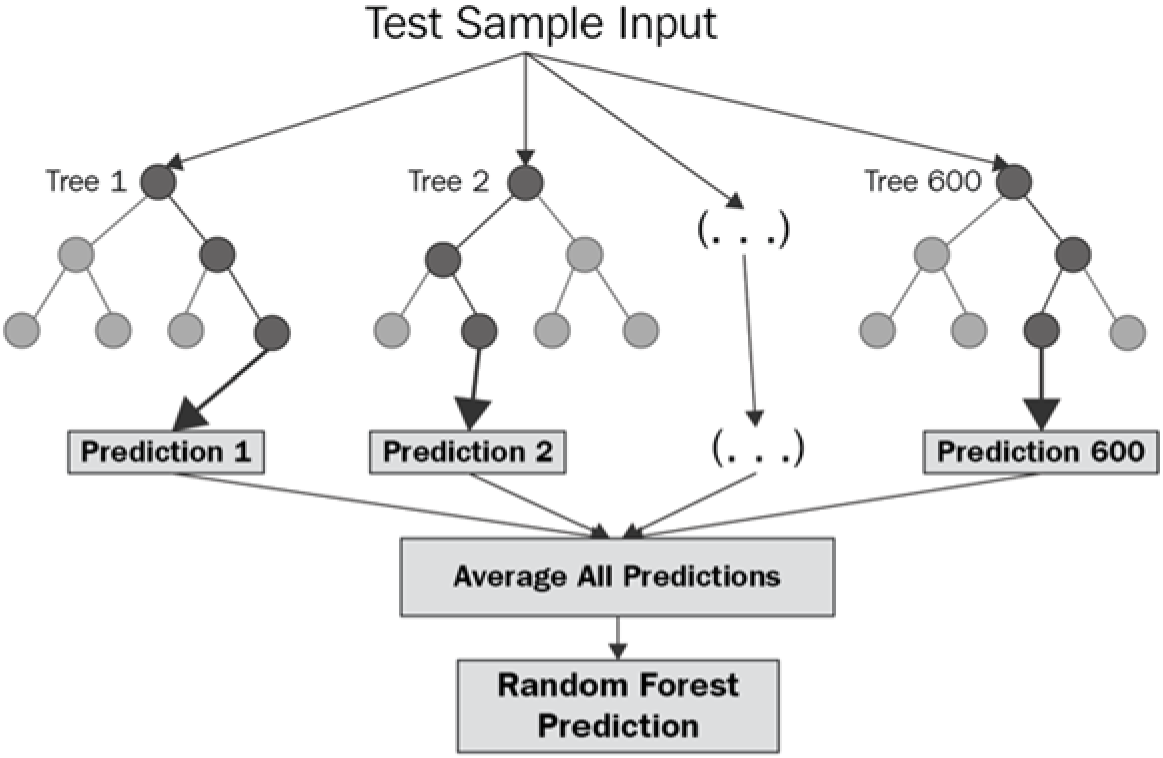


Figure 6.1: Random Forest (Source: Medium.com)[1]

6.3.2 Support Vector Classifier

It is a type of supervised machine learning algorithm, that provides analysis of data for classification and regression.The goal of the SVM algorithm is, to create the best line or decision boundary ,that can segregate n-dimensional space into classes, so that we can easily put the new data point ,in the correct category in the future. This decision boundary is called a hyperplane.The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane.The SVM kernel is a function that takes low dimensional input space and transforms it into higher-dimensional space, i.e. it converts not separable problem to separable problem. It is mostly useful in non-linear separation problems. Simply put the kernel, it does some extremely complex data transformations then finds out the process to separate the data based on the labels or outputs defined.

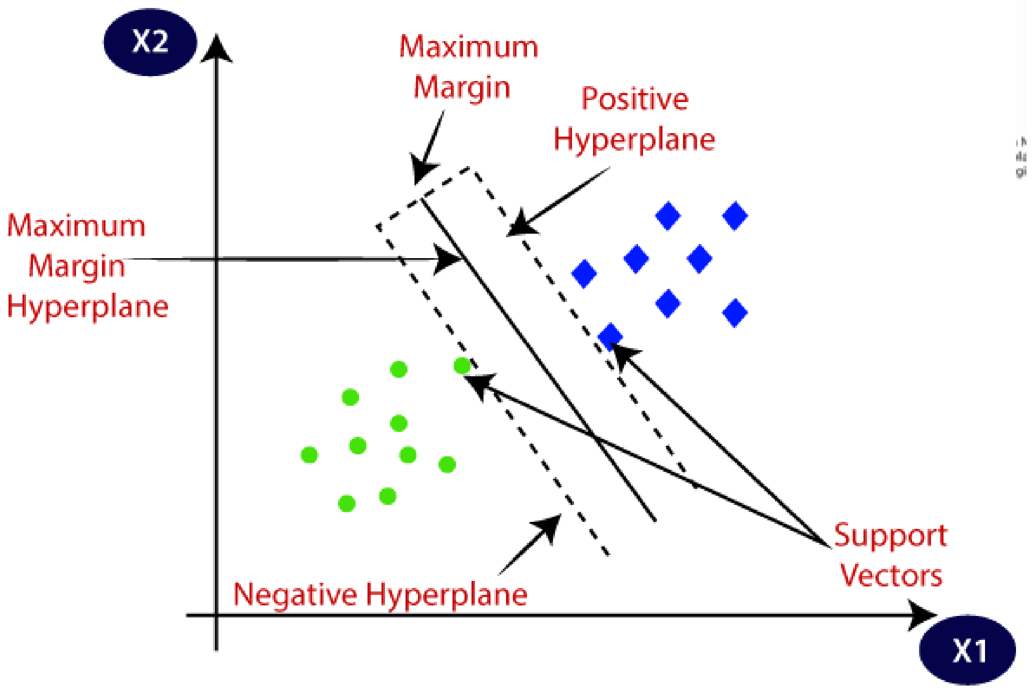
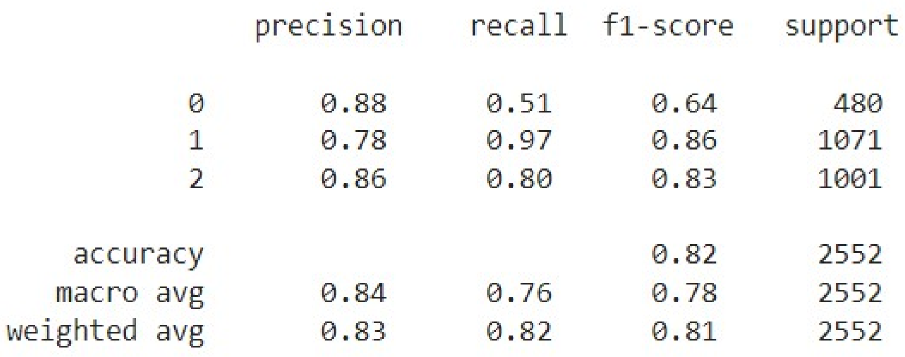


Figure 6.2: Support Vector Classifier (Source: Javapoint.com)[2]

6.4 Overview of Project Modules

- **Prediction using Random Forest Classifier**
  The module requires cleaned and preprocessed text as input of tweets for predicting the sentiment of the tweet. The module is fed over a pipeline which consists of a sequence consisting of CountVectorizer followed by TfidfTransformer and the classifier. Various parameters inside the classifier were tried out and the best set of hyperparameters were selected. An accuracy of 81.94% was achieved during training. The detailed accuracy of the model is shown in the below figure 6.3.

- **Prediction using Support Vector Classifer**
  Similar to the above module, SVC is fed over a pipeline by selecting different set of parameters. Initially feature extraction was done using only CountVectorizer and was later compared with TfidfVectorizer to get the best results. An accuracy of 75.86% was achieved.



```
              precision    recall  f1-score   support

           0       0.88      0.51      0.64       480
           1       0.78      0.97      0.86      1071
           2       0.86      0.80      0.83      1001

    accuracy                           0.82      2552
   macro avg       0.84      0.76      0.78      2552
weighted avg       0.83      0.82      0.81      2552
```

Figure 6.3: Model Accuracy

CHAPTER 7

SOFTWARE TESTING

7.1 Types Of Testing

Software testing is an activity that helps you compare actual results with expected results. also verifies that the system is not faulty. It also helps identify errors or missing requirements that were part of the original requirements.

Following are the types of testing

1. Functional Testing
2. Non-Functional Testing

7.1.1 Functional Testing

Functional Testing helps us to test against the Functional Specifications.
It involves following steps :

- Identification of the function of the software.
- Create an input database from a function specification.
- Define the output according to the functional specification.
- Run the test cases
- Compare actual results with expected results.

7.1.2 Non-Functional Testing

Non-functional testing helps to test non-functional aspects of the System such as performance, usability, reliability, etc.
Characteristics of Non-Functional Tests :

- Must be measurable. That is, there should be no room for subjective characteristics.
- It ensures the quality attributes of the system.
- The system has been tested for reliability, usability, portability, flexibility and reusability.

7.2 Test Cases and Test Results

Test Case Id	**Description**	**Test Case I/p**	**Actual Output**	**Expected Output**	**Result**
001	Click On Any Non-Election State from Map	Maharashtra	No Transition Of Page	No Transition Of Page	Pass
002	Hover On Any State	Goa	Displays State Details	Displays State Details	Pass
003	Select Election-State	Uttar Pradesh	UP Result and Analysis	UP Result And Analysis	Pass
004	Select Any Political Party of State -Punjab	AAP	Sentiment Analysis Graph Date-wise	Sentiment Analysis Graph Date-wise	Pass
005	Hover back to Punjab Results using dropdown list	Results	Punjab Results and Analysis	Punjab Results and Analysis	Pass

Table 7.1: Test Cases And Result

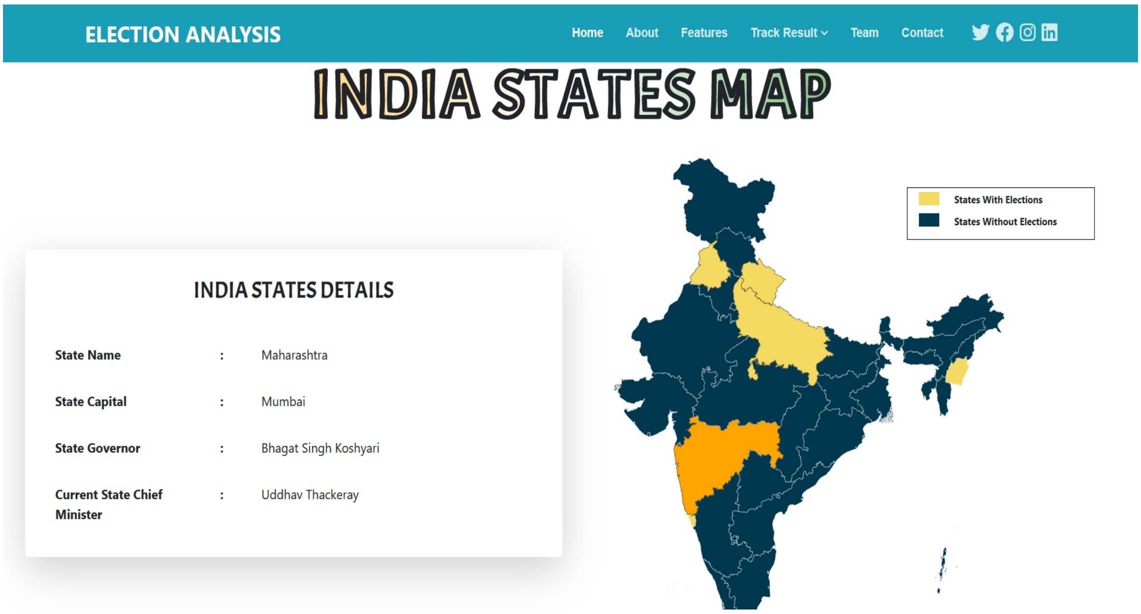
CHAPTER 8

RESULTS

8.1 Outcomes

- A user-friendly web app.
- Successfully Calculated Sentiments of each party tweets of dataset.
- Depending upon the rate of positive tweets , we compare different political parties and infer the winning party for 2022 election.

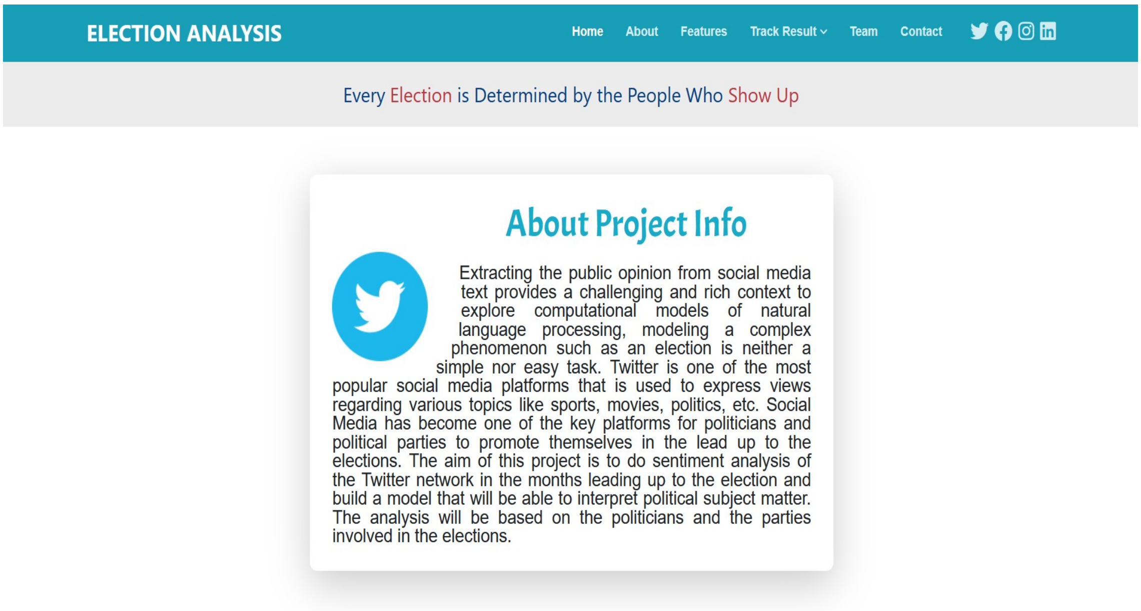
8.2 Screenshots



INDIA STATES DETAILS

State Name	:	Maharashtra
State Capital	:	Mumbai
State Governor	:	Bhagat Singh Koshyari
Current State Chief Minister	:	Uddhav Thackeray

Figure 8.1: Home Page



Every Election is Determined by the People Who Show Up

About Project Info

Extracting the public opinion from social media text provides a challenging and rich context to explore computational models of natural language processing, modeling a complex phenomenon such as an election is neither a simple nor easy task. Twitter is one of the most popular social media platforms that is used to express views regarding various topics like sports, movies, politics, etc. Social Media has become one of the key platforms for politicians and political parties to promote themselves in the lead up to the elections. The aim of this project is to do sentiment analysis of the Twitter network in the months leading up to the election and build a model that will be able to interpret political subject matter. The analysis will be based on the politicians and the parties involved in the elections.

Figure 8.2: About Page

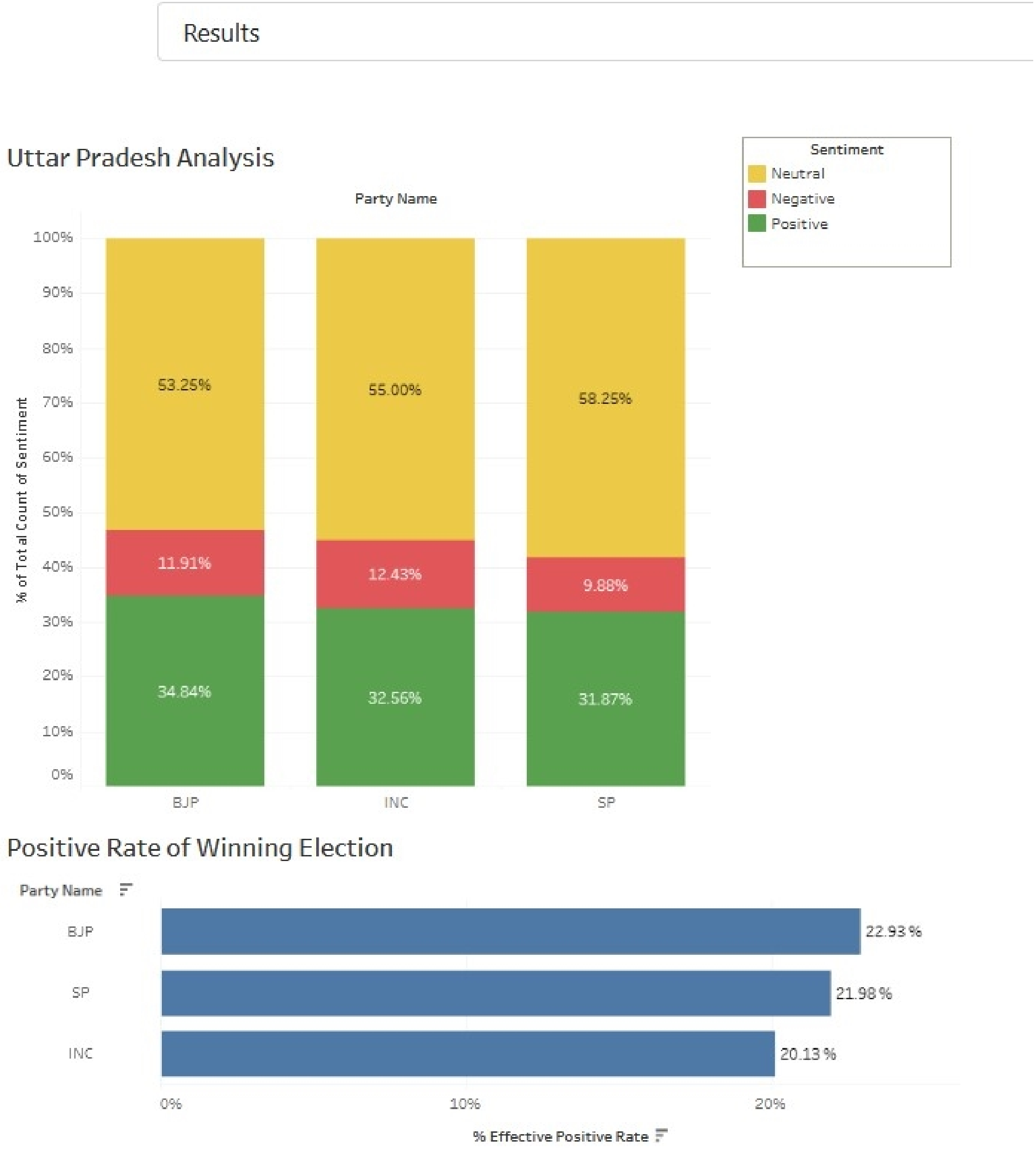
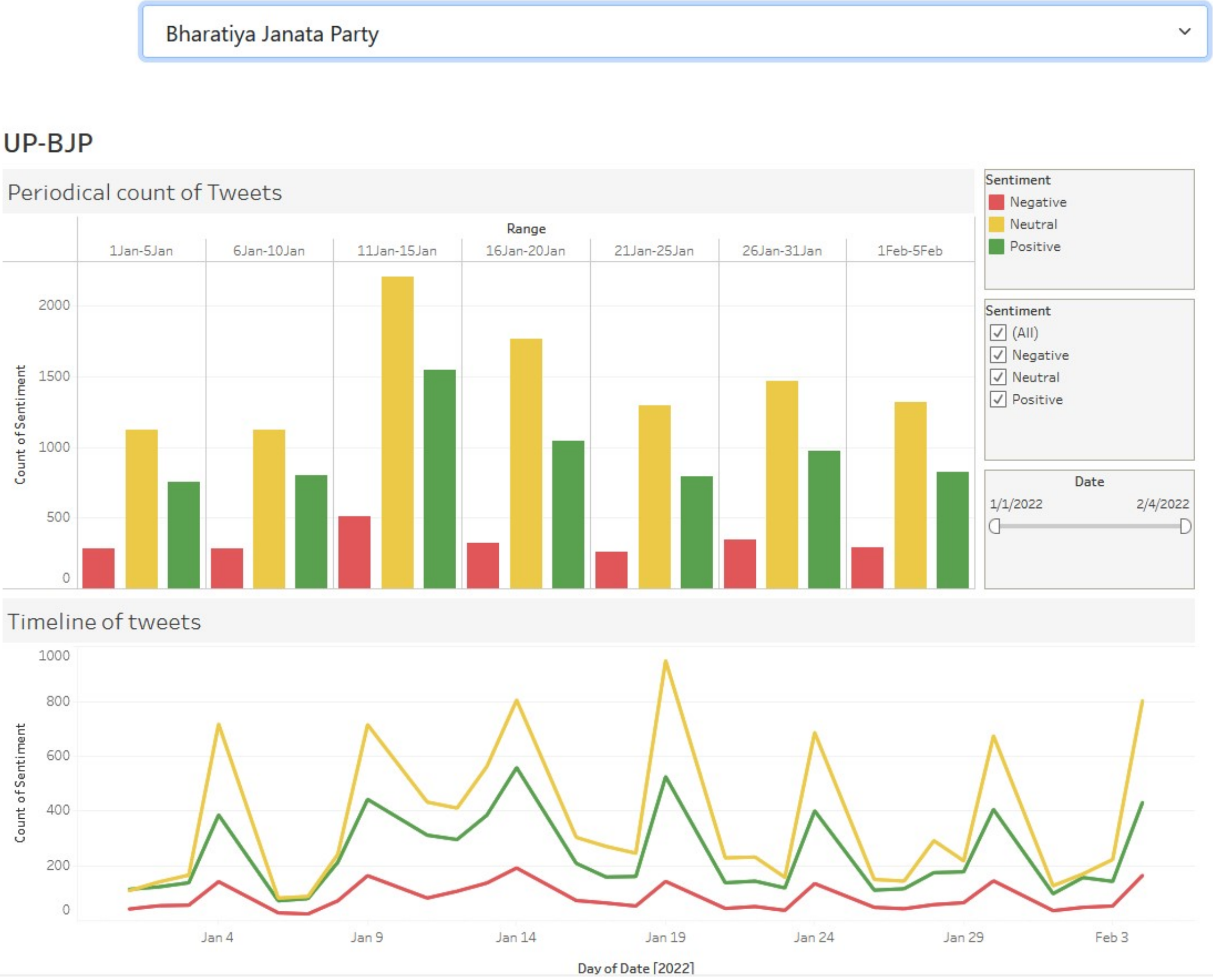


Figure 8.3: UP results

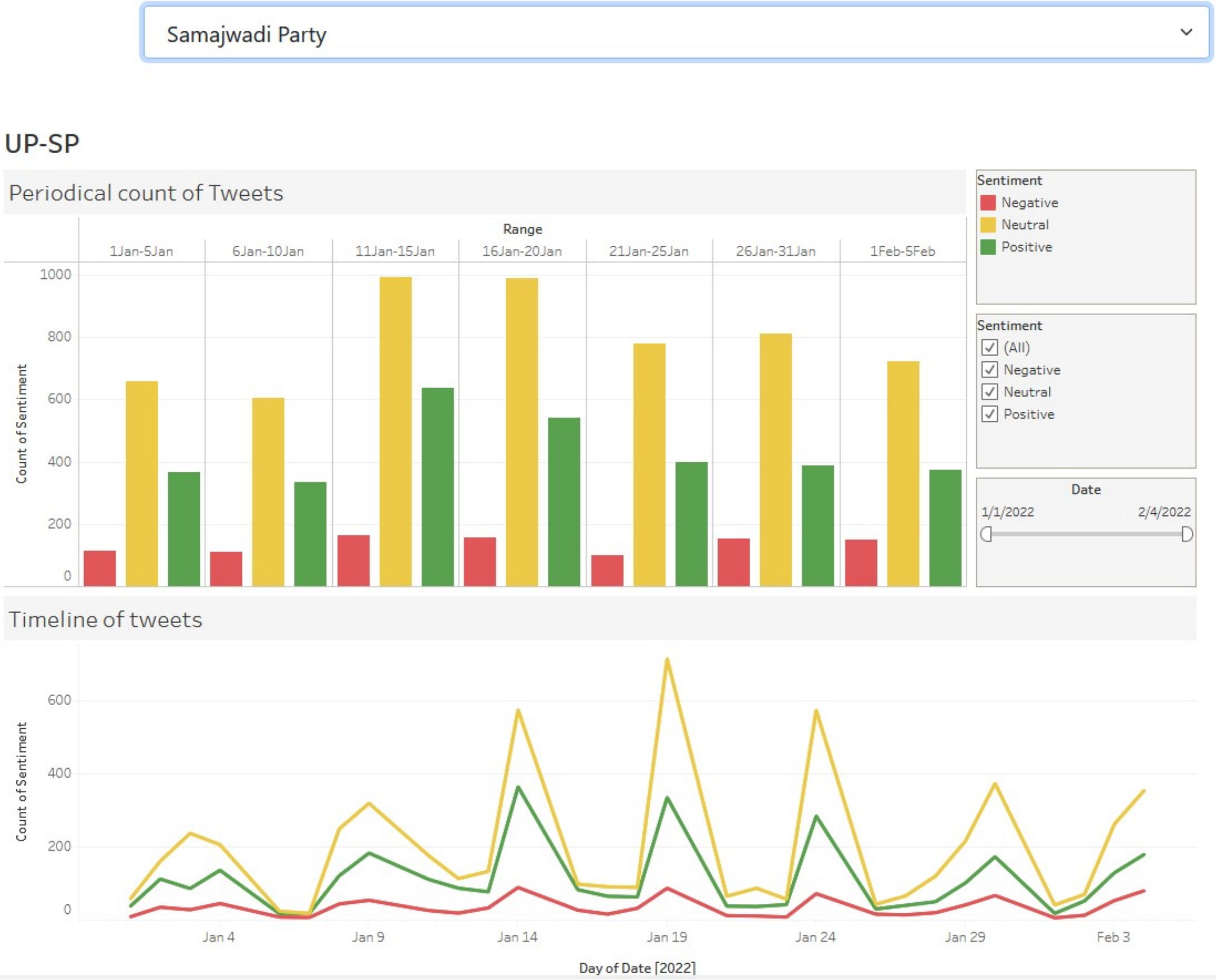


Bharatiya Janata Party

Sentiment
☑ (All)
☑ Negative
☑ Neutral
☑ Positive

Date
1/1/2022 2/4/2022

Figure 8.4: UP-BJP Analysis



Samajwadi Party

Figure 8.5: UP-SP Analysis

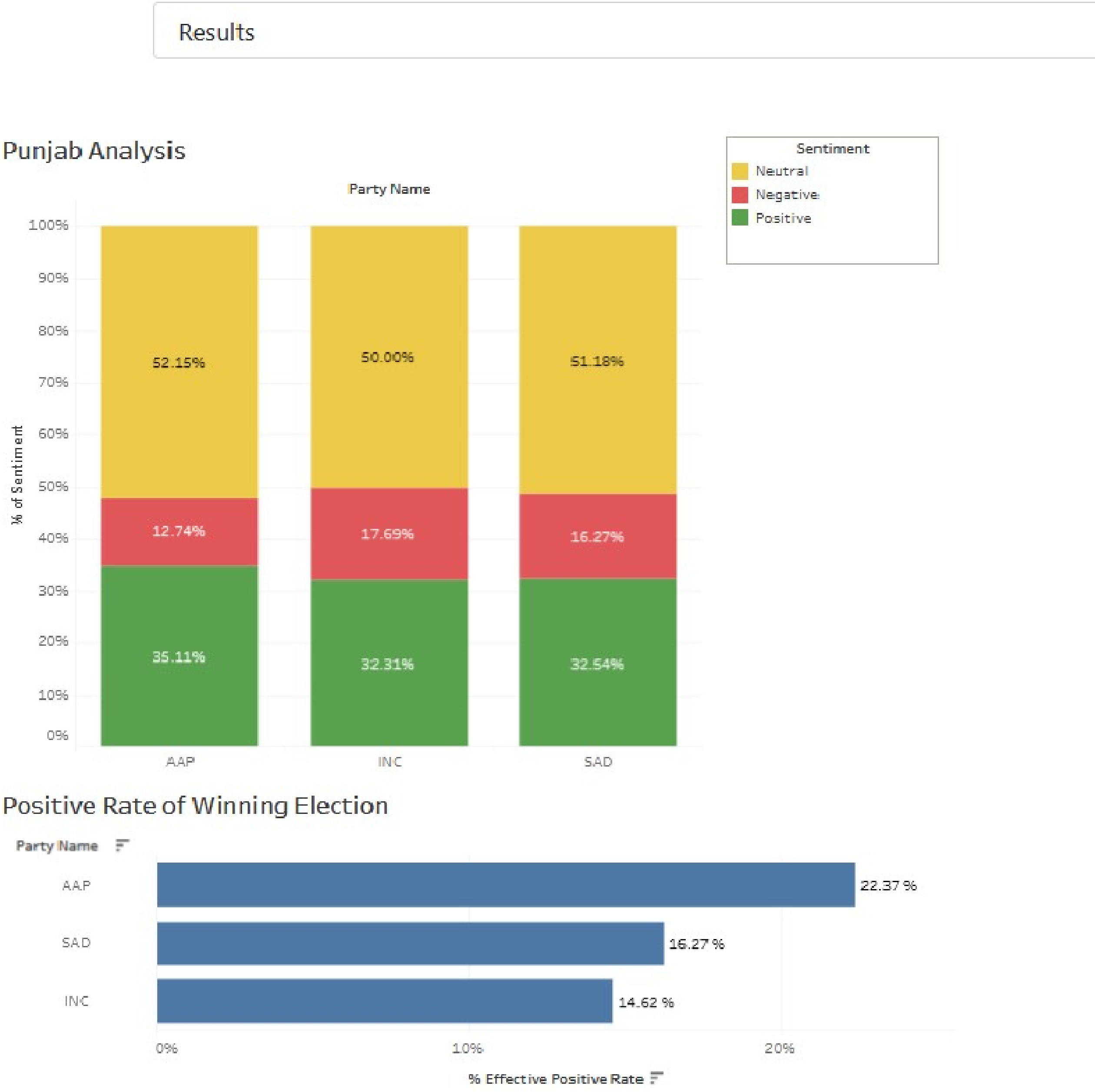


Figure 8.6: Punjab Results

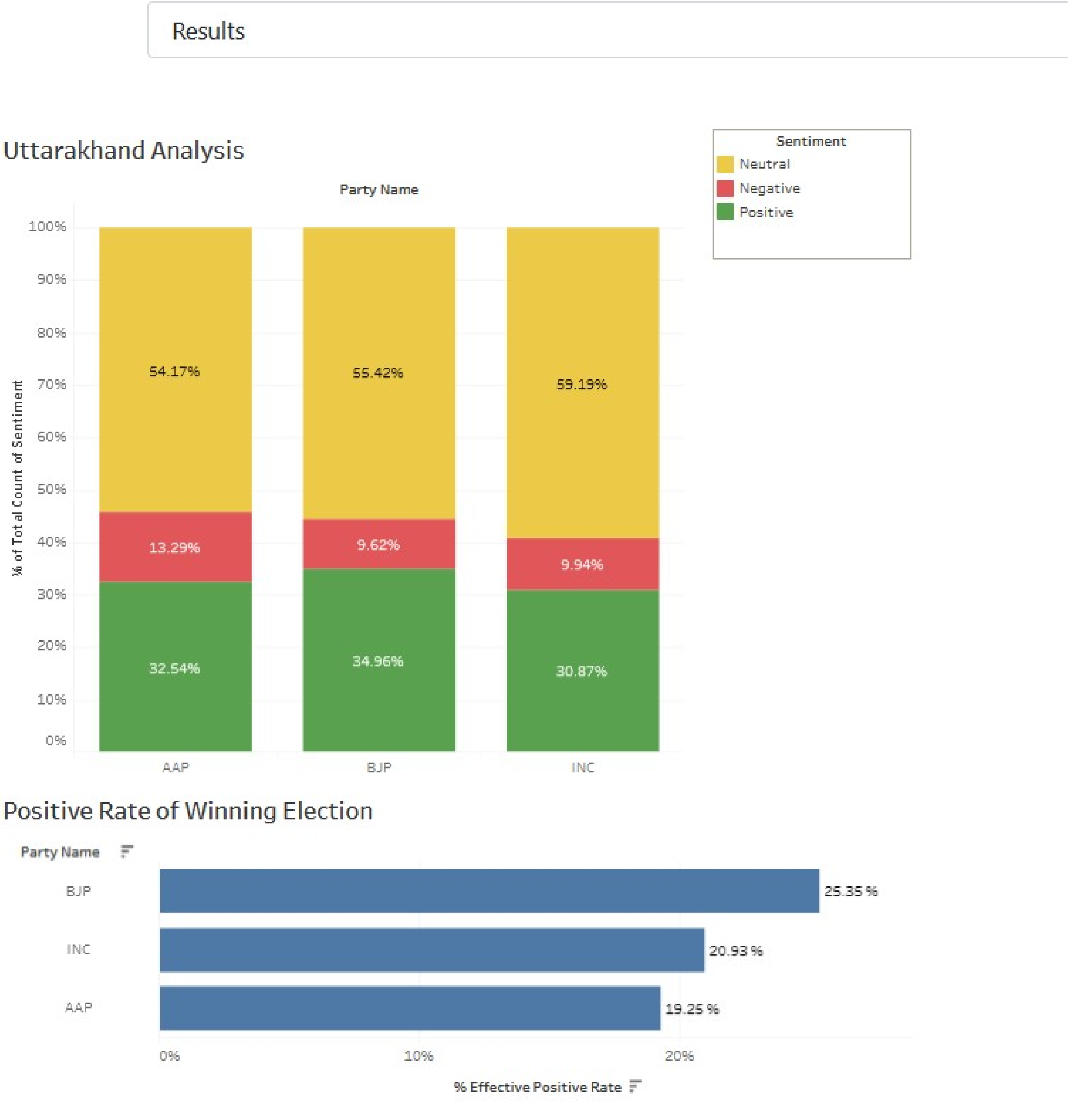


Figure 8.7: Uttarakhand Results

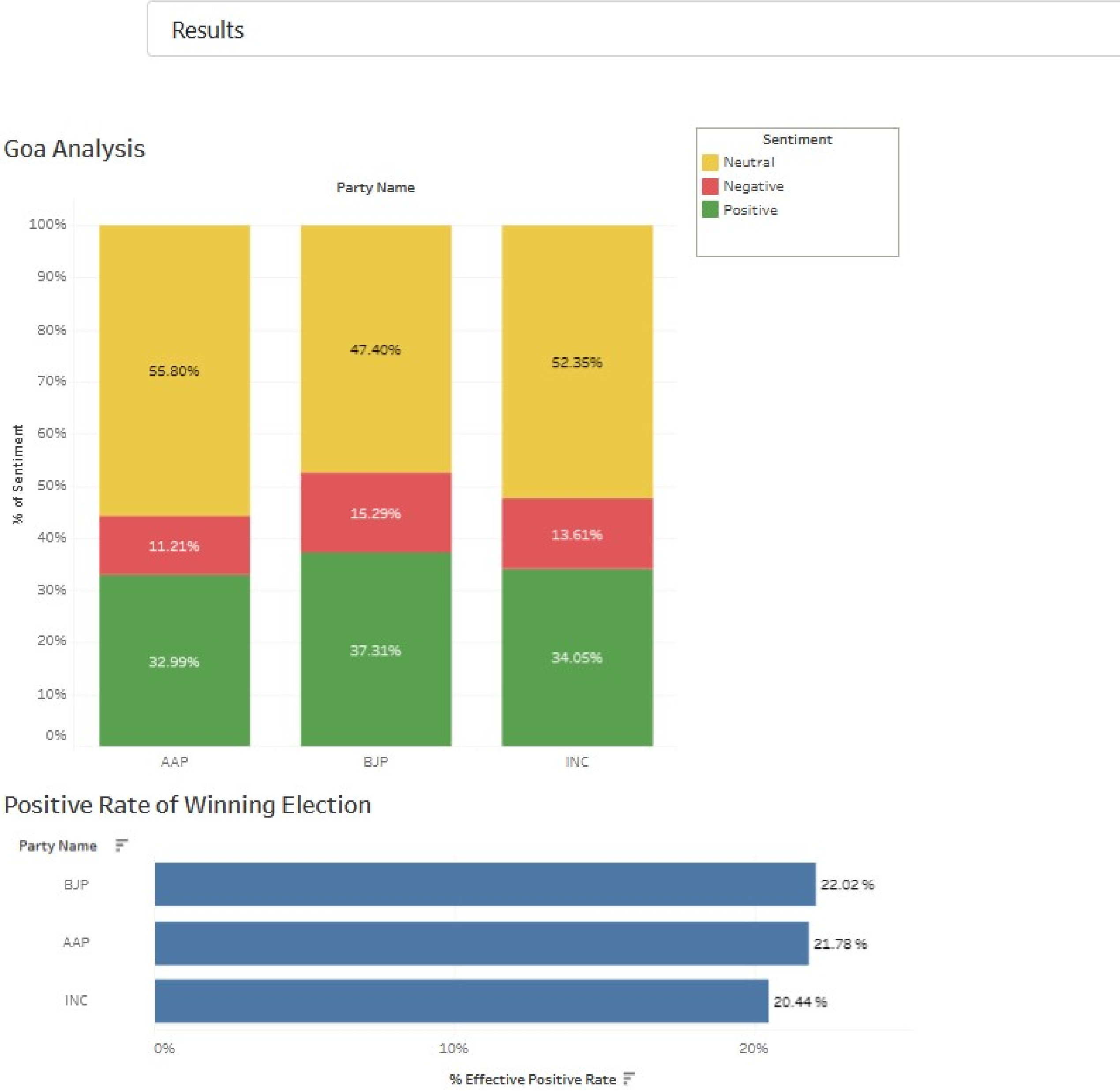


Figure 8.8: Goa Results

CHAPTER 9

OTHER SPECIFICATION

9.1 Advantages

- In a multi-billion dollar country, it is very difficult to predict an election situation if predictions are not available. They help people decide if their votes will go to someone who may really need them or use them rather than wasting on a candidate who has no chance of contesting prominent parties. It prevents the division of votes on a large scale.
- The new party that has never been a player in the area is those who are under their care and their efforts will not be seen if the pre-voting vote does not indicate a chance for them to get certain votes and a chance to get seats.
- The losing team gets a real check, with a chance to fight harder and prove the predictions are wrong. No wonder the promises are huge as the pre-election vote comes out.

9.2 Limitations

- Sarcasm was not detected in some sentences due to misuse of the semantics.
- Not everyone has access to social media where they can stand out from the crowd and express their support for one another.
- The twitter search api could retrieve data only of past 7 days.

9.3 Applications

- The application can be used by political parties to improve their campaigning strategies during the election period. It can be used by them as a part of social media analytics to study the trends of other political parties as well.
- User can make informed decision in voting by seeing the current trends of political parties.
- Political analyst and strategist can use this application as a long term plan for a political party to study the sentiments of people over a long time period and thus, making informed decisions.

CHAPTER 10

CONCLUSION

10.1 Conclusion

Observing the expanded use of social media platforms, this project concentrated on exploring of social platform (Twitter) as the chase for elections campaign. Understanding India to be one of the highest socially connected countries, having greater than 70% of its young generation below 35 years of age; Social platform plays essential part in young youth's life. The designed system will work upon the analysis of the Various states meeting election; taking a look at the impact of social platforms on the politics system of respective states political parties, found people can express their perspective more efficiently and openly.

10.2 Future Scope

- This work can further be extended on tweets of different regional languages of Indian states other than English to improve accuracy. Currently, regional languages supported by Twitter are Hindi, Gujrati, Marathi, Urdu, Tamil, Bengali, and Kannada.

- The proposed system does not consider geographical location of the tweet as a filter for state elections, as Twitter does not provide adequate information about user's location, thus, a general mood of entire blogosphere is considered to predict the election results.

APPENDIX A
MATHEMATICAL FORMULAE

1. **TF-IDF**

$$\text{TF(t,d)} = \frac{\text{Occurrence of t in d}}{\text{Total number of terms in d}}$$

$$\text{IDF(t)} = \log \frac{\text{Total number of documents}}{\text{Number of documents that contain t + 1}}$$

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

2. **Random Forest Classifier**

$$\text{Gini} = 1 \text{ - } \sum_{i=1}^{c} (p_i)^2$$

APPENDIX B
DETAILS OF THE PAPER

The proposed paper will discuss the chances of a political party winning the State Elections of March 2022, by studying their popularity over the social media(Twitter). A detailed timeline of tweets of top three political parties are visualized over the election period. Training data used consists of 'Tweets', 'Date', 'User', and 'Sentiment' as features. Sentiment has been labelled using VADER library. Testing dataset has been collected every five days, during the period of Jan-Feb 2022, using popular keywords and hashtags relating to different parties and leaders. The paper will also mention different cleaning and preprocessing steps used like CountVectorizer and TfidfTransformer. The performance of different ML models are compared and Random Forest Classifier performed the best. The model is applied over the test data and different graphs are analyzed. The results show that BJP has higher chances to win elections in Uttar Pradesh, Goa, Manipur and Uttarakhand.

The above paper can be published in International Conference on Data Science, Technology and Applications, and International Journal of Computer Engineering and Technology(IJCET).

APPENDIX C
PLAGIARISM REPORT

Document Information

Analyzed document	BE_Project_Group_64.pdf (D136267573)
Submitted	2022-05-12T07:45:00.0000000
Submitted by	Shital
Submitter email	sssupase@gmail.com
Similarity	0%
Analysis address	sssupase.pict@analysis.urkund.com

Sources included in the report

REFERENCES

[1] Afroz Chakure. Random forest classifier. https://miro.medium.com/max/1400/0*f_qQPFpdofWGLQqc.png, 2019. [Online; accessed 20-02-2022].

[2] Javapoint.com. Support vector machine. https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm.png. [Online; accessed 20-02-2022].

[3] Parul Sharma and Teng-Sheng Moh. Prediction of indian election using sentiment analysis on hindi twitter. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1966–1971. IEEE, 2016.

[4] Dr D Rajeswara Rao, S Usha, S Krishna, M Sai Ramya, G Charan, and U Jeevan. Result prediction for political parties using twitter sentiment analysis. *International Journal of Computer Engineering and Technology*, 11(4), 2020.

[5] Ferdin Joe John Joseph. Twitter based outcome predictions of 2019 indian general elections using decision tree. In *2019 4th International Conference on Information Technology (InCIT)*, pages 50–53. IEEE, 2019.

[6] Meng-Hsiu Tsai, Yingfeng Wang, Myungjae Kwak, and Neil Rigole. A machine learning based strategy for election result prediction. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1408–1410. IEEE, 2019.

[7] Jyoti Ramteke, Samarth Shah, Darshan Godhia, and Aadil Shaikh. Election result prediction using twitter sentiment analysis. In *2016 International Conference on Inventive Computation Technologies (ICICT)*, volume 1, pages 1–5. IEEE, 2016.