

# Abstract

---

One of the most important players in modern civilisation is social media. By providing consumers with real-time online information channels without the limits of time and space, social media is redefining existing information behaviour. This results in a large amount of unstructured data that can be mined. This opens up a lot of possibilities for data analysis challenges for scientists. In this project, the system will analyse general user tweets from an electoral standpoint. The system will investigate how users feel about the Indian election. Based on the users' tweets, the proposed system determines the feasibility of developing a classification model to identify the political orientation of Twitter users based on tweet content and other user-based features. Voting advise applications (VAAs) are online tools that are commonly used in determining which party/candidate to vote for during an election in countries such as Greece and Cyprus, but there is no similar application in India. The suggested system will include an internet platform for user recommendations and comparison of political viewpoints.

# Introduction

---

The 21st century is the century of Knowledge. For every answer people first use Google. The theme line is connecting people using Internet. When we talk about Internet along with knowledge we have Social sites are specially used for connecting world. People especially youth spend its lot of time on sites like Facebook, Twitter etc... Every current topic due to social media is supported with general public's reviews and comment which makes a clear picture of that particular topic to world. The credit goes to popularity of social media. According to Com. Score India is now at third position taking place of Japan becoming the world's third largest Internet user after China and United States. As per Internet and Mobile Association of India (IAMAI) in June 2014 India is having more than 243 million internet users; which is 23 % increase in count compared to year 2013. The use of social media is significantly complex. The key concerns with social media are security, privacy, and accuracy, but it has also provided new potential for user sharing, cooperation, and involvement. It serves as a forum for politics. Social media platforms such as Twitter, Facebook, and YouTube encourage people to participate in all political activities by sharing their opinions about parties and candidates. Elections have a significant impact on social media.

It is primarily used by the general public to stay up to date on political events, as well as Nilesh V. Alone Department of Computer Engineering, GES's R. H. Sapat College of Engineering, Savitribai Phule Pune University, parties/candidates to update their strategies and policies based on the valuable feedback provided by social media. The following is how this document is structured. Section II examines the existing mechanism for election-based analysis. Present the proposed system in Section III. The implementation details are provided in Section IV. Section V assesses the dataset and its results.

# Problem Definition

---

With all the real time data collected over period of time, the system will analyze and draw meaningful inferences from the collection of tweets. Proposed system will analyze tweets data from many perspectives to make meaningful inferences. Trend analysis, sentiment analysis volume analysis are major parts of proposed system. In trend analysis, system will try to find trending discussions, parties, personalities throughout the period of time. From literature, K-means is more suitable algorithm for clustering of tweet

data and to find trends. Volume analysis of tweets will give idea of popularity of particular topic or person over a period of time. Volume analysis with respect to geo-location and date will help to make certain conclusions. Sentiment analysis of tweets will help draw conclusion for political orientation of overall users respective to political parties, topics.

#### 1. Input

- Twitter tweets related to election
- A hash-tag List

#### 2. Output

- Volume Analysis
- Trend Analysis
- Sentiment Analysis

## Literature Review

---

Social media has become such an integral element of our civilisation. Social media is altering existing information behaviour by providing users with real-time online information conduits that are not limited by time or geography. This provides scientists with a huge amount of potential for data analysis challenges. Most earlier research used broad-brush methods, which resulted in limited analysis options.

Min Song, MeenChul Kim, and Yoo Kyung Jeong investigated a Twitter dataset for the 2012 Korean election by gathering real-time tweets. The relationship between topics extracted from tweets and related real-time events was identified and traced chronologically using the term co-occurrence retrieval technique.

Author Abhishekbhola used Twitter to research user attitudes about parties and candidates in India's 2014 general election. A dataset of 17.60 million tweets was evaluated in order to assess user, candidate, and party popularity depending on peak time, topic, or region. Using classification techniques, a sentiment analysis was carried out.

In nations such as Greece, voting advice applications (VAAs) are internet tools created for election purposes. It is growing in popularity, and they are assisting consumers in determining which party/candidate to vote for. It is based on a concept known as the community-based recommendations system. It allows users' political opinions to be compared and serves as a channel for user communication. This system provided several methods for recommending votes based on community input. In terms of prediction accuracy, the techniques were tested on five real VAA datasets. Lars Kaczmirek and his GESIS team gathered several features of communication systems using data from Facebook and Twitter. They matched social media data with local survey data and contributed fresh insights to social media by demonstrating how social media may be used during elections.

Based on the findings of this study, the German Longitudinal Election Study (GLES), a long-term research project is being developed that will evaluate the German federal elections in 2009, 2013, and 2017. The primary goal of this project is to track the German electoral process throughout time by gathering data from Twitter and Facebook concerning the German Bundestag elections.

## Proposed Work

---

The system's major purpose and difficulty is to analyse Twitter data for Indian elections in order to determine the impact of twitters on Indian or specific state elections. The suggested algorithm is divided into three major phases.

Our proposed system is an analysis system based on a mechanism that analyses User Tweets with Hashtags and Keywords. The suggested system captures tweets by using HashTags, which are just popular personalities/parties. The general public's attitude toward these parties can be analysed using tweets made on Twitter. Academics, journalists, and politicians have all praised twitter for its potential political worth.

Many politicians utilise this microblogging site to express themselves under a restriction of 140 characters. These tweets can be classified according to several rules, such as geo-location analysis, which analyses people's perceptions of a certain place and may aid parties in developing a winning strategy. The proposed system focuses on tweet collection to make volume analysis to and out of popular election days; a trend analysis to and a popular or trending party/candidate and a sentiment analysis to actually bifurcate the positive and negative tweets for the party/candidate so that making trend analysis on this tweets can help this party/candidate to act accordingly to improve their reputation at the same time it might help user to actually make a clear opinion.

This will be done in three stages. To summarise, the first step is to connect with Twitter and get the tweets. The second part is concerned with loading these tweets onto HDFS for future analysis, while the third phase is concerned with the actual analysis, which includes volume analysis, trend analysis, and sentiment analysis.

## Tools and Frameworks Used

---

### TensorFlow

TensorFlow is a complete open source machine learning platform. It features a rich, adaptable ecosystem of tools, libraries, and community resources that enable academics to push the state-of-the-art in ML while also allowing developers to quickly build and deploy ML-powered applications.

TensorFlow was created by researchers and engineers on Google's Machine Intelligence Research organization's Google Brain team to undertake machine learning and deep neural network research. The system is broad enough to be usable in a wide range of different fields.

TensorFlow offers robust Python and C++ APIs, as well as non-guaranteed backward compatibility with other languages.

### Tweepy

Tweepy is a library that provides abstractions over the Twitter REST API, allowing for simple access to the Twitter API, with clearer syntax and internal error handling, as well as pre-implemented functions and exceptions, making it easier to access and manage Twitter. It allows one to stream live tweets, search for a certain hashtag, view user timeline tweets, and post/delete tweets on your own timeline.

### Flask

Flask is a web framework and a Python module that allows you to easily create web applications. It has a tiny and easy-to-extend core: it's a microframework without an ORM (Object Relational Manager) or similar

functionality. It has a lot of cool features, such as url routing and a template engine. It is a WSGI web application framework.

A Web Application Framework, often known as a Web Framework, is a set of tools and modules that allow web application developers to construct applications without having to worry about low-level issues such as protocol, thread management, and so on.

Flask is a Python-based web application framework. It was created by Armin Ronacher, who led the Pocco team of international Python aficionados. Flask is built on the Werkzeug WSGI toolkit as well as the Jinja2 template engine. Both are Pocco creations. Flask, in contrast to the Django framework, is very Pythonic. Flask is simple to learn because it does not have a steep learning curve.

Furthermore, it is quite explicit, which improves reading. You only need a few lines of code to construct the "Hello World" app.

Although it is a microframework, this does not imply that your entire app should be contained within a single Python file. To handle complexity, you can and should use many files for larger programmes. The term "micro" refers to the Flask framework's simplicity while remaining expandable. Flask does not make any decisions for you, such as which database to use or whether to utilise an ORM.

Flask is one of the most popular web frameworks, which means it's up to date and cutting-edge. Its functionality can be simply expanded. It may be scaled up for more complicated applications.

## Numpy

NumPy is a Python library used for working with arrays and is an essential package for scientific computing with Python. It also has functions for working with linear algebra, the Fourier transform, and matrices. NumPy, which stands for Numerical Python, is a library that contains multidimensional array objects as well as a set of algorithms for manipulating those arrays. NumPy can conduct mathematical and logical operations on arrays. Lists in Python serve the same purpose as arrays, although they are slower to process. NumPy's goal is to provide array objects that are up to 50 times faster than ordinary Python lists. The array object in NumPy is named `ndarray`, and it comes with a slew of helper methods that make working with `ndarray` a breeze. Arrays are often utilised in data science, where speed and resources are critical. NumPy arrays, unlike lists, are stored in a single continuous location in memory, allowing programmes to access and modify them relatively fast. In computer science, this is referred to as locality of reference. This is the primary reason NumPy outperforms lists. It is also tuned to function with the most recent CPU architectures. NumPy is a Python library that is partially written in Python, but the majority of the elements that require fast computation are written in C or C++.

## Matplotlib

Matplotlib is a Python package that allows you to create static, animated, and interactive visualisations. Matplotlib makes simple things simple and difficult things possible.

A huge number of third party packages extend and build on the functionality of Matplotlib, including several higher-level plotting interfaces (seaborn, HoloViews, ggplot, ...), and a projection and mapping toolkit (Cartopy). Matplotlib is a community project maintained for and by its users. It is the result of development efforts by John Hunter (1968–2012) and the project's many contributors.

## Python

Python is a programming language that allows you to work faster and more efficiently integrate systems. Whether one is new to programming or an experienced developer, it's easy to learn and use Python. It features a simple syntax along with modern features to allow for a better programming experience. Python's convenience has made it the most popular language for machine learning and artificial intelligence.

## React.js

React is a JavaScript library for creating user interfaces that makes it simple to develop interactive UIs. Create simple views for each state of your application, and React will effectively update and render only the necessary components when your data changes. Declarative views make programmes more predictable and easier to debug. Create enclosed components that handle their own state, then combine them to create complicated UIs. Because component logic is done in JavaScript rather than templates, you can quickly pass rich data through your app while keeping state out of the DOM. React makes no assumptions about the rest of your technological stack, so you can create new features in React without rewriting current code.

## Node.js

Node.js® is a JavaScript runtime based on the V8 JavaScript engine in Chrome. Node.js is a scalable network application runtime that is built as an asynchronous event-driven JavaScript runtime. Many connections can be handled concurrently in a Node.js programme. This is in contrast to today's more popular concurrency approach, which employs OS threads. Thread-based networking is wasteful and complex to implement. Furthermore, because there are no locks, Node.js users are not concerned about deadlocking the process. Almost no function in Node.js does I/O directly, therefore the process never blocks unless the I/O is performed via synchronous methods from the Node.js standard library. Scalable systems are relatively feasible to design in Node.js since nothing blocks. Node.js is designed similarly to and influenced by systems such as Ruby's Event Machine and Python's Twisted. Node.js extends the event model even further. It exposes an event loop as a runtime construct rather than a library. In other systems, the event-loop is always started with a blocking call. In Node.js, HTTP is a first-class citizen, designed with streaming and low latency in mind. As a result, Node.js is well suited for use as the backbone of a web library or framework.

## Framer Motion

It is a production-ready motion library for React that utilizes the power behind Framer, a powerful prototyping tool for teams and is distributed under an open source license. A simple declarative syntax means you write less code. Less code means your codebase is easier to read and maintain. When animating between two separate components, Framer Motion will take care of everything in between.

Framer Motion offers more advanced listeners and also extends the basic set of React event listeners. Plus, designers who prototype in Framer can leverage the power of Framer Motion without touching a line of code, then handoff the animation values to be used 1:1 in production.

## TailwindCSS

TailwindCSS helps you to quickly create modern websites without ever leaving your HTML. A utility-first CSS framework that includes classes like `flex`, `pt-4`, `text-center`, and `rotate-90` that may be combined to create any design directly in your markup. Instead of polluting your stylesheets with random values, utility

classes allow you to work within the boundaries of a system. They make it simple to maintain consistency in colour schemes, spacing, font, shadows, and everything else that constitutes a well-engineered design system. Tailwind never encourages you to design the same site twice since it is so low-level. Even with the same colour palette and scaling scale, it's simple to design the same component in the next project with an entirely different look. Tailwind automatically removes any superfluous CSS while developing for production, resulting in the smallest CSS bundle possible. In fact, the majority of Tailwind projects deliver less than 10kB of CSS to the client. Because wrestling with a slew of sophisticated media queries in your CSS is difficult, Tailwind allows you to construct responsive designs directly in your HTML. Put a screen size in front of any utility class and watch it apply at a certain breakpoint. Do you want to style something while hovering? Stick hover: at the start of the class you'd like to add. Works for focus, active, disabled, focus-within, focus-visible, and even our own fancy states like group-hover. If you're constantly repeating the same utilities, all you have to do is extract them into a component or template partial, and presto — you've got a single source of truth where you can make changes in one spot. If you prefer to go old school, you can utilise Tailwind's @apply directive to extract recurring utility patterns into custom CSS classes simply by copying and pasting the list of class names. Don't want to be one of those websites that causes people to go blind when they open it on their phone at 2 a.m.? Enable dark mode in your configuration file, then append dark: to any colour utility to use it when dark mode is enabled. This method is applicable to background colours, text colours, border colours, and even gradients. Tailwind comes with a carefully built set of defaults, but virtually everything may be modified, from the colour palette to the spacing scale to the box shadows to the mouse cursor. Create your own design system in the tailwind.config.js file, then let Tailwind translate it into your own unique CSS framework.

## JQuery

jQuery is a lightweight, feature-rich JavaScript library. It simplifies HTML page navigation and manipulation, event handling, animation, and Ajax with an easy-to-use API that works across a wide range of browsers. jQuery has altered the way millions of people write JavaScript by combining versatility and extensibility.

## Lodash

Lodash is a modern JavaScript utility library that provides modularity, performance, and additional features. Lodash simplifies JavaScript by removing the tedium of dealing with arrays, numbers, objects, texts, and so on. Lodash's modular methods are excellent for iterating through arrays, objects, and strings, manipulating and testing data, and creating composite functions.

# Project Overview

---

## Dataset

- The dataset for the system is the election related downloaded tweets.
- For Sentiment analysis positive and negative thesaurus will be provided.

The system will generate trend analysis and sentiment analysis. The tweets downloaded from twitter database are processed with help of user defined functions. These filtered tweets will be used as input for various analysis modules generating the trend or sentiment analysis.

## Web App

The web app allows for easy collection and visualisation of tweets, analysis of the sentiment of tweets, and the sentiment of a user's tweets with a user-friendly graphical interface.

The tweets are fetched from the Twitter API on a hashtag or user basis. There are three main categories of tweets that are analysed: Tweets coming from hashtags containing the political party names, tweets from the political leaders, and tweets that contain the hashtags for a particular political event or controversy. The tweets are fetched using Tweepy, and then passed over to the React frontend for visualisation, where they are presented in a list showing the tweet text, the tweet's author, the tweet's posting date and time, and the number of likes and retweets the tweet has.

Then they can be easily sent over for sentiment analysis or trend analysis using two respective buttons.

## Trend Analysis

The trend analysis is done using the Tweepy module, that fetches the number of tweets made about the topics in last 24 hours by using a Cursor object and traversing through the tweets. Since Cursor arranges the tweets chronologically, as soon as we hit the 24 hour mark, we have collected all the tweets made in the last 24 hours. It then uses Matplotlib and Seaborn to plot the data. The data is then returned to the frontend for visualisation.

## Sentiment Analysis

The sentiment analysis is done using the Sentiment Analysis module, which uses the TensorFlow library to analyse the sentiment of the tweets. It is trained on the IMDB Sentiment Database, which is a large corpus of movie reviews to serve as a thesaurus for sentiment-defining keywords and traits.

The analysed tweets are then sent back to the frontend, where the user can view the total number of positive and negative tweets made. The user can optionally view sentiment of each tweet fetched in a list containing cards consisting of Tweet text and Sentiment of the Tweet alongside the Tweet author, Tweet date and time and number of Upvotes and Retweets.

## System Configuration

---

The models were trained and tested on three systems with the following choice of hardware:

### **macOS 12.1 Monterey**

MacBook Pro (16-inch, 2019)

Intel Core i9 9980HK @ 2.9 GHz

AMD Radeon Pro 5500XT

Intel UHD Graphics 630

16 GB 2667 MHz DDR4 Dual-Channel RAM

### **Windows 11 Pro Build 22509**

Asus ROG Strix Scar 3

Intel Core i7-9750H @ 2.6 GHz

NVIDIA GeForce GTX 1660 Ti 6GB

Intel UHD Graphics 630

16 GB 2800 MHz DDR4 Dual-Channel RAM

**Windows 10 Pro 21H2**

Custom Built

Intel Core i7 6800K @ 2.7 GHz

Nvidia GeForce GTX 1080 8GB

32 GB 3000 MHz DDR4 Dual-Channel RAM

## Results

---

### Trend Analysis

#### Parties

Party	Number of Tweets
BJP	5,879
Congress	3,221
NCP	2,791
Shivsena	4,111
Samajwadi Party	1,277
BSP	821

#### Leaders

Leader	Number of Tweets
Narendra Modi	13
Rahul Gandhi	12
Akhilesh Yadav	3
Arvind Kejriwal	1
Sonia Gandhi	0

#### Including Mentions:

Leader	Number of Tweets
Narendra Modi	2348
Rahul Gandhi	1231
Akhilesh Yadav	122
Arvind Kejriwal	1324
Sonia Gandhi	867



## Controversies

Party	Number of Tweets
BJP	5,879
Congress	3,221
NCP	2,791
Shivsena	4,111
Samajwadi Party	1,277
BSP	821

Controversy	Number of Tweets
kisanandolan	32213
caa	182
farmlaw	12982
covid	2188932
vaccine	5621278

## Sentiment Analysis

### Parties

Party	Number of Tweets	Positive	Negative
BJP	5,879	1133	4746
Congress	3,221	2119	1102
NCP	2,791	2133	658
Shivsena	4,111	2486	1625
Samajwadi Party	1,277	998	279
BSP	821	233	588

### Leaders

Leader	Number of Tweets	Positive	Negative
Narendra Modi	13	12	1
Rahul Gandhi	12	9	3
Akhilesh Yadav	3	2	1
Arvind Kejriwal	1	0	1

Leader	Number of Tweets	Positive	Negative
Sonia Gandhi	0	0	0

#### Including Mentions:

Leader	Number of Tweets	Positive	Negative
Narendra Modi	2348	1291	1057
Rahul Gandhi	1231	797	434
Akhilesh Yadav	122	72	50
Arvind Kejriwal	1324	710	614
Sonia Gandhi	867	124	743

#### Controversies

Controversy	Number of Tweets	Positive	Negative
kisanandolan	32213	2122	30091
caa	182	57	125
farmlaw	12982	288	12694
covid	2188932	848771	1340161
vaccine	5621278	54,92,059	129219

## Conclusions

---

We created a web app that streams live tweets related to indian election and politics, and analyses them using the Sentiment Analysis module. The app allows the user to view the total number of positive and negative tweets made, as well as the sentiment of each tweet fetched in a list containing cards consisting of Tweet text and Sentiment of the Tweet alongside the Tweet author, Tweet date and time and number of Upvotes and Retweets. The user can also view the total number of positive and negative tweets made for each party, leader and controversy.

We also looked at the different technologies and tools used to build the app, and how they were used alongside each other to solve a problem. We also looked at the different systems used to train and test the models, and how they were used to analyse the tweets.

## References

---

**Tweepy Docs:** <https://tweepy.readthedocs.io/en/>

**TensorFlow:** <https://www.tensorflow.org/>

**IMDB Dataset:** <https://www.kaggle.com/jamesmcguigan/imdb-dataset>

**Create-React-App** <https://create-react-app.dev/docs/getting-started/>

**TailwindCSS** <https://tailwindcss.com/docs/>

**Framer Motion** <https://framer.com/motion/docs/>