2023 4th International Conference for Emerging Technology (INCET)
Belgaum, India. May 26-28, 2023

Predicting The General Election 2024 Using ML And Data Analytics

1st Pradyumna Parida
School of Engineering & Technology
Sharda University
Greater Noida, India
parida.bapunu@gmail.com

2nd Sneha Sinha
School of Engineering & Technology
Sharda University
Greater Noida, India
sneha1002sinha@gmail.com

3rd Arun Prakash Agrawal
School of Engineering & Technology
Sharda University
Greater Noida, India
arunpragrawal@gmail.com

4th RaghuRaj Singh Yadav
School of Engineering & Technology
Sharda University
Greater Noida, India
raghuraj.yadav@sharda.ac.in

2023 4th International Conference for Emerging Technology (INCET) | 979-8-3503-3575-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/INCET57972.2023.10170638

***Abstract*— The world today is full of emerging technologies and one of the most rising fields is Artificial Intelligence. As a result of increased computation power, numerous new machine learning algorithms have emerged. We are working on experimental research which aims to predict the election results of the 2024 Indian General Elections/Lok Sabha Election which will we contested for 545 Lok Sabha Seats. To parametrize computations such as coalitions and swings on all seats, Data Analysis would be employed. The swing parameters are then calculated using machine learning techniques like Linear Regression, Naive Bayes, Random Forest, Time Series, etc. using prior election data specific to relevant seats. To reckon emotions important to the elections, a huge corpus of recent publications and/or twitter tweets can be used. We may be able to apply swings to the vote shares of each party in each constituency and produce viable projections with the necessary biases based on subjective data. Various parameters will be taken in account to thoroughly analyze the data to get the right swings and the mindset of the respondent to know what their needs are, reasons to switch and are they satisfied with the current candidate and whether their conditions are met or not et cetera. Our vision is to detect the popular trends to get the insights into existing potential voters. Determine voter's response to promote cross opportunities. Attract, Retain & Grow improving decision making & better drive decisions.**

Keywords: Election Forecasting, Machine Learning, Swings, Data Analysis, Linear Regression, Lok Sabha.

I. INTRODUCTION

A. General Elections 2024

General elections are due to be held in India to constitute the 18th Lok Sabha (the Lower House of the Indian Parliament), which consists of 543 seats elected by universal adult suffrage. To win a seat, a political party's candidate needs to get the numerically highest votes for that constituency. In order to form the government, a political party needs to gather at least 272 winning candidates. Failing to do so will require them to make coalitions with other parties.

The two major coalitions in the contest are NDA and UPA, with BJP and INC leading them respectively. Currently, the BJP-led NDA is the ruling party.

B. What are coalitions?

Coalitions simply means when political parties cooperate in elections in order to share power. There are two kinds of coalitions:

a) Post-poll alliance: If two or more political parties have arithmetically enough seats after the election to form a majority, it is called a post-poll alliance. This takes place in order to form the government when no party has absolute majority (272+) alone.

b) Pre-poll alliance: If two or more political parties choose to contest separate seats before the election, in order to avoid clashing of votes (crossvoting), it is called a pre-poll alliance. For example, if Party B and Party C contest separately to beat a more powerful Party A: their votes may be cancelled out by each other, and Party A may sweep by getting majorities in each seat. However, if Party B and Party C contest in different seats and convince their bases to vote for each other in them respective seats, it could prove very beneficial to both parties and could maximize their seats.

C. What are Swings?

The change in vote share from the previous election to the next by the political party is called the swing. It is denoted as a percentage value. On a seat, if a party had 100,000 votes in the first election, and 120,000 votes in the second election, it is counted as +20% swing. Similarly, a loss in vote share can be denoted as a negative percentage. It is generally seen the swings are uniform across seats for political par-ties throughout elections, especially in the same state.

D. Contestants for the 2024 General Election

Bhartiya Janata Party (BJP), the ruling party, is widely hailed as the favorite to attain the highest number of seats. Its coalition National Democratic Alliance (NDA) seeks to cross the 272 mark and form a government after the election polling. BJP's major opposition is the Indian National Congress (INC) which is part of a potentially larger United Progressive Alliance (UPA). In 2019, BJP secured 303 seats and INC secured a meagre 52 as illustrated in Fig 1. The seat-shares for each coalition are also shown as per Fig 2 and Fig 3.

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

Leader	Narendra Modi	Rahul Gandhi
Party	BJP	INC
Leader since	13 September 2013	16 December 2017
Leader's seat	Varanasi (*won*)	Wayanad (*won*) and Amethi (*lost*)
Last election	282 seats	44 seats
Seats won	303	52
Seat change	▲21	▲8
Popular vote	229,076,879	119,495,214
Percentage	37.36%	19.49%
Swing	▲6.36pp	▲0.18pp
Alliance	NDA	UPA
Alliance seats	353	91

Fig. 1. Results of the 2019 General Elections between BJP & INC, in which BJP formed government alone by securing 303 seats.

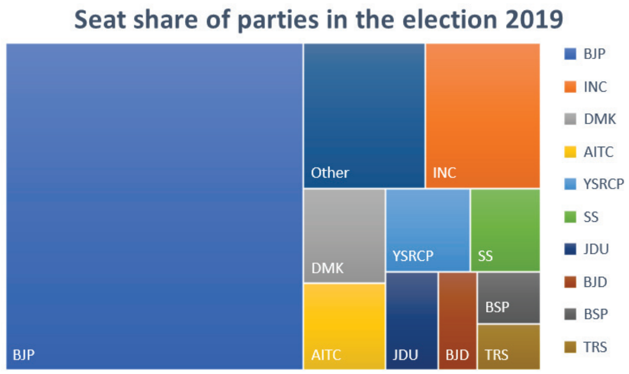


Fig. 2. Seat-shares of the 2019 General Elections between NDA, UPA and Others

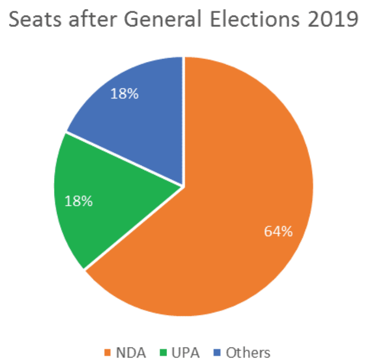


Fig. 3. Percentage seat-shares of the 2019 General Elections between NDA, UPA and Others

II. Literature Review

A. Literature Survey for General Election Forecasting

Facebook, Twitter, and other social media platforms have developed as digital elites with significant control over what is said in virtual worlds as a result of the recent rapid progress of digital technology. In the age of digital democracy, social perception and social influence in numerous complex social systems have quickly transformed. In this study, PL, Francia et al. carried out extensive empirical studies on social influence and social perception in relation to the Trump phenomenon from the views of individual experience, media attention, and general public interest. [1].

By creating networks of public attention using intricate networks based on Google search information, Francia et al. discovered that there are obvious links between Trump's social media activity and the attention of the mainstream media, demonstrating how digital platforms can significantly change social perception and social influence. Political events can always influence public attention. [1].

The management of complex social systems and the research of digital oligarchies should be pursued by more academics from various fields and locales. Cross-cultural research and theoretical modeling should be combined, and more scientific strategies should be offered to improve these complex social systems, in order to reveal the fundamental inequalities between nations and locales.

Every day, users read and publish short messages on a variety of topics on Twitter, a microblogging website. Many researchers are interested in political analysis using social media to better understand general public opinion and trends, particularly during election season. In this study, Singhal et al. presented a novel method for gauging public opinion and forecasting election outcomes using semantic and con-text-aware rules that crawled political tweets during India's general election. They then compared their suggested technique to the outcome of the election. The suggested guidelines are successful in assessing the sentiment of political tweets, ac-cording to experimental results [2].

One of the difficulties in analyzing political tweets is sarcasm, which tends to make the classifier less accurate. This example shows how sarcastic tweets, in which a positive attitude is followed by a tragic scenario, are handled. For a thorough analysis of the phrases, a dependency parsing tool that can identify connections between the words that make up the sentence should be used. Modifiers, Intensifiers, Divid-ers, Negations, Verbs, and Objects are the six entities that make up this category. We believe that these things are significant because they can have a considerable impact on the overall tone of the sentences.

Despite the importance of the term extraction methods and that several efforts have been devoted to improving them, they still have 4 main problems:

(i) noise and silence generation;

(ii) difficulty dealing with high number of terms;

(iii) human effort and time to evaluate the terms; and

(iv) still limited extraction results.

In this research, we explore a comprehensive feature set in a machine learning strategy to deal with these four fundamental issues in autonomous term extraction.

This study described some of the Twitter mining literature and provided several references to the Twitter mining literature that can be utilized to help with Twitter mining research. This research also investigated some of the methodology concerns that come up while mining Twitter, such as sentiment analysis. Furthermore, Mark et al. investigated a variety of various types of applications, some of which could be applied in other situations. Finally, the principles of touting were built upon in this study [3].

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

This paper could be taken in several different directions. First, data from other sources, such as Facebook, might be included into Twitter data mining. For example, for privacy legislation, O'Leary (2012) investigated some of the relationships be-tween the two data sources. Second, more Twitter-based applications that aren't included here due to scope could be investigated. Even though this work cites about 90 different sources, the literature is continually expanding. Third, other Twitter data applications are expected to emerge over time. Those applications could be studied in the future. Fourth, rather of focusing on Twitter alone, future research may look at Tumblr, Stocktwits, and other similar platforms. Finally, future studies could investigate some of mining's privacy implications [4].

New methodologies have become available as a result of the shortcomings of traditional survey techniques for assessing political climate and forecasting significant events like elections. Although computational linguistic methods have shown to be a practical alternative for social network analysis, several studies in these fields were conducted after the fact (post hoc). Three elections were held in Chile in 2017, the election year for the 2018–2022 period, which makes it the perfect setting for a case study on predicting these elections with social media as the main data source.

This paper describes the application of various supervised machine learning algorithms to conduct political sentiment analysis and predict the results of each election using Twitter data. These algorithms include Decision Trees, AdaBoost, Random Forest, Linear Support Vector Machines, and ensemble voting classifiers. To train the algorithms, experts manually annotate a training set to classify pragmatic attitude across tweets containing an account or a candidate's name. Then, a predictive set is gathered and an automatic categorization is carried out days before the election. The distribution of votes for each contender is then determined based on the tweets' positive sentiment.

Other labels could be used in the future to increase the effectiveness of predictive models. As a result, this gives a source of data that was not used in the current study. Another useful piece of information that can be gleaned from the complete tagged corpus is the usage of topic models to track the evolution of political discourse and Twitter user attitudes over time. Furthermore, while making predictions, it would be interesting to apply the topic model terms that were obtained and give them a higher weight.

Using current survey methods, spatially or temporally dense polling remains challenging and expensive. As a result, there have been more attempts to approximate various survey metrics using social media, however most of these approaches are still methodologically faulty.

Beauchamp et al. combined more than 100 million state-specific political tweets with 1,200 state-level polls from the 2012 presidential campaign in order to address these flaws. They also used a new linear regularization feature-selection method to modify the polls as a function of the Twitter text. Out-of-sample testing demonstrates that, when properly modified, Twitter-based measures track and somewhat predict opinion polls, can be expanded to un-polled states, and possibly even be used to The most foretelling textual components highlight the themes and occasions associated with opinion shifts, shed light on more general ideas of party variations in attention and information processing, and may be helpful for in-the-moment campaign planning. [6].

This study offers a machine learning model for forecasting Pakistan's general election results that is based on sentiment analysis. Voters choose their preferred party or candidate in a general election based on personal preferences. Social media had a significant role in Pakistan's general election campaign. We outline a five-step procedure that uses machine learning to assess the overall election results and deter-mine whether they were fair or unjust. The work comes to a close with a discussion of the sentiment analysis findings for general election predictions and approval in the actual world for Pakistan.

J. Prada provides an overview of the state-of-the-art in this particular field of study, concentrating its examination on the forecasting of the following five events: predictions of the results of political elections, criminal risk, disease transmission and syndromic surveillance, stock market movements, and National Football League (NFL) game results. [9].

Jaidka, Kokil, et al introduces and evaluates the robustness of different volumetric, sentiment, and social network approaches to predict the elections in three Asian countries – Malaysia, India, and Pakistan from Twitter posts. We find that predictive power of social media performs well for India and Pakistan but is not effective for Malaysia. Overall, we find that it is useful to consider the recency of Twitter posts while using it to predict a real outcome, such as an election result. Sentiment information mined using machine learning models was the most accurate predictor of election outcomes. Social network information is stable despite sudden surges in political discussions, for e.g., around elections-related news events. Methods combining sentiment and volume information, or sentiment and social network information, are effective at predicting smaller vote shares, for e.g., vote shares in the case of independent candidates and regional parties. We conclude with a detailed discussion on the caveats of social media analysis for predicting real-world out-comes and recommendations for future work [8].

Singh, Prabhsimran, et al takes into consideration the Twitter data related to the 2017 Punjab (a state of India) assembly elections and applies different social media analytic techniques on collected tweets to extract and unearth hidden but useful in-formation. In addition to this, we have employed machine learning algorithm to perform polarity analysis and have proposed a new seat forecasting method to accurately predict the number of seats that a political party is likely to win in the elections. Our results confirmed that Indian National Congress was likely to emerge winner and that in fact was the outcome, when results got declared [9].

Safiullah, Md, et al. states with the increase in popularity and growth in the use of social media, the present study aims at examining whether the use of social media (Twitter) influenced the 2014 General elections outcome. For this research, a total of 8,877,275 social media buzz for 100 days from January 01, 2014, to April 09, 2014, of 12 Indian political parties has been considered. The result indicates that social media buzz has a positive and significant impact on the outcome of General elections 2014.

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

III. Problem definition and objectives

A. Problem Definition

Elections have a crucial role in the functioning of a modern democracy. With a bar-rage of stakeholders at play, it is in the interest of many to have advance knowledge of political results. It also helps political parties strategize better and adopt to the public favorability.

B. Objectives

The objective of this research is to forecast the number of seats won by the two major alliances – NDA and UPA. In order to get a forecast, the results of the 2019 General Elections are taken as a base case, and data about vote shares in each constituency is tweaked with parameters such as updated coalitions and swings. Using data analysis libraries, the results are then computed. The swings can be forecasted using machine learning techniques based on data from sources such as past vote share or sentiments in corpuses of (social) media.

C. Challenges

The challenges faced in choosing such a research topic are numerous, mainly due to the unavailability of a proper data set. Indian government data is stored in a very outdated manner, such that it's hard to manipulate with modern frameworks. Also, news and social media has a level of volatility which is very difficult to overcome.

IV. Background

A. Data Analysis

Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains. In today's fields, data analysis is playing a role in making decisions more scientific and helping the business achieve effective operation [9]. Pandas is a Python library of rich data structures and tools for working with structured data sets common to a variety of fields. Structured data sets commonly arrive in tabular format, i.e., as a two-dimensional list of observations and names for the fields of each observation. Usually, an observation can be uniquely identified by one or more values or labels. Pandas gives us the capability to extract data from .csv files and generate data structures called Data Frames. These data structures help us handle complex data using simple variables. By doing this, pandas provide a solid foundation upon which a very powerful data analysis ecosystem can be established [10].

B. Machine Learning

Computer systems employ machine learning (ML), a scientific method and statistical model, to gradually improve their performance on a given task. Machine learning algorithms create a mathematical model using training data to make predictions or judgments without having to be explicitly programmed to do so. In applications like email filtering, network intrusion detection, and computer vision, when it is impossible to create an algorithm with precise instructions for carrying out the task, machine learning techniques are utilized. Computational statistics, which focuses on computer-assisted prediction and machine learning, are closely related [11]

1) Linear Regression

In statistics, a scalar response (or dependent variable) and one or more explanatory variables are modeled using a linear approach called linear regression (or independent variables). Simple linear regression is the situation where there is just one explanatory factor. Several linear regressions is the procedure used when there are multiple explanatory variables [12].

A predictive model can be fitted using linear regression to an observed data set of response and explanatory variable values. If extra explanatory variable values are gathered after creating such a model but without a corresponding response value, the fitted model can be used to predict the response.

Support vector machines (SVM) evaluate data used for classification and linear regression. They are supervised learning models with associated learning techniques. An SVM training method creates a model that classifies new examples into one of two categories based on a collection of training examples that have been identified as belonging to either one or the other of two categories [13].

Support Vector Regression (SVR) is a function estimation technique that uses the principles of SVM [14]. A multilayer perceptron (MLP) is a class of feedforward artificial neural network. An MLP consists of, at least, three layers of nodes: an in-put layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function.

MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable [15]. Sigmoid values are used to squeeze a range to a value between 0 and 1. If sigmoid nodes are used in the nodes of the neural network, it can be used for regression.

The numbers of model parameters that must be estimated as well as the degree of data unpredictability determine how well a time series model fits the data. The level of noise in the data and the number of parameters to be estimated, both affect the necessary sample size. Only if you have more observations than parameters can you estimate a model using least squares estimation or another non-regularized estimation technique. A fitted model may not, however, always be useful for forecasting, especially when the data is noisy [16].

2) Text Classification using Clustering

Text classification is an important foundation for Information Retrieval and Text Mining, the main task is assigning text document to one or more predefined categories according to its content and the labelled training samples [17].

KNN is a supervised learning predicted classification method and one of the most significant non-parametric algorithms in the field of pattern recognition. Without using any other information, the training samples alone build the classification rules for KNN. According to the K training samples that are the test sample's closest neighbors, the KNN classification algorithm guesses its category and assigns it to the category with the highest category probability [18].

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

3) Sentiment Analysis

The goal of sentiment analysis, a branch of natural language processing, is to extract both objective and subjective information from a sentence in natural language. Peo-ple are expressing their loves and dislikes about many topics on blogs, micro blogs, and social networking sites like Twitter and Facebook as the online community has grown. Short colloquial language can contain a wealth of information on human behavior that can be used in a variety of fields, including political science, opinion mining, and human computer interaction (HCI). Typically, a low score indicates a word's negative meaning, and a high score indicates a word's good connotation. Var-ious sets of words (e.g., for nouns, verbs, adjectives etc.) and grammatical rules are stored beforehand to assist in extracting the meaning of the tweet, and ultimately generate a sentiment score. Finally, the percentage of sentiment score for a political party can be considered as its corresponding vote share forecast [19].

V. PROPOSED MODEL

A. Description

The proposed model will be used to forecast the 2019 Lok Sabha Elections, relative to the 2014 elections. The coalitions are parametrized in the beginning, along with the expected swings in four types of seats – UPA or NDA which are incumbent or non-incumbent. Incumbent seats are those which were won by that party in the previous election. It is usually likelier that parties are going to lose votes in incumbent seats [3]. Once the coalitions are set, swings can be computed either heuristically or using ML techniques. After that, the forecasted seats would be outputted, and curated results would be displayed. Therefore, this model has two main components – the data analysis component which, using python data analysis libraries, is used to take in the coalitions and swings as parameters; and the swing forecast component using machine learning algorithms which can be estimated and passed as a parameter into the data analysis component.

B. Data Analysis Component

The data analysis component of the application will first import the election results of the previous election (2014) as base values. After that, the coalition participants of UPA and NDA will be passed on as parameters. Then swings would be ap-plied to each, and every constituency (nationally or state-wise) and coalitions votes would be counted together. These swings are going to be computed using the Machine Learning component of the system.

C. Machine Learning Component (Swing)

The machine learning component's task is to forecast the value of the swings, either nationally or for each state separately. To accomplish the task, machine learning could be applied in two ways – Regression or Sentiment Analysis.

1) Regression: The swing for a state could be determined by the vote shares of the past elections, and the patterns observed over time. For example, using past state and general elec-tion data (4-5 data points) we can deploy regression algorithms, and estimate swings state-wise. Based on these patterns, an approximate future vote share can be fore-casted for each party. A swing can thus be derived with respect to the latest vote shares. For example, the vote shares in Gujarat are illustrated Table 1. Further, a graph for these vote shares is shown in Fig 4.

TABLE I. RECENT VOTE SHARES IN GUJARAT NATIONAL AND STATE ELECTIONS

Election Held On	**INC**	**BJP**
2009 (General)	43.38%	46.52%
2012 (Assembly)	38.93%	47.85%
2014 (General)	33.45%	60.11%
2017 (Assembly)	41.40%	49.10%
Predicted Vote Share	**39%**	**50%**

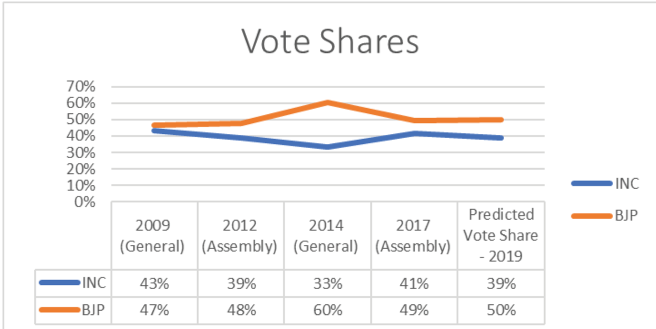


Fig. 4. Recent Vote Shares in Gujarat National and State Elections (Graph)

In the simplest case, we take average of these four vote shares to be the predicted vote share. Therefore, for BJP the vote share is approximately 50% and for INC it is 39%. Therefore, a swing of -10% and +6% are observed respectively, by subtracting the respective vote shares from the ones obtained in 2014 (since base election is 2014).

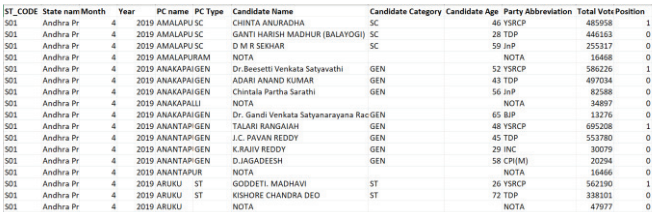
D. Sentiment Analysis

A dataset could be made by assigning tweets favorable to each political party and trained to form a classification model, that classifies a new tweet as favorable to a certain political party. With this model, it could take the tweet as input and output its favorability for a certain party. A corpus of, say, a million tweets could be extracted from the Twitter API, and then this model could be applied to each tweet. The percentage of tweets that are favorable to a certain political party would deter-mine their respective swings. Similarly, we can use a corpus of news articles or headlines to generate headlines that are positive or negative for certain political parties. Swings would be higher for those parties which have a higher percentage of positive headlines.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

A. Initial Data

The constituency-wise results of the 2014 General Election are considered as initial data. This is because



ST_CODE	State nam	Month	Year	PC name	PC Type	Candidate Name	Candidate Category	Candidate Age	Party Abbreviation	Total Vote	Position
S01	Andhra Pr	4	2019	AMALAPU	SC	CHINTA ANURADHA	SC	46	YSRCP	485958	1
S01	Andhra Pr	4	2019	AMALAPU	SC	GANTI HARISH MADHUR (BALAYOGI)	SC	28	TDP	446163	0
S01	Andhra Pr	4	2019	AMALAPU	SC	D M R SEKHAR	SC	59	JnP	255317	0
S01	Andhra Pr	4	2019	AMALAPURAM		NOTA			NOTA	16468	0
S01	Andhra Pr	4	2019	ANAKAPAI	GEN	Dr.Beesetti Venkata Satyavathi	GEN	52	YSRCP	586226	1
S01	Andhra Pr	4	2019	ANAKAPAI	GEN	ADARI ANAND KUMAR	GEN	43	TDP	497034	0
S01	Andhra Pr	4	2019	ANAKAPAI	GEN	Chintala Partha Sarathi	GEN	56	JnP	82588	0
S01	Andhra Pr	4	2019	ANAKAPALLI		NOTA			NOTA	34897	0
S01	Andhra Pr	4	2019	ANAKAPAI	GEN	Dr. Gandi Venkata Satyanarayana Rac	GEN	65	BJP	13276	0
S01	Andhra Pr	4	2019	ANANTAPI	GEN	TALARI RANGAIAH	GEN	48	YSRCP	695208	1
S01	Andhra Pr	4	2019	ANANTAPI	GEN	J.C. PAVAN REDDY	GEN	45	TDP	553780	0
S01	Andhra Pr	4	2019	ANANTAPI	GEN	K.RAJIV REDDY	GEN	29	INC	30079	0
S01	Andhra Pr	4	2019	ANANTAPI	GEN	D.JAGADEESH	GEN	58	CPI(M)	20294	0
S01	Andhra Pr	4	2019	ANANTAPUR		NOTA			NOTA	16466	0
S01	Andhra Pr	4	2019	ARUKU	ST	GODDETI. MADHAVI	ST	26	YSRCP	562190	1
S01	Andhra Pr	4	2019	ARUKU	ST	KISHORE CHANDRA DEO	ST	72	TDP	338101	0
S01	Andhra Pr	4	2019	ARUKU		NOTA			NOTA	47977	0

Fig. 5. Sample of data acquired from the results of 2019 General Elections

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

coalitions and swings are applied to each constituency individ-ually. For that, we must be cognizant about the results of each constituency.

It can be seen in Fig 5 that each constituency (identified by the combination of ‘State Name’ and ‘PC Name’) has various parties contesting, each of which are de-noted by a ‘Party Abbreviation’. Each party, correspondingly, has ‘Total Votes Polled’ and the respective positions of each party. Similarly, data of each constituency in the elections is stored in the .csv file.

B. Making Coalitions Using Data Analysis

We must utilize a Data Analysis library called Pandas to run various calculations on the.csv file. It is a high-performance, user-friendly, open-source library containing a selection of tools and data structures. As illustrated in Fig. 6 the system designed for the experiment requires that we pass an array of "Party Abbreviations" for each coalition (UPA and NDA) as parameters.

```
"""Coalitions in 2014"""
#NDACoalition = ["BJP","SHS","TDP","LJP","SAD","BLSP","SWP","AD","PMK","AINRC","NPF","NPEP"]
#UPACoalition = ["INC","NCP","RJD","IUML","JMM","KEC(M)","RSP"]

"""If BJP and INC were alone"""
NDACoalition = ["BJP"]
UPACoalition = ["INC"]

"""Expected Coalitions in 2019"""
#NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","PMK","NPEP","AINRC","ADMK", "AGP"]
#UPACoalition = ["INC", "AD", "RJD","NCP", "DMK","IUML","JD(S)","JKN","JMM","KEC(M)","RSP","BOPF"]

"""In percent. Incumbent swings are likelier to be lesser, as people want change."""
#ndaSwingIncumbent = 93
#ndaSwingNotIncumbent = 97
#upaSwingIncumbent = 103
#upaSwingNotIncumbent = 107

"""No swings"""
ndaSwingIncumbent = 100
ndaSwingNotIncumbent = 100
upaSwingIncumbent = 100
upaSwingNotIncumbent = 100
```

Fig. 6. Demonstration of parametrization of Swings and Coalitions in the system

We can choose the swings in addition to the relevant NDA and UPA coalitions. A "0 swing" is a swing of 100. (no change compared to previous election). A swing of 102 would be a +2% swing, whereas a swing of 97 would be a -3% swing. This method can be modified to give each seat a unique swing (not shown). However, for the sake of simplicity, we assume that the swings on the NDA incumbent and non-incumbent, as well as the UPA incumbent and non-incumbent, are distinct in this program. The seats that the party won in the primary are considered incumbent seats. Ordinarily, incumbent seats show a negative swing, while non-incumbent seats show a positive swing (due to change in voter mood).

After calculations on swings and alliances are completed, we can run the software to obtain detailed state-by-state figures on how many seats are retained or lost by each coalition.

```
FROM NDA SEATS (359):
Won by UPA: 67
Still Won by NDA: 292

TOTAL WON BY NDA IN 2014 REALITY
defaultdict(<class 'int'>, {'Andhra Pradesh': 3, 'Arunachal Pradesh': 1, 'Assam': 7, 'Bihar': 33, 'Goa': 2, 'Gujara
t': 26, 'Haryana': 7, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Karnataka': 17, 'Madhya Pradesh': 27, 'Maharashtr
a': 42, 'Meghalaya': 1, 'Nagaland': 1, 'Odisha': 1, 'Punjab': 6, 'Rajasthan': 25, 'Tamil Nadu': 39, 'Uttar Pradesh':
73, 'West Bengal': 2, 'Chattisgarh': 10, 'Jharkhand': 12, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandiga
rh': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})

WON BY UPA FROM NDA SEATS IN 2014 AFTER COALITION
defaultdict(<class 'int'>, {'Assam': 1, 'Bihar': 1, 'Karnataka': 2, 'Madhya Pradesh': 5, 'Maharashtra': 1, 'Punjab':
2, 'Uttar Pradesh': 50, 'West Bengal': 2, 'Chattisgarh': 2, 'Jharkhand': 1})

WON BY NDA FROM NDA SEATS IN 2014 AFTER COALITION
defaultdict(<class 'int'>, {'Andhra Pradesh': 3, 'Arunachal Pradesh': 1, 'Assam': 6, 'Bihar': 32, 'Goa': 2, 'Gujara
t': 26, 'Haryana': 7, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Karnataka': 15, 'Madhya Pradesh': 22, 'Maharashtr
a': 41, 'Meghalaya': 1, 'Nagaland': 1, 'Odisha': 1, 'Punjab': 4, 'Rajasthan': 25, 'Tamil Nadu': 39, 'Uttar Pradesh':
23, 'Chattisgarh': 8, 'Jharkhand': 11, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Na
gar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})
```

Fig. 7. A demonstration of the system's parametrization of swings and coalitions

```
TOTAL SEATS (486):
defaultdict(<class 'int'>, {'Andhra Pradesh': 21, 'Arunachal Pradesh': 2, 'Assam': 10, 'Bihar': 40, 'Haryana': 8, 'Ka
rnataka': 28, 'Kerala': 18, 'Madhya Pradesh': 29, 'Maharashtra': 48, 'Manipur': 2, 'Meghalaya': 2, 'Mizoram': 1, 'Pun
jab': 9, 'Tripura': 2, 'Uttar Pradesh': 80, 'West Bengal': 42, 'Chattisgarh': 11, 'Jharkhand': 14, 'Lakshadweep': 1,
'Goa': 2, 'Gujarat': 26, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Nagaland': 1, 'Odisha': 1, 'Rajasthan': 25, 'T
amil Nadu': 39, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Nagar Haveli': 1, 'Daman
& Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})

TOTAL NDA SEATS STATEWISE (FROM NDA AND UPA) ARE 296
defaultdict(<class 'int'>, {'Andhra Pradesh': 3, 'Arunachal Pradesh': 1, 'Assam': 6, 'Bihar': 36, 'Goa': 2, 'Gujara
t': 26, 'Haryana': 7, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Karnataka': 15, 'Madhya Pradesh': 22, 'Maharashtr
a': 41, 'Meghalaya': 1, 'Nagaland': 1, 'Odisha': 1, 'Punjab': 4, 'Rajasthan': 25, 'Tamil Nadu': 39, 'Uttar Pradesh':
23, 'Chattisgarh': 8, 'Jharkhand': 11, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Na
gar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})

TOTAL UPA SEATS STATEWISE (FROM UPA AND NDA) ARE 190
defaultdict(<class 'int'>, {'Andhra Pradesh': 18, 'Arunachal Pradesh': 1, 'Assam': 4, 'Bihar': 4, 'Haryana': 1, 'Karn
ataka': 13, 'Kerala': 18, 'Madhya Pradesh': 7, 'Maharashtra': 7, 'Manipur': 2, 'Meghalaya': 1, 'Mizoram': 1, 'Punja
b': 5, 'Tripura': 2, 'Uttar Pradesh': 57, 'West Bengal': 42, 'Chattisgarh': 3, 'Jharkhand': 3, 'Lakshadweep': 1})

NDA Vote Share: 42.304977260973686 | UPA Vote Share: 45.144515769852255 in NDA/UPA winning seats
```

Fig. 8. Detailed results showing the number of seats each party gained in each state, along with their respective vote percentages in key seats when the coalition was extended.

For instance, Fig. 7 shows each state-by-state seat that the NDA would lose from 2014 to 2019 and Fig. 8 shows each seat for both the NDA and the UPA.

1) Same 2014 coalitions, no swings

If the 2014 NDA and UPA coalitions remain in place and no adjustments are made to the 2014 outcomes. Figure 9 displays the coalition and swing parameters that are stored in the data analysis system.

```
"""Coalitions in 2014"""
NDACoalition = ["BJP","SHS","TDP","LJP","SAD","BLSP","SWP","AD","PMK","AINRC","NPF","NPEP"]
UPACoalition = ["INC","NCP","RJD","IUML","JMM","KEC(M)","RSP"]

"""No swings"""
ndaSwingIncumbent = 100
ndaSwingNotIncumbent = 100
upaSwingIncumbent = 100
upaSwingNotIncumbent = 100
```

Fig. 9. Parameters for same coalitions as 2014 with no swings

```
TOTAL SEATS (396):
defaultdict(<class 'int'>, {'Andhra Pradesh': 21, 'Arunachal Pradesh': 2, 'Assam': 10, 'Bihar': 38, 'Haryana': 8, 'Ka
rnataka': 26, 'Kerala': 12, 'Madhya Pradesh': 29, 'Maharashtra': 48, 'Manipur': 2, 'Meghalaya': 2, 'Mizoram': 1, 'Pun
jab': 9, 'Uttar Pradesh': 75, 'West Bengal': 6, 'Chattisgarh': 11, 'Jharkhand': 14, 'Lakshadweep': 1, 'Goa': 2, 'Guja
rat': 26, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Nagaland': 1, 'Odisha': 1, 'Rajasthan': 25, 'Tamil Nadu': 2,
'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT
OF Delhi': 7, 'Puducherry': 1})

TOTAL NDA SEATS STATEWISE (FROM NDA AND UPA) ARE 336
defaultdict(<class 'int'>, {'Andhra Pradesh': 19, 'Arunachal Pradesh': 1, 'Assam': 7, 'Bihar': 31, 'Goa': 2, 'Gujara
t': 26, 'Haryana': 7, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Karnataka': 17, 'Madhya Pradesh': 27, 'Maharashtr
a': 42, 'Meghalaya': 1, 'Nagaland': 1, 'Odisha': 1, 'Punjab': 6, 'Rajasthan': 25, 'Tamil Nadu': 2, 'Uttar Pradesh': 7
3, 'West Bengal': 2, 'Chattisgarh': 10, 'Jharkhand': 12, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigar
h': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})

TOTAL UPA SEATS STATEWISE (FROM UPA AND NDA) ARE 60
defaultdict(<class 'int'>, {'Andhra Pradesh': 2, 'Arunachal Pradesh': 1, 'Assam': 3, 'Bihar': 7, 'Haryana': 1, 'Karna
taka': 9, 'Kerala': 12, 'Madhya Pradesh': 2, 'Maharashtra': 6, 'Manipur': 2, 'Meghalaya': 1, 'Mizoram': 1, 'Punjab':
3, 'Uttar Pradesh': 2, 'West Bengal': 4, 'Chattisgarh': 1, 'Jharkhand': 2, 'Lakshadweep': 1})

NDA Vote Share: 45.96198756725868 | UPA Vote Share: 26.89944148735648 in NDA/UPA winning seats
```

Fig. 10. Results for same coalitions as 2014 and no swings

The fact that Fig. 10 properly predicts the number of seats won by the NDA and UPA as 336 and 60, respectively, shows that the data analysis system is functioning.

2) BJP & INC alone, no swings

If only the BJP and INC make up the NDA and UPA, respectively, and no swings are used, the 2014 results will stand as is. Its parameters are displayed in Fig. 11. Both parties would lose seats, as seen in Fig. 12, but the BJP still manages to form the government (at 273+) with 282 seats. INC is down to just 40 seats.

```
"""If BJP and INC were alone"""
NDACoalition = ["BJP"]
UPACoalition = ["INC"]

"""No swings"""
ndaSwingIncumbent = 100
ndaSwingNotIncumbent = 100
upaSwingIncumbent = 100
upaSwingNotIncumbent = 100
```

Fig. 11. Parameters for NDA and UPA having only BJP and INC alone respectively; with no swings

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

```
TOTAL SEATS (326):
defaultdict(<class 'int'>, {'Andhra Pradesh': 5, 'Arunachal Pradesh': 2, 'Assam': 10, 'Bihar': 24, 'Haryana': 8, 'Karnataka': 26, 'Kerala': 8, 'Madhya Pradesh': 29, 'Maharashtra': 25, 'Manipur': 2, 'Meghalaya': 1, 'Mizoram': 1, 'Punjab': 5, 'Uttar Pradesh': 73, 'West Bengal': 6, 'Chattisgarh': 11, 'Goa': 2, 'Gujarat': 26, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Odisha': 1, 'Rajasthan': 25, 'Tamil Nadu': 1, 'Jharkhand': 12, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7})

TOTAL NDA SEATS STATEWISE (FROM NDA AND UPA) ARE 282
defaultdict(<class 'int'>, {'Andhra Pradesh': 3, 'Arunachal Pradesh': 1, 'Assam': 7, 'Bihar': 22, 'Goa': 2, 'Gujarat': 26, 'Haryana': 7, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Karnataka': 17, 'Madhya Pradesh': 27, 'Maharashtra': 23, 'Odisha': 1, 'Punjab': 2, 'Rajasthan': 25, 'Tamil Nadu': 1, 'Uttar Pradesh': 71, 'West Bengal': 2, 'Chattisgarh': 10, 'Jharkhand': 12, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7})

TOTAL UPA SEATS STATEWISE (FROM UPA AND NDA) ARE 44
defaultdict(<class 'int'>, {'Andhra Pradesh': 2, 'Arunachal Pradesh': 1, 'Assam': 3, 'Bihar': 2, 'Haryana': 1, 'Karnataka': 9, 'Kerala': 8, 'Madhya Pradesh': 2, 'Maharashtra': 2, 'Manipur': 2, 'Meghalaya': 1, 'Mizoram': 1, 'Punjab': 3, 'Uttar Pradesh': 2, 'West Bengal': 4, 'Chattisgarh': 1})

NDA Vote Share: 45.9990093118977 | UPA Vote Share: 24.629035160137658 in NDA/UPA winning seats
```

Fig. 12. Results for BJP and INC alone with no swings

3) Extended 2019 coalitions, no swings

The NDA would still form government if it were to join the UPA and NDA in their expanded revised coalitions (Fig. 13) and no swings were used. However, UPA would increase its seat total from 60 to 190 by a huge margin (Fig 14). This demonstrates the importance of powerful coalitions, particularly for the UPA.

```
"""Expected Coalitions in 2019"""
NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","PMK","NPEP","AINRC","ADMK", "AGP"]
UPACoalition = ["INC", "AD", "RJD","NCP", "DMK","IUML","JD(S)","JKN","JMM","KEC(M)","RSP","BOPF"]

"""No swings"""
ndaSwingIncumbent = 100
ndaSwingNotIncumbent = 100
upaSwingIncumbent = 100
upaSwingNotIncumbent = 100
```

Fig. 13. Parameters for NDA and UPA being in their respective extended coalitions; with no swings.

```
TOTAL SEATS (486):
defaultdict(<class 'int'>, {'Andhra Pradesh': 21, 'Arunachal Pradesh': 2, 'Assam': 10, 'Bihar': 40, 'Haryana': 8, 'Karnataka': 28, 'Kerala': 18, 'Madhya Pradesh': 29, 'Maharashtra': 48, 'Manipur': 2, 'Meghalaya': 2, 'Mizoram': 1, 'Punjab': 9, 'Tripura': 2, 'Uttar Pradesh': 80, 'West Bengal': 42, 'Chattisgarh': 11, 'Jharkhand': 14, 'Lakshadweep': 1, 'Goa': 2, 'Gujarat': 26, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Nagaland': 1, 'Odisha': 1, 'Rajasthan': 25, 'Tamil Nadu': 39, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})

TOTAL NDA SEATS STATEWISE (FROM NDA AND UPA) ARE 296
defaultdict(<class 'int'>, {'Andhra Pradesh': 3, 'Arunachal Pradesh': 1, 'Assam': 6, 'Bihar': 36, 'Goa': 2, 'Gujarat': 26, 'Haryana': 7, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Karnataka': 15, 'Madhya Pradesh': 22, 'Maharashtra': 41, 'Meghalaya': 1, 'Nagaland': 1, 'Odisha': 1, 'Punjab': 4, 'Rajasthan': 25, 'Tamil Nadu': 39, 'Uttar Pradesh': 23, 'Chattisgarh': 8, 'Jharkhand': 11, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})

TOTAL UPA SEATS STATEWISE (FROM UPA AND NDA) ARE 190
defaultdict(<class 'int'>, {'Andhra Pradesh': 18, 'Arunachal Pradesh': 1, 'Assam': 4, 'Bihar': 4, 'Haryana': 1, 'Karnataka': 13, 'Kerala': 18, 'Madhya Pradesh': 7, 'Maharashtra': 7, 'Manipur': 2, 'Meghalaya': 1, 'Mizoram': 1, 'Punjab': 5, 'Tripura': 2, 'Uttar Pradesh': 57, 'West Bengal': 42, 'Chattisgarh': 3, 'Jharkhand': 3, 'Lakshadweep': 1})

NDA Vote Share: 42.304977260973686 | UPA Vote Share: 45.144515769852255 in NDA/UPA winning seats
```

Fig. 14. Results for extended coalitions of NDA and UPA; with no swings

4) Extended 2019 coalitions, arbitary swings

For the purposes of illustration, we can heuristically estimate some possible swings from prior elections, which are displayed in Table II. We assume that NDA will gain 2% votes and suffer a 10% vote loss in each incumbent seat. Additionally, UPA will win 7% votes and lose 5% in each existing seat.

TABLE II. ARBITRARY SWINGS TO BE APPLIED FOR NDA AND UPA

	Incumbent	Non-incumbent
NDA	-11% swing (93)	+1% swing (103)
UPA	-3% swing (95)	+5% swing (105)

```
"""Expected Coalitions in 2019"""
NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","PMK","NPEP","AINRC","ADMK", "AGP"]
UPACoalition = ["INC", "AD", "RJD","NCP", "DMK","IUML","JD(S)","JKN","JMM","KEC(M)","RSP","BOPF"]

ndaSwingIncumbent = 89.0
ndaSwingNotIncumbent = 101.0
upaSwingIncumbent = 97.0
upaSwingNotIncumbent = 105.0
```

Fig. 15. Parameters for NDA and UPA being in their respective extended coalitions; with arbitrary swings.

```
TOTAL SEATS (486):
defaultdict(<class 'int'>, {'Andhra Pradesh': 21, 'Arunachal Pradesh': 2, 'Assam': 10, 'Bihar': 40, 'Haryana': 8, 'Karnataka': 28, 'Kerala': 18, 'Madhya Pradesh': 29, 'Maharashtra': 48, 'Manipur': 2, 'Meghalaya': 2, 'Mizoram': 1, 'Punjab': 9, 'Tripura': 2, 'Uttar Pradesh': 80, 'West Bengal': 42, 'Chattisgarh': 11, 'Jharkhand': 14, 'Lakshadweep': 1, 'Goa': 2, 'Gujarat': 26, 'Himachal Pradesh': 4, 'Jammu & Kashmir': 3, 'Nagaland': 1, 'Odisha': 1, 'Rajasthan': 25, 'Tamil Nadu': 39, 'Uttarakhand': 5, 'Andaman & Nicobar Islands': 1, 'Chandigarh': 1, 'Dadra & Nagar Haveli': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1})

TOTAL NDA SEATS STATEWISE (FROM NDA AND UPA) ARE 242
defaultdict(<class 'int'>, {'Andhra Pradesh': 3, 'Assam': 4, 'Bihar': 35, 'Goa': 1, 'Gujarat': 24, 'Haryana': 5, 'Himachal Pradesh': 3, 'Jammu & Kashmir': 1, 'Karnataka': 9, 'Madhya Pradesh': 21, 'Maharashtra': 37, 'Meghalaya': 1, 'Nagaland': 1, 'Odisha': 1, 'Punjab': 2, 'Rajasthan': 22, 'Tamil Nadu': 31, 'Uttar Pradesh': 10, 'Chattisgarh': 7, 'Jharkhand': 9, 'Uttarakhand': 4, 'Chandigarh': 1, 'Daman & Diu': 1, 'NCT OF Delhi': 7, 'Puducherry': 1, 'Manipur': 1})

TOTAL UPA SEATS STATEWISE (FROM UPA AND NDA) ARE 244
defaultdict(<class 'int'>, {'Andhra Pradesh': 18, 'Arunachal Pradesh': 2, 'Assam': 6, 'Bihar': 5, 'Haryana': 3, 'Karnataka': 19, 'Kerala': 18, 'Madhya Pradesh': 8, 'Maharashtra': 11, 'Manipur': 1, 'Meghalaya': 1, 'Mizoram': 1, 'Punjab': 7, 'Tripura': 2, 'Uttar Pradesh': 70, 'West Bengal': 42, 'Jharkhand': 5, 'Lakshadweep': 1, 'Goa': 1, 'Gujarat': 2, 'Himachal Pradesh': 1, 'Jammu & Kashmir': 2, 'Rajasthan': 3, 'Tamil Nadu': 8, 'Chattisgarh': 4, 'Uttarakhand': 1, 'Andaman & Nicobar Islands': 1, 'Dadra & Nagar Haveli': 1})

NDA Vote Share: 38.6931666619056 | UPA Vote Share: 46.125338850091694 in NDA/UPA winning seats
```

Fig. 16. Results for extended coalitions of NDA and UPA; with arbitrary swings

According to Fig. 16, the UPA competes strongly against the NDA and barely manages to win the most seats. If we can correctly map the relevant swings to each district using this model, we can predict the 2019 General Election results based on the 2014 Elections with accuracy.

C. Determination of Swings using Machine Learning

We can utilize WEKA's Machine Learning Forecast capabilities to identify swings based on historical data. Given that different states have vastly varied sentiments toward a political party, it seems sense to calculate the swings of each state separate-ly. Therefore, it will be illustrated in the example below how future swings might be anticipated for Rajasthan using the linear regression algorithm. The vote shares for both parties are based on information from five previous General and Assembly (G/A) elections (Table III). The result is a.arff dataset, as seen in Fig. 17

TABLE III. TABLE 6.2 VOTE SHARES IN RAJASTHAN IN GENERAL (G) AND ASSEMBLY (A) ELECTIONS

Rajasthan	Vote Shares	
Election Held On	**INC**	**BJP**
2008 (Assembly)	**36.82%**	**34.27%**
2009 (General)	**47.19%**	**36.57%**
2013 (Assembly)	**33.07%**	**45.17%**
2014 (General)	**30.40%**	**50.90%**
2018 (Assembly)	**39.30%**	**38.80%**
2019 (General)	**38.80%**	**43.74%**

```
@relation Rajasthan

@attribute year date 'yyyy'
@attribute INCshare numeric
@attribute BJPshare numeric

@data
2008,    36.82,   34.27
2009,    47.19,   36.57
2013,    33.07,   45.17
2014,    30.40,   50.90
2018,    39.30,   38.80
```

Fig. 17. arff dataset for Vote Shares in Rajasthan to be used in WEKA

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

```
=== Future predictions from end of training data ===
Time  INCshare BJPshare
2008     36.82    34.27
2009     47.19    36.57
2010     43.66    38.72
2011     40.13    40.87
2012      36.6    43.02
2013     33.07    45.17
2014      30.4     50.9
2015    32.625   47.875
2016     34.85    44.85
2017    37.075   41.825
2018      39.3     38.8
2019*  40.5839  39.6074
```

Fig. 18. Results after using Linear Regression Algorithm on yearly basis for Rajasthan

The swings for 2019 in relation to 2014 would therefore be the difference — 10% for UPA (INC) and -11% for NDA, as seen in Fig. 18. (BJP). Using our data analysis model and these swings, we may predict the number of seats in Rajasthan as follows:

```
"""Expected Coalitions in 2019"""
NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","PMK","NPEP","AINRC","ADMK", "AGP"]
UPACoalition = ["INC", "AD", "RJD","NCP", "DMK","IUML","JD(S)","JKN","JMM","KEC(M)","RSP","BOPF"]

ndaSwingIncumbent = 89.0
ndaSwingNotIncumbent = 88.0
upaSwingIncumbent = 110.0
upaSwingNotIncumbent = 110.0

stateName = "Rajasthan"
```

Fig. 19. Parameters for extended coalitions of NDA and UPA; Linear Regression swings applied for 2019.

```
TOTAL SEATS (25):
defaultdict(<class 'int'>, {'Rajasthan': 25})

TOTAL NDA SEATS STATEWISE (FROM NDA AND UPA) ARE 21
defaultdict(<class 'int'>, {'Rajasthan': 21})

TOTAL UPA SEATS STATEWISE (FROM UPA AND NDA) ARE 4
defaultdict(<class 'int'>, {'Rajasthan': 4})

NDA Vote Share: 50.0377600796027 | UPA Vote Share: 36.81297145231752 in NDA/UPA winning seats
```

Fig. 20. Results of seat-share in Rajasthan after applying Linear Regression swings.

Applying these swings (Fig. 19) to our data analysis model, we discovered that the NDA wins 21 of the 25 seats in Rajasthan while the UPA only receives 4 of the 25. (Fig 20). As an alternative, we compute the findings depicted in Fig. 21 using the Multilayer Perceptron Algorithm (with Sigmoid Function).

```
=== Future predictions from end of training data ===
Time  INCshare BJPshare
2008     36.82    34.27
2009     47.19    36.57
2010     43.66    38.72
2011     40.13    40.87
2012      36.6    43.02
2013     33.07    45.17
2014      30.4     50.9
2015    32.625   47.875
2016     34.85    44.85
2017    37.075   41.825
2018      39.3     38.8
2019*  44.2391  33.4776
2020*  53.1935  25.8563
```

Fig. 21. Results after using Multilayer Perceptron Algorithm on yearly basis for Rajasthan

Therefore, the swings for 2019 in comparison to 2014 would be +14% for UPA (INC) and -17% for NDA (BJP).

```
"""Expected Coalitions in 2019"""
NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","PMK","NPEP","AINRC","ADMK", "AGP"]
UPACoalition = ["INC", "AD", "RJD","NCP", "DMK","IUML","JD(S)","JKN","JMM","KEC(M)","RSP","BOPF"]

ndaSwingIncumbent = 83.0
ndaSwingNotIncumbent = 83.0
upaSwingIncumbent = 114.0
upaSwingNotIncumbent = 114.0

stateName = "Rajasthan"
```

Fig. 22. Parameters for extended coalitions of NDA and UPA; Multilayer Perceptron Algorithm swings applied for 2019.

On applying these swings (Fig 22) to our data analysis model, we observe that in Rajasthan, NDA gets 19/25 seats and UPA gets 6/25 seats (Fig 23).

```
TOTAL SEATS (25):
defaultdict(<class 'int'>, {'Rajasthan': 25})

TOTAL NDA SEATS STATEWISE (FROM NDA AND UPA) ARE 19
defaultdict(<class 'int'>, {'Rajasthan': 19})

TOTAL UPA SEATS STATEWISE (FROM UPA AND NDA) ARE 6
defaultdict(<class 'int'>, {'Rajasthan': 6})

NDA Vote Share: 46.66442793940476 | UPA Vote Share: 38.151624959674514 in NDA/UPA winning seats
```

Fig. 23. Results of seat-share in Rajasthan after Multi-layer Perceptron swings applied.

D. *Curated Results*

A recapitulated table of results previously obtained is shown in Table IV and Table V

TABLE IV. NDA AND UPA CURATED SEAT AND VOTE SHARE IN VARIOUS SCENARIOS

	NDA		UPA	
	Seats	**Vote Share**	**Seats**	**Vote Share**
Same coalitions as 2014; no swings (actual results)	**336**	**45%**	**60**	**26%**
BJP and INC alone; no swings	**282**	**45%**	**44**	**24%**
Extended coalitions; no swings	**296**	**42%**	**190**	**45%**
Extended coalitions; arbitrary swings	**242**	**38%**	**244**	**46%**

TABLE V. RAJASTHAN'S BJP AND INC CURATED SEATS AND SWINGS AFTER APPLYING MACHINE LEARNING ALGORITHMS

	BJP		INC	
	Seats	**Swings**	**Seats**	**Swings**
No Swing	25	0%	0	0%
Linear Regression Swing	21	11%	4	10%
Multi-layer Perceptron Swing	19	17%	6	14%

VII. CONCLUSION

Even while it can be very challenging and requires painstaking detail, it is nevertheless possible to predict something as complex as the Lok Sabha elections with the correct methodology and information.

Additionally, it would demonstrate how important subjects like artificial intelligence and data science are for political scientists and other interested parties to do research in order to draw relevant conclusions from data that is generally disorganized.

It may be possible to forecast elections with respectable accuracy by using the data from the previous election as a benchmark for the following one and then applying appropriate parametrized algorithms to it.

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.

References

[1] Francia, Peter L. "Free media and Twitter in the 2016 presidential election: The unconventional campaign of Donald Trump." Social Science Computer Review 36.4 (2018): 440-455.

[2] Singhal, Kartik, Basant Agrawal, and Namita Mittal. "Modeling Indian general elections: sentiment analysis of political Twitter data." Information Systems Design and Intelligent Applications. Springer, New Delhi, 2015. 469-477.

[3] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

[4] O'Leary, Daniel E. "Twitter mining for discovery, prediction and causality: Applications and methodologies." Intelligent Systems in Accounting, Finance and Management 22.3 (2015): 227-247.

[5] Rodríguez, Sebastián, et al. "Forecasting the Chilean electoral year: Using twitter to predict the presidential elections of 2017." International Conference on Social Computing and Social Media. Springer, Cham, 2018.

[6] Beauchamp, Nicholas. "Predicting and interpolating state ‐ level polls using Twitter textual data." American Journal of Political Science 61.2 (2017): 490-503.

[7] Ali, Haider, et al. "Deep learning-based election results prediction using Twitter activity." Soft Computing (2021): 1-9.

[8] Hasan, Ali, et al. "Machine learning-based sentiment analysis for twitter accounts." Mathematical and Computational Applications 23.1 (2018): 11.

[9] J. Prada, "Predicting with Twitter," in Proc. Eur. Conf. Social Media, 2015, p. 10.

[10] F. J. J. Joseph, "Twitter Based Outcome Predictions of 2019 Indian General Elections Using Decision Tree," 2019 4th International Conference on Information Technology (InCIT), 2019, pp. 50-53, doi: 10.1109/INCIT.2019.8911975.

[11] V. Kagan, A. Stevens, and V. S. Subrahmanian, "Using twitter sentiment to forecast the 2013 pakistani election and the 2014 indian election," IEEE Intell. Syst., vol. 30, no. 1, pp. 2–5, 2015.

[12] G. Jagdev and A. Kaur, "Analyzing and scripting indian election strategies using big data via Apache Hadoop framework," 2016 5th International Conference on Wireless Networks and Embedded Systems (WECON), 2016, pp. 1-9, doi: 10.1109/WECON.2016.7993431.

[13] Zhou, Z., Serafino, M., Cohan, L. et al. Why polls fail to predict elections. J Big Data 8, 137 (2021)

[14] Kennedy C, Hartig H. Response rates in telephone surveys have resumed their decline. Pew Research Center ;2019.

[15] Ravi, Kumar, and Vadlamani Ravi. "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-based systems 89 (2015): 14-46.

[16] Jaidka, Kokil, et al. "Predicting elections from social media: a three-country, three-method comparative study." Asian Journal of Communication 29.3 (2019): 252-273.

[17] Singh, Prabhsimran, et al. "Can twitter analytics predict election outcome? An insight from 2017 Punjab assembly elections." Government Information Quarterly 37.2 (2020): 101444.

[18] Safiullah, Md, et al. "Social media as an upcoming tool for political marketing effectiveness." Asia Pacific Management Review 22.1 (2017): 10-15.

[19] Ceron, Andrea, Luigi Curini, and Stefano M. Iacus. "Using sentiment analysis to monitor electoral campaigns: Method matters—evidence from the United States and Italy." Social Science Computer Review 33.1 (2015): 3-20.

Authorized licensed use limited to: K K Wagh Inst of Engg Education and Research. Downloaded on September 30,2023 at 07:35:42 UTC from IEEE Xplore. Restrictions apply.