Forecasting the 2019 Lok Sabha Election using Machine Learning and Data Analysis

UE20CS312
Data Analytics

Team Mates :

Name : Adarsh Subhas Nayak
SRN : PES1UG20CS620

Name : Harsh
SRN : PES1UG20EC078

ABSTRACT

With the advent of increased computational power, various new forecasting techniques have emerged for a plethora of applications. One such area is the forecasting of competitive elections, which are the hallmark of modern democracy, and being able to foreshadow who wins the elections is a tantalizing skill that has garnered significant scientific attention [1].

This seminar seeks to predict the election results of the General Election, 2019 in India; which will be contested for 545 seats in the Lok Sabha.

Major focus would be given to a potential extended alliances that may be the deciding factor in the contest.

Data Analysis would be used to parametrize computations such as coalitions and swings on all the seats. After that, Machine Learning algorithms such as Linear Regression can be utilized to calculate the aforementioned swing parameters using past election data specific to the relevant seats.

A large corpus of recent articles and/or twitter tweets can be used to calculate moods relevant to the elections. With the appropriate biases based on subjective data, we may be able to apply swings to the vote shares of each party in each constituency, and obtain viable forecasts.

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 GENERAL ELECTIONS 2019

General elections are due to be held in India to constitute the 17th Lok Sabha -- the lower house of the Indian parliament, which consists of 545 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 273 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, but appropriate coalitions by UPA may spell danger for the former in 2019.

1.2 WHAT ARE COALITIONS?

Coalitions simply means when political parties co-operate 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 (273+) 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 (cross-voting), 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 their respective seats, it could prove very beneficial to both parties and could maximize their seats.

1.3 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 parties throughout elections, especially in the same state.

1.4 CONTESTANTS FOR THE 2019 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 273 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 2014, BJP secured 282 seats and INC secured a meagre 44 as illustrated in Fig 1.1. The seat-shares for each coalition are also shown in Fig 1.2 and Fig 1.3.

Party	BJP	INC
Alliance	NDA	UPA
Leader since	10 June 2013	May 2004
Leader's seat	Varanasi Vadodara *(Vacated)*	Amethi
Last election	116	206
Seats won	282[1][3]	44[1][3]
Seat change	▲166	▼162
Popular vote	171,660,230	106,935,942
Percentage	31.34%[4]	19.52%[4]
Swing	▲12.5%	▼9.03%

Fig. 1.1 Results of the 2014 General Elections between BJP & INC, in which BJP formed government alone by securing 282 seats.

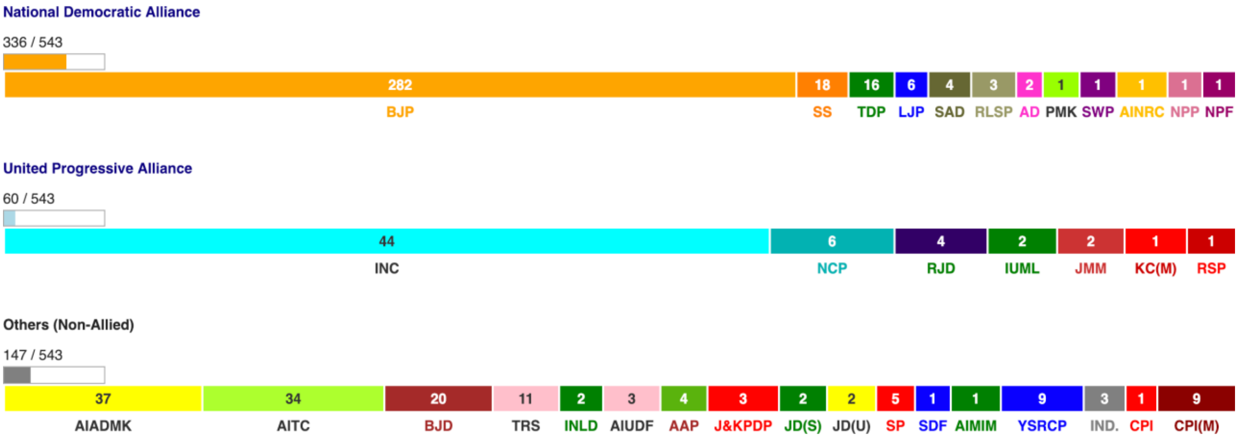

Fig. 1.2 Seat-shares of the 2014 General Elections between NDA, UPA and Others

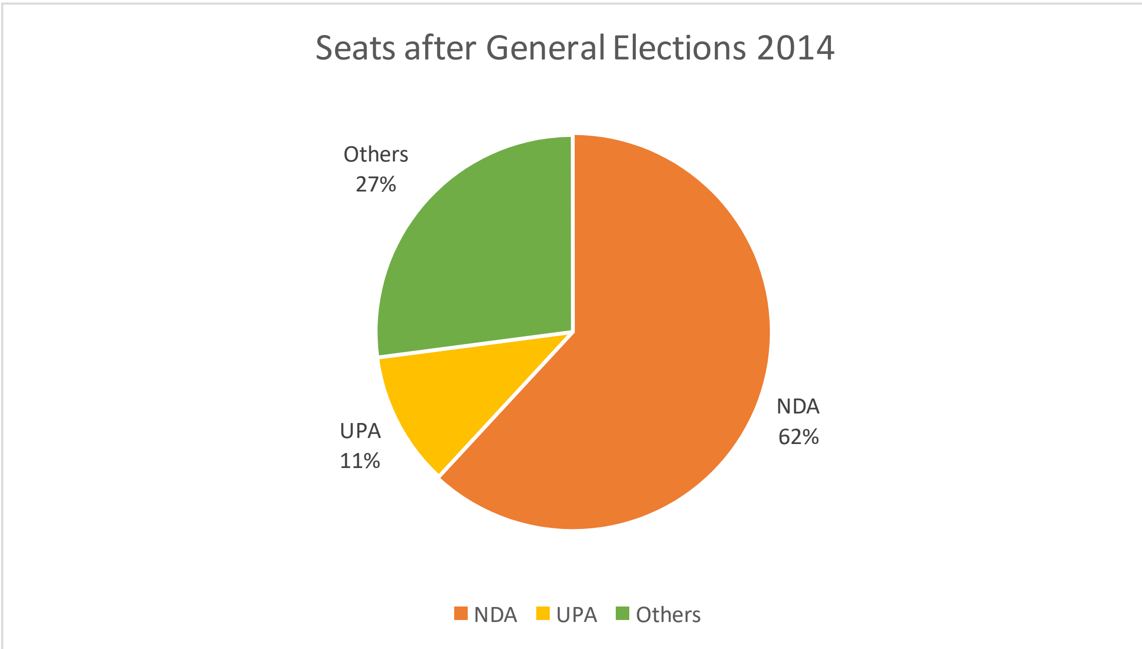

Fig. 1.3 Percentage seat-shares of the 2014 General Elections between NDA, UPA and Others

CHAPTER 2

LITERATURE REVIEW

Due to the requirement of detailed data pertaining to the Indian General Elections, it was first necessary to learn the basis of how it's data is acquired. Techniques such as Optical Character Recognition (OCR), Web Scraping, PDF Conversion, Parsing and Fuzzy Name Matching are used to enter, reformat and clean data found in a miscellany of sources [2]. Constituency level data is then amalgamated with national level data and useful datasets are produced, on which data analysis may be applied.

Then, it was imperative to review the different techniques used by other researchers to forecast elections. Data such as economic variables, incumbency, state ideology, previous election results, third party candidates, regional variables, and biographical information may be used to create advanced models. Also their relative importance to each other is considered, and weights are applied to them based on past empirical data. For example, it was observed that incumbent seats see a fall in vote share for the ruling party [3]. Estimates for the coefficients used in the structural models are generated by applying the model to as many past elections as are available to learn what effect the variables have had in the past. Predicting the next election therefore becomes a simple matter of inserting the relevant values for the political and economic indicators as they stand in the election year and then multiplying them with the coefficients that previous elections tell us provide the best fit [1]. However, the above research was conducted on democracies that are of a different nature than the Indian democratic system which is of a parliamentary form.

Other factors such as Social Media sentiments were looked into. Mainly, papers had discussed about how a corpus of Twitter tweets can be used to forecast elections by seeing the overall favourability of the political party. The corpus is generated by downloading through the use of an Application Programming Interface (API) provided by Twitter itself. Thus, geo-location data for the tweets is also available. Then, search terms are collected for various parties. For example, UK independence party is also called 'ukip', so that is a search term that needs to be filtered when considering data for the aforementioned party. The tweets are filtered through issue search terms such as Economic (deficit, economy, business, austerity, budget, debt, borrowing, gdp, unemployment, job) or Education (education, tuition, school, university, universities, apprenticeship, childcare, teachers, uni). Then opinion polls are used to allocate importance to each issue based on what respondents consider the most important [4]. However, there are arguments that electoral forecasting is partly flawed and there is no definite way to predict with certainty using Twitter. Critical factors include the lack of inclusion of incumbency, flaws in sentiment analysis classifiers, rumours or propaganda, lack of consideration of demographics and ethnicity, and self-selection bias [5]. A common system which will be able to solve different problems like sarcasm, conjunction and implicit negation was proposed by using a hybrid approach of Lexicon Based and Rule Based Sentiment Analysis. Sentiment scores of words can be extracted using APIs available online [6].

Studies about news data were also reviewed. In one such study, media bias in different news outlets was calculated using machine learning classification techniques. Each word was given an *indicativeness* value based on the number of times a news outlet would use that term. Their conservative or liberal nature can be determined by the frequency of words they use the most, and various KNN clusters were generated for their classification [7].

Lastly, predictive models used in other fields were also considered. One such field is the Stock Market. The Stock Market can be correlated to the political voting system by researching the similarity in short-term stock prices and votes. Various techniques such as Linear Regression and Support Vector Regression were discussed to forecast future stock prices, and a success rate of 70% was produced [8]. However, it should be noted that a typical stock market has significantly more time data points than an electoral system, as elections happen only once in 4-5 years. Therefore, due to the lack of availability of enough time data points, it is hard to achieve high level of accuracy using the aforementioned techniques.

CHAPTER 3

PROBLEM DEFINITION AND OBJECTIVES

3.1 PROBLEM DEFINITION

Elections have a crucial role in the functioning of a modern democracy. With a barrage 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.

3.2 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 2014 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 aforementioned 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.

3.3 MOTIVATION AND CHALLENGES

As an ardent follower of politics for almost half a decade, it seemed befitting for me to choose a topic such as this one, and apply various concepts from Computer Science to the political sphere. Very little to no research has been done for encapsulating machine learning and computational data analysis concepts in election forecasting. While there are findings from a mathematical and statistical point of view, most of them do not incorporate the technological intelligence aspects. 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.

CHAPTER 4

BACKGROUND

4.1 DATA ANALYSIS

Data analysis is a process of inspecting, cleansing, transforming and modeling 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 through the use of simple variables. By doing this, pandas provides a solid foundation upon which a very powerful data analysis ecosystem can be established [10].

4.2 MACHINE LEARNING

Machine Learning (ML) is the scientific of algorithm and statistical models that computer systems use to progressively improve their performance on a specific task. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers [11].

4.2.1 Linear Regression

In statistics, linear regression is a linear approach to model the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression [12].

Linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a forecast of the response.

Support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data used for classification and linear regression. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other [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 input 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.

To fit a time series model, it depends on the number of model parameters to be estimated and the amount of randomness in the data. The sample size required increases with the number of parameters to be estimated, and the amount of noise in the data. Using least squares estimation, or some other non-regularized estimation method, it is possible to estimate a model only if you have more observations than parameters. However, there is no guarantee that a fitted model will be any good for forecasting, especially when the data is noisy [16].

4.2.2 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 its content and the labelled training samples [17].

KNN is one of the most important non-parametric algorithms in pattern recognition field and it's a supervised learning predictable classification algorithm. The classification rules of KNN are generated by the training samples themselves without any additional data. KNN classification algorithm predicts the test sample's category according to the K training samples which are the nearest neighbours to the test sample, and judges it to that category which has the largest category probability [18].

4.2.3 Sentiment Analysis

Sentiment analysis is a field of natural language processing which focuses on extraction of objective and subjective information from a natural language sentence. With the boom of online community people are expressing their likes and dislikes towards different subjects in blogs, microblogs and social networking sites like Twitter and Facebook. Analyzing these expressions of short colloquial text can yield vast information about the behaviour of the people that can be helpful in many other subjects like Political Science, opinion extraction and Human Computer Interaction (HCI).

Usually, a negative score signifies negative connotation and a positive score signifies positive connotation of the word. Various 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].

CHAPTER 5

METHODOLOGY

In this research, experimental analysis will be carried out in order to propose a model that can help in forecasting of elections methods.

1. Firstly, current election forecasting techniques would be researched and analyzed. This includes identifying the various types of data sets and techniques that previous papers have discussed in order to forecast elections relevant to their countries.
2. Then, fields such as Data Analysis and Machine Learning would be studied in order to decide the relevance of their respective technologies that can applied to the given task.
3. Using Data Analysis libraries, a system would be devised that would parametrize manipulations to past election data. These parameters would then be derived using Machine Learning algorithms.
4. Finally, the results would be curated to propose a model for forecasting.

CHAPTER 6

PROPOSED MODEL

6.1 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 particular 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.

6.2 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 applied 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.

6.3 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.

6.3.1 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 election 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 forecasted 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 6.1. Further, a graph for these vote shares is shown in Fig 6.1.

Gujarat Vote Shares		
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)	41.4%	49.1%
Predicted Vote Share	39%	50%

Table 6.1 Recent Vote Shares in Gujarat National and State Elections

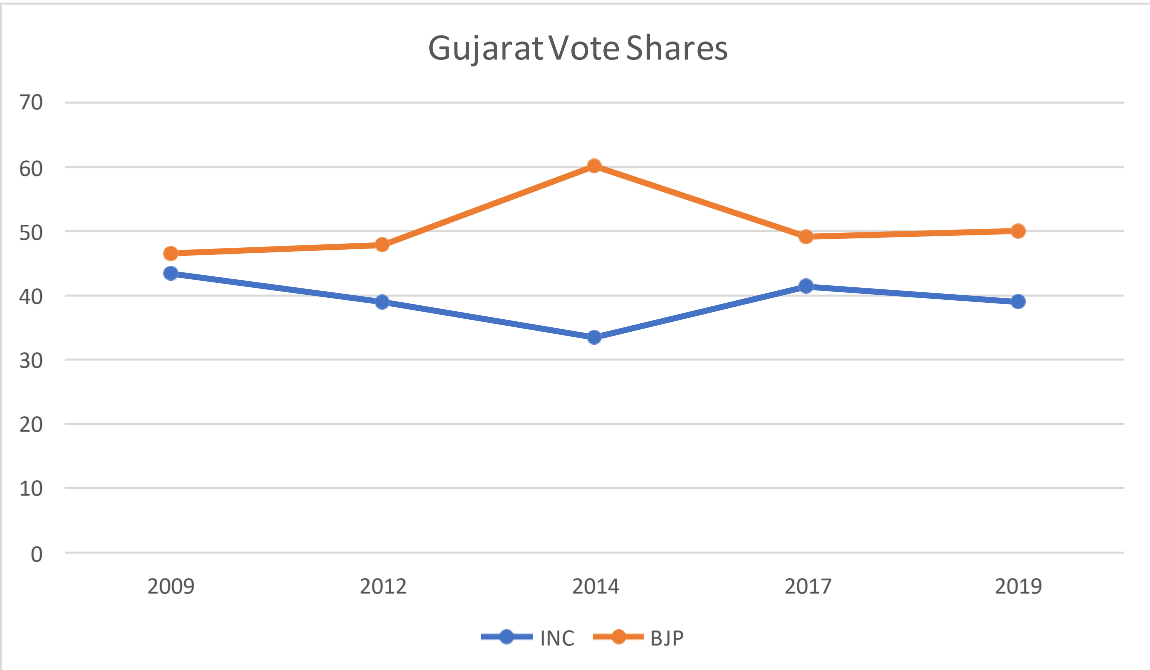

Fig 6.1 Recent Vote Shares in Gujarat National and State Elections (Graph)

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).

6.3.2 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 determine 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.

CHAPTER 7

EXPERIMENTAL RESULTS AND ANALYSIS

7.1 INITIAL DATA

The constituency-wise results of the 2014 General Election, is considered as *initial* data. This is because coalitions and swings are applied to each constituency individually. For that, we must be cognizant about the results of each constituencies.

ST_CODE	State name	Year	PC name	Candidate Name	Candidate Category	Party Abbreviation	Total Votes Polled	Position
S24	Uttar Pradesh	2014	Sultanpur	FEROZE VARUN GANDHI	GEN	BJP	410348	1
S24	Uttar Pradesh	2014	Sultanpur	PAWAN PANDEY	GEN	BSP	231446	2
S24	Uttar Pradesh	2014	Sultanpur	SHAKEEL AHMED	GEN	SP	228144	3
S24	Uttar Pradesh	2014	Sultanpur	AMEETA SINGH	GEN	INC	41983	4
S24	Uttar Pradesh	2014	Sultanpur	VARUN GANDHI	GEN	IND	14021	5
S24	Uttar Pradesh	2014	Sultanpur	SHAILENDRA PRATAP SING	GEN	AAAP	5835	6
S24	Uttar Pradesh	2014	Sultanpur	BRIJESH KUMAR	SC	IND	5752	7
S24	Uttar Pradesh	2014	Sultanpur	None of the Above	NULL	NOTA	5412	8
S24	Uttar Pradesh	2014	Sultanpur	VIJAY	GEN	IND	4630	9
S24	Uttar Pradesh	2014	Sultanpur	GIRISH LAL	GEN	SBSP	3804	10
S24	Uttar Pradesh	2014	Sultanpur	DR. NAFEES AHMAD	GEN	RaIP	3381	11
S24	Uttar Pradesh	2014	Sultanpur	PITAMBAR NISHAD	GEN	PMSP	2737	12

Fig 7.1 Sample of data acquired from the results of 2014 General Elections

It can be seen in Fig 7.1 that each constituency (identified by the combination of 'State Name' and 'PC Name') has various parties contesting, each of which are denoted 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.

7.2 MAKING COALITIONS USING DATA ANALYSIS

In order to perform various computations on the *.csv* file, we must make use of a Data Analysis library called *Pandas*. It is an open-source, high performance, easy to use library with a variety of data structures and tools to choose from.

In the system devised in the experiment, we need to pass an array of 'Party Abbreviations' for each coalition (UPA and NDA) as parameters to the system, as shown in Fig 7.2.

```
"""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","AD","PMK","NPEP","AINRC","ADMK"]
#UPACoalition = ["INC","BSP","SP","RLD","RJD","AITC","NCP", "TDP", "DMK","IUML","JD(S)","CPI","CPM", "JKN","JMM","

"""In percent. Incumbent swings are likelier to be lesser, as people want change."""
# ndaSwingIncumbent = 90.0
# ndaSwingNotIncumbent = 102.0
# upaSwingIncumbent = 95.0
# upaSwingNotIncumbent = 107.0

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

Fig. 7.2 Demonstration of parametrization of Swings and Coalitions in the system

Along with deciding the respective NDA and UPA coalitions, we can also select the swings. A swing of 100 is called a 0 swing (no change compared to previous election). Similarly a swing of 102 would be a +2% swing and a swing of 97 would be a -3% swing. This system can be tweaked in order to provide custom swings to each seat (not shown). However, for simplicity, in this program we assume that the swings on NDA incumbent & non-incumbent, and UPA incumbent & non-incumbent are unique. Incumbent seats are those won by the party in the base election. Usually, a negative swing is observed on incumbent seats and positive swing is observed on non-incumbent seats (due to change in voter mood).

Upon running the program, we can get detailed state-wise results on how many seats are retained or lost by each alliance, after computations on swings and alliances are performed.

```
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. 7.3 Seats NDA would retain and lose, state-wise, if extended coalitions were formed by NDA and UPA.

```
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. 7.4 Detailed results of how many state-wise seats NDA and UPA won, along with respective vote shares in relevant seats after extended coalition.

For example, Fig 7.3 indicates each state-wise seat that NDA would lose from 2014 to 2019 and Fig 7.4 indicates each state-wise seat for both NDA and UPA.

7.2.1 Same 2014 coalitions, no swings

If coalitions stay same as 2014 NDA and UPA; and no swings are applied to 2014 results. The parameters for coalitions and swings, that are coded into the Data Analysis system are shown in Fig 7.5.

```
"""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. 7.5 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. 7.6 Results for same coalitions as 2014 and no swings

Fig 7.6 is indicative that the system for Data Analysis is working, as it correctly predicts the number of seats won by NDA and UPA as 336 and 60 respectively.

7.2.2 BJP & INC alone, no swings

If NDA consists of only BJP, and UPA consists of only INC; and no swings are applied are applied to 2014 results. Parameters for it are shown in Fig 7.7. It is clear (in Fig 7.8) that both parties would lose seats, however, BJP still manages to form government (at 273+) with 282 seats. INC is reduced to a meagre 40 seats.

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

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

Fig. 7.7 Parameters for NDA and UPA having only BJP and INC alone respectively; with no swings

```
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. 7.8 Results for BJP and INC alone with no swings

7.2.3 Extended 2019 coalitions, no swings

If NDA and UPA were to be part of their extended updated coalitions (Fig 7.9), and no swings were applied, the NDA would still form government. However, UPA would gain by a massive margin from 60 to 190 seats (Fig 7.10). This indicates the power of strong coalitions, especially for the UPA.

```
"""Expected Coalitions in 2019"""
NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","AD","PMK","NPEP","AINRC","ADMK"]
UPACoalition = ["INC","BSP","SP","RLD","RJD","AITC","NCP", "TDP", "DMK","IUML","JD(S)","CPI","CPM", "JKN","JMM","KI

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

Fig. 7.9 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. 7.10 Results for extended coalitions of NDA and UPA; with no swings

7.2.4 Extended 2019 coalitions, arbitrary swings

For the sake of demonstration, we can intuitively assume some likely swings, as observed heuristically from previous elections, which are shown in Table 7.1. We assume that NDA will lose 10% votes in every incumbent seat, and gain 2% votes. Further, UPA will lose 5% in every incumbent seat, and gain 7% votes.

	Incumbent	**Non-incumbent**
NDA	-10% swing (90)	+2% swing (102)
UPA	-5% swing (95)	+7% swing (107)

Table 7.1 Arbitrary swings to be applied for NDA and UPA

```
"""Expected Coalitions in 2019"""
NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","AD","PMK","NPEP","AINRC","ADMK"]
UPACoalition = ["INC","BSP","SP","RLD","RJD","AITC","NCP", "TDP", "DMK","IUML","JD(S)","CPI","CPM", "JKN","JMM","KE

"""In percent. Incumbent swings are likelier to be lesser, as people want change."""
ndaSwingIncumbent = 90.0
ndaSwingNotIncumbent = 102.0
upaSwingIncumbent = 95.0
upaSwingNotIncumbent = 107.0
```

Fig. 7.11 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, '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 242
defaultdict(<class 'int'>, {'Andhra Pradesh': 3, 'Assam': 4, 'Bihar': 35, 'Goa': 1, 'Gujarat': 24, 'Haryana': 5, 'Him
achal Pradesh': 3, 'Jammu & Kashmir': 1, 'Karnataka': 9, 'Madhya Pradesh': 21, 'Maharashtra': 37, 'Meghalaya': 1, 'Na
galand': 1, 'Odisha': 1, 'Punjab': 2, 'Rajasthan': 22, 'Tamil Nadu': 31, 'Uttar Pradesh': 10, 'Chattisgarh': 7, 'Jhar
khand': 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, 'Karn
ataka': 19, 'Kerala': 18, 'Madhya Pradesh': 8, 'Maharashtra': 11, 'Manipur': 1, 'Meghalaya': 1, 'Mizoram': 1, 'Punja
b': 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. 7.12 Results for extended coalitions of NDA and UPA; with arbitrary swings

As shown in Fig 7.12, the UPA gives NDA a very tough competition and scrapes by to attain the highest number of seats. Using this model, if we can properly map the appropriate swings to each constituency, we can determine accurately the results of the 2019 General Elections based on the 2014 Elections.

7.3 DETERMINATION OF SWINGS USING MACHINE LEARNING

To determine swings using past data, we can use Machine Learning Forecast tools provided in WEKA. It is logical to calculate the swings of each state independently, as different states have completely different attitudes towards a political party.

Therefore, in the following example, it will be demonstrated how future swings can be predicted using Linear Regression algorithm for Rajasthan. Data from 5 previous General and Assembly (G/A) elections held is taken as the vote shares of both parties (Table 7.2). A *.arff* dataset is produced as shown in Fig 7.13.

Rajasthan Vote Shares		
Election Held On	**INC**	**BJP**
2008 (A)	36.82%	34.27%
2009 (G)	47.19%	36.57%
2013 (A)	33.07%	45.17%
2014 (G)	30.40%	50.90%
2018 (A)	39.3%	38.8%

Table 7.2 Vote Shares in Rajasthan in General (G) and Assembly (A) elections

```
@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, 38.8
```

Fig. 7.13 *.arff* dataset for Vote Shares in Rajasthan to be used in WEKA

```
=== 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. 7.14 Results after using Linear Regression Algorithm on yearly basis for Rajasthan

Therefore, as observed in Fig 7.14, the swings for 2019 in comparison to 2014 would be the difference -- +10% for UPA (INC) and -11% for NDA (BJP). Applying these swings to our Data Analysis Model, we can forecast the number of seats in Rajasthan as:

```
"""Expected Coalitions in 2019"""
NDACoalition = ["BJP", "SHS","JD(U)","LJP","NPF","SAD","BLSP", "SWP","AD","PMK","NPEP","AINRC","ADMK"]
UPACoalition = ["INC","BSP","SP","RLD","RJD","AITC","NCP", "TDP", "DMK","IUML","JD(S)","CPI","CPM", "JKN","

"""In percent. Incumbent swings are likelier to be lesser, as people want change."""
ndaSwingIncumbent = 89
ndaSwingNotIncumbent = 88
upaSwingIncumbent = 110
upaSwingNotIncumbent = 110

stateName = "Rajathan"
```

Fig. 7.15 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. 7.16 Results of seat-share in Rajasthan after Linear Regression swings applied

Upon applying these swings (Fig 7.15) to our data analysis model, we observed that in Rajasthan, NDA gets 21/25 seats and UPA gets 4/25 seats (Fig 7.16).

Alternatively, using Multilayer Perceptron Algorithm (with Sigmoid Function), we compute results as shown in Fig 7.17.

```
=== 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. 7.17 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","AD","PMK","NPEP","AINRC","ADMK"]
UPACoalition = ["INC","BSP","SP","RLD","RJD","AITC","NCP", "TDP", "DMK","IUML","JD(S)","CPI","CPM", "JKN","JMM","

"""In percent. Incumbent swings are likelier to be lesser, as people want change."""
ndaSwingIncumbent = 83
ndaSwingNotIncumbent = 83
upaSwingIncumbent = 114
upaSwingNotIncumbent = 114

stateName = "Rajathan"
```

Fig. 7.18 Parameters for extended coalitions of NDA and UPA; Multilayer Perceptron Algorithm swings applied for 2019.

On applying these swings (Fig 7.18) to our data analysis model, we observe that in Rajasthan, NDA gets 19/25 seats and UPA gets 6/25 seats (Fig 7.19).

```
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. 7.19 Results of seat-share in Rajasthan after Multi-layer Perceptron swings applied

Using this model, swings for all states can be predicted using state-specific election data. However, due to constraints on time it is not feasible to experimentally determine swings for each state in this report.

7.4 CURATED RESULTS

A recapitulated table of results previously obtained is shown in Table 7.3 and Table 7.4.

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

Table 7.3 NDA and UPA curated seat and vote share in various scenarios

	BJP		INC	
No Swing	25 seats	0% swing	0 seats	0% swing
Linear Regression Swing	21 seats	-11% swing	4 seats	+10% swing
Multi-layer Perceptron Swing	19 seats	-17% swing	6 seats	+14% swing

Table 7.4 Rajasthan's BJP and INC curated seats and swings after applying Machine Learning algorithms

CHAPTER 8

CONCLUSION

Although forecasting something as complex as the Lok Sabha elections can be very difficult and requires meticulous detail, it is still viable given the right methodology and data.

This research may produce results that will pave the way for a new kind of forecasting, and may produce a shift from relatively outdated polling techniques. It would also prove that Artificial Intelligence and Data Science are very useful fields for political scientists and various stakeholders to research about in order to come to fruitful conclusions using relatively unorganized data.

Using the previous election data as a reference for the next, and then applying apt parametrized computations on it can be viable in forecasting elections with considerable accuracy.

REFERENCES

[1] D. Walther, "Picking the winner(s): Forecasting elections in multiparty systems," 2014.

[2] F. R. Jensenius, "Studying Indian Politics with Large-scale Data," 2017.

[3] P. Hummel, "Fundamental Models for Forecasting Elections," 2014.

[4] P. Burckhardt, "Tweet as a Tool for Election Forecast: UK 2015 General Election as an Example," 2016.

[5] D. Gayo-Avello, "A Balanced Survey on Election Prediction using Twitter Data," 2012.

[6] B. Agarwal, "Sentiment Analysis of Politcal Twitter Data," 2015.

[7] J. M. Shola, "Predicting Media Bias in Online News," 2016.

[8] A. Zheng, "Using AI to Make Predictions in Stock Market," 2014.

[9] B. S. Xia, "Review of business intelligence through data analysis," 2015.

[10] W. McKinney, "pandas: a Foundational Python Library for Data Analysis and Statistics," 2011.

[11] C. Bishop, "Pattern Recognition and Machine Learning," 2006.

[12] D. A. Freedman, "Statistical Models: Theory and Practice," 2009.

[13] V. N., "Support-vector networks," 1995.

[14] A. J. Smola, "A Tutorial on Support Vector Regression," 2003.

[15] Rumelhart, "Learning Internal Representations by Error Propagation," 1986.

[16] R. Hyndman, "Fitting models to short time series," 2014.

[17] X. Xin, "Advances in Machine Learning Based Text Categorization," 2006.

[18] Z. Yong, "An Improved KNN Text Classification Algorithm Based on Clustering," *2009.*

[19] K. Sanghal, "Modeling Indian General Elections: Sentiment Analysis of Political Twitter Data," 2015.