BANGALORE UNIVERSITY

JNANA BHARATHI CAMPUS, BENGALURU – 560056

DEPARTMENT OF COMPUTER SCIENCE AND APPLICATONS

A project report entitled

"Analysing and Predicting Election Outcomes"

Submitted By

Praveen M R

P03NK22S126023

Under the Guidance of

Dr. Muralidhara.B.L

DEPARTMENT OF COMPUTER SCIENCE AND APPLICATIONS

TOWARDS

MINI PROJECT

PRESCRIBED BY THE BANGALORE UNIVERSITY

FOR THE III SEMESTER MASTER OF COMPUTER SCIENCE

FOR THE ACADEMIC YEAR

2023-2024

BANGALORE UNIVERSITY

JNANA BHARATHI CAMPUS, BENGALURU – 560056

DEPARTMENT OF COMPUTER SCIENCE AND APPLICATONS

CERTIFICATE

This is to certify that **Praveen M R** Register No: **P03NK22126023** have satisfactorily completed the project work entitled "**Analysing and Predicting Election Outcomes**" for partial fulfilment of the requirement for the award of the degree Master of Computer Application (MCA) awarded by Bangalore University for the year 2023-24.

Dr. Muralidhara.B.L
GUIDE

Dr. Muralidhara.B.L
HOD

Examiners:

1)

2)

DECLARATION

Praveen M R, student of III semester MCA of **Department of Computer Science and Application, Bangalore University,** hereby declare that the dissertation entitled "**Analysing and Predicting Election Outcomes**" has been independently carried out by me at Department of Computer Applications, Bangalore University and submitted during the academic year 2023-2024. Further the content presented in the report is a genuine and authentic work, created solely by me.

To our knowledge this dissertation has not been submitted to any other college or university or published at any prior to this.

Place: Bangalore **Praveen M R**

Date:

ACKNOWLEDGEMENT

I am extremely grateful for the immense help and support I received from my project guide, **Dr. Muralidhara.B.L,** during the development of my project titled "Analysing and Predicting Election Outcome." Their guidance, expertise, and encouragement have played a major role in shaping the direction and success of our project. Their valuable insights and advice have been instrumental in helping us navigate through the challenges and complexities of analysing election data.

I would also like to express my heartfelt appreciation to my friends and colleagues who have been there for me throughout this project. Their unwavering support and collaborative efforts have greatly enhanced the quality of our work. Their diverse perspectives and ideas have enriched our project and made it more comprehensive.

Furthermore, I would like to extend my thanks to all the individuals who have participated in our project by providing their valuable time and data. Without their cooperation and involvement, our analysis and predictions would not have been possible. Their contributions have been integral in gathering the necessary information and ensuring the accuracy of our findings.

I am truly grateful for the opportunity to work on this project, and I want to express my deep appreciation to everyone who has been involved. Their guidance, support, and collaboration have made this project an enriching and rewarding experience.

Praveen M R

ABSTRACT

This project aims to analyse and predict election outcomes by utilizing data analysis techniques and machine learning algorithms. The objective is to gain insights into the factors influencing election results and develop a predictive model to forecast future election outcomes.

The project begins by collecting comprehensive data on various variables, such as voter demographics, present issues in their locality, and historical election results. This dataset is then processed and cleaned to ensure accuracy and consistency.

Machine learning algorithms, such as logistic regression, decision trees, and random forests, are employed to build predictive models. These models are trained using historical election data and tested for accuracy. The predictive models aim to forecast election outcomes based on the identified factors and provide insights into the likelihood of success for different parties.

The project includes visualizations and interactive tools to aid in interpreting and presenting analysis results. These visualizations help in understanding underlying patterns and trends in election data, assisting in strategic decision-making and campaign planning.

Overall, this project helps to understand patterns and trends in electoral processes, enabling to make informed decisions and predictions regarding election outcome.

Table of contents

CHAPTER 1

INTRODUCTION

Predicting election outcomes is a crucial aspect of understanding and anticipating the results of democratic processes. In our case, we are focusing on predicting the election outcome for the Kolar constituency. To make this prediction, we have employed the power of machine learning techniques. To make our prediction system more accessible and user-friendly, we have utilized the Streamlit library. This library helps us create a visually appealing and interactive interface for users to interact with our prediction system. It allows users to visualize data, and explore the election outcome for the Kolar constituency, By combining the power of machine learning with the simplicity of the Streamlit frontend, our prediction system aims to provide an easy-to-use tool for analysing and predicting the election outcome in the Kolar constituency.

we meticulously crafted questionnaires centered around critical issues, Congress guarantees, the Ram Mandir construction, EWS reservation, and more. These questionnaires served as the bedrock of your dataset, capturing voters' sentiments and preferences. Our dataset comprises historical voting records, demographic features (age, gender, occupation, income, education level), and contextual variables. Rigorous data preprocessing ensures our models make informed predictions. Excel became our canvas as we applied filters, handled missing values, and ensured data consistency. Mapping categorical variables and scaling numerical features were essential steps. our data was polished and ready for analysis.

We'll harness the power of logistic regression to model probabilities, decision trees to visualize decision boundaries, and random forests to combine multiple predictions. These ensemble methods allow us to capture the nuances of voter behavior.

CHAPTER 2

HYPOTHESIS

Based on the literature review, we propose the following hypotheses:

- Female voters prefer the congress party due to its focus on women-centric schemes like shakti and Gruha Lakshmi.
- Voters choose the congress party due to the guaranteed benefits provided by the congress government.
- Congress party attract Muslim voters due to concerns related to the hijab and the concept of a Hindu Rashtra
- Hindu voters tend to vote for the BJP due to its stance on the construction of Ram Mandir in Ayodhya.
- The failure of the Modi government in controlling essential commodity(LPG gas, milk, mustard oil, petrol, diesel) will effect the voting behaviour of low-income and middle-class households.
- Unemployment is a critical concern for the youth population and influence their voting decisions.
- poor and middle-class families like Gruha Jyoti and Gruha Lakshmi scheme and influence their voting decisions
- The Karnataka BJP's decision to rejig the EWS(economically weaker section) quota and allocate part of it to Vokkaligas and Lingayats may impact voting preferences.
- Youth voters are inclined to support the BJP due to its positions on foreign policy decisions and international relations.

CHAPTER 3

Data Collection

To understand the current situation of elections in Kolar district, we reviewed the following data sources:

1. Newspapers: Newspapers play a crucial role in understanding and reviewing elections for your election prediction mini project. Newspapers extensively cover election campaigns, providing updates on rallies, speeches, and candidate profiles. They analyse poll results, helping readers gauge the political climate and candidate popularity. Newspapers conduct debates and interviews with candidates, allowing voters to understand their stances on critical issues. We Extracted information on election-related updates, candidate profiles, party manifestos, and campaign developments. Newspapers provided insights into local issues relevant to Kolar district, such as unemployment and job creation, prise rise in LPG gas, diesel and petrol, cost of milk, household groceries and other issues like caste which was playing critical issue ,communities etc and voter sentiments. newspapers serve as valuable sources of information, shaping public opinion and providing insights into electoral dynamics. Incorporating newspaper data into your election prediction project can enhance its accuracy and relevance.

2. Social Media: twitter and Facebook are powerful platforms for monitoring real-time conversations related to elections. Trends in social media conversation may provide early signals about shifts in public opinion. Social media platform allow the candidate to showcase their attributes and connect with the voters. Social media data help identify key political topics, gauge the impact of debates, and assess candidate popularity.
   Data collection
   Feature Identification and Questionnaire Preparation

   Based on our data review, we identified key features for election prediction:
   - Party Manifestos and Guarantees: Considered promises made by political parties like Congress's five guarantees and BJP's promises. Designed questions to assess voters' awareness and preferences regarding these promises.
   - Local issues: Explored issues specific to Kolar district like unemployment and job creation, prise rise in LPG gas, diesel and petrol, cost of milk, household groceries and other issues like caste which was playing critical issue , communities etc. and Incorporated questions related to voters' priorities and concerns.

Based on these features we prepared the questionnaires. Which are

1. Age
2. What is your family size?
3. What is your marital status?
4. Gender
5. What's your highest level of education?
6. What is your current occupation?
7. Mother tongue?
8. What is your approximate monthly household income?
9. Religion
10. What is your caste?
11. What is your assembly constituency?
12. Which guarantee scheme are you a beneficiary of?
13. To what extend has the SHAKTI scheme influenced women's empowerment in your opinion?
14. Has the ANNABHAGYA scheme shown measurable results in reducing poverty rates?
15. Has the YUVANIDHI scheme demonstrated positive outcomes in terms of economic growth?
16. Has GRUHAJYOTHI scheme been helpful for your house maintainability?
17. Has the GRUHALAKSHMI scheme shown measurable results in terms of women's empowerment and household well-beings in community?
18. What is your view on prime minister modi's ability to manage partnerships with major global powers through his foreign policy?
19. How would you perceive the impact of 10% reservation for EWS(economically weaker section) on equal opportunities in education and employment?
20. How do you perceive the issue of price rise in petrol, diesel in your locality?
21. How has the rise in LPG gas, milk, mustard oil prices affected your household budget?
22. How would you rate the Modi government's performance in addressing unemployment concerns?
23. Are you feeling happy with the construction of ram mandir and with the PM's active participation in that?
24. Is there any possibility of India becoming a "HINDU RASHTRA" due to the construction of ram mandir?
25. How would you rate Siddaramaiah's performance as chief minister?
26. How would you rate Narendra modi's performance as prime minister?
27. Whom do you want to be the next chief minister?
28. Whom do you want to be the next prime minister?
30. Who benefits from JDS-BJP alliance?
31. Is there discrimination in releasing funds by the modi government to the state government?
32. Do you support separate south india?
33. Do you think south Indians paying more tax than north Indians?
34. Do you think south Indians paying more tax than north Indians will affect the upcoming elections?
35. What are the electoral issues in india?

36.When considering candidates in an election, do factors such as community/caste affiliation play a role in influencing your voting decisions?
37.To whom you vote in last assembly elections?
38.Based on the five guarantees provided by the congress government, would you vote for this government in the upcoming elections?
39.Based on the construction of ram mandir ,EWS reservation and foreign policy, would you vote for BJP government?
40.To whom you will vote in upcoming elections?

- These Questionnaires allow us to directly collect information from voters. By asking specific questions, we gain insights into their preferences, opinions, and voting intentions.
- These Questionnaires help us to gather demographic information about voters. This includes: Age, gender, education level, income, occupation, ethnicity and Also the assembly constituency.
- Through these questionnaires, voters can express their opinions about specific candidates.
- These Questionnaires help identify which issues are most critical to them.
- Combine questionnaire responses with other features (historical voting data, demographics, etc.) to build predictive models.
- Machine learning algorithms can learn patterns from questionnaire-based data to predict election outcomes.

Challenges and Considerations:

- ❖ Bias: we Ensured that these questionnaires are unbiased and do not favor any specific group or party.
- ❖ Sample Size: we Collected a representative sample size to generalize findings.
- ❖ Timing: Administer questionnaires during relevant periods (e.g., election campaigns).
- ❖ Question Design: Craft clear, concise, and relevant questions. We made these questionnaires based on hypotheses we made.
- ❖ Privacy: Respect respondents' privacy and anonymity.

questionnaires provide valuable data for our election prediction models. They bridge the gap between voters' opinions and quantitative analysis, contributing to more accurate predictions.

CHAPTER 4

METHODOLOGY

4.1 Dataset Overview

Our dataset on Indian elections is a comprehensive collection of information that covers various aspects of the political landscape. It includes details about candidates, their party affiliations, demographic factors, local issues, and benefits. In total, the dataset consists of 443 records.

When we look at the gender distribution within the dataset, we have 264 male candidates, 177 female candidates, and 2 transgender candidates. In addition to the gender distribution and the total number of records in our dataset, we have also collected data on various age groups. This allows us to analyze the political landscape across different generations.

Furthermore, our dataset aims to provide a comprehensive representation of the electorate in Kolar by covering all castes. By including information on caste, we can gain insights into the political preferences and dynamics within different communities in the region.

4.2 Data Cleaning and Preprocessing

Preprocessing ensures that the data is well-structured and ready for machine learning algorithms.

Let's see the step by step in detail:

i. **Translation**

This step involves converting text from one language to another. Google Translation (or any other translation service) helps bridge the language gap, making the data consistent.

our survey responses are in two languages (Kannada and English). To analyse them together, we need a common language. So the data in google sheets can be translated using formula GOOGLETRANSLATE which can be using to detect other language and translate it. for each text entry in Kannada, send it to the google translation API with the source language (Kannada) and the target language (English).

Formula → ***=GOOGLETRANSLATE(cell_address, ""kn"", ""en"")***

ii. **Mapping**

mapping language (creating a correspondence between different values or categories. It involves mapping categorical features to binary values(0 or 1).If we have multiple categories, can create binary columns for each category.

Using the formula,

=IF(cell_address="category", 1, 0)

This assigns a value of 1 if category matches and 0 otherwise.

iii. **Substitution**

This substitute function used to replace specific text within a cell.We use substitution function to replace the string with another. Syntax is

=SUBSTITUTE(text/cell_address, old_text, new_text)

iv. **Scaling**

Scaling ensures that numerical features are on a similar scale. It prevents bias due to different units. We can transform data in excel to specific range to make it more comparable or interpretable. We used a simple formula,

=IF(cell_address="text1" , "numeric _value1")

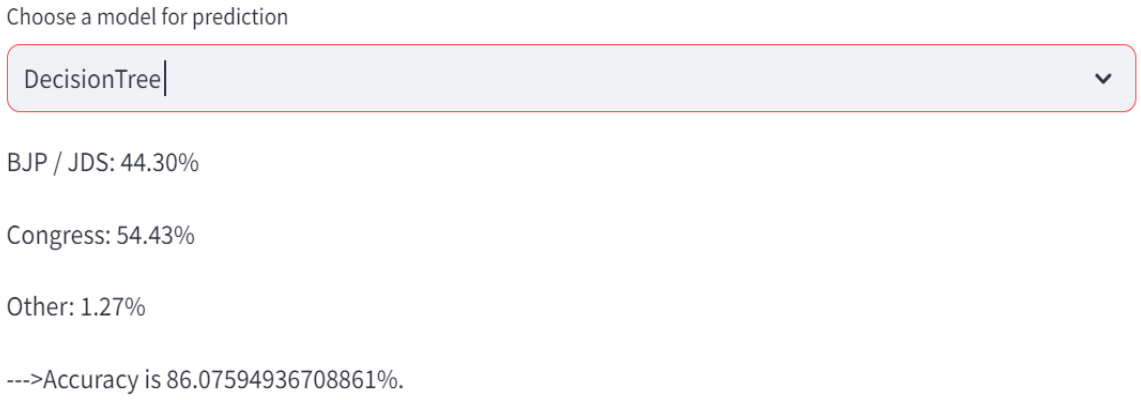
v. **Splitting the dataset**

Splitting a dataset into training and test sets is a fundamental step in machine learning. The goal is to create two distinct subsets of data. One for training a model and the other for evaluating its performance. Training set is used to train the machine learning model. Validation(test) set is used to evaluate the final model's performance. We splits the data into training and test sets in the ratio of 90:10.

4.3 Model selection and training

1. Decision Tree

Our Decision Tree model exhibited accuracy of 86%. The model effectively captured patterns within the data, showcasing its robust predictive capabilities.



A decision tree is a flowchart-like structure used to make decisions or predictions. It consists of nodes representing decisions or tests on attributes, branches representing the outcome of these decisions, and leaf nodes representing final outcomes or predictions. Each internal node corresponds to a test on an attribute, each branch corresponds to the result of the test, and each leaf node corresponds to a class label or a continuous value.

How Decision Trees Work?

The process of creating a decision tree involves:

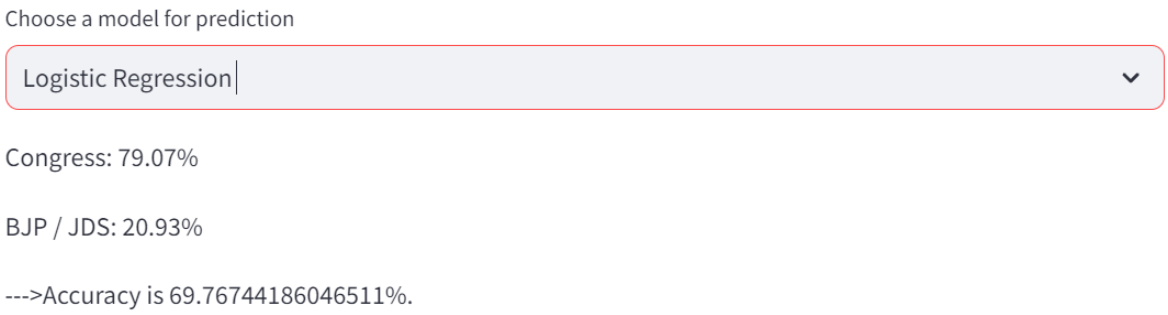
1. Selecting the Best Attribute: Using a metric like Gini impurity, entropy, or information gain, the best attribute to split the data is selected.

2. Splitting the Dataset. The dataset is split into subsets based on the selected attribute.

3. Repeating the Process. The process is repeated recursively for each subset, creating a new internal node or leaf node until a stopping criterion is met (e.g., all instances in a node belong to the same class or a predefined depth is reached).

2. Logistic regression

Logistic regression analysis studied the association between a categorical dependent variable and a set of independent variables. The name logistic regression is used when the dependent variable has only two values, such as 0 and 1 or yes or no. the name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values, such as Married, single, divorced, or widowed. Although the type of data used the dependent variable is different from that of multiple regression, the practical use of the procedures is similar. Logistic regression competes with discriminant analysis as a method for analysing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does. This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of the fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help to determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.



It has given the accuracy of 69.76%.

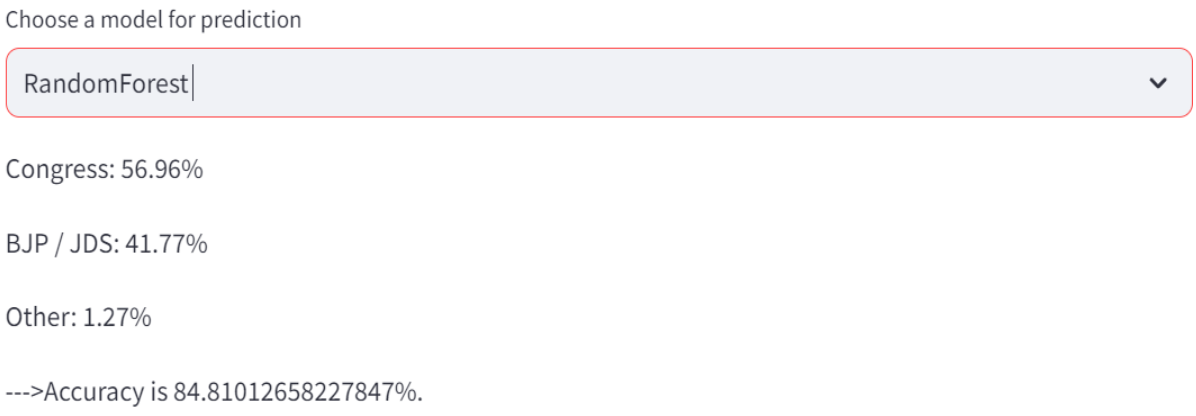
3. Random Forest

Random Forest or random decision forests algorithm is a powerful tree learning technique in Machine Learning It works by creating a number of Decision Trees during the training phase. Each tree is constructed using a random subset of the data set to measure a random subset of features in each partition. This randomness introduces variability among individual trees, reducing the risk of overfitting and improving overall prediction performance. In prediction, the algorithm aggregates the results of all trees, either by voting (for classification tasks) or by averaging (for regression tasks) This collaborative decision-making process, supported by multiple trees with their insights, provides an example stable and precise results. Random-forest are widely used for classification and regression functions, which are known for their ability to handle complex data, reduce overfitting, and provide reliable forecasts in different environments.

The random Forest algorithm works in several steps which are discussed below:

- Ensemble of decision tress
- Random feature selection
- Bootstrap aggregating or bagging
- Decision making and voting

Random forests are frequently used as “Blackbox” models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.



It has given the accuracy of 84.8%.

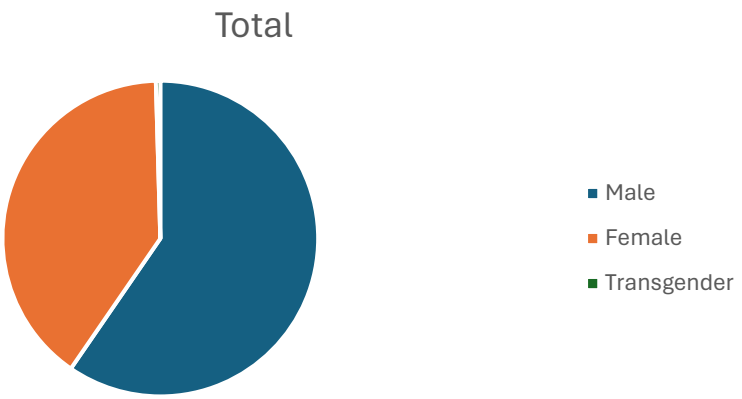
CHAPTER 5

Analysis of Voter Preferences and Trends

Demographic Data Distribution:

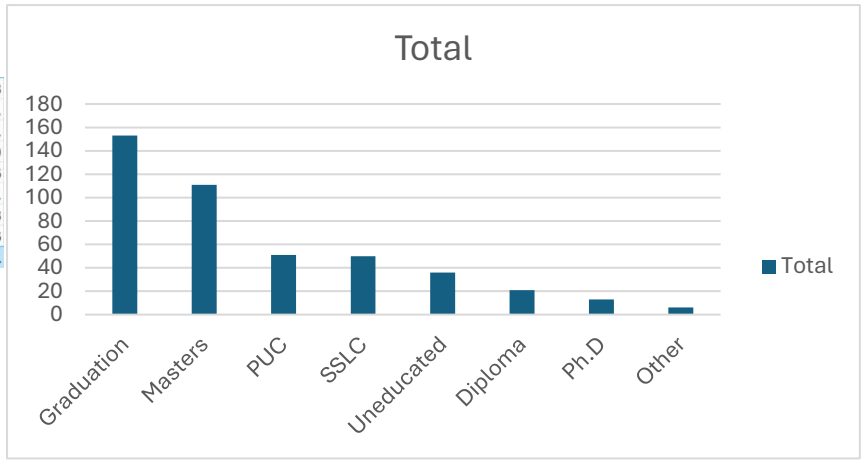
1. **Gender**

Gender	Count of Gender
Male	264
Female	177
Transgender	2
Grand Total	**443**



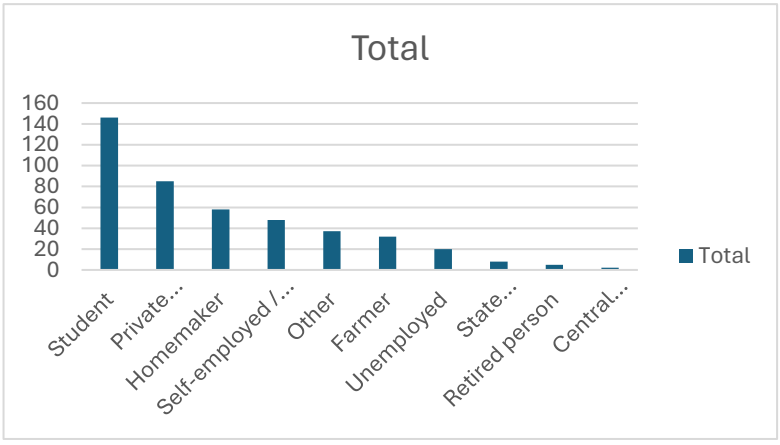
2. **Educational qualification**

Graduation	153
Masters	111
PUC	51
SSLC	50
Uneducated	36
Diploma	21
Ph.D	13
Other	6
Grand Total	**441**



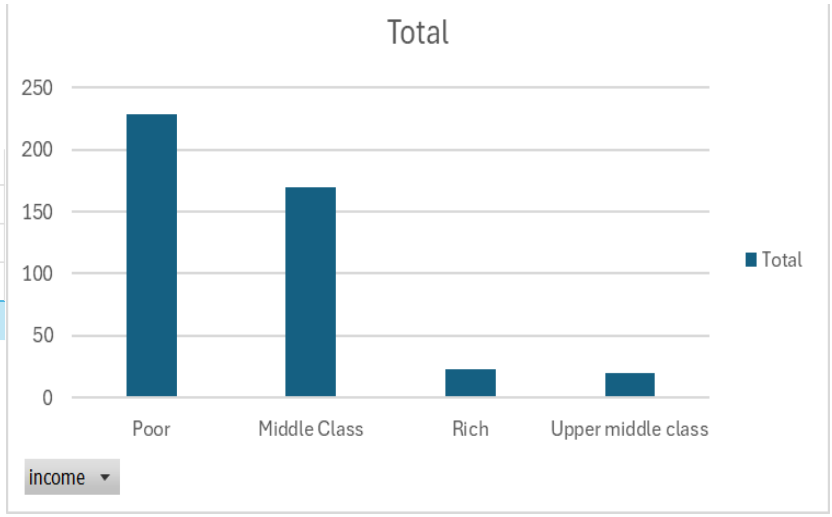
3. **Occupation**

Student	146
Private company or organization employee	85
Homemaker	58
Self-employed / Entrepreneur	48
Other	37
Farmer	32
Unemployed	20
State government employee	8
Retired person	5
Central government employee	2
Grand Total	**441**



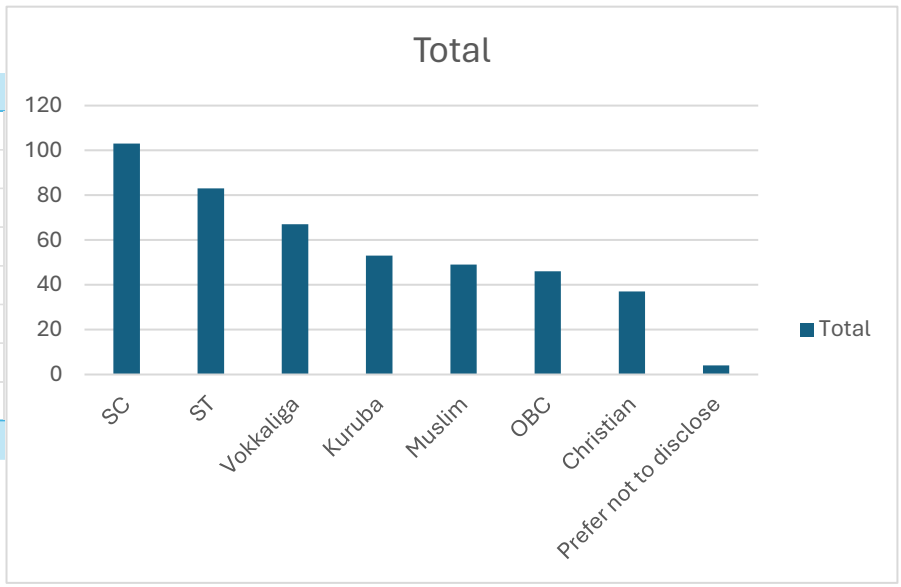
4. Income

Poor	228
Middle Class	170
Rich	23
Upper middle class	20
Grand Total	**441**



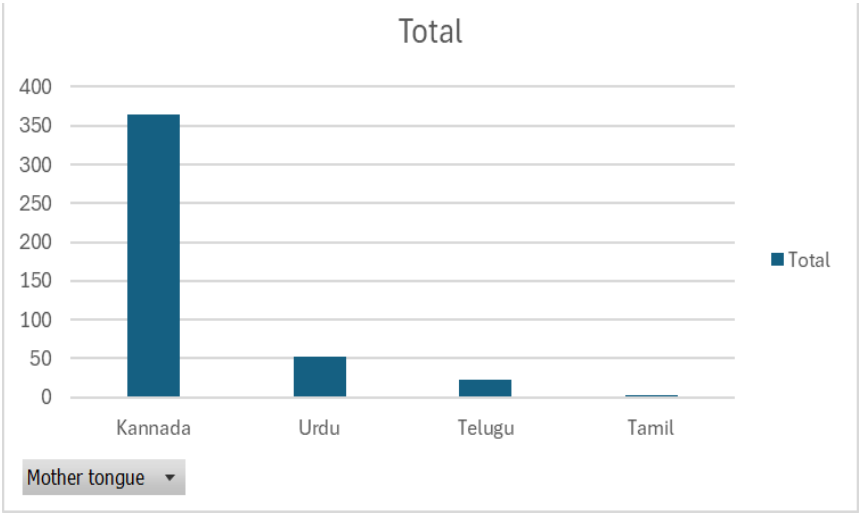
5. Caste

Caste	Count of Field1
SC	103
ST	83
Vokkaliga	67
Kuruba	53
Muslim	49
OBC	46
Christian	37
Prefer not to disclose	4
Grand Total	**442**



6. Mother tongue

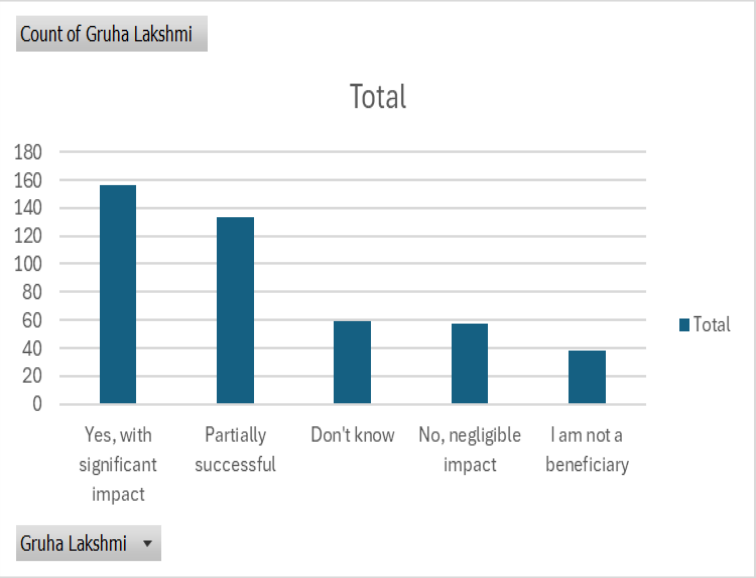
Mother tongue	Count
Kannada	364
Urdu	52
Telugu	22
Tamil	3
Grand Total	**441**



People's opinions

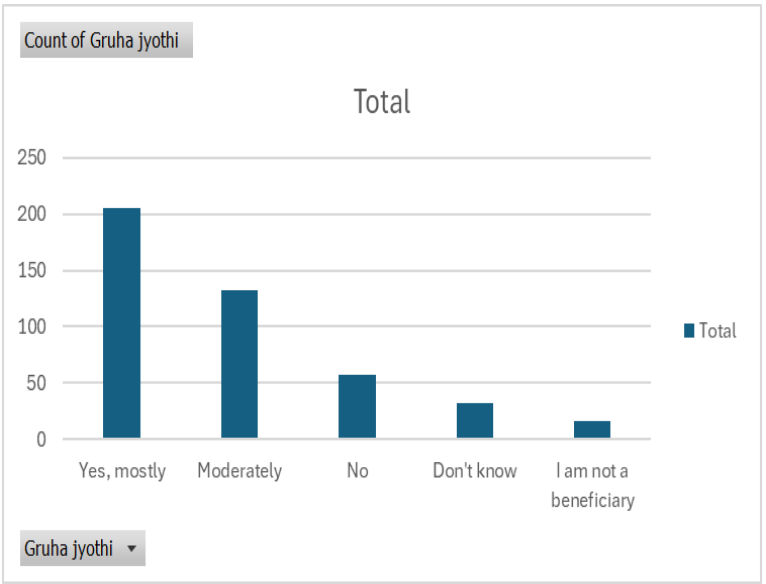
1. Has the GRUHALAKSHMI scheme shown measurable results in terms of women's empowerment and household well-beings in community?

Gruha Lakshmi	Count of Gruha Lakshmi
Yes, with significant impact	156
Partially successful	133
Don't know	59
No, negligible impact	57
I am not a beneficiary	38
Grand Total	**443**



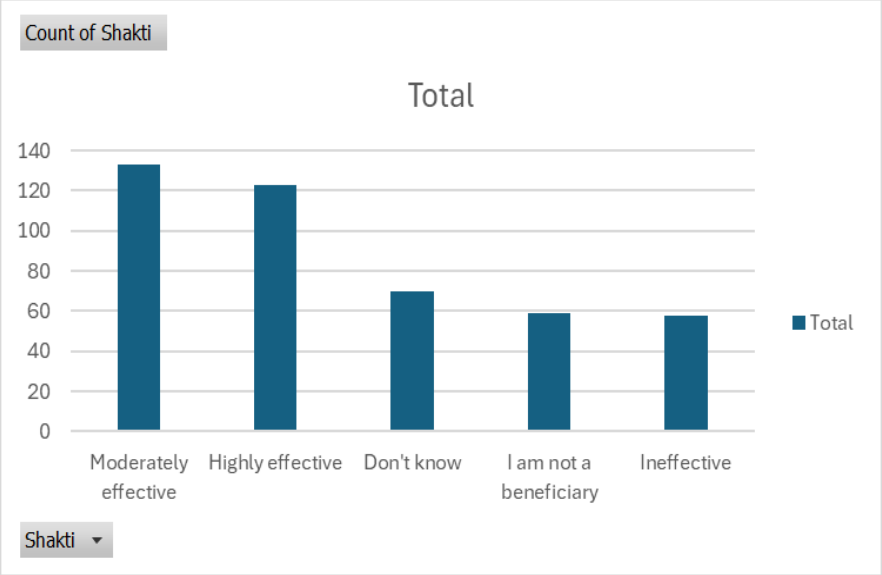
2. Has GRUHAJYOTHI scheme been helpful for your house maintainability?

Gruha jyothi	Count of Gruha jyothi
Yes, mostly	205
Moderately	133
No	57
Don't know	32
I am not a beneficiary	16
Grand Total	**443**



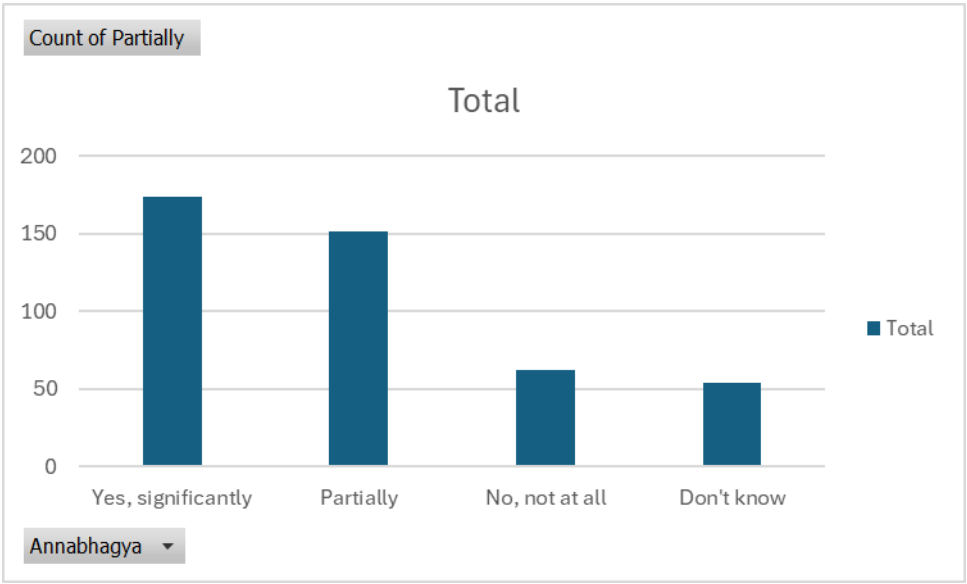
3. To what extend has the SHAKTI scheme influenced women's empowerment in your opinion?

Shakti	Count of Shakti
Moderately effective	133
Highly effective	123
Don't know	70
I am not a beneficiary	59
Ineffective	58
Grand Total	**443**



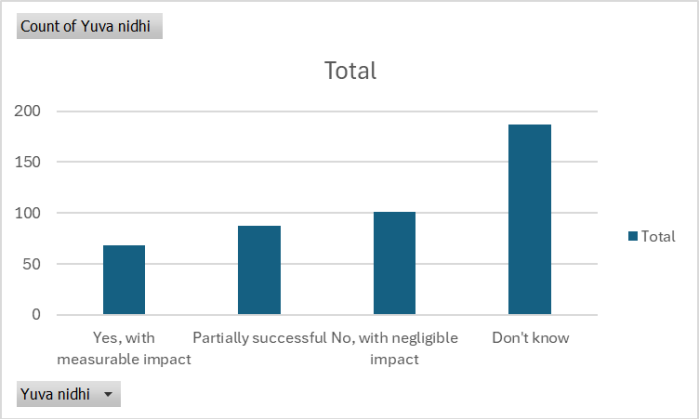
4. Has the ANNABHAGYA scheme shown measurable results in reducing poverty rates?

Annabhagya	Count of Partially
Yes, significantly	174
Partially	151
No, not at all	62
Don't know	54
Grand Total	**441**



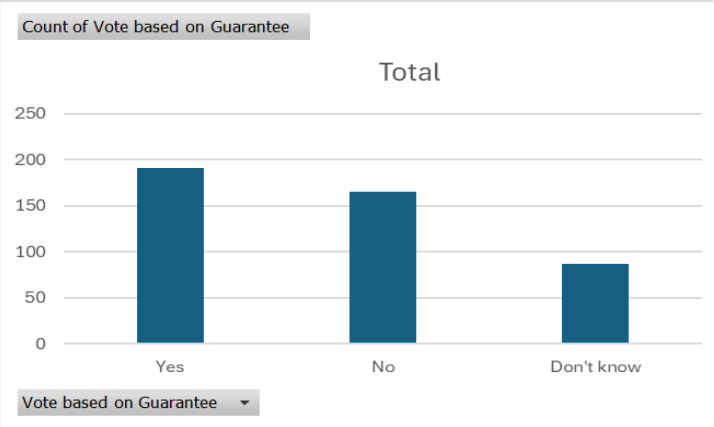
5. Has the YUVANIDHI scheme demonstrated positive outcomes in terms of economic growth?

Yuva nidhi	Count of Yuva nidhi
Yes, with measurable impact	68
Partially successful	87
No, with negligible impact	101
Don't know	187
Grand Total	**443**



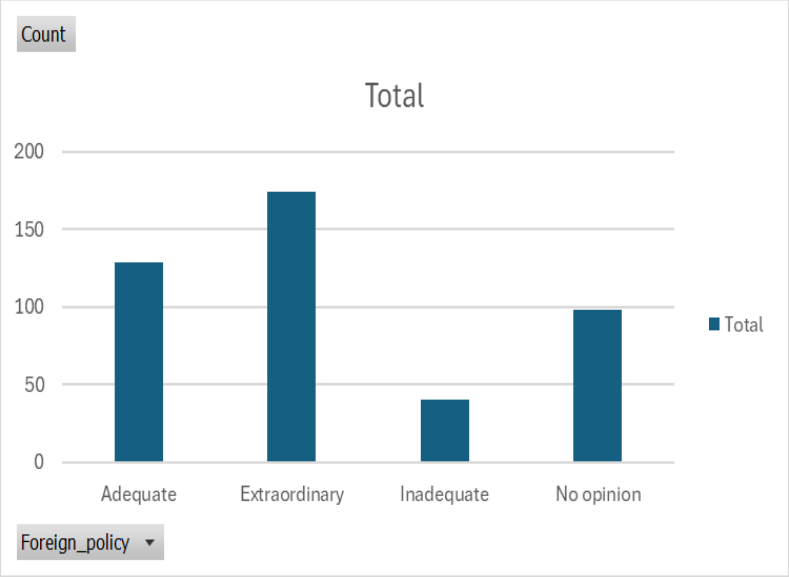
6. Vote based on Guarantee

Vote based on Guarantee	Count of Vote based on Guarantee
Yes	191
No	165
Don't know	87
Grand Total	**443**



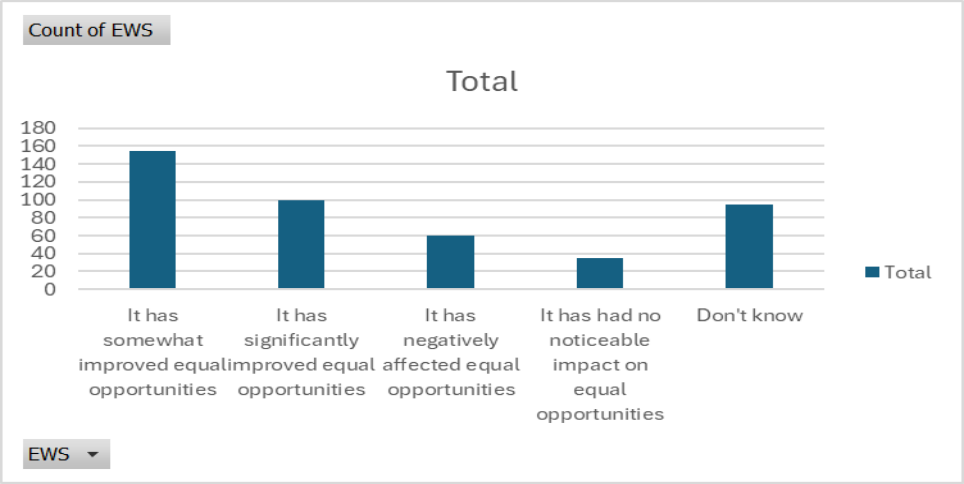
7. What is your view on prime minister modi's ability to manage partnerships with major global powers through his foreign policy?

Foreign_pol	Count
Adequate	129
Extraordinary	174
Inadequate	40
No opinion	98
Grand Total	**441**



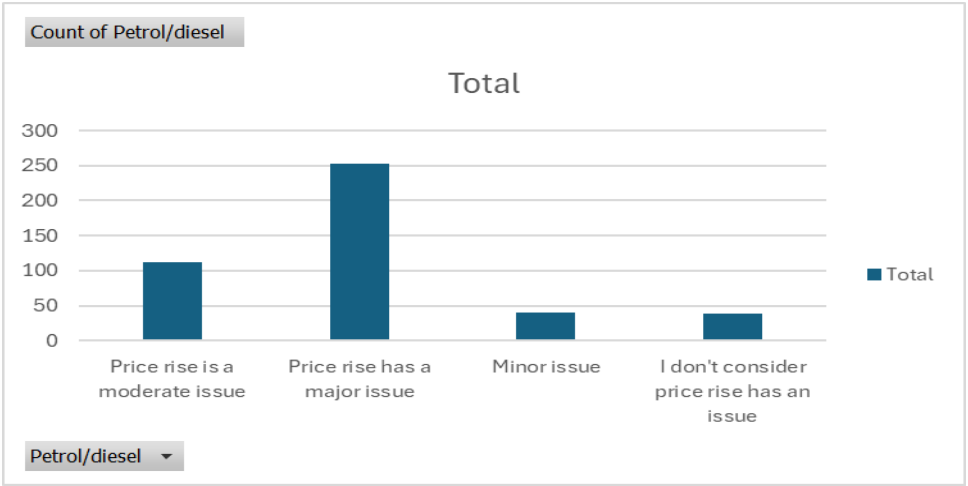
8. How would you perceive the impact of 10% reservation for EWS(economically weaker section) on equal opportunities in education and employment?

EWS	Count of EWS
It has somewhat improved equal opportunities	154
It has significantly improved equal opportunities	99
It has negatively affected equal opportunities	60
It has had no noticeable impact on equal opportunities	35
Don't know	95
Grand Total	**443**



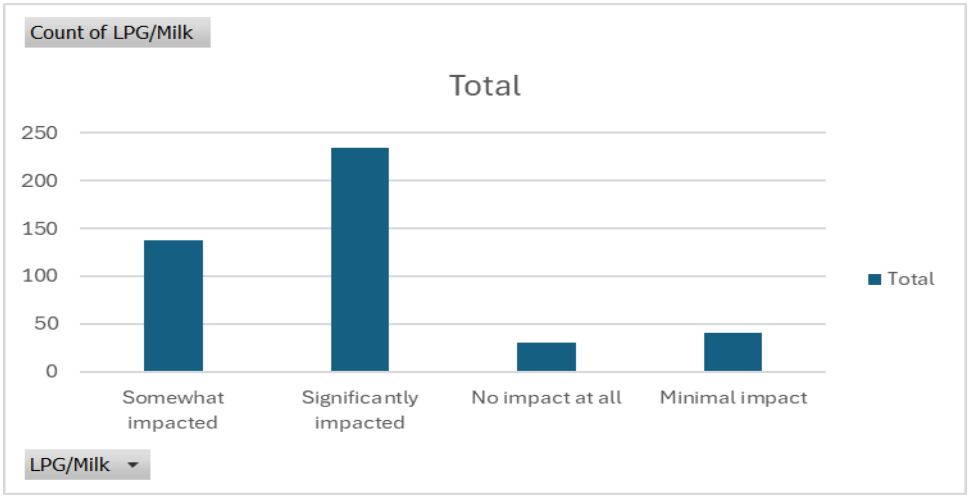
9. How do you perceive the issue of price rise in petrol, diesel in your locality?

Petrol/diesel	Count of Petrol/diesel
Price rise is a moderate issue	112
Price rise has a major issue	253
Minor issue	40
I don't consider price rise has an issue	38
Grand Total	**443**



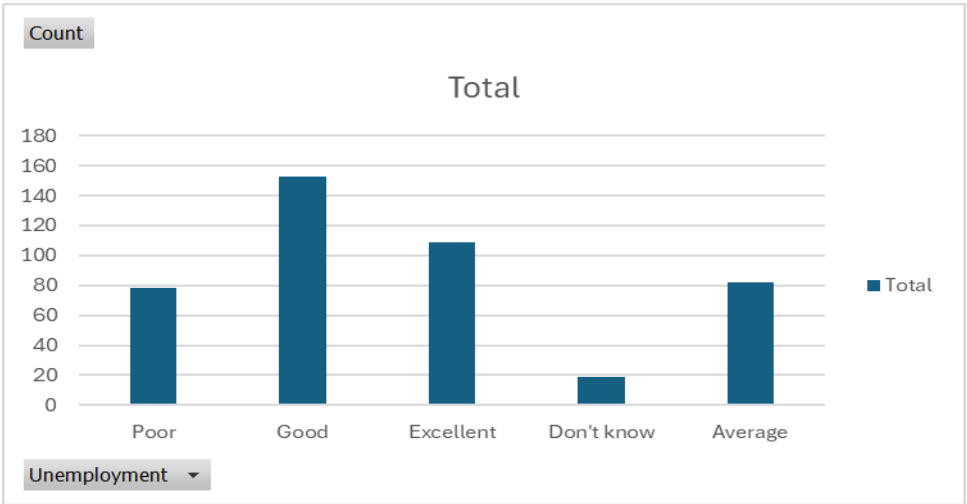
10. How has the rise in LPG gas, milk, mustard oil prices affected your household budget?

LPG/Milk	Count of LPG/Milk
Somewhat impacted	138
Significantly impacted	234
No impact at all	30
Minimal impact	41
Grand Total	**443**



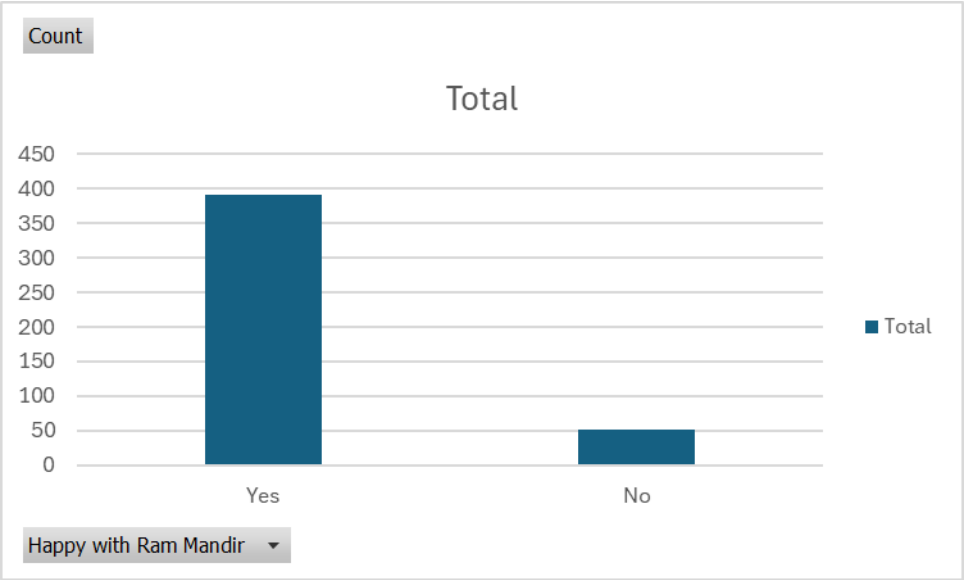
11. How would you rate the Modi government's performance in addressing unemployment concerns?

Unemployment	Count
Poor	78
Good	153
Excellent	109
Don't know	19
Average	82
Grand Total	**441**



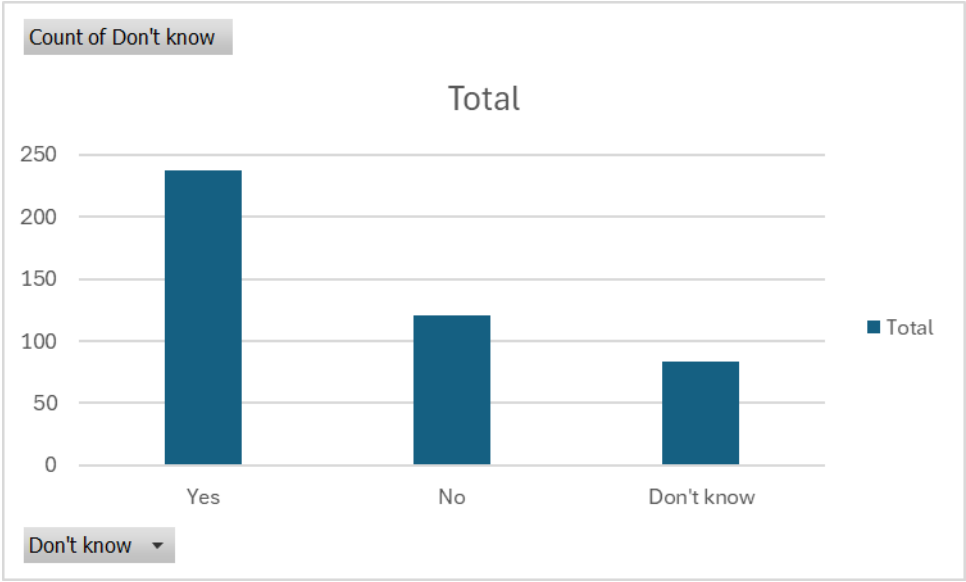
12. Are you feeling happy with the construction of ram mandir and with the PM's active participation in that?

Happy with Ram Mandir	Count
Yes	392
No	51
Grand Total	**443**



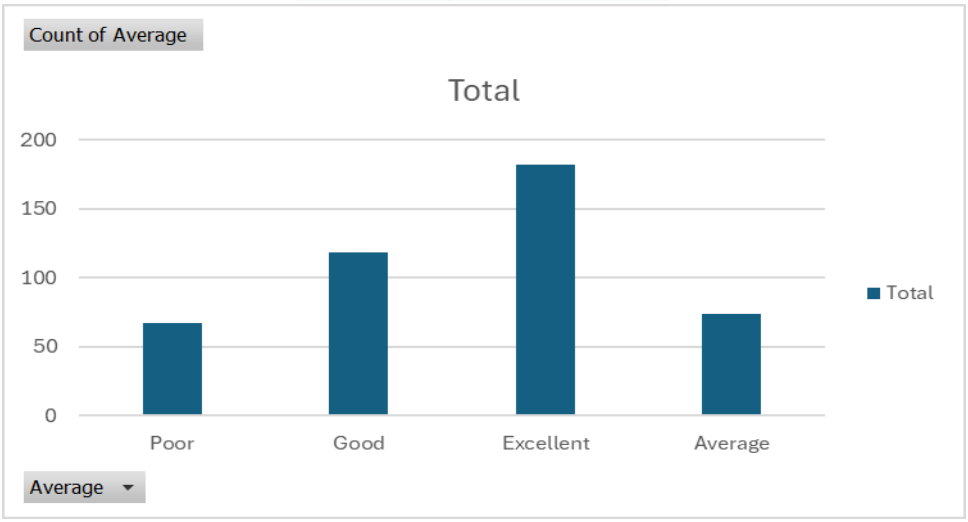
13. Is there any possibility of India becoming a "HINDU RASHTRA" due to the construction of ram mandir?

Don't know	Count of Don't know
Yes	237
No	121
Don't know	83
Grand Total	**441**



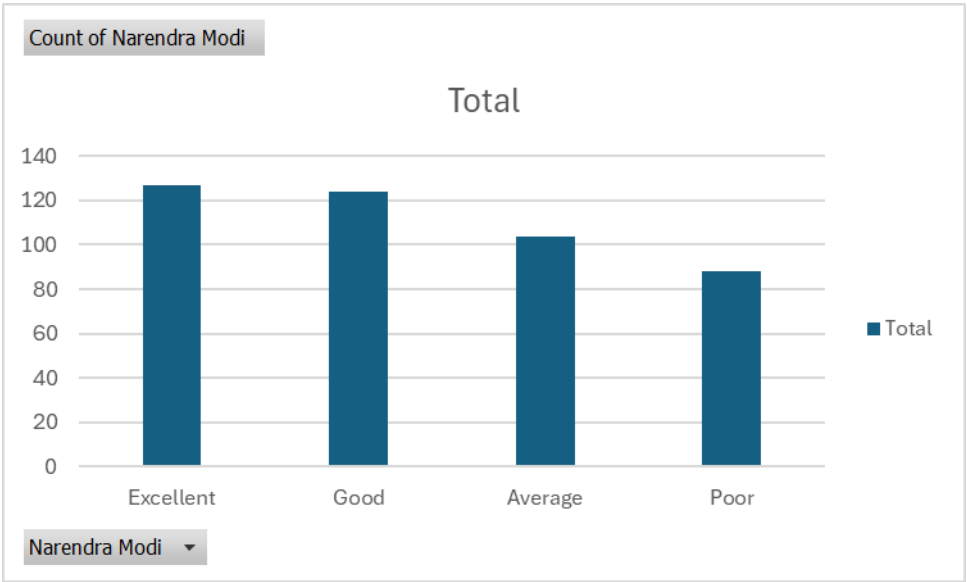
14. How would you rate Siddaramaiah's performance as chief minister?

Average	Count of Average
Poor	67
Good	118
Excellent	182
Average	74
Grand Total	**441**



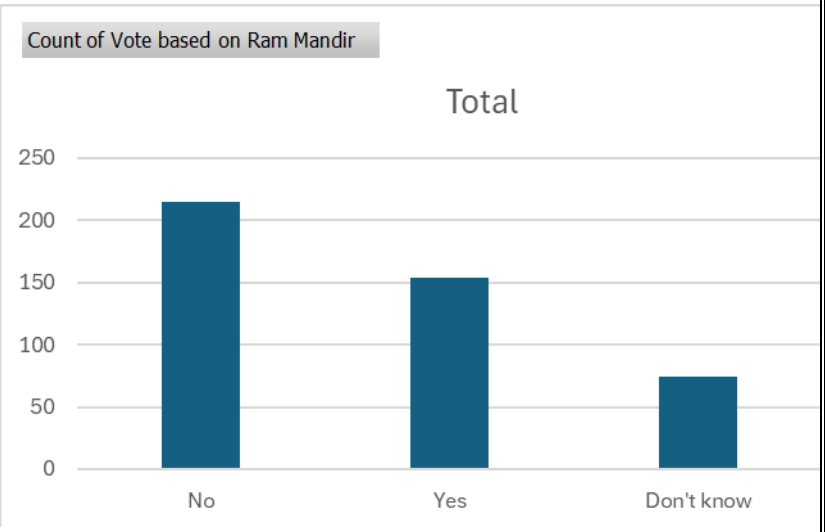
15. How would you rate Narendra modi's performance as prime minister?

Narendra Modi	Count of Narendra Modi
Excellent	127
Good	124
Average	104
Poor	88
Grand Total	**443**



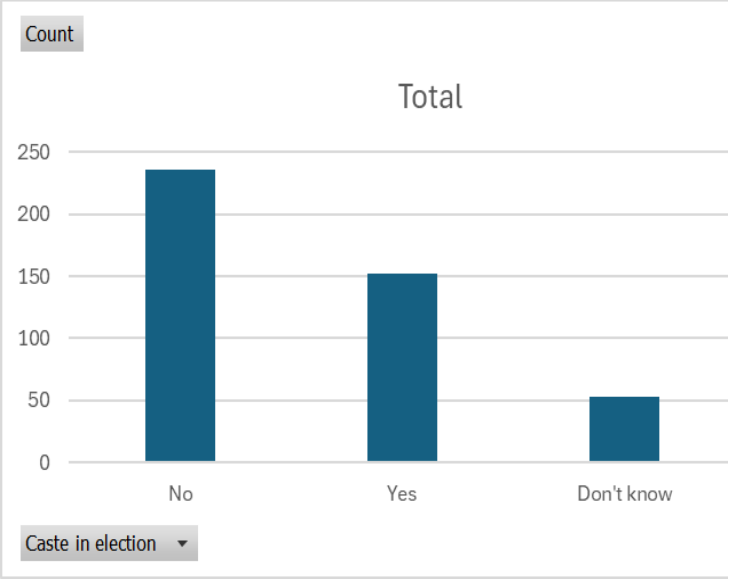
16. Vote based on Ram Mandir

Vote based on Ram Mandir	Count of Vote based on Ram Mandir
No	215
Yes	154
Don't know	74
Grand Total	**443**



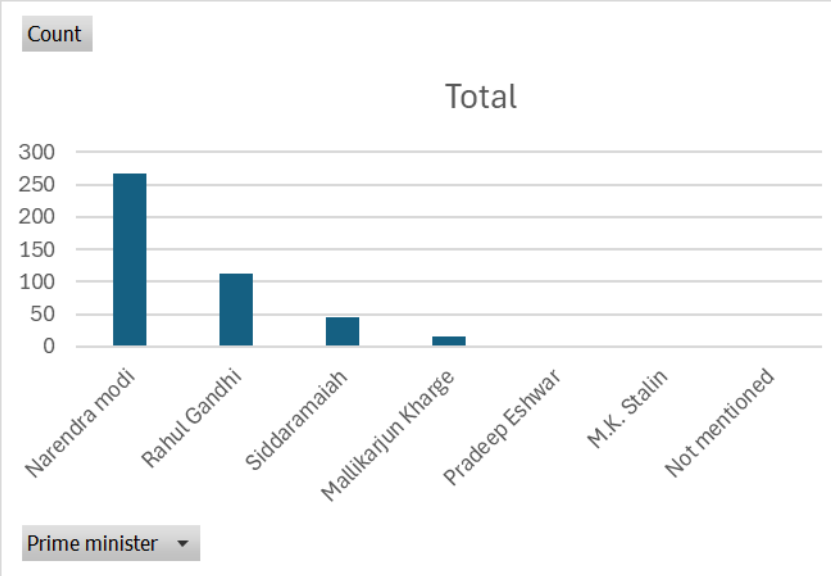
17. Caste plays a major role in elections

Caste in electio	Count
No	236
Yes	152
Don't know	53
Grand Total	**441**



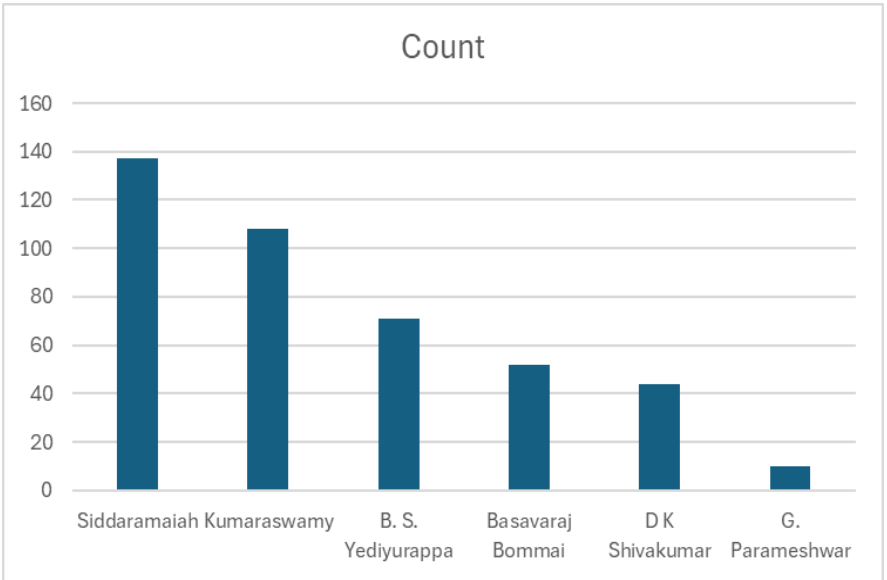
18. better choice for prime minister

Prime minister	Count
Narendra modi	268
Rahul Gandhi	112
Siddaramaiah	45
Mallikarjun Kharge	15
Pradeep Eshwar	1
M.K. Stalin	1
Not mentioned	1
Grand Total	**443**



19. better choice for chief minister

Chief minister	Count
Siddaramaiah	137
Kumaraswamy	108
B. S. Yediyurappa	71
Basavaraj Bomm	52
D K Shivakumar	44
G. Parameshwar	10



CHAPTER 6

Evaluation

6.1 Performance of the models

Below is the output provided when tested with any random tuple from the data frame. Further we have also provided an option to test the models, with newly provided data.

```
Accuracy of DecisionTree classification is 86.07594936708861%.
BJP / JDS: 44.30%
Congress: 54.43%
Other: 1.27%
```

```
-------->Accuracy of RandomForest is 84.81012658227847%.
Congress: 56.96%
BJP / JDS: 43.04%
```

```
The accuracy of the Logistic Regression model is 55.69620253164557%.
Congress: 81.01%
BJP / JDS: 17.72%
Other: 1.27%
```

Our project demonstrates commendable performance across multiple machine learning algorithms. The Decision Tree model yields accurate predictions, offering a clear and interpretable decision-making process. The Random Forest algorithm further improves accuracy, leveraging the strength of ensemble learning to handle diverse data patterns.

6.2 Hypothesis Statement and machine learning analysis results

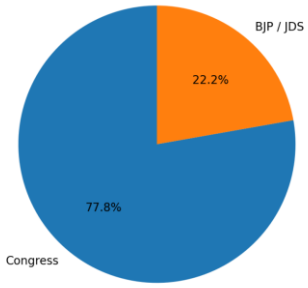
Hypothesis 1: Female voters prefer the congress party due to its focus on women-centric schemes like shakti and Gruha Lakshmi.

RandomForest

Congress: 77.78%

BJP / JDS: 22.22%

--->Accuracy is 88.88888888888889%.



Female voters prefer the Congress party due to its focus on schemes like Shakthi and Gruha Lakshmi. The random forest model, with an accuracy of 88.8%, considering features as both shakti and gruha lakshmi schemes predicts that the Congress party receives approximately 77.78% of the female votes, while the BJP / JDS receives approximately 22.2% of the female votes. Based on the data and observations, it can be concluded that our assumption is correct, and our hypothesis has been proven successful

Hypothesis 2: Voters choose the congress party due to the guaranteed benefits provided by the congress government.

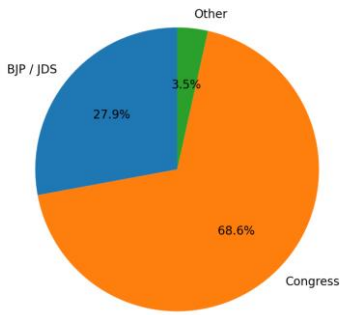
RandomForest

BJP / JDS: 27.91%

Congress: 68.60%

Other: 3.49%

--->Accuracy is 84.88372093023256%.



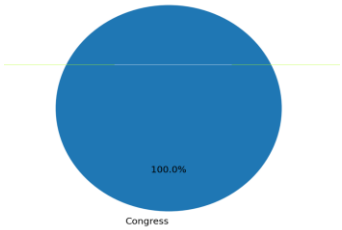
The random forest model, with an accuracy of 84.8%, considering features as all five guarantees, predicts that the Congress party receives approximately 68.6% of the votes, while the BJP / JDS receives approximately 27.9% of the votes, and others then these parties has 3.49% of the votes, Based on the data and observations, it can be concluded that our assumption "Voters choose the congress party due to the guaranteed benefits provided by the congress government" is correct, and our hypothesis has been proven successful

Hypothesis 3: Congress party attract Muslim voters due to concerns related to the hijab and the concept of a Hindu Rashtra, and Social Welfare Policies advocated by congress.

```
RandomForest

Congress: 100.00%

--->Accuracy is 80.0%.
```



The random forest model, with an accuracy of 80%, considering features Hijab, Hindu Rastra and there voting preference, predicts that the Congress party receives approximately 100% of the votes, Based on the data and observations, it can be concluded that our assumption is correct, and our hypothesis has been proven successful

Hypothesis 4: Hindu voters tend to vote for the BJP due to its stance on the construction of Ram Mandir in Ayodhya.

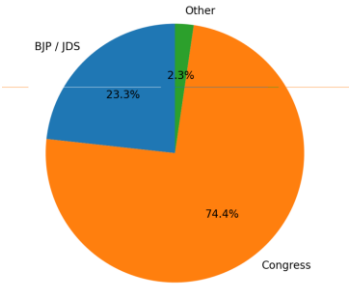
```
RandomForest

BJP / JDS: 23.26%

Congress: 74.42%

Other: 2.33%

--->Accuracy is 75.5813953488372%.
```



The random forest model, with an accuracy of 75.5%, considering features as Religion, Vote based on Ram mandir and there voting preference, predicts that the Congress party receives approximately 74.42% of the votes, while the BJP / JDS receives approximately 23.26% of the votes, and others then these parties has 2.33% of the votes, Based on the data and observations, it can be concluded that our assumption is not correct,

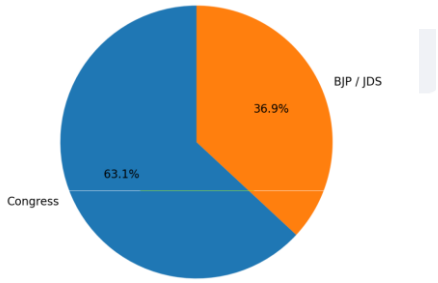
Hypothesis 5: The failure of the Modi government in controlling essential commodity (LPG gas, milk, mustard oil, petrol, diesel) will effect the voting behaviour of low-income and middle-class households.

DecisionTree

Congress: 63.10%

BJP / JDS: 36.90%

--->Accuracy is 63.095238095238095%.



The Decision tree model, with an accuracy of 63%, considering features as price rise, monthly income and there voting preference, predicts that the Congress party receives approximately 63.10%% of the votes, while the BJP / JDS receives approximately 36.9% of the votes, Based on the data and observations, it can be concluded that our assumption is correct, as most of the middle-class and poor people choose congress party.

Hypothesis 6: Unemployment is a critical concern for the youth population and influence their voting decisions.

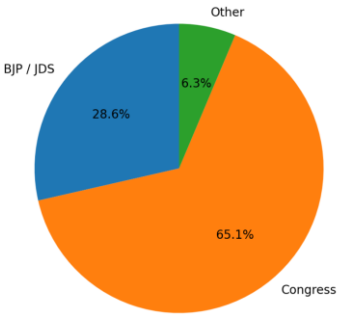
RandomForest

BJP / JDS: 28.57%

Congress: 65.08%

Other: 6.35%

--->Accuracy is 73.01587301587301%.



The random forest model, with an accuracy of 73%, considering features as age, education, monthly income, last time vote and there voting preference, predicts that the Congress party receives approximately 65%% of the votes, while the BJP / JDS receives approximately 28.5% of the votes, and others then these parties has 6.35% of the votes, Based on the data and observations, it can be concluded that our assumption is correct, as most of the choose congress party considering the unemployment.

Hypothesis 7: poor and middle-class families like Gruha Jyoti and Gruha Lakshmi scheme and influence their voting decisions.

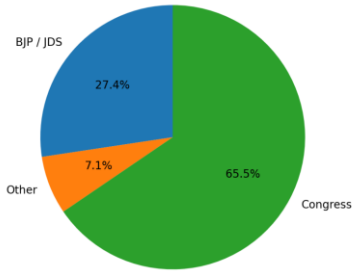
RandomForest

BJP / JDS: 27.38%

Other: 7.14%

Congress: 65.48%

--->Accuracy is 77.38095238095238%.



The random forest model, with an accuracy of 77%, considering features as Gruha Jyoti and Gruha Lakshmi scheme, and there voting preference, predicts that the Congress party receives approximately 65.4%% of the votes, while the BJP / JDS receives approximately 27.38% of the votes, and others then these parties has 7.14% of the votes, Based on the data and observations, it can be concluded that our assumption is correct, as most of the choose congress party considering the schemes provided by congress government.

Hypothesis 8: The Karnataka BJP's decision to rejig the EWS(economically weaker section) quota and allocate part of it to Vokkaligas and Lingayats may impact voting preferences.

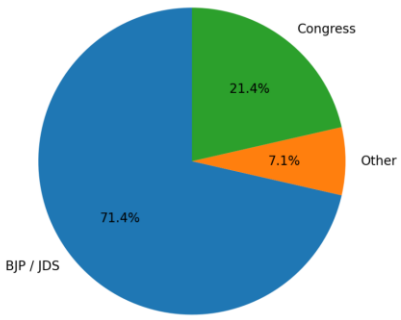
RandomForest

BJP / JDS: 71.43%

Other: 7.14%

Congress: 21.43%

--->Accuracy is 78.57142857142857%.



The random forest model, with an accuracy of 77%, considering features as EWS, caste, Monthly income and there voting preference, predicts that the BJP / JDS party receives approximately 71.4%% of the votes, while the Congress receives approximately 21.43% of the votes, and others then these parties has 7.14% of the votes, Based on the data and observations, it can be concluded that our assumption is correct, as most of the choose BJP / JDS party considering the decision to rejig the EWS(economically weaker section) quota and allocate part of it to Vokkaligas and Lingayats.

Hypothesis 9: Youth voters are inclined to support the BJP due to its positions on foreign policy decisions and international relations.

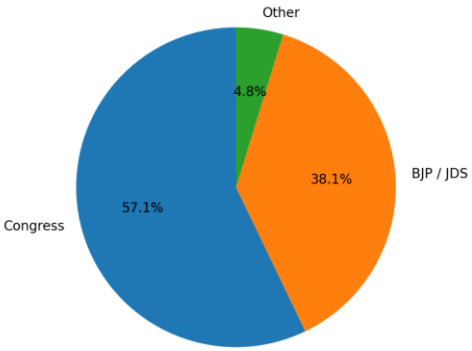
RandomForest

Congress: 57.14%

BJP / JDS: 38.10%

Other: 4.76%

--->Accuracy is 82.53968253968253%.



The random forest model, with an accuracy of 82.5%, considering features as foreign policy and there voting preference, predicts that the BJP / JDS party receives approximately 38.1% of the votes, while the Congress receives approximately 57.14% of the votes, and others then these parties has 4.76% of the votes, Based on the data and observations, it can be concluded that our assumption is not correct.

6.3 Projection of Final Output

Based on the hypotheses and the available data, I will now project the final output.

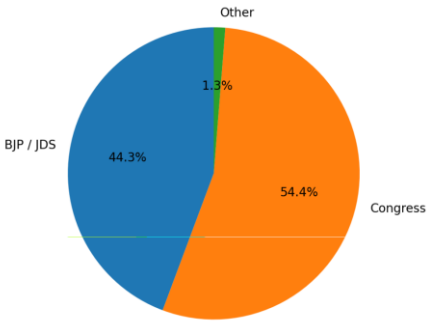
DecisionTree

BJP / JDS: 44.30%

Congress: 54.43%

Other: 1.27%

--->Accuracy is 86.07594936708861%.



Considering the result we getting from Decision tree model with the accuracy of 86%, congress may get 54.43% of votes in Kolar, BJP / JDS party may get 44.3% and other may get 1.27% of vote share, based on these data we can conclude that in 2024 elections congress will win in Kolar constitution.

CHAPTER 7

SAMPLE CODE

```
# Import necessary libraries

import pandas as pd

from sklearn.metrics import accuracy_score

from collections import Counter

import matplotlib.pyplot as plt

# Create a DataFrame with your data

data = pd.read_excel("EP_data.xlsx", sheet_name="Train_data")

data_test=pd.read_excel("EP_data.xlsx", sheet_name="Test_data1")

# Define features and target

features_train = data[['Gruha lakshmi0','Gruha Lakshmi1','Shakthi0','Shakti1','Gruha
jyothi0','Gruha jyothi1','Annabhagya0','Annabhagya1','Yuva nidhi0','Yuva nidhi1','Vote based on
Guarantee1','Last Vote1']]

target_train = data['Vote']

data.shape

features_test=data_test[['Gruha lakshmi0','Gruha Lakshmi1','Shakthi0','Shakti1','Gruha
jyothi0','Gruha jyothi1','Annabhagya0','Annabhagya1','Yuva nidhi0','Yuva nidhi1','Vote based on
Guarantee1','Last Vote1']]

target_test=data_test['Vote']

#Logistic Regression model

# Initialize the Logistic Regression model

from sklearn.linear_model import LogisticRegression

model = LogisticRegression()

# Train the model
```

```
all_guarantee_LR=model.fit(features_train, target_train)
# Make predictions
predictions = model.predict(features_test)
print(predictions )
# Calculate the accuracy of the model
accuracy = accuracy_score(target_test, predictions)
# Print the accuracy
print(f"---->The accuracy of the Logistic Regression model is {accuracy * 100}%.")
Counter(predictions)

#Decision Tree Classifier model
from sklearn.tree import DecisionTreeClassifier
# Create a Decision Tree Classifier object
dtree = DecisionTreeClassifier()
# Train the Decision Tree Classifier
all_guarantee_tree=dtree.fit(features_train, target_train)
# Predict the response for the test dataset
target_pred = dtree.predict(features_test)
# Calculate the accuracy of the model
accuracy = accuracy_score(target_test, target_pred)
# Print the accuracy
print(f"Accuracy of DecisionTree classification is {accuracy * 100}%.")
Counter(target_pred)
```

```
# Random Forest Classifier model

from sklearn.ensemble import RandomForestClassifier

# Create a Random Forest Classifier object

clf = RandomForestClassifier(n_estimators=100)

# Train the Random Forest Classifier

all_guarantee = clf.fit(features_train,target_train)

# Predict the response for the test dataset

target_pred = clf.predict(features_test)

print(target_pred)

# Calculate the accuracy of the model

accuracy = accuracy_score(target_test, target_pred)

# Print the accuracy

print(f"-------->Accuracy of RandomForest is {accuracy * 100}%.")

Counter(target_pred)

#model creation

import pickle

pickle_out=open("all_guarantee_LR.pkl","wb")

pickle.dump(all_guarantee_LR,pickle_out)
```

CHAPTER 8

Screenshots

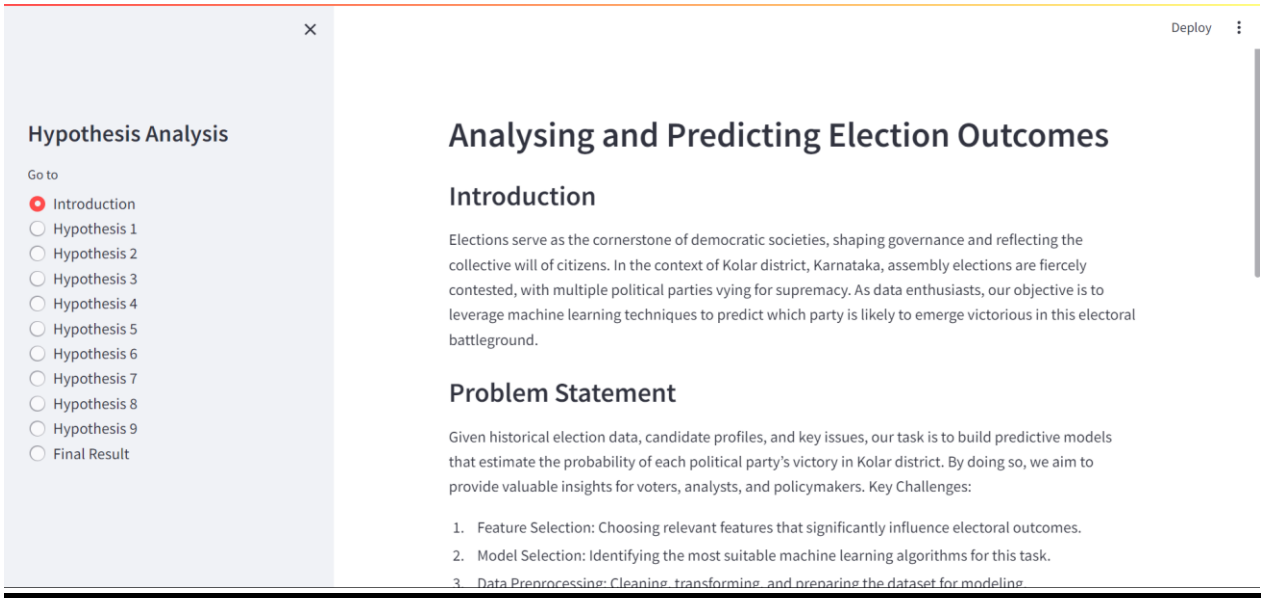


Fig: Main page

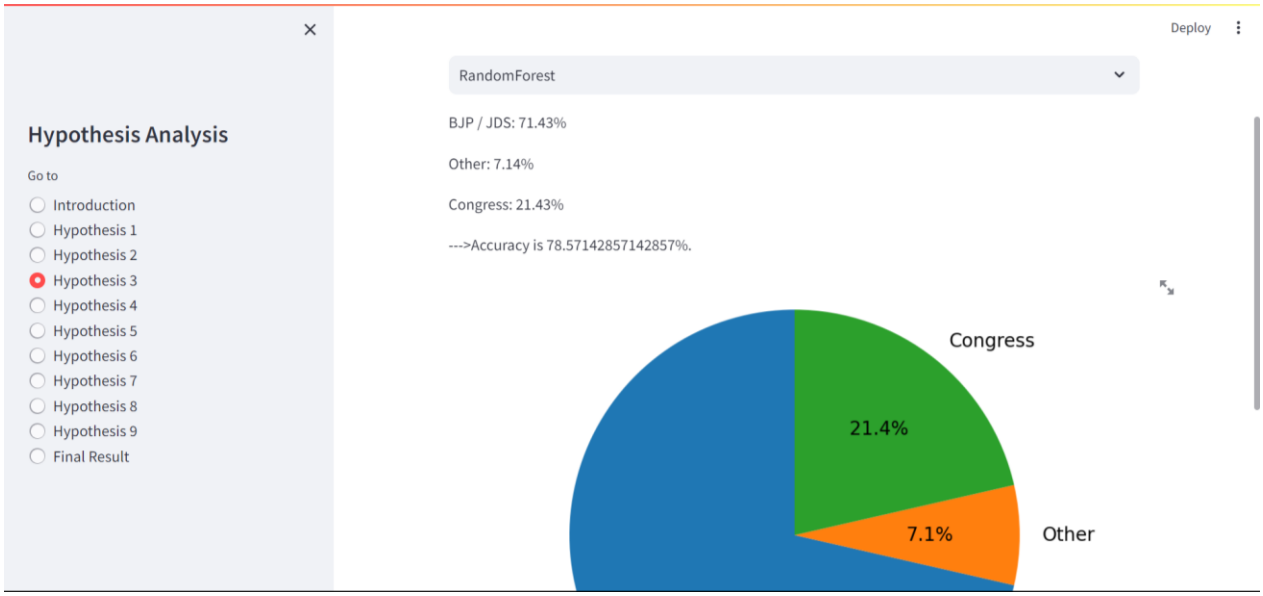


Fig: Hypothesis Analysis page

CHAPTER 9

CONCLUSION AND FUTURE SCOPE

In conclusion, predicting election results is a complex task that involves analysing various factors and utilizing advanced techniques. By examining historical data, demographic factors, opinions, machine learning models, and external factors, one can develop more informed predictions about election outcomes. Considering the result we getting from Decision tree model with the accuracy of 86%, congress may get 54.43% of votes in Kolar, BJP / JDS party may get 44.3% and other may get 1.27% of vote share, based on these data we can conclude that in 2024 elections congress will win in Kolar constitution.

However, it is important to remember that predicting election results is not an exact science, and uncertainties and variables can always impact the outcome. Therefore, it is crucial to approach election predictions with caution and consider multiple sources of information to get a comprehensive understanding of the political landscape. we embarked on a comprehensive journey through the stages of data collection, analysis, and model training to forecast election outcomes. Our methodology involved designing and distributing questionnaires to gather relevant data, which served as the foundation for our predictive models. By utilizing a machine learning model implemented in Python language. I leveraged the power of Streamlit libraries to create a user-friendly interface for visualizing and interacting with the prediction results. Additionally, I utilized the Pickle library to save and load the trained model for future use.

FUTURE SCOPE

The system that has been proposed can be made more reliable and efficient by implementing other machine learning algorithms along with the ones that already have been implemented so that intrusion can be detected easily.

CHAPTER 10

REFERENCES

- A Machine Learning Based Strategy for Election Result Prediction, Publisher: IEEE
- Predicting Elections Using Twitter Data Analysis, by Saurabh Roy
- Feature Engineering for Election Result Prediction (in Python), by Thamindu Dilshan Jayawickrama
- https://towardsdatascience.com/feature-engineering-for-election-result-prediction-python-943589d89414
- https://github.com/roysaurabh1308/Election-Result-Prediction/blob/master/README.md
- https://chat.openai.com/domain_migration?next=https%3A%2F%2Fchatgpt.com%2F
- https://www.bing.com/chat?q=Microsoft+Copilot&FORM=hpcodx