ANALYSIS OF EXPECTATIONS OF COMMON PEOPLE FOR CANDIDATE IN PARLIAMENT

ANEKANT EDUCATION SOCIETY'S

TULJARAM CHATURCHAND COLLAGE OF ARTS, SCIENCE, AND COMMERCE, BARAMATI

(AUTONOMOUS)

A PROJECT REPORT ON

"ANALYSIS OF EXPECTATIONS OF COMMON PEOPLE FOR CANDIDATE IN PARLIAMENT"

SUBMITTED TO

DEPEARTMENT OF STATISTICS

BY

MISS. NISHA DADA HOLE (16692)

UNDER THE GUIDANCE OF

MRS. SHITAL B. CHOUDHAR

Assistant Professor

Department of Statistics

YEAR 2024-25

ANEKANT EDUCATION SOCIETY'S

TULJARAM CHATURCHAND COLLEGE OF ARTS, COMMERCE AND SCIENCE

BARAMATI-413102

CERTIFICATE

This is to certify that partial fulfilment of curriculum MSc-II College students Ms. Nisha Dada Hole have successfully completed the research project work in the statistics entitled "Analysis of Expectations of Common People for Candidate in Parliament." prescribed by Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati during academic year 2024-2025.

This is record of Bonafide work carried out by them under my supervision and guidance.

Mrs. Shital B. Choudhar		**Head**
Project Guide	**Examiner**	**Department of Statistics**

Date :

Place : Baramati

ACKNOWLEDGEMENT

We are satisfied with the completion of the research project work entitled " Expectations of Common People for candidate in Parliament." At Department of Statistics of " Tuljaram Chaturchand College, Baramati" during the academic year 2024-2025.

We would like to take this opportunity to thank Mrs. Shital B. Choudhar for her invaluable advice and assistance in helping us complete this project. And we also want to thank our Principal Prof. Avinash S. Jagtap, for providing us with the required facilities, constant support, and encouragement.

We also appreciate Prof. Vikas C. Kakade, Head of Statistics Department, for providing the required facilities.

We are grateful to all non-teaching staff in the statistics department for providing laboratory at any time during the project, as well as to our fellow department college students for their support and encouragement.

INDEX

ABSTRACT

India's journey towards becoming a globally developed nation by 2047 requires capable political leadership that aligns with the expectations of its citizens. With the implementation of delimitation legislation in 2025, increasing the number of parliamentary seats, the role of political parties and aware voters in shaping governance becomes critical. This study explores the expectations of common people regarding the qualities, skills, and attributes they seek in candidates who will represent them in Parliament.

Using a dataset collected from 520 respondents through a structured questionnaire, we conducted an extensive Exploratory Data Analysis (EDA) to understand demographic trends and key patterns in voter preferences. Chi-square tests for independence were performed to identify significant associations between demographic factors and political expectations. Machine Learning (ML) models were built to predict voter preferences based on various attributes, providing insights into the most influential factors driving electoral choices. Additionally, sentiment analysis was conducted on open-ended responses to gauge public sentiment regarding the political system and candidate expectations.

The findings highlight the critical role of candidate education, governance experience, transparency, and inclusivity in voter decision-making. The results emphasize the importance of informed voting and the need for political parties to adopt strategic candidate selection and training mechanisms. By integrating statistical analysis with machine learning and sentiment analysis, this study provides a data-driven perspective on the evolving expectations of Indian voters, contributing to a more accountable and visionary political framework for the country's future.

Keywords:

Voter Expectations, Political Representation, Electoral Preferences, Public Opinion, Democratic Participation, Governance, Representative Democracy.

MOTIVATION

In a democratic society, the relationship between elected representatives and the public plays a fundamental role in shaping governance and policy outcomes. Members of Parliament are expected to represent the voices of the people, yet there often exists a disconnect between what citizens expect and what their representatives deliver. This research is driven by the need to understand and bridge that gap by exploring the expectations that common people hold for their parliamentary representatives.

With increasing political awareness and access to information, citizens today are more vocal and participative in the democratic process. However, there is still a lack of systematic understanding of what people truly prioritize in their leaders. Are they looking for integrity, development, accessibility, or issue-based policymaking? By conducting a structured analysis of public opinion, this study seeks to provide insights that can inform political communication, candidate selection, and policy focus.

The motivation behind this research stems from a desire to contribute meaningfully to democratic strengthening. By giving voice to the electorate’s concerns and expectations, this study hopes to support a more accountable and responsive parliamentary system. Moreover, it serves as a platform for encouraging greater civic engagement, fostering transparency, and ultimately promoting better governance.

INTRODUCTION

The India approaches its centenary of independence in 2047, the nation envisions a future where it emerges as a fully developed country on the global stage. A significant step towards this transformation is the implementation of delimitation legislation in 2025, which will increase the number of parliamentary seats to between 750 and 800. This expansion will shape the governance framework for the next 25 years, making it crucial to ensure that political leaders possess the necessary calibre, skills, and vision to steer India toward international development. The question that arises is whether the elected representatives will have the competence required to guide the country in this new era. To achieve this, political parties must take deliberate steps in selecting, training, and supporting candidates who can effectively lead India toward its goal of becoming a globally influential nation.

The role of political parties in this process is paramount. They serve as the primary gatekeepers of leadership, responsible for selecting individuals who will represent the country both nationally and internationally. The selection of candidates should not be based merely on political loyalty or popularity but on merit, governance experience, and ethical integrity. Parties must implement structured training programs that equip leaders with essential knowledge in policy-making, economic development, global diplomacy, and effective communication. Additionally, fostering inclusivity by promoting women, minorities, and young leaders will ensure a diverse and dynamic political landscape. Transparency and accountability in governance should also be emphasized, as they help build trust among the electorate and strengthen democratic values. Moreover, politicians must have a clear vision for India's future and embrace innovative approaches to address economic, technological, and social challenges.

However, the responsibility of shaping a capable leadership does not rest solely on political parties. The role of aware voters is equally crucial in this transformative period. In today's digital age, the electorate is more informed, educated, and analytically equipped than ever before. Voters have the power to elect leaders who not only have the calibre to represent India globally but also align with their expectations and aspirations. Voter awareness directly influences the quality of leadership that emerges from the electoral process, making civic education a key aspect of strengthening democracy.

For this reason, voter education initiatives should be enhanced to ensure that citizens are well-informed about the qualifications, policies, and ideologies of candidates. Civic education programs must focus on providing a deeper understanding of the political system, voter rights, and responsibilities. Easy access to credible and comprehensive information about candidates should be ensured through transparent media coverage, public debates, and digital platforms. Encouraging active political engagement through voter registration drives, community discussions, and participation campaigns will further strengthen democracy. Additionally, promoting critical thinking skills will help voters differentiate between genuine and misleading information, enabling them to make well-informed choices.

The purpose of this study is to examine the expectations of common people regarding the candidates who will lead India in the coming years. By analyzing public opinions on the qualities, skills, and attributes that voters prioritize in politicians, this research aims to provide insights into how political parties can better align their candidate selection and training processes with the aspirations of the electorate. Furthermore, it seeks to explore the role of voter awareness in shaping political leadership and the impact of informed voting on governance quality. Understanding these dynamics will contribute to building a stronger democratic framework that supports India's vision of becoming a globally developed nation by 2047.

REVIEW OF LITERATURE

1) Evaluating the Role of Media in Shaping Political Opinions & Political Preferences: A Statistical Critical Research Analysis

Author: [Bikram Kumar Sahu]
Year of Publication: [March 2025]

Introduction

The research paper Evaluating the Role of Media in Shaping Political Opinions investigates the influence of media on public perception and political decision-making. In an era of digital connectivity, media channels, including television, newspapers, and social media, significantly impact voters' opinions. This literature review critically examines the study, its theoretical framework, methodology, and statistical analysis while placing it within the broader research landscape.

Theoretical Background

The paper draws upon existing theories of mass communication and political psychology, particularly agenda-setting theory, framing theory, and cultivation theory. Agenda-setting theory posits that media influences which issues are deemed important by the public. Framing theory suggests that the way news is presented shapes audience interpretation. Cultivation theory argues that prolonged media exposure alters individuals' perception of reality. These theoretical perspectives provide a robust foundation for analyzing media's impact on political beliefs.

Research Methodology

The study employs a quantitative approach, utilizing survey data to assess public attitudes towards political entities and policies. The sample size is adequately large, ensuring statistical power and generalizability. Respondents were asked questions regarding their media consumption habits, trust in different news sources, and the extent to which media influenced their voting decisions. The survey design appears methodologically sound, incorporating both closed-ended and Likert-scale questions to quantify political opinion shifts.

Statistical Analysis

The research utilizes various statistical techniques to analyze the collected data:

Descriptive Statistics: The study provides means, standard deviations, and frequency distributions of key variables, offering insights into general media consumption trends.

Correlation Analysis: A Pearson correlation test is employed to assess the relationship between media exposure and political opinion formation. Results indicate a moderate positive correlation, suggesting that increased media consumption correlates with stronger political biases.

Regression Analysis: The paper presents a multiple regression model to predict political leanings based on media consumption. Independent variables include media type, frequency of engagement, and trust level. The model demonstrates statistical significance ($p < 0.05$) and an R-squared value indicating a moderate to strong explanatory power.

Chi-square Tests: Used to examine associations between categorical variables, such as political affiliation and preferred media sources. Results suggest significant differences in media trust among political groups.

Findings and Discussion

The findings reinforce existing literature on media influence in politics. The study confirms that individuals who consume partisan media tend to develop more polarized views. Furthermore, it highlights the growing role of social media as a primary news source, particularly among younger demographics. One of the paper's key contributions is its examination of misinformation and media trust, showing that skepticism towards mainstream media correlates with a higher likelihood of consuming alternative news sources.

Strengths and Limitations

Strengths:

A comprehensive dataset with a diverse respondent base.
Use of multiple statistical techniques to ensure robust analysis.
Integration of theoretical perspectives with empirical findings.

Limitations:

Self-reported data may introduce response biases.
The cross-sectional nature of the study limits causality claims.
Potential underrepresentation of certain demographic groups.

Conclusion

The research paper makes a valuable contribution to understanding how media shapes political attitudes. Its empirical findings align with established theories, reinforcing the significance of media literacy and responsible journalism. Future studies could explore longitudinal effects and cross-cultural comparisons to enhance our understanding of media's evolving role in politics.

2) How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science

Author: [Gary King, *New York University*]

Year of Publication: [2022]

Introduction

The research paper How Not to Lie with Statistics critically examines the misuse and misinterpretation of statistical data in various fields, including media, business, and research. The study highlights how statistics, when presented inaccurately or selectively, can mislead audiences and distort facts. This literature review evaluates the paper's theoretical approach, research methodology, statistical analysis, and its significance in the field of data interpretation.

Theoretical Background

The paper is grounded in statistical theory, particularly focusing on:

Descriptive Statistics and how they can be manipulated to mislead conclusions.

Sampling Bias and its role in producing unreliable results.

Misinterpretation of Graphs and Visuals, showing how visual representations can distort data.

Correlation vs. Causation, explaining why statistical relationships do not always indicate a cause-and-effect link.These concepts provide a strong foundation for understanding the ethical application of statistical methods.

Research Methodology

The study employs a critical analysis approach, reviewing multiple cases where statistical data has been misrepresented. The methodology includes:

Case studies from media, marketing, and political surveys.

Examination of real-world examples where statistical fallacies have influenced public perception.

A review of published reports and their use of misleading statistics.The research is qualitative in nature, focusing on identifying patterns in statistical misrepresentation rather than conducting new experiments or surveys.

Statistical Analysis

The paper provides several statistical techniques to illustrate manipulation tactics:

Descriptive Statistics: Demonstrates how measures of central tendency (mean, median, mode) can be selectively used to support different conclusions.

Regression Analysis: Highlights instances where regression models are misused to create misleading relationships between variables.

Sampling Bias: Discusses how non-representative samples can lead to skewed results and incorrect generalizations.

Probability Theory: Examines common probability errors that mislead interpretations of risk and statistical significance.

Misleading Visual Representations: Analyzes the impact of truncated graphs, altered scales, and misleading charts on audience perception.

Findings and Discussion

The findings of the paper emphasize that statistical misrepresentation is often intentional, used to sway opinions in politics, business, and media. Key conclusions include:

Misuse of averages to distort economic and social data.

Selective sampling techniques that favor a particular narrative.

Overreliance on correlation without considering external influencing factors.

The ethical responsibility of statisticians and data analysts to ensure accurate representation of data.

Strengths and Limitations

Strengths:

Clear and engaging writing style that simplifies complex statistical concepts.

Provides real-world examples to demonstrate statistical manipulation.

Raises awareness about the ethical use of statistics.

Limitations:

Lacks quantitative data to support some claims.

Does not delve deeply into advanced statistical methodologies.

Primarily focuses on media and business applications, with limited discussion on scientific research misinterpretation.

Conclusion

The paper How Not to Lie with Statistics is an essential contribution to statistical literacy, shedding light on how data can be manipulated for different agendas. It serves as a cautionary guide for researchers, journalists, and policymakers to critically assess statistical claims before accepting them as fact. Future research could further explore quantitative methods to measure the extent of statistical misrepresentation across different fields.

3) One Nation One Election : Analyzing the Impact on Indian Polity

Author: [Ekta Basoya ,CCS university]

Year of Publication: [September 2023]

Introduction

The research paper One Nation One Election explores the concept of synchronizing elections for the Lok Sabha and State Assemblies in India. It discusses the historical context, constitutional provisions, political feasibility, and economic implications of implementing simultaneous elections across the country. This literature review critically examines the study's theoretical foundation, methodology, statistical analysis, and its broader implications in the Indian political system.

Theoretical Background

The paper is based on constitutional and legal frameworks governing elections in India. It discusses key aspects such as:

Article 83 and Article 172 of the Indian Constitution, which define the terms of Lok Sabha and State Assemblies.

The Representation of the People Act, 1951, which governs election procedures.

Election Commission's recommendations and their role in electoral reforms.

Previous cases of simultaneous elections in India, such as those held until 1967 before political instability disrupted the cycle.

Research Methodology

The study adopts a mixed-method approach, incorporating:

Historical analysis of past elections in India and their impact on governance.

Comparative study of election models from other democratic nations.

Economic assessment of election expenditures using financial reports.

Survey-based public opinion analysis on voter perception towards simultaneous elections.

Statistical Analysis

The paper employs various statistical techniques to analyze the feasibility and impact of One Nation One Election:

Election Cost Analysis: Data on election expenses from multiple sources (e.g., Election Commission reports) is used to estimate the potential savings from holding simultaneous elections.

Voter Turnout Trends: A comparative study is conducted using past election data to analyze whether synchronizing elections would affect voter participation.

Impact on Governance: Statistical comparisons are made between single-phase and multi-phase election governance efficiency, using key performance indicators such as policy implementation timelines.

Public Opinion Polls: Surveys and polling data are analyzed to gauge public support for One Nation One Election.

Findings and Discussion

The study highlights several key findings:

Economic Efficiency: Conducting a single nationwide election could significantly reduce costs related to security, logistics, and administrative expenses.

Governance Stability: Continuous elections lead to policy disruptions; a synchronized model could enhance decision-making efficiency.

Voter Turnout: The paper finds mixed evidence, with some data suggesting increased participation and others indicating possible voter fatigue.

Political Feasibility: Despite potential benefits, political challenges such as constitutional amendments and opposition from regional parties pose major hurdles.

Strengths and Limitations

Strengths:

Well-structured analysis covering constitutional, political, and economic aspects.

Use of statistical evidence to support claims.

Consideration of both benefits and challenges.

Limitations:

Limited empirical data on voter behavior in a simultaneous election scenario.

Requires further case studies from other federal nations for stronger comparative analysis.

Political resistance and regional dynamics are not analyzed in depth.

Conclusion

The research paper One Nation One Election provides a comprehensive examination of the potential advantages and challenges of synchronizing elections in India. While the statistical analysis strengthens its argument for cost efficiency and governance stability, additional empirical research and deeper political analysis are needed to assess its real-world feasibility. Future research could explore experimental models or pilot studies to evaluate the actual impact of implementing this reform.

4) Women’s Parliamentary Representation in Africa: The Impact of Democracy and Corruption on the Number of Female Deputies in National Parliaments.

Author : [Daniel Stockemer]

Introduction

This study examines the factors influencing women’s representation in African parliaments, focusing on the impact of democracy and corruption. While previous research has highlighted the importance of electoral systems and gender quotas, this study incorporates two less frequently analyzed factors—a country’s level of democracy and its corruption level. The paper argues that lower corruption levels correlate with higher percentages of women in parliament, while democracy does not necessarily enhance women’s representation.

Literature Review

The paper discusses key factors influencing women’s political representation in Africa:
Quota Provisions – Many African countries have introduced gender quotas to increase women’s parliamentary representation. For example, Burundi’s 2004 constitution mandates a 30% quota for women in parliament.
Electoral System Type – Proportional representation (PR) systems are linked to higher numbers of female deputies. Studies suggest that PR systems provide more opportunities for women compared to majoritarian systems.
Political Culture – Patriarchal norms limit women’s political engagement, with traditional values and gender inequalities serving as significant barriers.
Corruption and Governance – Countries with better corruption control tend to have higher female representation, as fairer systems provide more opportunities for women to access political power.

Methodology and Statistical Analysis

The study employs quantitative regression analysis to test the impact of democracy and corruption on women’s representation in 44 African countries. The model includes:

Key Independent Variables:
Democracy Level: Measured using an index that considers procedural and substantive democracy.
Corruption Control: Captures how public power is exercised for private gain, including both petty and grand corruption.
Quota Provisions and PR Systems: Examines their influence on female parliamentary representation.
Dependent Variable: The percentage of women in national parliaments.

Key Findings

The results suggest:
Lower corruption levels are associated with higher female representation in parliaments.
Democratic transitions in Africa have not consistently improved women's access to political power.
Quota provisions and PR electoral systems positively impact women's representation.
Political culture and economic development have little direct effect on the number of female deputies.
Excluding Angola (a corrupt country with nearly 38% female representation), the difference in female parliamentarians between countries with above-average corruption and below-average corruption exceeds 20%.

Conclusion and Recommendations

The research provides an in-depth analysis of women's representation in 44 African countries, concluding that:
Quota systems and PR electoral frameworks remain the most effective institutional mechanisms for increasing female representation.
Democracy alone does not guarantee higher women's representation, as patriarchal norms and political barriers persist.
Corruption control plays a significant role, as fairer political systems enable greater access for women.
The study calls for further research on political culture's influence on women's parliamentary participation and advocates for stronger gender-based political reforms.

5) When politics is not just a man's game: Women's representation and political engagement.

Author : [Jeffrey A. Karp , Susan A. Banducci]

This study investigates the impact of women's representation in national legislatures on women's political engagement and attitudes about the political process across 35 countries, using data from the Comparative Study of Electoral Systems (CSES). Previous research has highlighted institutional and cultural factors influencing women's representation, but the effects of such representation on political engagement remain less understood. While women generally show lower political interest and engagement than men, some argue that the presence of women as candidates and officeholders can stimulate political participation among women.

Women's Descriptive Representation

The study finds that the election of women representatives enhances women's political knowledge, interest, engagement, and discussion. This aligns with existing research suggesting that descriptive representation strengthens political support and engagement among minority groups. However, the representation of women varies significantly across the CSES sample, with an average of 22% of lower house members being women. Nordic countries have the highest proportion of women in national parliaments.

Political Attitudes by Gender

Men are generally more politically engaged and more satisfied with the political process than women. However, analysis across the 35 sampled countries reveals that sex-based differences in political engagement and attitudes vary. For instance, in Albania, women are 8% less likely than men to have engaged in at least one political activity. Significant gender differences in political attitudes exist in about half of the sampled countries, particularly concerning views on parties and leaders. Women are generally less likely to identify a party leader who represents their views.

Cross-Level Interaction and Political Engagement

The study examines the interaction between the proportion of women in parliament and the gender of the respondent. Previous research has shown that women's lower political engagement persists even when controlling for socio-economic status. While descriptive representation does not significantly influence political engagement, it does appear to enhance positive political attitudes. In two out of four models tested, the proportion of women in parliament has a positive and significant effect on political attitudes.

Discussion

The underrepresentation of women in political office has been widely studied, with scholars advocating for increased descriptive representation. The visibility of women in politics has symbolic mobilization effects, encouraging political engagement among historically underrepresented groups. The study finds that across most countries, a gender gap in political engagement remains, though in a few cases, the gap is reversed.

Despite low overall representation, there are no significant differences between men and women in political engagement and attitudes in most cases. Some scholars suggest that increased political representation of women may reduce or even reverse these gender differences. However, findings are consistent with previous research by Lawless, which found no evidence that the presence of women in Congress influenced political engagement, efficacy, or trust among respondents.

Scholars investigating the relationship between women's political engagement and gendered political contexts argue that mechanisms such as role models and gender cues play a role. The presence of women in politics signals that political participation is socially acceptable for women, potentially reducing gender disparities in political engagement.

OBJECTIVES

- To evaluate India’s new parliament after delimitation law 2025 implemented and calibre of new parliamentarians to face challenges of new world.
- To explore the ways for more proactive role of women in parliament after passing of Nari Shakti Vandan Adhiniyam 2023 which gives 33% reservation to women in Lok Sabha.
- To identify key aspects to be considered while allocating ministerial portfolios.
- To analyze public suggestions for improving the political system.
- Measures to be taken to groom the parliamentarians to be future ready, according to Indian Voters.

METHODOLOGY

We started our project by discussing on various subjects that could understands as project topic. After shortlisting the topics, we finalized name of project. To initiate our study, we began by thoroughly reading relevant research papers related to our topic. The literature reviews helped us gain a deeper understanding of the subject matter and identify gaps in existing research.

Subsequently, we engaged in discussions with our project guide to clarify our thoughts and refine our research objectives. These conversations provided valuable insights and helped us develop a clear direction for our study. With a solid grasp of the subject, we proceeded to create a questionnaire that would effectively gather data for our research. The questionnaire was carefully crafted to ensure that the questions were clear, concise, and relevant to our research objectives.

Data Collection :

To gather data for our project, we employed a direct, person-to-person approach using Google Forms. This method enabled us to collect primary data from a diverse range of respondents.

Respondent Demographics:

Our project targeted individuals across various age groups, with a specific focus on those within the 18-80 age bracket. This broad age range allowed us to capture a comprehensive understanding of the perspectives and opinions of people from different generations.

Data Collection Process:

The data collection process involved:

1. Creating a Google Form to collect responses
2. Directly approaching individuals within the specified age range
3. Requesting their participation in the survey
4. Ensuring respondents understood the purpose and scope of the project
5. Collecting and recording responses through the Google Form.

Data Analysis :

Exploratory Data Analysis (EDA):

By conducting Exploratory Data Analysis, we obtain a better understanding of our dataset before applying any models or statistical tests. It involves summarizing, visualizing, and identifying patterns in data.

Text Analysis on Public Suggestions:

In our dataset, there is a column containing suggestions for improving the political system. Each individual has shared their opinion, and we plan to perform text analysis on this data. This analysis will help us generate a word cloud, highlighting the most frequently mentioned terms and key themes. Additionally, we will conduct sentiment analysis to classify opinions as positive, negative, or neutral, providing deeper insights into public perception.

Chi- Square Test :

The Chi-Square Test 22nalyses relationships between categorical variables like voter demographics and expectations. It helps determine statistical dependence, making it ideal for survey-based political analysis.

Model Building:

We used supervised learning techniques because our data had labelled outcomes, making it suitable for predictive modelling.

- **Logistic Regression**
- **Undersampling**
- **Adaboost**
- **Xgboost**
- **Random Forest**
- **Gradient Boost**

We started with Logistic Regression as our first model to create a baseline for comparison. While it helped us understand basic patterns, its simplicity pushed us to try more advanced models for better results.

During our data exploration, we found that one class of data was much larger than the others, which could lead to unfair predictions. To fix this, we used undersampling, which made sure all classes were equally represented. This step helped improve the fairness of our models.

Next, we used Adaboost to improve classification by focusing on the data points that were misclassified earlier. Then came XGBoost, known for being fast and effective with large datasets. We added regularization techniques to stop it from overfitting the data. Random Forest followed, combining multiple decision trees to give stable and reliable predictions. Lastly, we used Gradient Boosting, which improved the predictions step by step by learning from previous mistakes.

To make sure our models were strong, we checked their performance using metrics like accuracy, sensitivity, specificity, and F1-score. This process helped us pick the best models that worked well with the data and gave reliable results.

Tools Used Here:

- Python
- Microsoft Excel
- Microsoft Word

TERMINOLOGY

Chi-Square Test:

The hypothesis, that one factor is independent of the other or not can be tested by the chi-square test. The hypothesis,

H_0: The two factors are independent of each other
H_1: The two factors are not independent of each other

Can be tested by statistic.

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim \chi^2_{(m-1)(n-1)} \quad \text{..........} \ (\mathrm{A})$$

Statistic χ^2 has (m-1)(n-1) d.f. in formula (A), E_{ij} is the expected frequency corresponding to $(i,j)^{th}$ cell observed frequency O_{ij}. Under the null hypothesis H_0, the expected frequency.

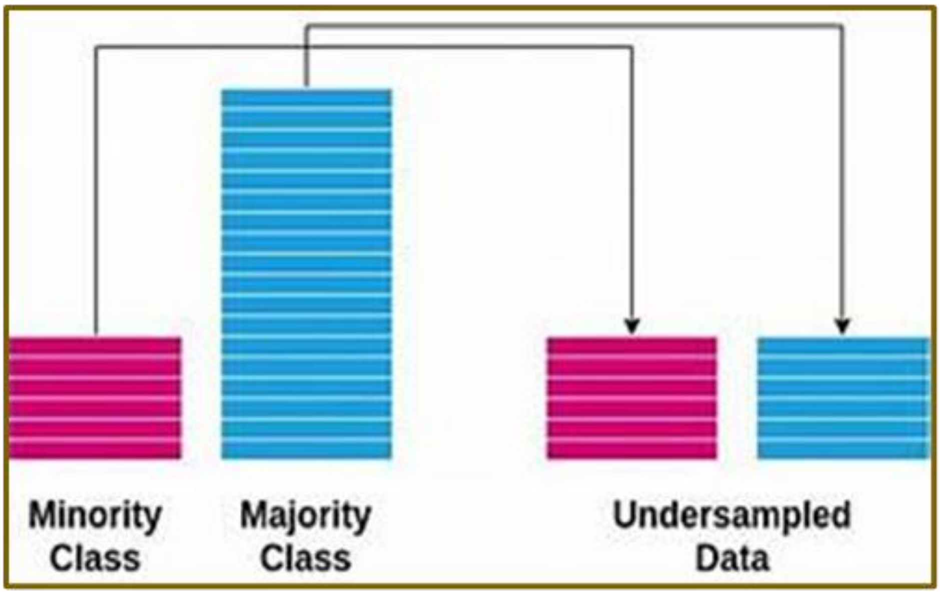
$$E_{ij} = \frac{i^{th} \text{ row total} \times \text{column total}}{\text{sample size}} \quad \text{or} \quad E_{ij} = \frac{R_i \cdot C_j}{n}$$

Once the expected frequencies are calculated, the value of chi-square is obtained with the help of chi-square is obtained with the help of formula (A). To make a decision about H_0 Calculated value of chi-square is compared with the table value if chi-square for (m-1)(n-1) d.f. and α l. o. s.

If Cal$\chi^2 \geq$ tab$\chi^2_{(m-1)(n-1)}$ reject H_0.
This means that two factors are independent.

Undersampling Technique:

Undersampling is a data balancing technique commonly used in machine learning when dealing with imbalanced datasets. In such datasets, one class (the majority class) contains significantly more samples than the other (the minority class). This imbalance can cause models to become biased, leading them to predict the majority class more often, and overlook the minority class. To address this, undersampling is applied by randomly removing instances from the majority class until it matches the number of samples in the minority class, resulting in a balanced dataset.



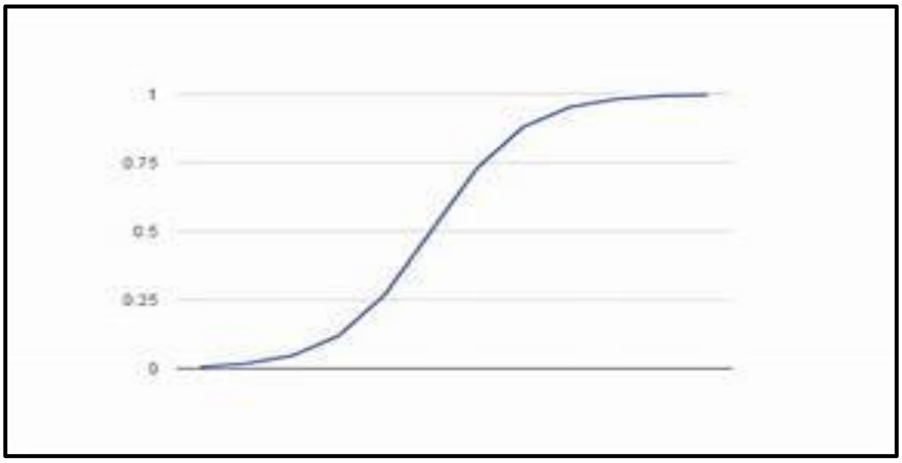
Logistic Regression :

Logistic regression is named for the function used at the core of the method, the logistic function.

The logistic function, also called the sigmoid function, was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

1/(1+e^-value)

Where e is the base of the natural logarithms (Euler's number or the EXP() function in your spreadsheet) and value is the actual numerical value that you want to transform. Below is a plot of the numbers between -5 and 5 transformed into the range 0 and 1 using the logistic function.



Logistic regression uses an equation as the representation, very much like linear regression.

Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). A key difference from linear regression is that the output value being modeled is a binary value (0 or 1) rather than a numeric value.

Logistic regression equation:

y=e^(b0+b1*x)/(1+e^(bo+b1*x))

Where y is the predicted output, b0 is the bias or intercept term and b1 is the coefficient for the single input value (x). Each column in the input data has an associated b coefficient (a constant real value) that must be learned from the training data.

Random Forest :

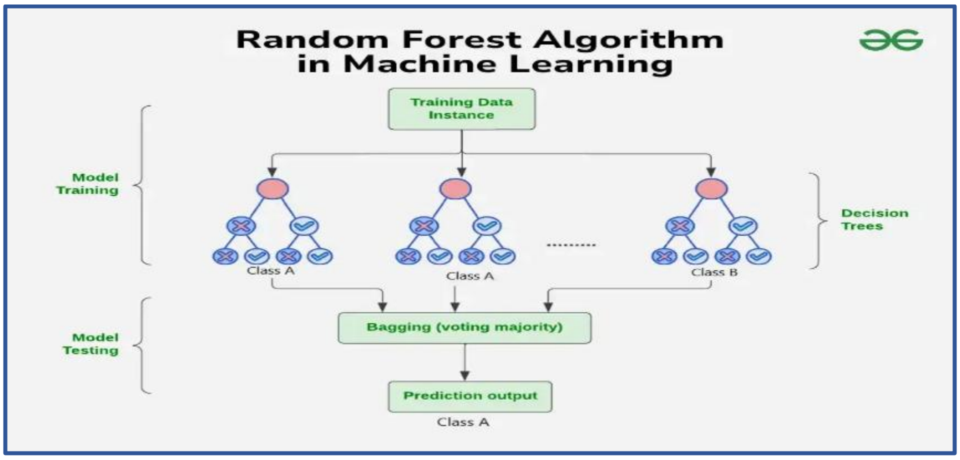
A Random Forest is a collection of decision trees that work together to make predictions.

Understanding Intuition for Random Forest Algorithm

Random Forest algorithm is a powerful tree learning technique in Machine Learning to make predictions and then we do voting of all the tress to make prediction. They are widely used for classification and regression task.

- It is a type of classifier that uses many decision trees to make predictions.
- It takes different random parts of the dataset to train each tree and then it combines the results by averaging them. This approach helps improve the accuracy of predictions. Random Forest is based on ensemble learning.

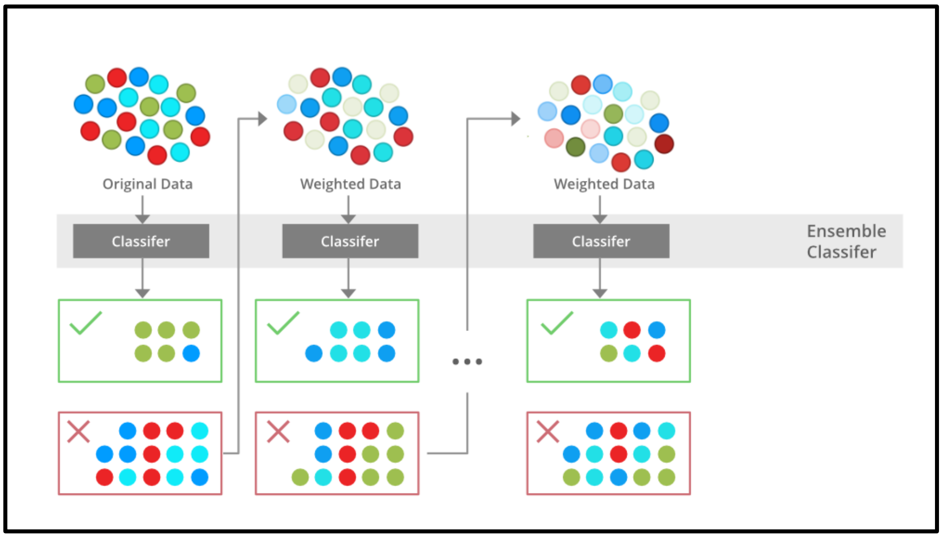
Imagine asking a group of friends for advice on where to go for vacation. Each friend gives their recommendation based on their unique perspective and preferences (decision trees trained on different subsets of data). You then make your final decision by considering the majority opinion or averaging their suggestions (ensemble prediction).



Adaptive Boosting :

AdaBoost means Adaptive Boosting and it is a is a powerful ensemble learning technique that combines multiple weak classifiers to create a strong classifier. It works by sequentially adding classifiers to correct the errors made by previous models giving more weight to the misclassified data points.

In this article we will learn to implement AdaBoost algorithm from scratch. By making it from scratch we will have a deep understanding of how AdaBoost works and key principles behind it.



Gradient Boosting :

Gradient Boosting is a ensemble learning method used for classification and regression tasks. It is a **boosting** algorithm which combine multiple weak learner to create a strong predictive model. It works by sequentially training models where each new model tries to correct the errors made by its predecessor.

In gradient boosting each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent. In each iteration the algorithm computes the gradient of the loss function with respect to the predictions and then trains a new weak model to minimize this gradient. The predictions of the new model are then added to the ensemble and the process is repeated until a stopping criterion is met.

Extreme Gradient Boosting (XgBoost) :

Traditional machine learning models like decision trees and random forests are easy to interpret but often struggle with accuracy on complex datasets. XGBoost, short for **eXtreme Gradient Boosting**, is an advanced machine learning algorithm designed for efficiency, speed, and high performance.

XGBoost is an optimized implementation of **Gradient Boosting** and is a type of **ensemble learning** method. Ensemble learning combines multiple weak models to form a stronger model.

- XGBoost uses **decision trees** as its base learners combining them sequentially to improve the model's performance. Each new tree is trained to correct the errors made by the previous tree and this process is called boosting.
- It has **built-in parallel processing to train models on large datasets quickly**. XGBoost also supports customizations allowing users to adjust model parameters to optimize performance based on the specific problem.

Accuracy:

The proportion of correct predictions to the total number of predictions. It is calculated as:

$$\textbf{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Sensitivity:

Also known as the true positive rate or recall, it is a measure of a model's ability to correctly identify positive instances out of all actual positive instances. It is particularly useful in binary classification problems, where one class is considered the "positive" class and the other is the "negative" class. Sensitivity is calculated as:

$$\textbf{Sensitivity} = \frac{TP}{(TP + FN)}$$

TP (True Positives) is the number of correctly predicted positive instances.

FN (False Negatives) is the number of instances incorrectly predicted as negative.

Specificity:

A measure of a model's ability to correctly identify negative instances out of all actual negative instances. Like sensitivity, specificity is commonly used in binary classification problems. Specificity is calculated as:

$$\textbf{Specificity} = \frac{TP}{(TN + FP)}$$

TN (True Negatives) is the number of correctly predicted negative instances.

FP (False Positives) is the number of instances incorrectly predicted as positive.

P-value:

If $p \leq 0.05$: This suggests that the results are statistically significant, meaning that the observed effect is unlikely to have occurred by chance alone. It allows for rejecting the null hypothesis in favor of the alternative hypothesis.

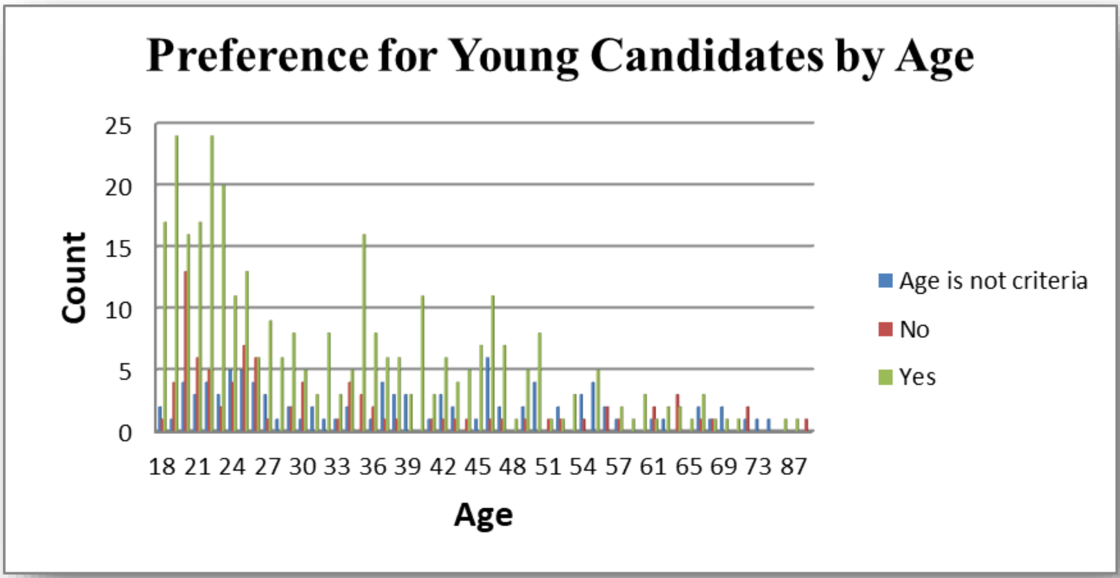
If $0.05 < p \leq 0.1$: This indicates marginal significance, suggesting there may be some evidence for the alternative hypothesis, though not as strong as when $p \leq 0.05$.

If $p > 0.1$: This suggests that the results are not statistically significant, and one fails to reject the null hypothesis, indicating the observed effect could reasonably be due to chance.

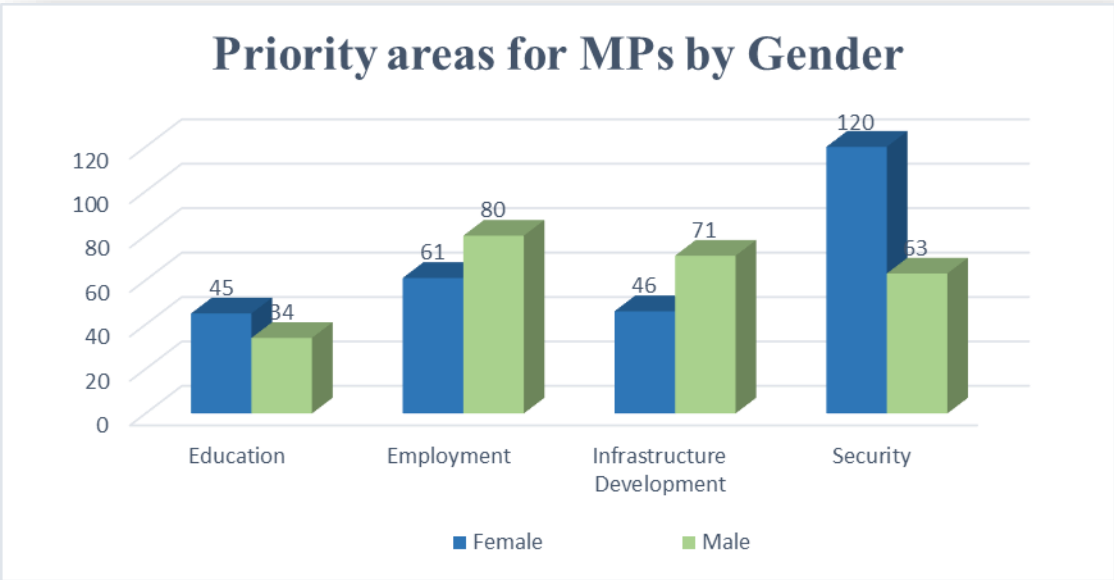
Text Analysis:

Text Analysis, also known as text mining or text data mining, is the process of extracting meaningful information from unstructured textual data. It involves techniques from linguistics, statistics, and machine learning to identify patterns, trends, and insights within large volumes of text. The primary goal is to convert raw text into structured data that can be analyzed and used for decision-making. Key steps in text analysis include text preprocessing (such as removing stop words, stemming, and tokenization), feature extraction (like Bag of Words or TF-IDF), and applying models for tasks like classification, sentiment analysis, topic modeling, or clustering.

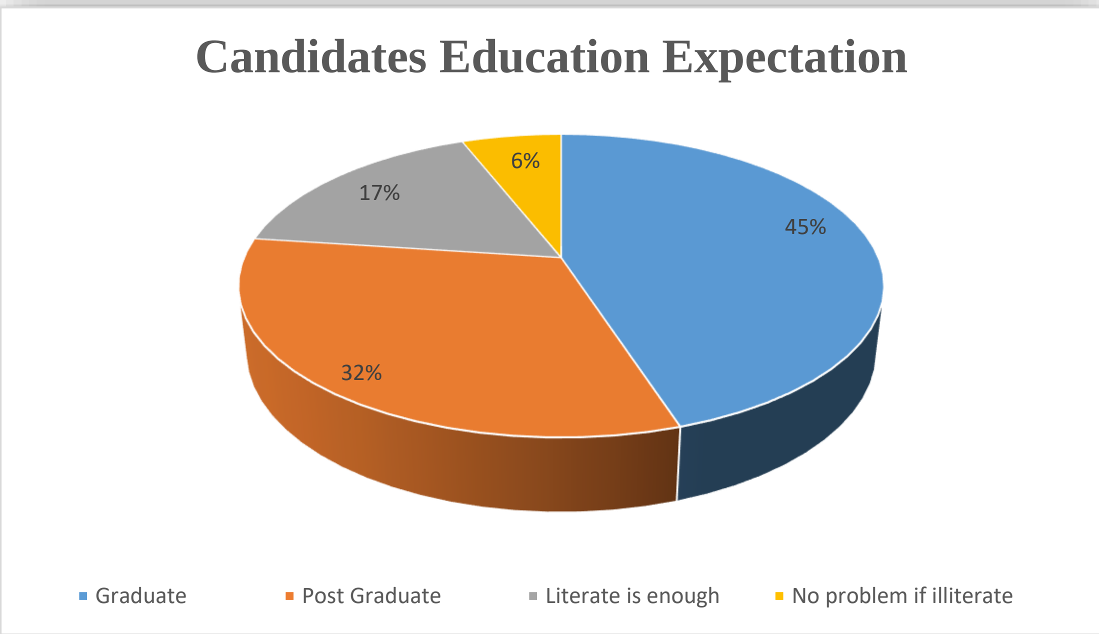
EXPLORATORY DATA ANALYSIS



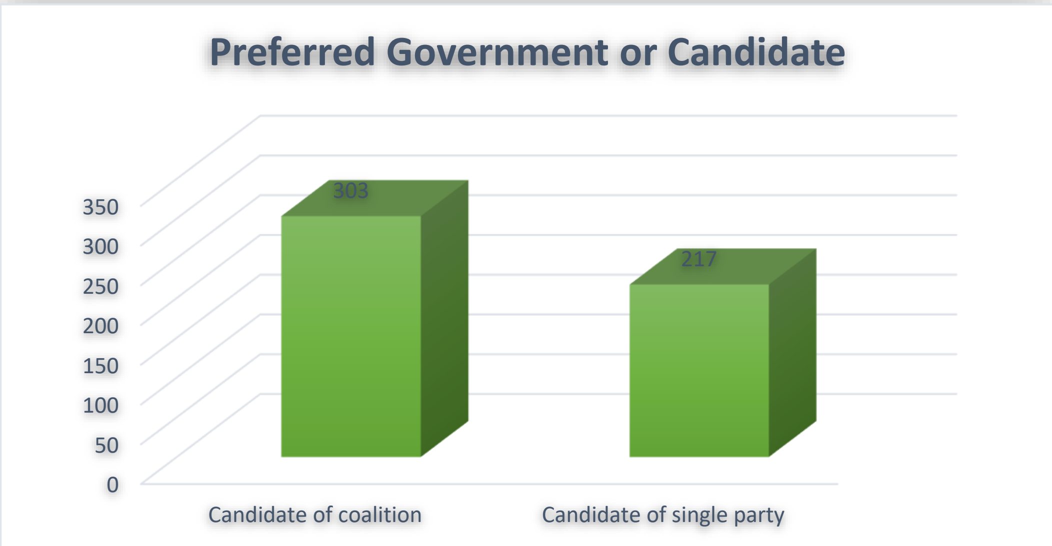
Most people prefer young candidates in Parliament. This trend is especially strong among younger age groups (18-40 years old).



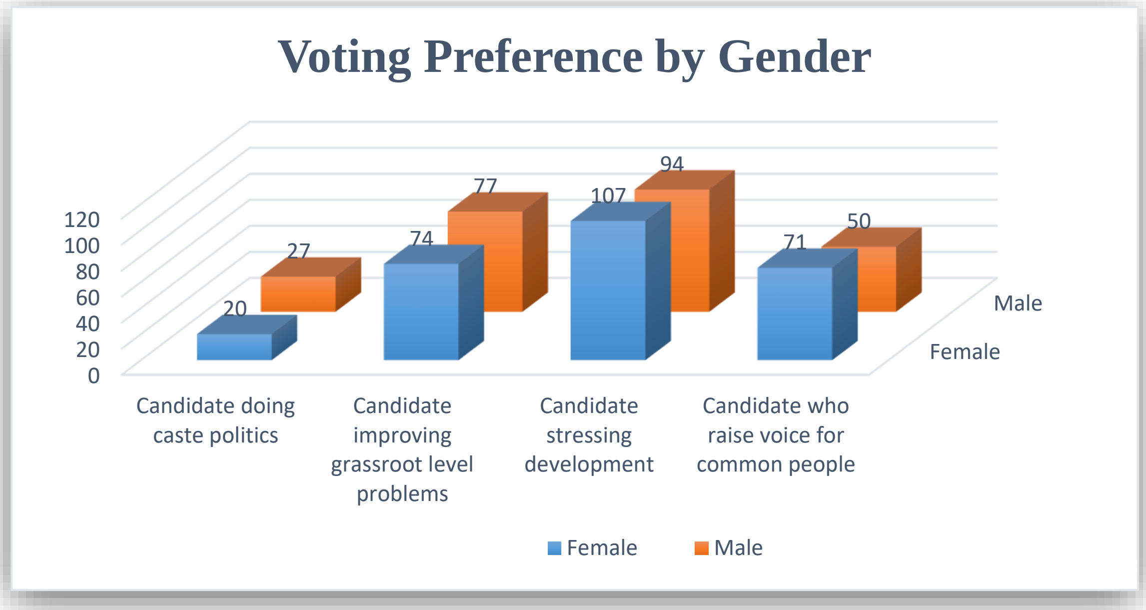
From the joint bar graph we say that 120 females chose security as their most important concern, compared to 63 males. This implies that women are more concerned about matters related to security and safety, such as crime, public protection, or the protection of women.80 males consider employment a important focus area, compared to 61 females. This indicates that men are more concerned about job opportunities and economic growth



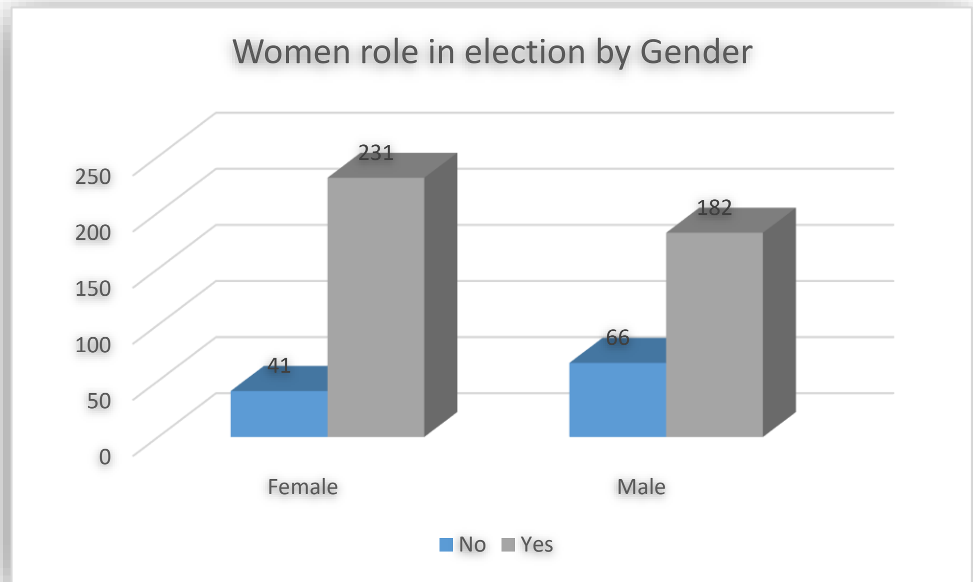
Most people expect well-educated candidates in Parliament. 45% expect a graduate candidate, and 32% expect a postgraduate candidate. A smaller group (17%) believes basic literacy is enough, while only 6% are okay with illiterate candidates. This suggests that education is a important factor for voters when choosing their leaders.



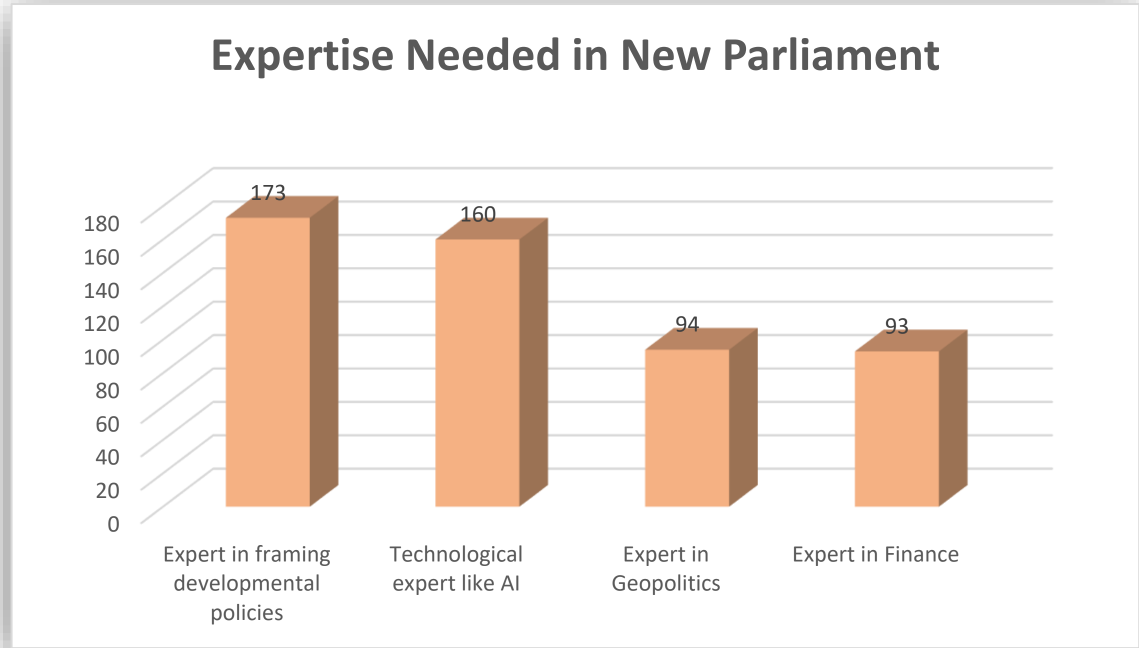
Most people (303 respondents) prefer a candidate from a coalition government, while 217 choose a single-party candidate. This suggests that many voters believe coalition governments provide better representation.



The majority of both male and females prefer candidates who focus on development (107 females, 94 males). The second most preferred choice is candidates working on grassroot-level problems (74 females, 77 males). This means that people want leaders who work on real issues like development and problem-solving .



A majority of both men and women believe that womens involvement in elections is important. 231 women and 182 men responded with "Yes." This suggests that most people support womens involvement in politics .



Interpretation : The graph suggests that most people want experts in framing developmental policies (173) and technology like AI (160) in the new Parliament. This shows a preference for leaders who can drive progress and innovation.

Statistical Analysis

CHI-SQUARED TEST :

I) Gender of the voter v/s Expectation about candidate in parliament

1)

H0 : Gender of the voter and their opinions about women Empowerment in Parliament are independent.

V/s

H1 : Gender of the voter and their opinions about women Empowerment in Parliament are dependent.

Chi-Square Statistic: 3.4961

P-Value: 0.0615

Significance Level (α): 0.05

Conclusion :

Here , P – value = 0.0615 > 0.05 (alpha) , Hence we may fail to reject H0. Therefore, we conclude that the gender of the voter has no significant effect on their opinions about women empowerment in parliament.

CONTINGENCY TABLE :

	Women Empowerment in Parliament	
Gender	**No**	**Yes**
Female	**108**	**164**
Male	**78**	**170**

Conclusion :

From the contingency table we say that 164 Females Satisfied with measures taken for woman empowerment in parliament and 170 Males Satisfied with measures taken for woman empowerment in parliament. Therefore, we conclude that the gender of the voter has no significant effect on their opinions about women empowerment in parliament.

2)
H0 : Gender of the voter and their opinion about womens role in Elections are independent.

V/s

H1 : Gender of the voter and their opinion about womens role in Elections are dependent.

Chi-Square Statistic: 9.8752

P-Value: 0.0016

Significance Level (α): 0.05

Conclusion :
Here , P – value = 0.0016 < 0.05 (alpha) , Hence we may reject H0. Therefore, we conclude that the gender of the voter has significant effect on their opinions about Womens Role in Elections.

CONTINGENCY TABLE :

	Womens Role in Elections	
Gender	**No**	**Yes**
Female	**41**	**231**
Male	**66**	**182**

Conclusion :
From the contingency table we say that the number of females (231) who think womens involvement is important in parliament is higher than the number of males (182) who think the same . Therefore, we conclude that the gender of the voter has significant effect on their opinions about Womens Role in Elections.

3)

H0 : Gender of the voter and their opinion about candidates education expectations are independent.

V/s

H1 : Gender of the voter and their opinion about candidates education expectations are dependent.

Chi-Square Statistic: 2.5731

P-Value: 0.4622

Significance Level (α): 0.05

Conclusion :

Here , P value = 0.4622 > 0.05 (α) , Hence we may fail to reject H0. Hence we conclude that there is no considerable correlation between the voters gender and expectation regarding the education of the candidate. This implies that both men and women share similar opinions about candidate education.

CONTINGENCY TABLE :

	Candidates Education Expectations			
Gender	**Graduate**	**Literate is enough**	**No problem if illiterate**	**Post Graduate**
Female	**116**	**48**	**14**	**94**
Male	**118**	**40**	**17**	**73**

Conclusion :

From the contingency table we say that there is no significant difference of female and male voters in opinion about candidate education expectations. Hence we conclude that there is no considerable correlation between the voters gender and expectation regarding the education of the candidate.

4)

H0 : Gender of the voter and their opinion on portfolio Specialization are independent.

V/s

H1 : Gender of the voter and their opinion on portfolio Specialization are dependent.

Chi-Square Statistic: 4.1011

P-Value: 0.0428

Significance Level (α): 0.05

Conclusion :

Here , P value = 0.0428 < 0.05 (α) , Hence we may reject H0. Therefore we conclude that there is a significant relationship between a voters gender and their opinion on portfolio Specialization.

CONTINGENCY TABLE :

	Portfolio Specialization	
Gender	**Must belong to the field of department he heads**	**No problem if he is unaware of department he heads**
Female	**210**	**62**
Male	**171**	**77**

Conclusion :

From the contingency table, we can see that 210 female voters and 171 male voters believe that a candidate must belong to the field of the department they head. Hence we conclude that there is a significant relationship between a gender of the voter and their opinion on whether a candidate should be specialized in the field of their department.

5)

H0 : Gender of the voter and their opinion about preference for young candidates are independent.

V/s

H1 : Gender of the voter and their opinion about preference for Young Candidates are dependent.

Chi-Square Statistic: 9.3743

P-Value: 0.0092

Significance Level (α): 0.05

Conclusion :

Here , P value = 0.0092 < 0.05 (α) , Hence we may reject H0. Therefore we conclude that gender of the voter and their preference for Young Candidates are dependent . This means that male and female voters have different views on whether political parties should prefer younger candidates while giving election tickets.

CONTINGENCY TABLE :

	Preference for Young Candidates		
Gender	**Age is not criteria**	**No**	**Yes**
Female	**56**	**34**	**182**
Male	**42**	**56**	**150**

Conclusion :

From the contingency table we say that more female voters expect political parties should prefer younger candidates while giving election tickets as compared to the male voters . This means that male and female voters have different views on whether political parties should prefer younger candidates while giving election tickets.

6)

H0 : Gender of the voter and their opinion about new central ministry Departments are independent.

V/s

H1 : Gender of the voter and their opinion about new central ministry Departments are dependent.

Chi-Square Statistic: 3.0041

P-Value: 0.2226

Significance Level (α): 0.05

Conclusion :

Here , P value = 0.2226 > 0.05 (α) , Hence we may fail to reject H0. Therefore we conclude that gender of the voter and their opinion about new central ministry Departments are independent .

CONTINGENCY TABLE :

	New Central Ministry Departments		
Gender	**Happy with current system**	**No**	**Yes**
Female	**64**	**57**	**151**
Male	**52**	**68**	**128**

Conclusion :

From the contingency table we say that 151 female voters want to add the new central ministry department and 128 male voters want to add the new central ministry department . Therefore we say that there is no significant difference of female and male voters in opinion about new central ministry department.

7)

H0 : Gender of the voter and their opinion about focus on rural development are independent.

V/s

H1 : Gender of the voter and their opinion about focus on rural development are dependent.

Chi-Square Statistic: 0.8975

P-Value: 0.3434

Significance Level (α): 0.05

Conclusion :

Here , P value = 0.3434 > 0.05 (α) , Hence we may fail to reject H0. Therefore we conclude that gender of the voter and their opinion about focus on rural development are independent.

CONTINGENCY TABLE :

	Focus on Rural Development	
Gender	**No**	**Yes**
Female	**43**	**229**
Male	**48**	**200**

Conclusion :

From the contingency table we say that 229 female voters want MPs to focus on rural development and 200 male voters want MPs to focus on rural development. Therefore we say that there is no significant difference between the opinions of female and male voters regarding whether MPs should focus on rural development.

8)

H0 : Gender of the voter and their opinion about Priority areas for MPs are independent.

V/s

H1 : Gender of the voter and their opinion about priority areas for MPs are dependent.

Chi-Square Statistic: 26.1358

P-Value: 8.9325e-06

Significance Level (α): 0.05

Conclusion :

Here , P value = 8.9325e-06 < 0.05 (α) , Hence we may reject H0. Therefore we say that gender of the voter and their opinion about priority areas for MPs are dependent.

CONTINGENCY TABLE :

	Priority areas for MPs			
Gender	**Education**	**Employment**	**Infrastructure Development**	**Security**
Female	**45**	**61**	**46**	**120**
Male	**34**	**80**	**71**	**63**

Conclusion :

From the contingency table we say that 120 females want MPs should focus on security and 80 males want MPs should focus on employment. Therefore we say that male and female voters have different opinions about priority areas for MPs.

9)

H0 : Gender of the voter and their opinion about increase women MPs for womens issues are independent.

V/s

H1 : Gender of the voter and their opinion about increase women MPs for womens issues are dependent.

Chi-Square Statistic: 2.2973

P-Value: 0.1295

Significance Level (α): 0.05

Conclusion :

Here , P value = 0.1295 > 0.05 (α) , Hence we may fail to reject H0. Therefore we say that gender of the voter and their opinion about increase women MPs for womens issues are independent.

CONTINGENCY TABLE :

	Increase Women MPs for Womens Issues	
Gender	**No**	**Yes**
Female	56	216
Male	66	182

Conclusion :

From the contingency table we say that 216 female voters want increase women MPs in parliament for womens issues and 182 male voters want increase women MPs in parliament for womens issues .Therefore we say that there is no significant difference of female and male voters in opinion about increase women MPs in parliament for womens issues.

10)
H0: Gender of the voter and their opinion about policy maker training program are independent.

V/s

H1: Gender of the voter and their opinion about policy maker training program are dependent.

Chi-Square Statistic: 1.4252

P-Value: 0.2325

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.2325 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that gender of the voter and their opinion about policy maker training program are independent.

CONTINGENCY TABLE :

	Policy Maker Training Program	
Gender	**No**	**Yes**
Female	**53**	**219**
Male	**60**	**188**

Conclusion :

From the contingency table we say that 219 female voters think the government should organize a training program for policy makers every year and 188 male voters think the government should organize a training program for policy makers every year. Therefore, we can say that there is no significant difference between male and female voters in their opinion about the government should organize a training program for policy makers every year.

11)
H0 : Gender of the voter and their opinion about polity as a compulsory subject are independent.

V/s

H1 : Gender of the voter and their opinion about polity as a compulsory subject are dependent.

Chi-Square Statistic: 0.0109

P-Value: 0.9167

Significance Level (α): 0.05

Conclusion:
Here, P value = 0.9167 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that gender of the voter and their opinion about polity as a compulsory subject are independent.

CONTINGENCY TABLE :

	Polity as a Compulsory Subject	
Gender	**No**	**Yes**
Female	**58**	**214**
Male	**51**	**197**

Conclusion :
From the contingency table we say that 214 female voters think polity should be a compulsory subject in education and 197 male voters think polity should be a compulsory subject in education. Therefore, we can say that there is no significant difference between male and female voters in their opinion about the polity should be a compulsory subject in education.

12)

H0 : Gender of the voter and their opinion about avoid nepotism in ticket allocation are independent.

V/s

H1 : Gender of the voter and their opinion about avoid nepotism in ticket allocation are dependent.

Chi-Square Statistic: 3.8163

P-Value: 0.0507

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.0507 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that gender of the voter and their opinion about avoid nepotism in ticket allocation are independent.

CONTINGENCY TABLE :

	Avoid Nepotism in Ticket Allocation	
Gender	**No**	**Yes**
Female	**45**	**227**
Male	**59**	**189**

Conclusion :

From the contingency table we say that 227 female voters want to avoid nepotism in ticket allocation and 189 male voters want to avoid nepotism in ticket allocation. Therefore, we can say that there is no significant difference between male and female voters in their opinion about the avoiding nepotism in ticket allocation.

13)
H0 : Gender of the voter and their voting preference are independent.

V/s

H1 : Gender of the voter and their voting preference are dependent.

Chi-Square Statistic: 4.4894

P-Value: 0.2132

Significance Level (α): 0.05

Conclusion :
Here, P value = 0.2132 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that gender of the voter and their voting preference are independent.

CONTINGENCY TABLE :

	Your Voting Preference			
Gender	**Candidate doing caste politics**	**Candidate improving grassroot level problems**	**Candidate stressing development**	**Candidate who raise voice for common people**
Female	**20**	**74**	**107**	**71**
Male	**27**	**77**	**94**	**50**

Conclusion :
From the contingency table we say that 107 female voters prefer the candidate who focuses on development and 94 male voters prefer the candidate who focuses on development. 74 female voters prefer the candidate who works on improving grassroot level problems and 77 male voters prefer the candidate who works on improving grassroot level problems. Therefore, we can say that there is no significant difference between male and female voters and their voting preferences.

14)
H0 : Gender of the voter and their preference for government or candidate are independent.

V/s

H1 : Gender of the voter and their preference for government or candidate are dependent.

Chi-Square Statistic: 0.7901

P-Value: 0.37406

Significance Level (α): 0.05

Conclusion :
Here, P value = 0.37406 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that gender of the voter and their preference for government or candidate are independent.

CONTINGENCY TABLE :

	Preferred Government or Candidate	
Gender	**Candidate of coalition**	**Candidate of single party**
Female	**153**	**119**
Male	**150**	**98**

Conclusion :
From the contingency table we say that 153 female voters prefer candidates from a coalition government and 150 male voters prefer candidates from a coalition government . 119 female voters prefer candidates from a single party and 98 male voters prefer candidates from a single party . Therefore, we can say that there is no significant difference between male and female voters and their preference for government or candidate.

15)

H0 : Gender of the voter and their opinion about what type of expertise needed in new parliament are independent.

V/s

H1 : Gender of the voter and their opinion about what type of expertise needed in new parliament are dependent.

Chi-Square Statistic: 5.9013

P-Value: 0.1165

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.1165 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that gender of the voter and their opinion about what type of expertise needed in new parliament are independent.

CONTINGENCY TABLE :

	Expertise Needed in New Parliament			
Gender	**Expert in Finance**	**Expert in Geopolitics**	**Expert in framing developmental policies**	**Technological expert like AI**
Female	**46**	**40**	**94**	**92**
Male	**47**	**54**	**79**	**68**

Conclusion :

From the contingency table we say that 92 female voters prefer the candidate who expert in technology and 68 male voters prefer the candidate who expert in technology. 94 female voters prefer the candidate who expert in framing developmental policies and 79 male voters prefer the candidate who expert in framing developmental policies. Therefore, we can say that there is no significant difference between male and female voters and their opinion about what type of expertise needed in new parliament.

Aim : To check there is association between gender of the voter and their expectation about candidate in parliament .

To test :

H0 : Gender of the voter and their expectation about candidate in parliament is independent .

v/s

H1 : Gender of the voter and their expectation about candidate in parliament is dependent .

Decision Criteria :

If p value < α = 0.05 then we may reject H0 at 5% l.o.s. otherwise accept it .

The Chi square test performed for all attributes as follows

Results :

Sr. No.	Attributes	P - Value	Decision Criteria	Conclusion
1	Women Empowerment in Parliament	0.0615	Failed to reject H0	Gender of the voter has no significant effect on their opinions about women empowerment in parliament.
2	**Womens Role in Elections**	**0.0016**	**Reject H0**	**Gender of the voter has significant effect on their opinions about Womens Role in Elections.**
3	Candidates Education Expectations	0.4622	Failed to reject H0	Gender of the voter and Opinion about Candidates Education Expectations are independent.
4	**Portfolio Specialization**	**0.0428**	**Reject H0**	**Gender of the voter and their opinion on portfolio Specialization are dependent.**
5	**Preference for Young Candidates**	**0.0092**	**Reject H0**	**Gender of the voter and their preference for Young Candidates are dependent .**
6	New Central Ministry Departments	0.2226	Failed to reject H0	Gender of the voter and their opinion about new central ministry departments are independent .
7	Focus on Rural Development	0.3434	Failed to reject H0	Gender of the voter and their opinion about focus on rural development are independent.
8	**Priority areas for MPs**	**8.9325 e-06**	**Reject H0**	**Gender of the voter and their opinion about priority areas for MPs are dependent.**
9	Increase Women MPs for Womens Issues	0.1295	Failed to reject H0	Gender of the voter and their opinion about increase women MPs for womens issues are independent.

10	Policy Maker Training Program	0.2325	Failed to reject H0	Gender of the voter and their opinion about policy maker training program are independent.
11	Polity as a Compulsory Subject	0.9167	Failed to reject H0	Gender of the voter and their opinion about polity as a compulsory subject are independent.
12	Avoid Nepotism in Ticket Allocation	0.0507	Failed to reject H0	Gender of the voter and their opinion about avoid nepotism in ticket allocation are independent.
13	Your Voting Preference	0.2132	Failed to reject H0	Gender of the voter and their voting preference are independent.
14	Preferred Government or Candidate	0.3740	Failed to reject H0	Gender of the voter and their preference for government or candidate are independent.
15	Expertise Needed in New Parliament	0.1165	Failed to reject H0	Gender of the voter and their opinion about what type of expertise needed in new parliament are independent.

Interpretation :
From the chi square tests we conclude that gender of the voter affects the opinion about following attributes 1] Womens Role in Elections 2] Portfolio Specialization 3] Preference for Young Candidates 4] Priority areas for MPs

II) Age of the voter v/s Expectation about candidate in parliament

1)
H0 : Age of the voter and their opinion about women empowerment in parliament are independent.

V/s

H1 : Age of the voter and their opinion about women empowerment in parliament are dependent.
Chi-Square Statistic: 2.5821
P-Value: 0.6299

Significance Level (α): 0.05

Conclusion :
Here, P value = 0.6299 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that age of the voter and their opinion about women empowerment in parliament are independent.

CONTINGENCY TABLE :

	Women Empowerment in Parliament	
Age group	**No**	**Yes**
0	**100**	**169**
1	**45**	**89**
2	**32**	**51**
3	**8**	**24**
4	**1**	**1**

Conclusion :
From the contingency table, we can say that more youngsters are satisfied with the measures taken for womens empowerment in Parliament compared to voters from other age groups. Therefore, we can conclude that there is no significant difference in opinions among voters of different age groups.

2)
H0 : Age of the voter and their opinion about womens role in Elections are independent.

V/s

H1 : Age of the voter and their opinion about womens role in Elections are dependent.

Chi-Square Statistic: 3.4627
P-Value: 0.4835

Significance Level (α): 0.05

Conclusion :
Here, P value = 0.4835 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that age of the voter and their opinion about women's role in Elections are independent.

CONTINGENCY TABLE :

	Womens Role in Elections	
Age group	**No**	**Yes**
0	**60**	**209**
1	**25**	**109**
2	**13**	**70**
3	**8**	**24**
4	**1**	**1**

Conclusion :
From the contingency table, we can say that 209 voters from 1st age group think women's involvement is important in parliament and 109 voters from 2nd age group think women's involvement is important in parliament. Therefore, we conclude that the age of the voter has no significant effect on their opinions about Women's Role in Elections.

3)
H0 : Age of the voter and their opinion about candidates education expectations are independent.

V/s

H1 : Age of the voter and their opinion about candidates education expectations are dependent.

Chi-Square Statistic: 50.4763

P-Value: 1.1520e-06

Significance Level (α): 0.05

Conclusion :
Here, P value =1.1520e-06 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about candidates education expectations are dependent.

CONTINGENCY TABLE :

	Candidates Education Expectations			
Age Group	Graduate	Literate is enough	No problem if illiterate	Post Graduate
0	118	34	12	105
1	63	25	4	42
2	35	25	7	16
3	16	4	8	4
4	2	0	0	0

Conclusion :
From the contingency table, we can say that 118 voters from 1st age group can expect graduate level candidate in parliament and 63 voters from 2nd age group can expect graduate level candidate in parliament . Therefore, we conclude that the age of the voter has significant effect on their opinions about candidates education expectations.

4)

H0 : Age of the voter and their opinion about portfolio Specialization are independent.

V/s

H1 : Age of the voter and their opinion about portfolio Specialization are dependent.

Chi-Square Statistic: 12.9854
P-Value: 0.01134

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.01134 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about portfolio Specialization dependent.

CONTINGENCY TABLE :

	Portfolio Specialization	
Age group	**Must belong to the field of department he heads**	**No problem if he is unaware of department he heads**
0	**211**	**58**
1	**97**	**37**
2	**55**	**28**
3	**17**	**15**
4	**1**	**1**

Conclusion :
From the contingency table, we can say that 211 voters from 1st age group believe that a candidate must belong to the field of the department they head and 97 voters from 2nd age group believe that a candidate must belong to the field of the department they head. Therefore, we conclude that the age of the voter has significant effect on their opinions about portfolio Specialization.

5)
H0 : Age of the voter and their opinion about preference for young candidates are independent.

V/s

H1 : Age of the voter and their opinion about preference for young candidates are dependent.

Chi-Square Statistic: 26.4255
P-Value: 0.0008

Significance Level (α): 0.05
Conclusion :

Here, P value = 0.0008 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about preference for young candidates are dependent.

CONTINGENCY TABLE :

	Preference for Young Candidates		
Age group	**Age is not criteria**	**No**	**Yes**
0	**38**	**55**	**176**
1	**24**	**16**	**94**
2	**26**	**9**	**48**
3	**10**	**9**	**13**
4	**0**	**1**	**1**

Conclusion :

From the contingency table we say that 176 voters from 1st age group can expect political parties should prefer younger candidates while giving election tickets and 94 voters from 2nd age group expect political parties should prefer younger candidates while giving election tickets. Therefore, we conclude that the age of the voter has significant effect on their opinions about preference for young candidates.

6)
H0 : Age of the voter and their opinion about new central ministry departments are independent.

V/s

H1 : Age of the voter and their opinion about new central ministry departments are dependent.

Chi-Square Statistic: 16.4570

P-Value: 0.0362

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.0362 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about new central ministry departments are dependent.

CONTINGENCY TABLE :

	New Central Ministry Departments		
Age Group	**Happy with current system**	**No**	**Yes**
0	**53**	**57**	**159**
1	**31**	**34**	**69**
2	**23**	**20**	**40**
3	**9**	**12**	**11**
4	**0**	**2**	**0**

Conclusion :

From the contingency table we say that 159 voters from 1st age group want to add the new central ministry department and 69 voters from 2nd age group want to add new central ministry department. Therefore, we conclude that the age of the voter has significant effect on their opinions about new central ministry departments.

7)
H0 : Age of the voter and their opinion about focus on rural development are independent.

V/s

H1 : Age of the voter and their opinion about focus on rural development are dependent.

Chi-Square Statistic: 9.8545

P-Value: 0.0429

Significance Level (α): 0.05

Conclusion :
Here, P value = 0.0429 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about focus on rural development are dependent.

CONTINGENCY TABLE :

	Focus on Rural Development	
Age Group	**No**	**Yes**
0	**46**	**223**
1	**25**	**109**
2	**13**	**70**
3	**5**	**27**
4	**2**	**0**

Conclusion :
From the contingency table we say that 223 voters from 1st age group want to MPs should focus on rural development and 109 voters from 2nd age group want to MPs should focus on rural development. Therefore, we conclude that the age of the voter has significant effect on their opinion regarding whether MPs should focus on rural development.

8)

H0 : Age of the voter and their opinion about priority areas for MPs are independent.

V/s

H1 : Age of the voter and their opinion about priority areas for MPs are dependent.

Chi-Square Statistic: 22.4786

P-Value: 0.0324

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.0429 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about priority areas for MPs are dependent.

CONTINGENCY TABLE :

	Priority areas for MPs			
Age Group	**Education**	**Employment**	**Infrastructure Development**	**Security**
0	**44**	**82**	**50**	**93**
1	**22**	**38**	**29**	**45**
2	**9**	**19**	**23**	**32**
3	**3**	**2**	**14**	**13**
4	**1**	**0**	**1**	**0**

Conclusion :

From the contingency table we say that 93 voters from 1st age group want MPs should focus on security and 82 voters from 2nd age group want MPs should focus on employment. Therefore, we conclude that the age of the voter has significant effect on their opinion about priority areas for MPs.

9)

H0 : Age of the voter and their opinion about increase women MPs for womens issues are independent.

V/s

H1 : Age of the voter and their opinion about increase women MPs for womens issues are dependent.

Chi-Square Statistic: 2.5327

P-Value: 0.6387

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.6387 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that age of the voter and their opinion about increase women MPs for womens issues are independent.

CONTINGENCY TABLE :

	Increase Women MPs for Womens Issues	
Age Group	**No**	**Yes**
0	**64**	**205**
1	**32**	**102**
2	**16**	**67**
3	**10**	**22**
4	**0**	**2**

Conclusion :

From the contingency table we say that 205 voters from 1st age group want increase women MPs in parliament for womens issues and 102 voters from 2nd age group want increase women MPs in parliament for womens issues. Therefore, we conclude that the age of the voter has no significant effect on their opinion about increase women MPs for womens issues in parliament.

10)

H0 : Age of the voter and their opinion about policy maker training program are independent.

V/s

H1 : Age of the voter and their opinion about policy maker training program are dependent.

Chi-Square Statistic: 4.7666

P-Value: 0.3120

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.3120 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that age of the voter and their opinion about policy maker training program are independent.

CONTINGENCY TABLE :

	Policy Maker Training Program	
Age group	**No**	**Yes**
0	**56**	**213**
1	**30**	**104**
2	**15**	**68**
3	**11**	**21**
4	**1**	**1**

Conclusion :

From the contingency table we say that 104 voters from 1st age group think the government should organize a training program for policy makers every year and 68 voters from 2nd age group think the government should organize a training program for policy makers every year. Therefore, we conclude that the age of the voter has no significant effect on their opinion about increase women MPs for womens issues in parliament.

11)
H0 : Age of the voter and their opinion about polity as a compulsory subject are independent.

V/s

H1: Age of the voter and their opinion about polity as a compulsory subject are dependent.

Chi-Square Statistic: 11.0629

P-Value: 0.0258

Significance Level (α): 0.05

Conclusion :
Here, P value = 0.0258 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about polity as a compulsory subject are dependent.

CONTINGENCY TABLE :

	Polity as a Compulsory Subject	
Age group	**No**	**Yes**
0	**56**	**213**
1	**34**	**100**
2	**8**	**75**
3	**10**	**22**
4	**1**	**1**

Conclusion :

From the contingency table we say that 213 voters from 1^{st} age group think polity should be a compulsory subject in education and 100 voters from 2^{nd} age group think polity should be a compulsory subject in education. Therefore, we conclude that the age of the voter has significant effect on their opinion about polity as a compulsory subject in education .

12)

H0 : Age of the voter and their opinion about avoid nepotism in ticket allocation are independent.

V/s

H1 : Age of the voter and their opinion about avoid nepotism in ticket allocation are dependent.

Chi-Square Statistic: 5.1516

P-Value: 0.2720

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.2720 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that age of the voter and their opinion about avoid nepotism in ticket allocation are independent.

CONTINGENCY TABLE :

	Avoid Nepotism in Ticket Allocation	
Age group	**No**	**Yes**
0	**59**	**210**
1	**21**	**113**
2	**14**	**69**
3	**9**	**23**
4	**1**	**1**

Conclusion :

From the contingency table we say that 210 voters from 1st age group want to avoid nepotism in ticket allocation and 113 voters from 2nd age group want to avoid nepotism in ticket allocation. Therefore, we conclude that the age of the voter has no significant effect on their opinion about avoid nepotism in ticket allocation.

13)

H0 : Age of the voter and their voting preference are independent.

V/s

H1 : Age of the voter and their voting preference are dependent.

Chi-Square Statistic: 22.1476

P-Value: 0.0358

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.0358 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their voting preference are dependent.

CONTINGENCY TABLE :

	Your Voting Preference			
Age group	**Candidate doing caste politics**	**Candidate improving grassroot level problems**	**Candidate stressing development**	**Candidate who raise voice for common people**
0	**36**	**73**	**96**	**64**
1	**7**	**38**	**60**	**29**
2	**2**	**31**	**33**	**17**
3	**2**	**9**	**10**	**11**
4	**0**	**0**	**2**	**0**

Conclusion :

From the contingency table we say that 96 voters from 1^{st} age group prefer the candidate who focuses on development. 73 voters from 1^{st} age group prefer the candidate who works on improving grassroot level problems. Therefore, we conclude that the age of the voter has significant effect on their voting preference

14)

H0 : Age of the voter and their preference for government or candidate are independent.

V/s

H1 : Age of the voter and their preference for government or candidate are dependent.

Chi-Square Statistic: 0.8721

P-Value: 0.9285

Significance Level (α): 0.05

Conclusion :

Here, P value = 0.9285 > 0.05 (α), Hence we may fail to reject H0. Therefore we say that age of the voter and their preference for government or candidate are independent.

CONTINGENCY TABLE :

	Preferred Government or Candidate	
Age group	**Candidate of coalition**	**Candidate of single party**
0	**159**	**110**
1	**80**	**54**
2	**46**	**37**
3	**17**	**15**
4	**1**	**1**

Conclusion :

From the contingency table we say that 159 voters from 1^{st} age group prefer candidates from a coalition government. 110 voters from 1^{st} age group prefer candidates from a single party. Therefore, we conclude that the age of the voter has no significant effect on their preference for government or candidate.

15)
H0 : Age of the voter and their opinion about what type of expertise needed in new parliament are independent.

V/s

H1 : Age of the voter and their opinion about what type of expertise needed in new parliament are dependent.

Chi-Square Statistic: 22.9922

P-Value: 0.0277

Significance Level (α): 0.05

Conclusion :
Here, P value = 0.0277 < 0.05 (α), Hence we may reject H0. Therefore we say that age of the voter and their opinion about what type of expertise needed in new parliament are dependent.

CONTINGENCY TABLE :

	Expertise Needed in New Parliament			
Age group	**Expert in Finance**	**Expert in Geopolitics**	**Expert in framing developmental policies**	**Technological expert like AI**
0	**38**	**48**	**81**	**102**
1	**33**	**25**	**42**	**34**
2	**15**	**15**	**33**	**20**
3	**6**	**6**	**16**	**4**
4	**1**	**0**	**1**	**0**

Conclusion :

From the contingency table we say that 34 voters from 1st age group prefer the candidate who expert in technology. 42 voters from 1st age group prefer the candidate who expert in framing developmental policies. Therefore, we conclude that the age of the voter has significant effect on their opinion about what type of expertise needed in new parliament.

Aim : To check there is association between age of the voter and their expectation about candidate in parliament .

To test :

H0 : Age of the voter and their expectation about candidate in parliament is independent .

v/s

H1 : Age of the voter and their expectation about candidate in parliament is dependent.

Decision Criteria :

If p value < $\alpha = 0.05$ then we may reject H0 at 5% l.o.s. otherwise accept it .

The Chi square test performed for all attributes as follows

Results :

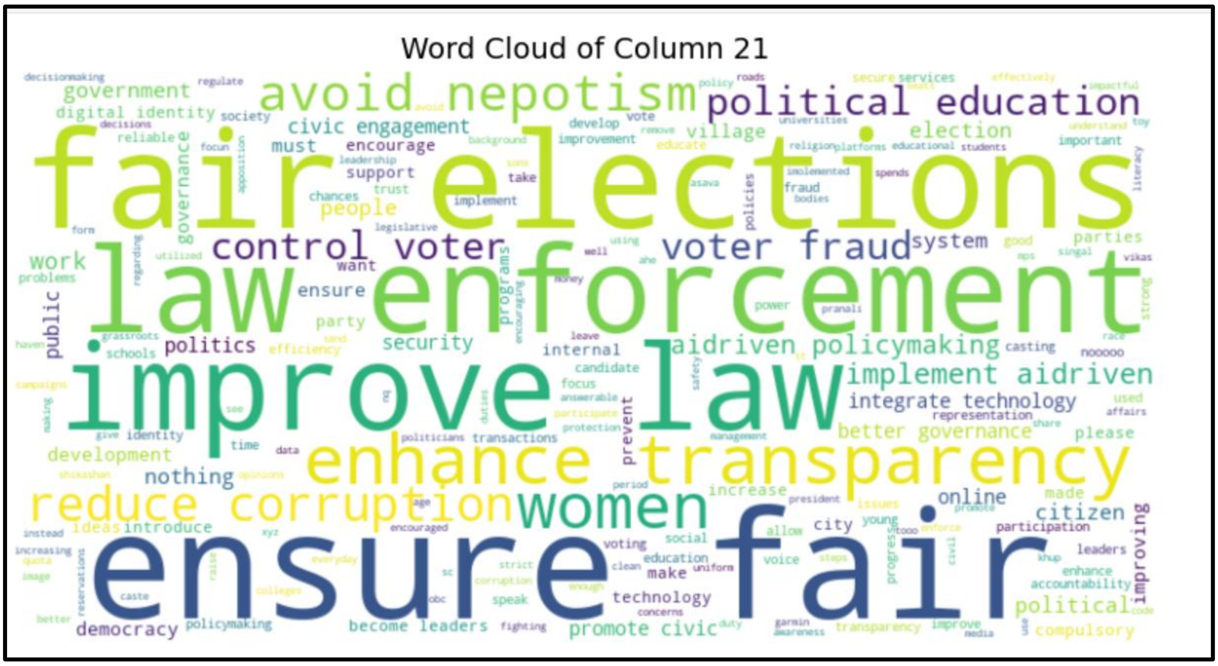
Sr. No.	Attributes	P - Value	Decision Criteria	Conclusion
1	Women Empowerment in Parliament	0.6299	Failed to reject H0	Age of the voter and their opinion about women empowerment in parliament are independent.
2	Womens Role in Elections	0.4835	Failed to reject H0	Age of the voter and their opinion about womens role in Elections are independent.
3	**Candidates Education Expectations**	**1.1520e-06**	**Reject H0**	**Age of the voter and their opinion about candidates education expectations are dependent.**
4	**Portfolio Specialization**	**0.01134**	**Reject H0**	**Age of the voter and their opinion about portfolio Specialization dependent.**
5	**Preference for Young Candidates**	**0.0008**	**Reject H0**	**Age of the voter and their opinion about preference for young candidates are dependent.**
6	**New Central Ministry Departments**	**0.0362**	**Reject H0**	**Age of the voter and their opinion about new central ministry departments are dependent.**
7	**Focus on Rural Development**	**0.0429**	**Reject H0**	**Age of the voter and their opinion about focus on rural development are dependent.**
8	**Priority areas for MPs**	**0.0324**	**Reject H0**	**Age of the voter and their opinion about priority areas for MPs are dependent.**

9	Increase Women MPs for Womens Issues	0.6387	Failed to reject H0	Age of the voter and their opinion about increase women MPs for womens issues are independent.
10	Policy Maker Training Program	0.3120	Failed to reject H0	Age of the voter and their opinion about policy maker training program are independent.
11	**Polity as a Compulsory Subject**	**0.0258**	**Reject H0**	**Age of the voter and their opinion about polity as a compulsory subject are dependent.**
12	Avoid Nepotism in Ticket Allocation	0.2720	Failed to reject H0	Age of the voter and their opinion about avoid nepotism in ticket allocation are independent.
13	**Your Voting Preference**	**0.0358**	**Reject H0**	**Age of the voter and their voting preference are dependent.**
14	Preferred Government or Candidate	0.9285	Failed to reject H0	Age of the voter and their preference for government or candidate are independent.
15	**Expertise Needed in New Parliament**	**0.0277**	**Reject H0**	**Age of the voter and their opinion about what type of expertise needed in new parliament are dependent.**

Interpretation :

From the chi square tests we conclude that age of the voter affects the opinion about following attributes 1] Candidates Education Expectations 2] Portfolio Specialization 3] Preference for Young Candidates 4] New Central Ministry Departments 5] Focus on Rural Development 6] Priority areas for MPs 7] Polity as a Compulsory Subject 8] Expertise Needed in New Parliament.

Text Analysis :



Interpretation :

The word cloud shows that people want fair elections, strict law enforcement, and more transparency in the political system. Many are concerned about corruption, voter fraud, and nepotism and believe that strong laws can help fix these issues. There is also interest in using technology and AI to improve elections and prevent fraud. People think better education and civic awareness can help citizens make informed decisions.

Overall, the message is clear: people want a fair, honest, and modern political system that works for everyone.

Model Building :

Logistic Regression :

Modelling Outline :

Y : 1 if a person wants to avoid Nepotism in Ticket Allocation.

0 if person don't wants to avoid Nepotism in Ticket Allocation.

X1 : Gender

X2 : Age

X4 : Responsible Voter

X6 : Women's role in election.

X8 : Portfolio Specialization.

X9 : Preferance for young candidates.

X10 : New central ministry departments

X12 : Priority areas for MPs

X14 : Policy maker training program

X15 : Polity as a compulsory subject.

X17 : Voting preference

X18 : Preferred Goveronment or Candidate.

X20 : Expertise needed in new parliament.

Logistic Regression :

Confusion Matrix and Statistics :

		Predicted	
		Don't wants to avoid nepotism (0)	Wants to avoid nepotism (1)
Actual	Don't wants to avoid nepotism (0)	6	19
	Wants to avoid nepotism (1)	13	118

Logistics	**Values**
Accuracy	0.7949
Sensitivity	0.90
Specificity	0.24
f1 Score	0.88

Accuracy (0.7949) : The accuracy of the model is 0.7949, indicating that approximately 79.49% of the instances were correctly classified by the model.

Sensitivity (0.90) : Sensitivity measures the proportion of actual positive cases that were correctly identified by the model. In this case, the sensitivity of the model is 0.90, indicating the approximately 90% of the actual positive cases were correctly identified by the model. But here the model is biased toward predicting the majority class, so it often predicts class 1 correctly.

Specificity (0.24) : Specificity measures the proportion of actual negative cases that were correctly identified by the model, In this case, the specificity of the model 0.24, indicating that approximately 24% of the actual negative cases were correctly identified by the model. Class imbalance negatively affects specificity, especially when the minority class is underrepresented in the training data. In such cases, the model becomes biased toward predicting the majority class and often fails to correctly identify negative instances (i.e., the minority class). As a result, specificity—which measures the model's ability to correctly classify actual negatives—tends to be very low. This means that many true negative cases are misclassified as positives, reducing the reliability of the model in detecting the minority class.

F1 Score (0.88) : The f1 score is the harmonic mean of precision and recall. It is a measure of a test's accuracy that considers both the precision and the recall of the test to compute the score. In this case, the f1 score of the model is 0.88 indicating approximately 88% which balances the trade-off between precision and recall.

Handling Class Imbalancing :

Undersampling :

Value Counts of actual dataset :

Avoid Nepotism in Ticket Allocation	Value Counts
Yes (1)	416
No (0)	104

Value Counts after Undersampling :

Avoid Nepotism in Ticket Allocation	Value Counts
Yes (1)	104
No (0)	104

The original dataset exhibited significant class imbalance, with class 1 being heavily overrepresented compared to class 0. To address this issue and reduce the model's bias toward the majority class, undersampling was applied to randomly reduce the number of instances in the majority class to match that of the minority class.

After undersampling, both classes contain 104 instances, ensuring a balanced dataset. This step was essential to improve the model’s ability to correctly classify the minority class, thereby enhancing specificity and overall fairness in model evaluation. While undersampling reduces the size of the dataset, it helps prevent the model from ignoring underrepresented outcomes, making the classification results more reliable and interpretable.

1) Logistic Regression :

Confusion Matrix and Statistics :

		Predicted	
		Don't wants to avoid nepotism (0)	Wants to avoid nepotism (1)
Actual	Don't wants to avoid nepotism (0)	22	10
	Wants to avoid nepotism (1)	6	25

Logistics	**Values**
Accuracy	0.7460
Sensitivity	0.81
Specificity	0.69
f1 Score	0.75

Interpretation:

Accuacy : The accuracy of the model is 74.60%, indicating that around 74.60% of the predictions made by the model are correct.

Sensitivity : Sensitivity also known as true positive rate, is 0.81. This indicates that out of all people wishing to avoid nepotism, the model correctly identifies 81% of them. It is a good measure of how well model captures the people wants to avoid the nepotism.

Specificity : Specificity also known as true negative rate, is 0.69,. This means that out of all people don't wants to avoid nepotism, the model correctly identifies 69% of them.

F1-Score : The F1-Score, which is harmonic mean of precision and recall, is 0.75. It provides a balance between precision and recall, with higher values indicating better performance.

2) Random Forest :

Confusion Matrix and Statistics :

		Predicted	
		Don't wants to avoid nepotism (0)	Wants to avoid nepotism (1)
Actual	Don't wants to avoid nepotism (0)	22	10
	Wants to avoid nepotism (1)	6	25

Logistics	**Values**
Accuracy	0.7777
Sensitivity	0.71
Specificity	0.84
f1 Score	0.78

Interpretation:

Accuacy : The accuracy of the model is 77.77%, indicating that around 77.77% of the predictions made by the model are correct.

Sensitivity : Sensitivity also known as true positive rate, is 0.71. This indicates that out of all people wishing to avoid nepotism, the model correctly identifies 71% of them. It is a good measure of how well model captures the people wants to avoid the nepotism.

Specificity : Specificity also known as true negative rate, is 0.84. This means that out of all people don't wants to avoid nepotism, the model correctly identifies 84% of them.

F1-Score : The F1-Score, which is harmonic mean of precision and recall, is 0.78. It provides a balance between precision and recall, with higher values indicating better performance.

3) Adaptive Boosting :

Confusion Matrix and Statistics :

		Predicted	
		Don't wants to avoid nepotism (0)	Wants to avoid nepotism (1)
Actual	Don't wants to avoid nepotism (0)	22	10
	Wants to avoid nepotism (1)	6	25

Logistics	**Values**
Accuracy	0.7301
Sensitivity	0.77
Specificity	0.69
f1 Score	0.73

Interpretation:

Accuacy : The accuracy of the model is 73.01%, indicating that around 73.01% of the predictions made by the model are correct.

Sensitivity : Sensitivity also known as true positive rate, is 0.77. This indicates that out of all people wishing to avoid nepotism, the model correctly identifies 77% of them. It is a good measure of how well model captures the people wants to avoid the nepotism.

Specificity : Specificity also known as true negative rate, is 0.69,. This means that out of all people don't wants to avoid nepotism, the model correctly identifies 69% of them.

F1-Score : The F1-Score, which is harmonic mean of precision and recall, is 0.73. It provides a balance between precision and recall, with higher values indicating better performance.

4) Gradient Boosting :

Confusion Matrix and Statistics :

		Predicted	
		Don't wants to avoid nepotism (0)	Wants to avoid nepotism (1)
Actual	Don't wants to avoid nepotism (0)	25	7
	Wants to avoid nepotism (1)	7	24

Logistics	**Values**
Accuracy	0.7777
Sensitivity	0.77
Specificity	0.78
f1 Score	0.78

Interpretation:

Accuacy : The accuracy of the model is 77.77%, indicating that around 77.77% of the predictions made by the model are correct.

Sensitivity : Sensitivity also known as true positive rate, is 0.77. This indicates that out of all people wishing to avoid nepotism, the model correctly identifies 77% of them. It is a good measure of how well model captures the people wants to avoid the nepotism.

Specificity : Specificity also known as true negative rate, is 0.78. This means that out of all people don't wants to avoid nepotism, the model correctly identifies 78% of them.

F1-Score : The F1-Score, which is harmonic mean of precision and recall, is 0.78. It provides a balance between precision and recall, with higher values indicating better performance.

5) eXtreme Gradient Boosting :

Confusion Matrix and Statistics :

		Predicted	
		Don't wants to avoid nepotism (0)	Wants to avoid nepotism (1)
Actual	Don't wants to avoid nepotism (0)	22	10
	Wants to avoid nepotism (1)	6	25

Logistics	**Values**
Accuracy	0.7460
Sensitivity	0.74
Specificity	0.75
f1 Score	0.75

Interpretation:

Accuacy : The accuracy of the model is 74.60%, indicating that around 74.60% of the predictions made by the model are correct.

Sensitivity : Sensitivity also known as true positive rate, is 0.74. This indicates that out of all people wishing to avoid nepotism, the model correctly identifies 74% of them. It is a good measure of how well model captures the people wants to avoid the nepotism.

Specificity : Specificity also known as true negative rate, is 0.75,. This means that out of all people don't wants to avoid nepotism, the model correctly identifies 75% of them.

F1-Score : The F1-Score, which is harmonic mean of precision and recall, is 0.75. It provides a balance between precision and recall, with higher values indicating better performance.

Table of Statistical Models and their Accuracy, Sensitivity, Specificity, F1 Score :

Model	Accuracy	Sensitivity	Specificity	F1- Score
Logistic Regression	0.7460	0.81	0.69	0.75
Random Forest	0.7777	0.71	0.84	0.78
Adaptive Boosting	0.7301	0.77	0.69	0.73
Gradient Boosting	0.7777	0.77	0.78	0.78
Extreme Gradient Boosting	0.7460	0.74	0.75	0.75

CONCLUSION

- The 2025 delimitation law will reshape India's parliamentary landscape, increasing representation and redefining political boundaries. To meet the demands of a rapidly changing world, the calibre of new parliamentarians must reflect qualities like education, technological awareness, ethical leadership, and vision for national development. This transition offers a historic opportunity to build a stronger, smarter, and more future-ready Parliament that truly represents the aspirations of modern India.

- The Nari Shakti Vandan Adhiniyam, 2023 is a crucial step toward gender-inclusive governance. To ensure meaningful participation, women must be supported through leadership training, mentorship, and equal opportunities. A bias-free political environment will enable them to lead effectively and contribute to national progress.

- The study highlights that educational background, domain expertise, and leadership skills are key factors in effective portfolio allocation. Public opinion favours ministers who are professionally qualified and aligned with their assigned responsibilities. These insights reinforce the need for merit-based and strategic decision-making in government formation.

- The word cloud analysis highlights key public suggestions for political reform, emphasizing the need for fair elections, strong law enforcement, increased transparency, and reduced corruption. Other notable themes include promoting political education, avoiding nepotism, and encouraging citizen engagement. These insights reflect a collective demand for a more accountable, inclusive, and transparent political system.

- To meet the expectations of Indian voters, future leaders must be equipped with strong educational backgrounds, technological awareness, ethical values, and a deep understanding of governance. Political parties should adopt structured training programs and transparent candidate selection processes, while voters must remain informed and engaged. Together, through better leadership and responsible citizenship, we can build a stronger, smarter, and more inclusive Parliament for tomorrow's India.

SCOPE

- By sending proper representative it avoids chaos in portfolio distribution.
- Strong and stable government with representatives is in highest need to guard all the interest and protect India from all the odds.
- Future ready MPs to tackle new and modern challenges**.**
- MPs equipped with new skills, knowledge can make changes in organization's like ASIAN, BRICS, SCO which directly affects economies and everyday life of common people of all the participating nations.
- MPs with zeal solve complicated issues with neighbour countries like border disputes and terrorism.
- When India, a country containing high percentage of population is lead by smart representatives it opens doors of safe and growing opportunities for future generations in secured environment.

LIMITATIONS

- The study is based on data collected from 520 respondents, which may not fully represent the views of all Indian voters.
- Sentiment analysis of open-ended responses may miss the depth or emotion behind certain comments due to language or expression limitations.
- The study focuses mainly on expectations from political candidates, and does not evaluate the actual performance of current politicians or the outcomes of specific policies.
- The data reflects public opinion at a specific point in time. Given the dynamic nature of politics, public expectations and sentiments may evolve rapidly due to new policies, events, or leadership changes.

REFERENCES

Research Book & Journal:

1) Daniel T. Larose, (2006). Data Minig Methods and Models, Wile-Interscience.

2) Chance Journal – Nguyen- Khang Pham, Jean-Hugues Chauchat, Jean Dumains.

Research papers:

1) Evaluating the Role of Media in Shaping Political Opinions & Political Preferences: A Statistical Critical Research Analysis (Bikram Kumar Sahu)

2) How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science (Gary King, *New York University*)

3) One Nation One Election : Analyzing the Impact on Indian Polity (Ekta Basoya ,CCS university)

4) Using Statistics as a Tool in Political Research. The case of Electoral Behaviour (Theodore Chadjipadelis)

5) Using Graphs Instead of Tables in Political Science (Jonathan P. Kastellec and Eduardo L. Leoni)

APPENDIX

Questionnaire :

1. Gender
Male
Female
2. Age
3. First Time Voter?
Yes
No
4. Wish to be responsible voter in shaping new parlimentarians
Yes
No
5. Satisfied with measures taken for woman empowerment in Parliament
Yes
No
6. Do you think that, women's candidate involvement in Election is Crucial?
Yes
No
7. Expectation about Education background of candidate
Post Graduate
Graduate
Literate is enough
No problem if illiterate
8. Expectation about speciality in field while allotting portfolio
Must belong to the field of department he heads
No problem if he is unaware of department he heads
9. Do you wish to give preference to youngsters while allocating candidature ticket by political party?
Yes
No
Age is not Criteria
10. After 2025, there will be delimitation implementor So number of MP may raise to 700 in parliament. Do you want to add new departments too in Central Ministry?
Yes
No
Happy with current system.
11. Do you want MPs to discuss more about rural development and more such schemes to be launched?
Yes
No
12. Which area you want MPs should focus?
Security

Employment
Infrastructure Development
Education

13. Do you feel that more women MPs can represent women related issues in parliament compare to men, so wish to increase number of women MPs?

Yes
No

14. Do you think that government should frame a training programme to be conducted for our policy maker every year?

Yes
No

15. Do you feel government should make Polity as compulsory subject so that yougsters can understand political system and choose right candidate to represent them?

Yes
No

16. Do you feel Nepotism should be avoided while allocating ticket to candidates?

Yes
No

17. Which candidate will you vote?

Candidate doing caste politics.
Candidate stressing development.
Candidate improving grassroot level problems.
Candidate who raise voice for common people.

18. Which government/ candidate will you prefer?

Candidate of coalition
Candidate of single party

19. Will you encourage your friends and relatives to vote to make stronger and stable government?

Yes
No

20. With changing scenario which expertise you wish to see in new parliament?

Technological expert like AI.
Expert in Geopolitics.
Expert in Finance.
Expert in framing developmental policies.

21. Any suggestion to improve political system.

DATA DICTIONARY

1. Gender
    - Male = 1
    - Female = 0
2. Age : Age in years
3. First time voter
    - Yes = 1
    - No = 0
4. responsible voter
    - Yes = 1
    - No = 0
5. Women Empowerment in Parliament
    - Yes = 1
    - No = 0
6. Womens Role in Elections
    - Yes = 1
    - No = 0
7. Candidates Education Expectations
    - Post Graduate = 3
    - Graduate = 0
    - Literate is enough = 1
    - No problem if illiterate = 2
8. Portfolio Specialization
    - Must belong to the field of department he heads = 0
    - No problem if he is unaware of department he heads = 1
9. Preference for Young Candidates
    - Yes = 2
    - No = 1
    - Age is not Criteria =0
10. New Central Ministry Departments
    - Yes = 2
    - No = 1
    - Happy with current system = 0
11. Focus on Rural Development
    - Yes = 1
    - No = 0
12. Priority areas for MPs
    - Security = 3
    - Employment = 1
    - Infrastructure Development = 2

- Education = 0

13. Increase Women MPs for Womens Issues
    - Yes = 1
    - No = 0
14. Policy Maker Training Program
    - Yes = 1
    - No = 0
15. Polity as a Compulsory Subject
    - Yes = 1
    - No =0
16. Avoid Nepotism in Ticket Allocation
    - Yes = 1
    - No = 0
17. Your Voting Preference
    - Candidate doing caste politics. = 0
    - Candidate stressing development. = 2
    - Candidate improving grassroot level problems. = 1
    - Candidate who raise voice for common people. = 3
18. Preferred Government or Candidate
    - Candidate of coalition = 0
    - Candidate of single party = 1
19. Encouraging Voting for Stability
    - Yes = 1
    - No = 0
20. Expertise Needed in New Parliament
    - Technological expert like AI. = 3
    - Expert in Geopolitics. = 1
    - Expert in Finance. = 0
    - Expert in framing developmental policies = 2
21. Any suggestion to improve political system

Python-Code

Chi-Square test

```
import numpy as np
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
data=pd.read_csv(r"C:\Users\DISHA_COMPUTER\Desktop\dplotpy.csv")
data
data.isnull().sum()
#from sklearn.preprocessing import LabelEncoder
#le = LabelEncoder()

#for col in data.columns:
   #if col != 'Age_group':
      #data[col] = le.fit_transform(data[col])
print("Columns available:", data.columns.tolist())
from scipy.stats import chi2_contingency
# Example1: Chi-square test between 'Gender' and Women_Empowerment _in_Parliament
contingency_table = pd.crosstab(data['Gender'], data["Women_Empowerment
_in_Parliament"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example2: Chi-square test between 'Gender' and Womens_Role_in_Elections
contingency_table = pd.crosstab(data['Gender'], data["Womens_Role_in_Elections"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example3: Chi-square test between 'Gender' and Candidates_Education_Expectations
contingency_table = pd.crosstab(data['Gender'], data["Candidates_Education_Expectations"]
)
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example4: Chi-square test between 'Gender' and Portfolio_Specialization
contingency_table = pd.crosstab(data['Gender'], data["Portfolio_Specialization"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
```

```
# Example5: Chi-square test between 'Gender' and Preference_for_Young_Candidates
contingency_table = pd.crosstab(data['Gender'], data["Preference_for_Young_Candidates"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example6: Chi-square test between 'Gender' and New_Central_Ministry_Departments
contingency_table = pd.crosstab(data['Gender'], data["New_Central_Ministry_Departments
"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example7: Chi-square test between 'Gender' and Focus_on_Rural_Development
contingency_table = pd.crosstab(data['Gender'], data["Focus_on_Rural_Development"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example8: Chi-square test between 'Gender' and Priority_areas_for_MPs
contingency_table = pd.crosstab(data['Gender'], data["Priority_areas_for_MPs"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example9: Chi-square test between 'Gender' and
Increase_Women_MPs_for_Womens_Issues
contingency_table = pd.crosstab(data['Gender'],
data["Increase_Women_MPs_for_Womens_Issues"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example10: Chi-square test between 'Gender' and Policy_Maker_Training_Program
contingency_table = pd.crosstab(data['Gender'], data["Policy_Maker_Training_Program"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example11: Chi-square test between 'Gender' and Polity_as_a_Compulsory_Subject
contingency_table = pd.crosstab(data['Gender'], data["Polity_as_a_Compulsory_Subject"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
```

```
contingency_table
# Example12: Chi-square test between 'Gender' and Avoid_Nepotism_in_Ticket_Allocation
contingency_table = pd.crosstab(data['Gender'],
data["Avoid_Nepotism_in_Ticket_Allocation"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example13: Chi-square test between 'Gender' and Your_Voting_Preference
contingency_table = pd.crosstab(data['Gender'], data["Your_Voting_Preference"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example14: Chi-square test between 'Gender' and  Preferred_Government_or_Candidate
contingency_table = pd.crosstab(data['Gender'],
data["Preferred_Government_or_Candidate"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
# Example15: Chi-square test between 'Gender' and  Expertise_Needed_in_New_Parliament
contingency_table = pd.crosstab(data['Gender'],
data["Expertise_Needed_in_New_Parliament"] )
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
print(f"Chi-Square Statistic: {chi2}")
print(f"P-Value: {p_value}")
contingency_table
```

Sentiment Analysis :

```
pip install wordcloud
import pandas as pd
import re
import nltk
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from nltk.sentiment import SentimentIntensityAnalyzer
# Ensure necessary NLTK downloads
nltk.download('vader_lexicon')
# Load the dataset with proper encoding
df = pd.read_csv(r"C:\Users\Nisha\OneDrive\Desktop\MAIN FILE.csv", encoding="ISO-
8859-1")
# Extract column 21 (adjust index if needed)
column_index = 20  # Column index starts from 0
column_name = df.columns[column_index]
```

```
text_data = df[column_name].dropna().astype(str)  # Drop NaN values and ensure strings
# Function to clean text
def clean_text(text):
    text = re.sub(r'http\S+', '', text)  # Remove URLs
    text = re.sub(r'[^a-zA-Z\s]', '', text)  # Remove special characters & numbers
    return text.lower().strip()  # Convert to lowercase
# Apply text cleaning
text_data_cleaned = text_data.apply(clean_text)
# Generate Word Cloud
wordcloud = WordCloud(width=800, height=400, background_color="white").generate("
".join(text_data_cleaned))

# Display the Word Cloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.title("Word Cloud of Column 21", fontsize=14)
plt.show()
# Sentiment Analysis using VADER
sia = SentimentIntensityAnalyzer()

# Analyze sentiment for each sentence
sentiment_df = text_data_cleaned.apply(lambda text: sia.polarity_scores(text))
sentiment_df = pd.DataFrame(list(sentiment_df))
sentiment_df['text'] = text_data_cleaned.values  # Add original text back

# Categorize sentiment
sentiment_df['sentiment_label'] = sentiment_df['compound'].apply(
    lambda score: "Positive" if score > 0.05 else "Negative" if score < -0.05 else "Neutral"
)

# Display results
print(sentiment_df[['text', 'sentiment_label', 'compound']])
```

<u>Model Building :</u>

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
data = pd.read_excel("C:/Users/Prabhune/Desktop/The data.xlsx")
data.head()
data.info()
data.isnull().sum()
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
for col in data.columns:
```

```
    if col != 'X2':
        data[col] = le.fit_transform(data[col])
print(data.head())
X = data[["X1", "X2", "X4", "X6", "X8", "X9", "X10", "X12", "X14", "X15", "X17", "X18",
"X20"]]
Y = data[["X16"]]
```

Logistic Regression

```
import sklearn
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X,Y, test_size=0.3,random_state=42)
x_train.shape, x_test.shape, y_train.shape, y_test.shape
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.linear_model import LogisticRegression
lg = LogisticRegression()
lg.fit(x_train,y_train)
lg_p = lg.predict(x_test)
lg.prob=lg.predict_proba(x_test)
accuracy = accuracy_score(y_test, lg_p)
conf_matrix = confusion_matrix(y_test, lg_p)
class_report = classification_report(y_test, lg_p)

# print the evaluation results
print(f'Accuracy: {accuracy:.4f}')
print('Confusion Matrix:')
print(conf_matrix)
print('Classification Report:')
print(class_report)
```

Undersampling

```
from imblearn.under_sampling import RandomUnderSampler
undersampler = RandomUnderSampler(random_state=42)

# Apply undersampling
X_resampled, y_resampled = undersampler.fit_resample(X, Y)

# Create a new DataFrame with undersampled data
df_resampled = pd.DataFrame(X_resampled, columns=X.columns)
df_resampled['X16'] = y_resampled

# Check new class distribution
print(y_resampled.value_counts())
X_train, X_test, Y_train, Y_test = train_test_split(X_resampled, y_resampled, test_size=0.3,
random_state=42, stratify=y_resampled)

print(f"Training size: {X_train.shape}, Testing size: {X_test.shape}")
```

Logistic Regression

```
# Initialize the model
model = LogisticRegression(class_weight="balanced",random_state=42)

# Train the model
model.fit(X_train, Y_train.values.ravel())  # .ravel() converts Y to 1D array
# Predict on test data
Y_pred = model.predict(X_test)
# Accuracy Score
accuracy = accuracy_score(Y_test, Y_pred)
print(f"Accuracy: {accuracy:.4f}")

# Confusion Matrix
print("Confusion Matrix:")
print(confusion_matrix(Y_test, Y_pred))

# Classification Report
print("Classification Report:")
print(classification_report(Y_test, Y_pred))
```

AdaBoost

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import AdaBoostClassifier
# Initialize base estimator (weak learner)
base_estimator = DecisionTreeClassifier(max_depth=1)

# Initialize AdaBoost classifier
adaboost = AdaBoostClassifier(base_estimator=base_estimator, n_estimators=50,
learning_rate=1.0, random_state=42)

# Train the model
adaboost.fit(X_train, Y_train)

# Predict on the test set
y_pred = adaboost.predict(X_test)

# Evaluate model performance
#accuracy
accuracy = accuracy_score(Y_test, y_pred)
print("Accuracy:", accuracy)
# Confusion Matrix
print("Confusion Matrix:")
print(confusion_matrix(Y_test, Y_pred))
#classification report
```

```
print("Classification Report:\n", classification_report(Y_test, y_pred))
```

Xgboost

```
!pip install xgboost
import xgboost as xgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
import pandas as pd
# Initialize XGBoost classifier
xgb_model = xgb.XGBClassifier(n_estimators=100, learning_rate=0.1, max_depth=3,
random_state=42)

# Train the model
xgb_model.fit(X_train, Y_train)

# Predict on the test set
y_pred = xgb_model.predict(X_test)

# Evaluate model performance
accuracy = accuracy_score(Y_test, y_pred)
print("Accuracy:", accuracy)
print("Confusion Matrix:")
print(confusion_matrix(Y_test, Y_pred))
print("Classification Report:\n", classification_report(Y_test, y_pred))
```

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
# Initialize Random Forest classifier with improved parameters
rf_model = RandomForestClassifier(
    n_estimators=200,  # More trees to capture complexity
    max_depth=None,  # Allow trees to grow fully
    min_samples_split=2,  # Default, but can reduce to 2 if needed
    min_samples_leaf=1,  # Default, allows deep splits
    max_features='sqrt',  # Default for classification
    random_state=42
)

# Train the model
rf_model.fit(X_train, Y_train)

# Predict on the test set
y_pred = rf_model.predict(X_test)

# Evaluate model performance
accuracy = accuracy_score(Y_test, y_pred)
print("Accuracy:", accuracy)
```

```
print("Confusion Matrix:")
print(confusion_matrix(Y_test, Y_pred))
print("Classification Report:\n", classification_report(Y_test, y_pred))
```

Gradient Boosting

```
from sklearn.ensemble import GradientBoostingClassifier
# Initialize Gradient Boosting classifier
gboost_model = GradientBoostingClassifier(
    n_estimators=100,  # Number of boosting stages
    learning_rate=0.1,  # Step size shrinkage
    max_depth=3,  # Depth of each tree
    subsample=0.8,  # Fraction of samples used per tree
    random_state=42
)

# Train the model
gboost_model.fit(X_train, Y_train)

# Predict on the test set
y_pred = gboost_model.predict(X_test)

# Evaluate model performance
accuracy = accuracy_score(Y_test, y_pred)
print("Accuracy:", accuracy)
print("Classification Report:\n", classification_report(Y_test, y_pred))
```