



# TEAM ETA

## PROJECT TITLE: US ELECTION PREDICTION

# MEET OUR TEAM



**ZARNAB ZAFAR (TEAM LEAD)**

**WAJAHAT HUSSAIN**

**M. MUSTAFA SHAH**

**AYESHA MAJEED**

**AISHA KAMRAN**

**SYEDA FATIMA**

**MOEED AHMAD**



# INTRODUCTION

In contemporary political analysis, predictive modeling plays a crucial role in understanding and forecasting election outcomes. This report delves into the application of machine learning techniques to predict political party affiliations based on socio-economic and demographic data from the 2020 U.S. elections. The analysis encompasses data preprocessing, feature selection, model training, evaluation, and prediction using various classifiers.

# Methodology



## Data Collection

Gather relevant data from diverse sources.



## Preprocessing

Clean, normalize, and engineer data.



## Model Selection

Choose and train ML/DL algorithms for peak performance.

# Data Collection

The dataset used in this analysis comprises socio-economic and demographic indicators from all 51 states in the U.S. These indicators include unemployment rates, gender ratios, poverty levels, polling data for Democrats and Republicans, and the number of candidates running in each state. The dataset is sourced from a cleaned Excel file, ensuring that the data is structured and ready for analysis without missing values.



# Data Preprocessing

The initial step in the analysis involved loading the dataset into a pandas DataFrame and inspecting its structure. Columns such as state names and party affiliations were encoded using Label Encoding to facilitate numerical analysis. This preprocessing step ensures that categorical data, like state names and party affiliations, are transformed into a format suitable for machine learning models.



# Feature Engineering and Selection



Features were chosen based on their potential influence on political outcomes, including socio-economic indicators like unemployment rates and poverty levels, demographic factors such as gender ratios, and political sentiment gauged through polling percentages for Democrats and Republicans. Feature selection aimed to capture the most influential variables that could predict party affiliation accurately.

# Model Training and Evaluation

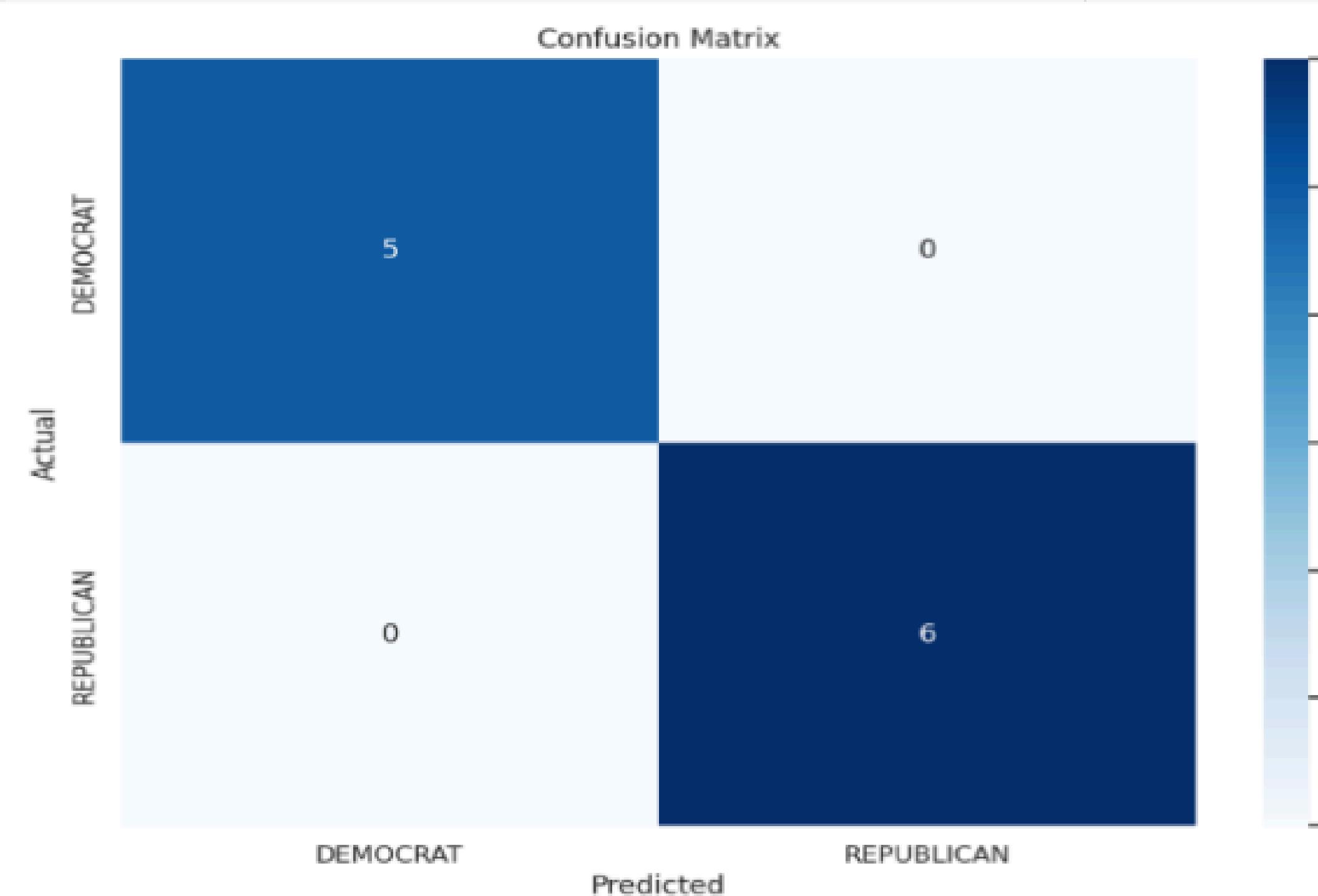


Several machine learning classifiers were trained and evaluated using the processed dataset:

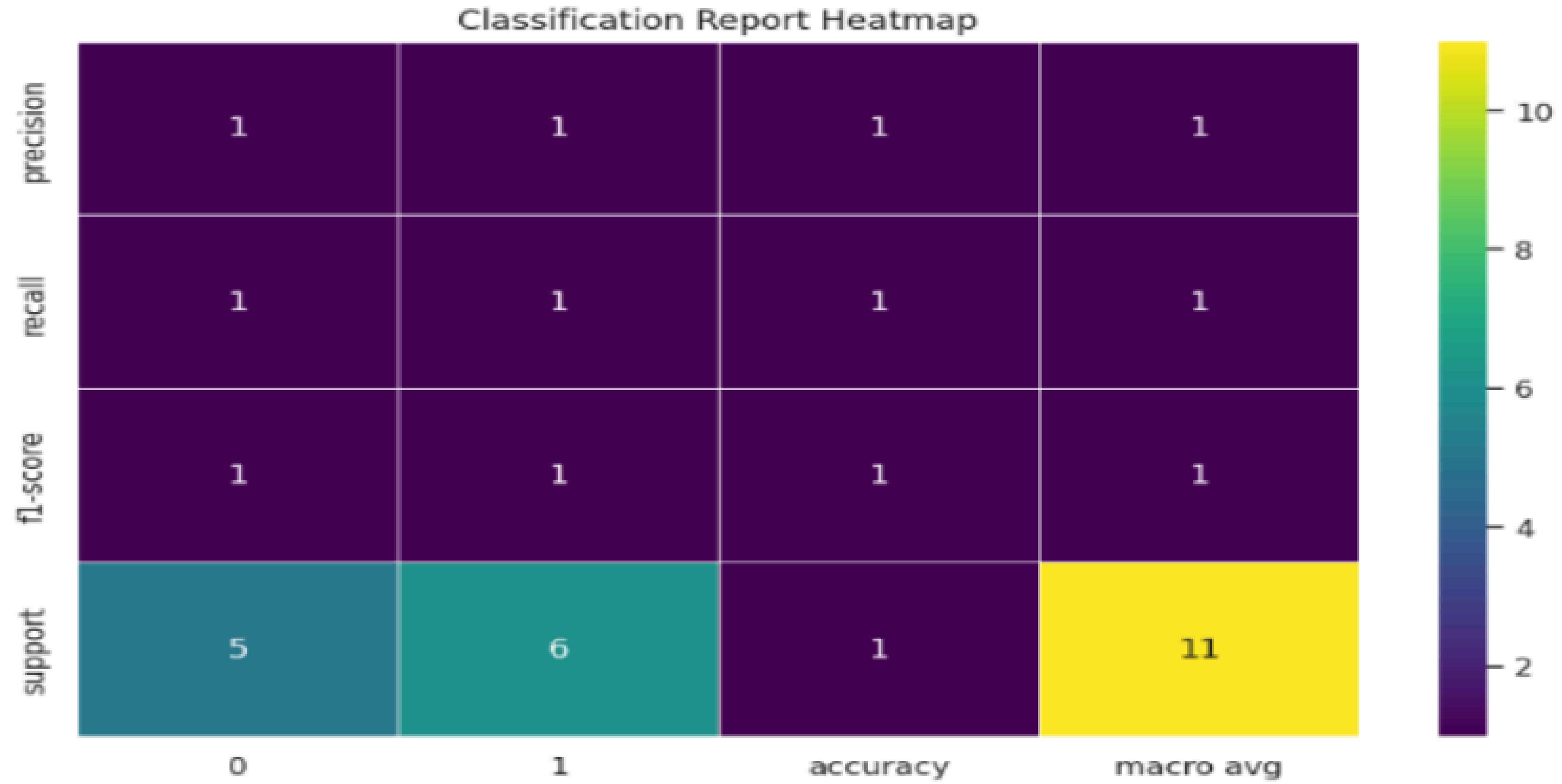
1. Decision Tree Classifier:
2. Random Forest Classifier:
3. XGBoost Classifier:
4. Voting Classifier (Ensemble):

# RESULTS:

## Confusion Matrix:



# Classification Report Heatmap:



# DEPLOYMENT

```
Enter value for Unemployment Rate: 5.9
Enter value for Sex ratio (males per 100 females): 93.7
Enter value for Below Poverty (%): 16.7
Enter value for Polling_Democrat(%): 37.8
Enter value for Polling_Republican(%): 57.4
Enter value for Candidates: 9
The predicted party for the given input is: REPUBLICAN
```

# CONCLUSION

In conclusion, this report highlights the application of machine learning techniques in predicting political party affiliations based on socio-economic and demographic factors. The ensemble model, comprising decision trees, random forests, and XGBoost classifiers, demonstrated robust performance in predicting party affiliations across states. The use of accurate data preprocessing, feature selection, and model evaluation ensures that the predictions are grounded in data-driven insights.

# FUTURE ANALYSIS

- **Feature Expansion:** Consider incorporating additional socio-economic and demographic variables that may influence political outcomes.
- **Model Tuning:** Optimize model parameters to improve accuracy and generalizability.
- **Real-Time Data Integration:** Integrate real-time data feeds to enhance model responsiveness and adaptability to dynamic political landscapes.

**THANK YOU**