

Summary

The purpose of this project is to build models to predict the outcome of the 2024 presidential election in Virginia and provide quantitative information about the precision of the prediction. The project leverages voting data from 2000 to 2020 from various sources including the IPUMS NHGIS dataset and data from the Virginia Geographic Information Network.

Our methodologies involve training machine learning models, linear regression, and decision trees to predict the likelihood of which party will win the majority for each county. In our project, we assess the effects of gerrymandering and how that can influence the overall result of the elections. Through our various visualizations, we uncover the relationship between certain demographics and voting patterns.

Our findings indicate that certain demographic features such as race, income, and population density are significant predictors of party preference. Additionally, other factors like proximity to urban centers affect voting behavior. The predictive models developed in this project provide valuable insight to the 2024 election results, and the analysis of the effect of gerrymandering reveals how the voting outcomes may be subjected to change.

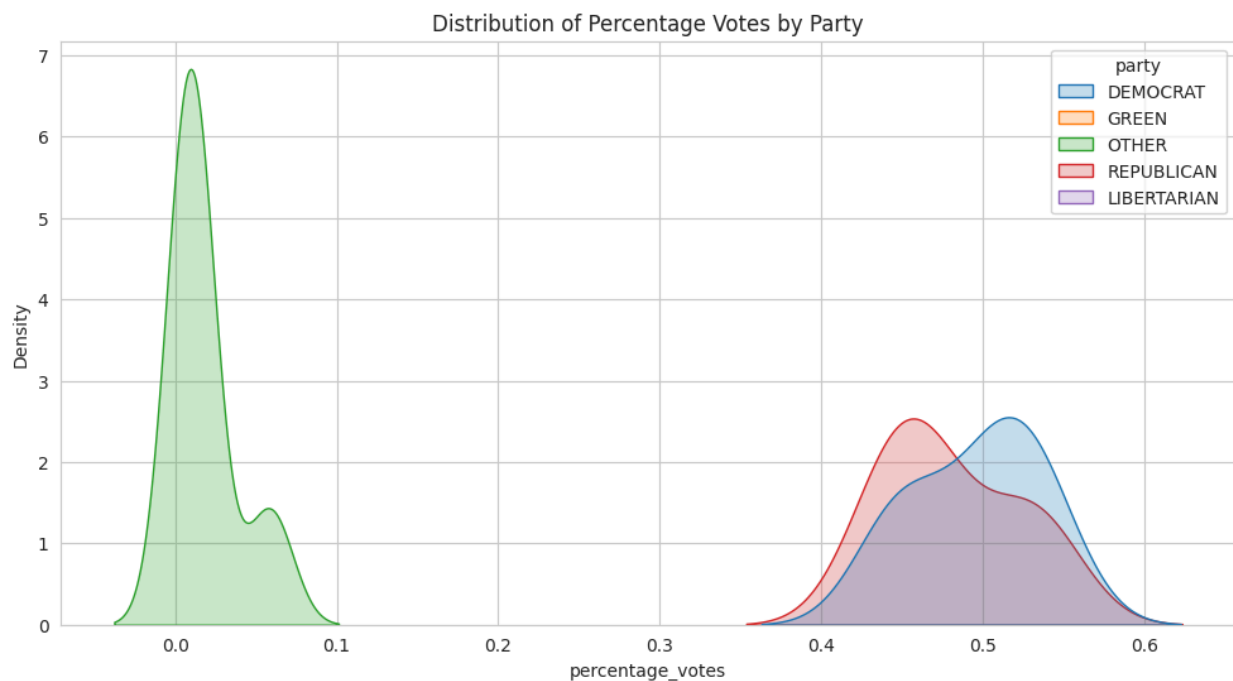
Data

This project uses voting data which includes historical data from Virginia elections from 2000 to 2020. Key variables used are year, indicating the year of the elections, county_name, indicating the name of the county, party, indicating the party affiliation, and totalvotes, indicating the number of votes received by a particular party in a county. No missing data points were found in this dataset. The dataset provides valuable insight to the historical voting patterns in Virginia.

The next dataset uses the NHGIS county data, which provides county level demographic information for every county in the United States. Some key variables include COUNTY, indicating the name of the country, QU1E001, indicating the median household incoming in the past year, QSQE002, indicating the white population, and QSQE003, indicating the black or African American population. Some more challenges with cleaning the data were naming conventions. There were some discrepancies in county names between datasets that needed to be changed before merging the set together. Next, the data was successfully merged combining geographic, demographic, and historical voting patterns through data wrangling.

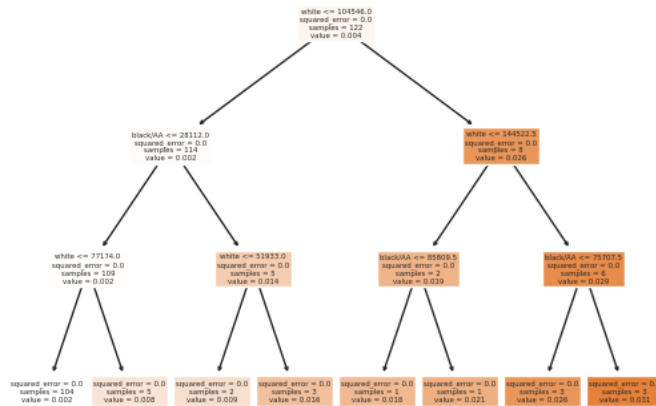
Results

The exploratory data analysis was crucial to uncovering key patterns and relationships in the data. The density plot displayed distinct voting trends for different parties over time. It showed that the Democratic Party had higher density around 0.55, while the Republican Party had higher density around 0.45. This suggests that the Democratic Party has won more times than the Republican party in Virginia.

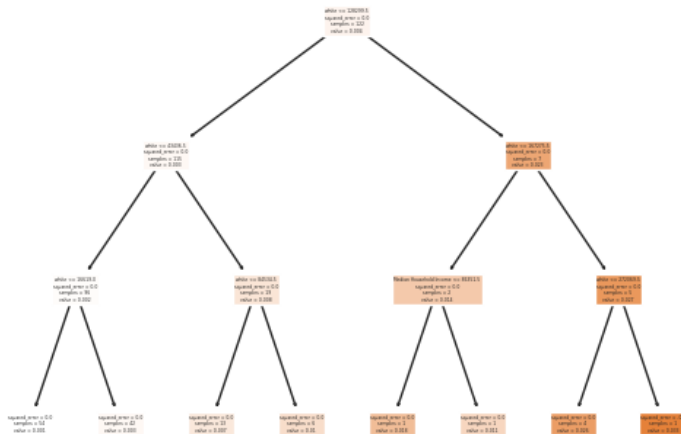


We developed and evaluated two predictive models to predict the winning party for each county in the 2024 elections. The first decision tree, Democratic support, used training data from years 2012 to 2016, and testing data was from 2020. The model achieved an R-squared value of 0.945. The graph shows the percentage of white and black/African American population and

median house income likely to vote Democratic. The model predicts a cumulative Democratic support of approximately 31.9% across Virginia counties.



The second decision tree, Republican support, used training data from years 2012 to 2016, and testing data was from 2020. The model achieved an R-squared value of 0.925. The graph shows the percentage of white and black/African American population and median house income likely to vote Republican. The model predicts a cumulative Republican support of approximately 34.3% across Virginia counties.



However, this may be subject to change when it comes to the effects of gerrymandering. Altering district boundaries can lead to significant changes to predicted outcomes, potentially flipping counties. Demographic and geographic features are significant predictors of party preferences. The models hold high confidence in predicting party preferences trends, but may be subject to change with any potential demographic shifts.

Conclusion

Our predictive model suggests that the 2024 Virginia elections are likely to follow similar patterns of previous elections, with Democrats securing urban and suburban counties and Republicans retaining majority in rural counties. Gerrymandering analysis indicates that these results can be flipped due to its effects. To improve the model more accurate and detailed demographics and voter behavior could be used.