

Boyu Zhang, Renjie Yao, Wensi Yang, Qinghan Yuan, Xiao Lu, Yuxiang Zhao

### Introduction

#### Motivation

Accurately predicting U.S. presidential election outcomes is challenging due to ever-changing political dynamics and diverse voter behaviors. Enhanced forecasting methods using advanced machine learning and visual analytics are vital. They not only improve campaign strategies and inform investment decisions but also promote voter participation by providing clear insights into electoral dynamics, thereby supporting a more informed and active democratic process.

#### Data

We utilized two datasets for our predictive model on U.S. Presidential Election outcomes. The first dataset includes economic and non-economic factors, sourced from established databases and detailed secondary research for scoring the impact of scandals, containing 19 rows covering elections from 1952 to 2024. The second dataset, used for visual analytics, consists of presidential primary polls, both current and historical, with around 30,000 records, enabling in-depth trend analysis of voter preferences over time.

### Approaches

#### ML Algorithms

Our predictive model blends economic and non-economic indicators like inflation, campaign spending, and scandal ratings, using Elastic Net for variable reduction and PCA to simplify data complexity. This method enhances prediction accuracy by effectively managing multicollinearity and improving model robustness, achieving substantial improvements in forecasting U.S. presidential election outcomes.

#### Interactive Visualizations

Our approach blends interactive maps with time-series charts to provide a nuanced view of voter trends. Clickable charts of map status cues track changes in polls, providing users with an engaging tool to discover political patterns over time and geographic location. This approach promotes deeper understanding through visual exploration.

### Experiments & Results

#### ML Algorithms

Our approaches were evaluated using key metrics such as the accuracy of binary election outcome predictions, the  $R^2$  value for goodness of fit, and cross-validation scores to ensure model robustness. Our model predicts that the vote share for the incumbent party (Democratic) is 50.54% for the upcoming election in 2024, slightly ahead of their competing party.

The initial Economy-Popularity model had limited predictive power with a negative  $R^2$ , but with the introduction of additional economic and non-economic variables and the application of PCA and Elastic Net for dimensionality reduction, we observed significant improvements. Our refined model demonstrated an  $R^2$  of 0.43 and 100% accuracy in binary classification for the winning party, a notable advance over previous methods.

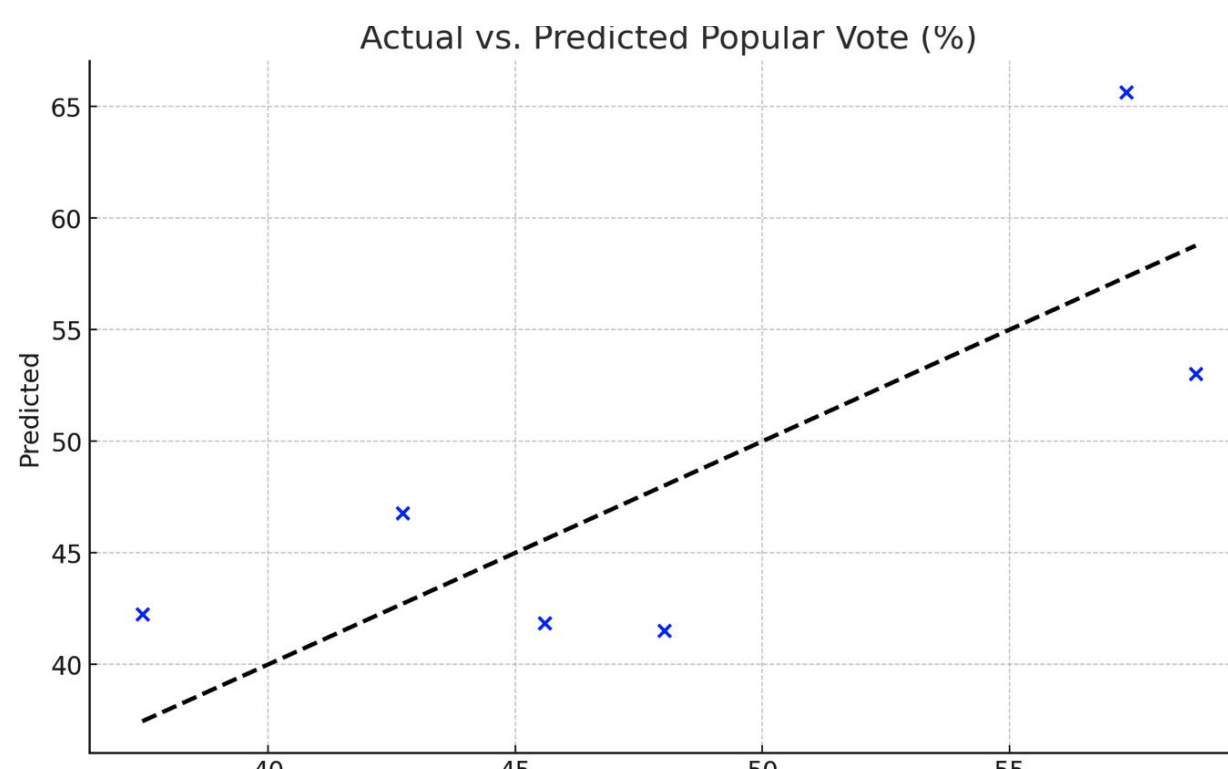


Figure 1: Prediction Result From PCA

The plot shows our PCA model's predicted popular vote percentages against actual outcomes. Data points close to the dashed trendline indicate accurate predictions, with a tendency for underprediction at higher vote percentages.

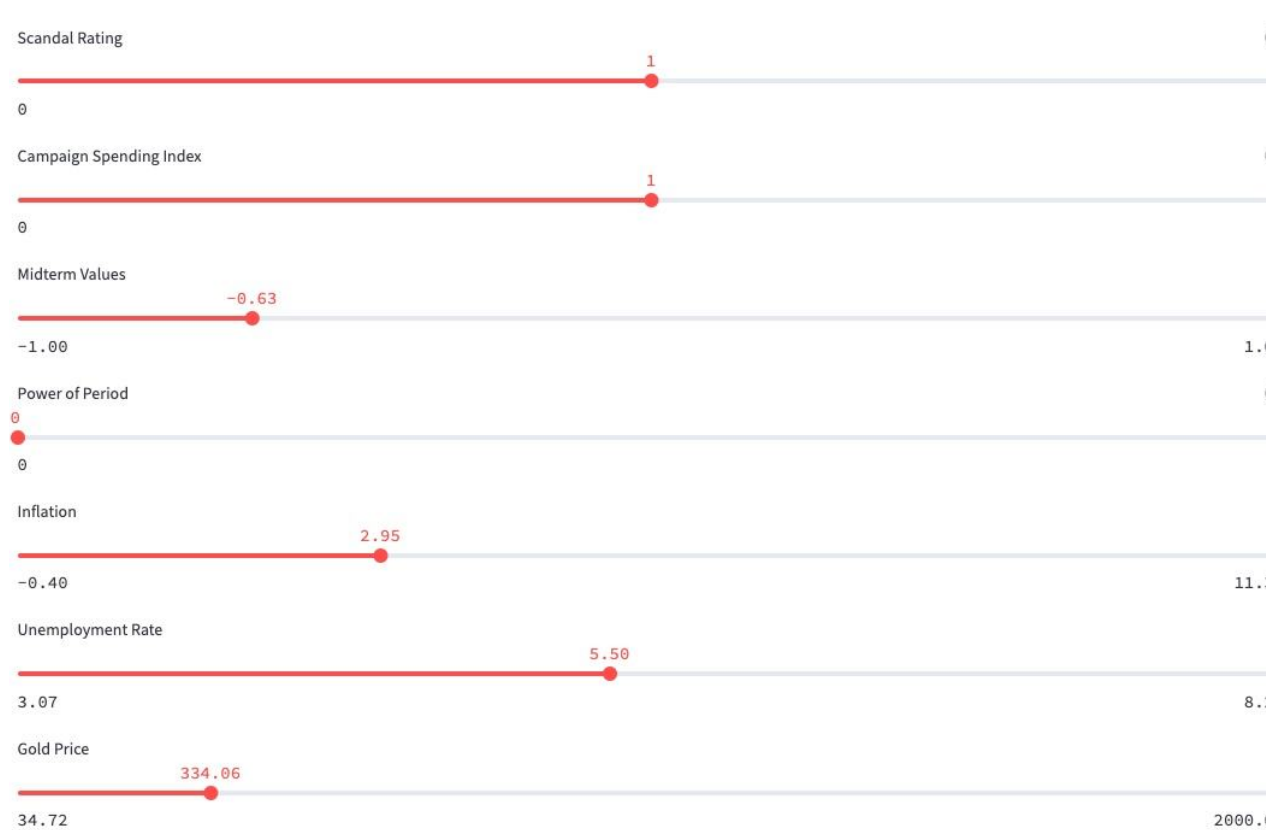


Figure 2: Predictor Interactive Dashboard

Our interactive dashboard allows users to dynamically adjust parameters, visually displaying their potential impact on election outcomes.

#### Interactive Visualizations

We integrated our work in a website, focusing on clarity, intuitiveness and the ability to provide insights. Users interacted through an interface that allowed them to visualize the election data through map, line chart, bar chart and table.

Users were able to seamlessly switch views between two candidates and drill down into time-series data by interacting with maps. Original data is also provided in the website, user can download the data to check more details by clicking on the downloading button. This interactive approach provides a more engaging and informative experience than static methods, allowing users to easily explore and understand complex data sets. The ability to filter data and dynamically update visualizations offers significant advantages over traditional static charts, enhancing user engagement and insight potential.

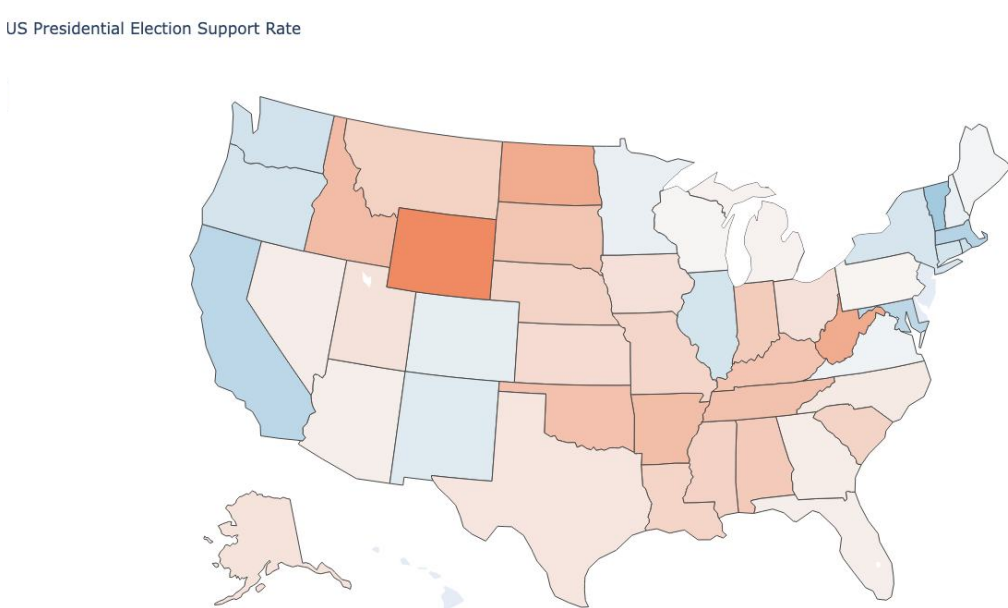


Figure 3: Interactive Map

Interactive map indicating the preference between Biden and Trump. Red means there is more support for Trump in that area and blue means there is more support for Biden.



Figure 4: Current President Approval Rate

Line chart showing the approval rate for the current president, red means "No" and blue means "Yes". We can see that starting from 2022, people are more dissatisfied with the current president.

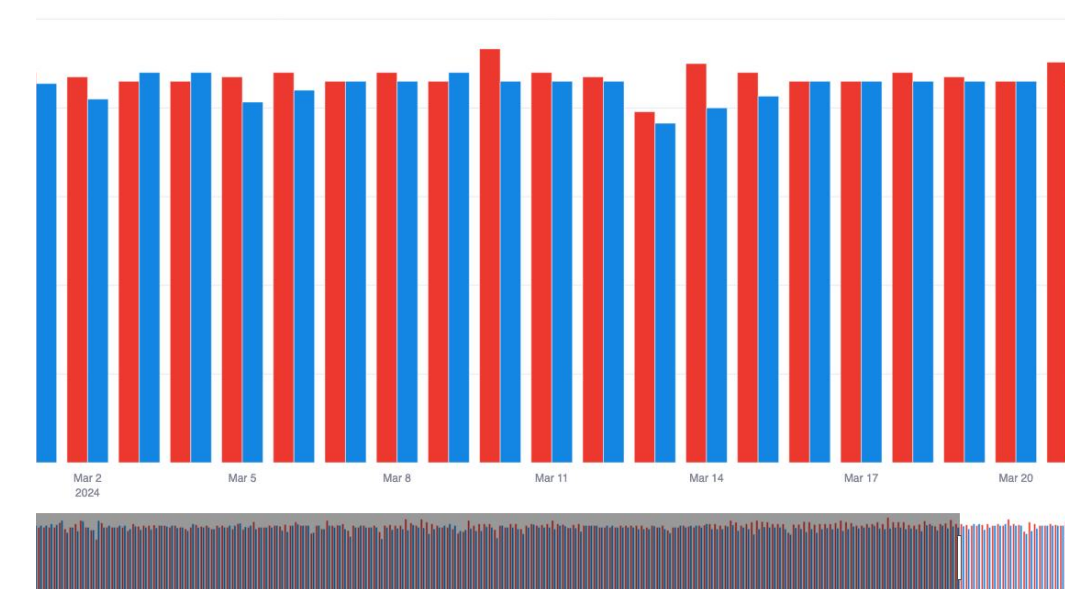


Figure 5: President Election Support Rate

Bar chart showing the support rate between Biden (blue) and Trump (red) in March 2024.

Table with columns: party, answer, candidate\_id, candidate\_name, pct. Includes a 'Download data as CSV' button.

	party	answer	candidate_id	candidate_name	pct
0	DEM	Biden	19,368	Joe Biden	44
1	REP	Trump	16,651	Donald Trump	43
2	DEM	Biden	19,368	Joe Biden	44
3	REP	Trump	16,651	Donald Trump	43
4	DEM	Biden	19,368	Joe Biden	44
5	REP	Trump	16,651	Donald Trump	43
6	DEM	Biden	19,368	Joe Biden	47.3
7	REP	Trump	16,651	Donald Trump	52.7
8	DEM	Biden	19,368	Joe Biden	49
9	REP	Trump	16,651	Donald Trump	51

Figure 6: Table with Download Button

Original data that shows support rate between the main candidates

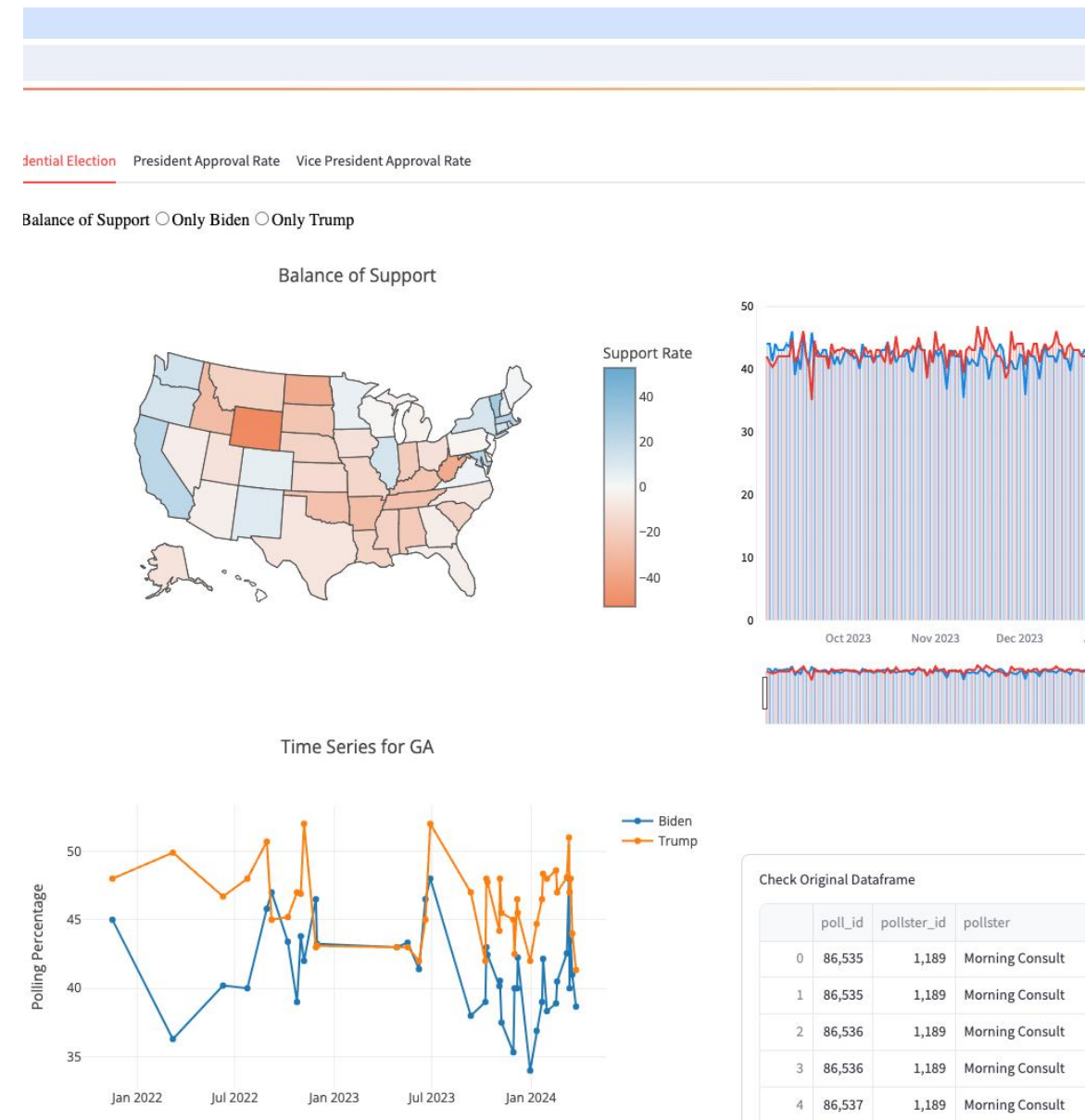


Figure 7: Website Design

Our website user interface that enables the user to explore with visualizations and prediction.