

ORIE 4741 Final Project

Fatima Al-Sammak (fma29) and Alexander Ellis (ase49)

GitHub Link: <https://github.com/fatimaalsammak/ORIE4741Project.git>

Question: Can we predict the voter turnout for a presidential election?

The primary dataset we will use to answer our question is the [National General Election VEP \(Voting Eligible Population\) Turnout Rates for 1789 - Present](#). It provides the VEP turnout rate for each Presidential election since 1789. This dataset will be what we use for our y_i parameters, or the *actual* voter turnout. We will join this dataset by year with other datasets that measure economic and political factors to build our x vector:

1. [Real Gross Domestic Product \(GDP\)](#), which is the United States GDP adjusted for inflation, measures the value of the goods the United States has sold each year. This data is only available starting from 1947.
2. [GINI Coefficient](#), which measures the economic inequality in the United States each year. This data is only available starting from 1963.
3. Incumbency, which is a binary variable indicating whether one of the candidates in the race is a sitting president at the time of the election. We will build this data by hand, which will make it easily available for all years since 1789.
4. [Primary election turnout](#), which is the percent of the voting eligible population that participates in primary elections. This data is only available starting from 2000.

Each of the first three factors is something that can have an impact on individuals' political involvement and, as a result, can be determinative of election turnout. The last factor, primary election turnout, might not necessarily be determinative of turnout but can be predictive as primary voting is one of the first available indicators of political involvement in an election cycle. Our study will involve teaching our algorithm to understand the relationship between each of these factors and the subsequent general election voter turnout.

Much of the mainstream news coverage today centers around politics and, in the years leading up to a presidential election (which is almost always) coverage of the election cycle is heavy. For example, recent clips from hosts across MSNBC's lineup show their clear interest in analyzing election and campaign events and in forecasting the upcoming election.

CNN takes this a step further, as they currently include an "[Election Center](#)" on their website, where they attempt to break down everything relevant about the upcoming 2024 presidential election. This page includes the "Road to 270," which is their model-based prediction of how electoral college votes will be distributed between the presumptive party nominees, Donald

Trump and Joe Biden, come election night. They also include a section that examines exit poll data they have collected from the primary elections that have been held so far and provides analysis of its significance towards the outcome of the general election in November.

It is evident that networks are interested in and collect primary data; we are recommending they take it a step further. The data available can be used to better understand and perhaps even predict election outcomes; forecasting election turnout would be a significant part of this endeavor.