

Matthew Burgess  
Capstone Project 1  
January 20, 2020

## **Data Science Capstone 1: Reliability of US Election Polls**

### **Problem Statement:**

In Presidential, Senatorial, and Congressional elections across the United States, all campaigns and media outlets rely on polling. Exit polls help determine decisions on election day. Polls of Iowa in presidential primaries help determine the front-runner for a given race. But how often do these polls reflect the actual outcome of elections? More importantly, how wrong are they? My goal is to create a model that predicts the error of a poll from the actual result.

### **Data Set:**

I plan on using data from the data news site FiveThirtyEight on polls. It is roughly 10000 rows by 25 columns from polls since 1998. Please see the link [here](#) to get to the CSV file. It not only includes data of polls of race by party, but also the year, partisan lean of both the pollster and the district, and real percentage of votes obtained by a candidate.

### **Application:**

Being able to reliably predict the percent error on a poll is arguably one of the single most important things for political campaigns. In knowing the error, polls can be determined as reliable or not, and percent error can be used to make key choices about the campaign moving forward.

### **Outcomes:**

The main outcomes from this project are to have a model that can predict the percent error of the poll from the actual outcome.