Self-Assessment Deliverable 8/23/2020

2020 Swing State Presidential Election Predictions

For our final project, we studied donation and voting data in six swing states (Arizona, Florida, North Carolina, Michigan, Pennsylvania and Wisconsin). Our team found this topic fascinating and of great importance. We had the combined interest of researching presidential election information. This was important to us, given the current state of affairs and all that is happening in our country with financial and health crises, during an election year. We thought it would be beneficial to study election information to see if we could learn anything about voting and how decisions are made. We decided to follow the money. Looking at donation data helped us understand a lot about Democratic and Republican parties per county and state in six historically swing states.

This project afforded me the opportunity to work in a team environment to accomplish one common goal. This was a huge yet rewarding task. My responsibilities revolved around the Machine Learning Model, technical pipelining, feature engineering, data cleaning, and sole responsibility of the python flask frontend dashboard. I identified the dataset as I was already familiar with this dataset from another project and knew that we could obtain all the donation records in the past 20 years. I knew that the dataset was rich – the six states we analyzed had over 15M records and in total there were over 85M records nationally. Working within the team we helped to identify the number of votes awarded to each presidential candidate since 2000 in each country across the country. We filtered out this dataset to just work with the six swing states of interest. After I identified the data source, Josh took off and built an ETL to download the files from the FEC site and upload them into the DB.

I spent a large amount of time architecting the solution to get the data in a fashion so that we could run a predictive model upon. As we started with records at the city level, the first step was to aggregate the donor data at a county level so we could compare it to the county level of votes for each candidate. I took this entire task and the aggregator took over 90 min to complete as it was not designed for efficiency but for accuracy and a row by row insertion process.

With the aggregated data in place along with demographics data that was keyed off the year and county plus the voting data segmented by election year, state, and county – we were finally set with all the data that we needed in order to run the models. I had the responsibility of running the full linear regression model to predict the number of red, blue, and other votes that were expected based off the 2016-current donation sets. I ran three models on six states producing 18 r2\_scores along with aic and bic to evaluate the models. After comparing Linear Regression, Random Forest, and Decision Tree models, our dataset yielded the best results with a Random Forest and thus we used that ultimately to run the final model to predict the Red or Blue results.

I then switched gears to write the entire dashboard with flask and SQLAlchemy integration to the PostGreSQL DB. I saved the files from the output of the jupyter notebook files to the directory on the flask app so that the server could organize and display all the results in one place. I then integrated the frontend to display the results of the voting predictions by the random forest model and showed the breakdown county by county but then also at a summary level.

I really had to step up my coding and integration game to get the results on a dashboard in time for the presentation. I really took ownership of the model and dashboard and dataset selection – essentially all the core functions of this final project. We collaborated, delegated, and communicated effectively as a team to deliver a great presentation in time.

The team worked very well together. We had protocols in place that allowed us to focus on specific aspects of the project. We met regularly, thru Zoom and communicated daily, via Slack. Now, 20 meetings, hundreds of hours of planning, preparation and coding as well as hundreds of Slack messages later, we were able to deliver our segments completed and on time. We worked together to complete tasks and follow thru on every part of the Mastery level section of the weekly rubrics. We usually agreed on the project and process. Where we didn’t agree, we discussed it and came to a resolution. Everyone brought a tremendous amount of skill and expertise to the table. Kanika is skilled in project planning and coordination, she helped keep us organized. Josh is very good at strategic planning, managing big datasets and technical details. Paul engineered our machine learning and flask protocols. I, Doris, was responsible for data and statistical analysis. We presented our project to the UCBE staff and students in a professional and timely fashion. We received positive reviews from our instructor, TA and fellow students. Completing the project was very rewarding!

A word of advance to a new cohort; communicate clearly and often. Set aside time to meet via Zoom or in-person. Assign tasks evenly and assist one another when necessary. Do not go it alone, the team mentality will be important.