Master Thesis Proposal

Class: 2021
AY: 2020-2021

1. What is your research topic?

Topic: Forecasting election results of federal and state-level elections in Germany using different machine learning techniques with macro-economic data as input.

Academic Fields:
1. Political Science:
-> Political behavior and voting
-> Election forecasting
2. Data Science:
-> Machine Learning
-> Artificial neural networks

2. What is your research question?

What role do changes in the over all economy of an election district play regarding the approval, i.e. the vote share, of an incumbent direct-candidate in the respective election district. To analyze the factors that actually have the biggest influence on the vote share of an incumbent candidate, several macro-economic variables are analyzed using different machine learning techniques (potentially including neural networks).

Studied objects:
- General elections from 1994-2017
- 299 electoral districts for general election in Germany
- Incumbent candidates in the respective districts
- Predicting the changes in vote share of incumbent direct candidates for the upcoming general election 2020

Dependent Variable: Normalized change in vote share of incumbent (or „change in candidate of the incumbent party“-dummy)

IV: Macro-economic data on district-level for all 299 districts from 1994-2017 (examples: ΔGDP, Δrent level, Δunemployment rate, Δ population, Δ R&D spending (public&private), economic development compared to neighboring districts, ΔProperty prices, Δ median income, Δinvestments)

Possible Controls (on election district level): average city size in district, population density, change of candidate own party, change of candidate other parties, former election results (direct and indirect vote, difference in vote share compared to other candidates), change in number of candidates running in the district, east/west state, gender of candidate, gender of most promising rival from other parties, academic title of candidate

Possible methods applied: Naive Bayes, Support Vector Machine, Random Forest, Linear and Logistic Regression

Alternative but very much related topic idea:
Identification of Swing States in Germany: Constituencies where the average change in the vote share of all parties is greater than the national average change in voteshare; clustering of these constituencies and identification of characteristics using machine learning techniques; analysis of the influence of macro-economic data in a second step.

3. Why would you like to explore this topic and question?

I think the topic and the research question are highly relevant because they answer the question to what degree macro-economic developments effect voting decision. The findings are important not only for the scientific community but also for parties and polticidal candidates searching reelection. Moreover, I find particularly interesting to assess not only who wins an election, but also who vote shares change even though there was no change in office. Additionally, I am very much interested in machine learning and deep learning techniques and look forward to deepen my knowledge in these areas by applying their techniques as part of my master thesis research.

4. What are the debates to which your dissertation will relate?

It's the economy, stupid! The most important debate my research will touch upon is the question to what degree economic developments influence election outcomes. It also relates to the debate which factors are most important regarding election decisions and wether the influence of economic factors is negligible or not. Moreover, it will relate to an ongoing discussion and research about the possibilities and best methods to predict election outcomes.

5. What is your proposed methodology?

My proposed methodology is mostly quantitate using analytical modeling.

To predict election outcomes and analyze the scope of the influence of certain economic indicators on the vote share (and also to potentially cluster different districts) I am planning on using different machine learning techniques such as Naive Bayes, Support Vector Machine, Random Forest, Linear and Logistic Regression. Especially for the forecasting part of the thesis I am also planning on using artificial neural networks.

6. What type of data do you need?

All of the data needed is public and most likely early accessible. It could be that the economic data is not available tailored to election districts (which problem could be solves with some data manipulation).

- German general elections results on district level from 1994-2017
- German general elections candidates on district level from 1994-2017
- Macro-economic data (several indicators) on district level from 1994-2020

7. Who are your proposed supervisors?

1st: Prof. Simon Munzert
2nd: Prof. Slava Jankin
3rd: Prof. Mark Kayser