Objective:

The objective of this project is to create a model (or models) through machine learning that will predict election results better than current models using county level data on past elections and demographic/economic factors.

Justification:  
This project should be completed because it could potentially create a more accurate and therefore useful model for predicting elections, and it develops data science skills that will be important to use in future research. Predicting elections could be useful in both monitoring local political sentiment and understanding how and why people vote the way they do based upon certain demographic factors.

Description:  
This project will be using machine learning techniques and a deep learning framework called TensorFlow to create the model(s) to identify what model(s) is/are superior or more effective. The models will be created by first sorting the data into usable categories. Then the model(s) will be trained. The training will occur by first generating a random “vector” from the variables in however many dimensions as there are variables, and then make a prediction based upon this vector. A residual between this and a real data value will be created, and the derivative of this (the gradient) will be used to adjust the model to create a more accurate vector. Over time, as the model is continually changed with respect to each variable, it will “learn” how to accurately predict values. For the purpose of this experiment, the dependent variable will be the final “loss” of the model after being trained (where a lower loss is better, as loss is essentially a residual) and the independent variables the economic and demographic data used, such as median income, racial makeup of the county, past results, etc.

Limitations:

While this project may not have traditional limitations, there is still an issue of computing power and the time taken to calculate these results. Oftentimes, large datasets can take vast amounts of time to computer answers (sometimes in the area of weeks) and this is certainly a limitation to the efficiency of the project, or depending on how long it takes, how much of the dataset/how many variables I can even use to create the model.

Feasibility Study  
**Resources currently available:**  
Personnel:  
I can talk to Mr. Writer about the subject and ask for advice, as well as my father and the team he runs at Verizon who work with TensorFlow and machine learning as their job.

Equipment:   
I already have a powerful computing laptop with GPU support that can help with processing the data and running/creating code.

Supplies:  
There are no supplies required.

Knowledge/skills:

I already know some elementary calculus and linear algebra which are necessary to understanding machine learning, and I also know how to program in Java.

**Resources that are still needed:**  
Personnel:  
I need to reach out to the authors of one of the studies I have, as they may be useful in both understanding machine learning as well as the modeling involved.

Equipment:

I need to access certain courses and books. As of right now, I need to access the following courses or resources (all of which are free):   
<https://www.codecademy.com/learn/learn-python>   
<http://codingbat.com/python>   
<https://www.coursera.org/specializations/mathematics-machine-learning>   
<https://www.coursera.org/learn/machine-learning>   
<https://www.coursera.org/specializations/aml>   
<https://www.kaggle.com/learn/machine-learning>

Supplies:

There are no other supplies needed.

Knowledge/skills:

I still need to learn some basic multivariable calculus to learn how to properly train the model. Additionally, I need to relearn some more linear algebra techniques regarding graphs, networks, and matrix operations in order to understand how to create models, and I need to learn Python and how to use its libraries to code it. Finally, I also need to learn how to create the models and machine learning algorithms.

Proposed Budget:

Any course I am currently considering is free, so at least as of right now there is no budget associated.

Safety Issues:

As this project essentially only has a laptop computer as a piece of equipment, there are no safety issues involved with this project.

Alternatives:

If the current structure of the project does not work out, it is possible that different neural network structures can be tried in order to alter the algorithm or how much coding will be required to create a basic model, and then proceed to add complexity from there.