This is the project proposal for Darwin Boersma, ,and , in which we would like to use probabilistic algorithms and methods in order to create an election predictor for this upcoming 2018 midterm.

\*\*\*\* I can add information about the data investigation I’ve done so far- Michael. I’ve been building a comprehensive pandas dataframe””

At one layer we will solve the problem using a neural network model that trains on 80 percent of the 2016 election data and then is tested against the other 20 percent of 2016 data to establish a baseline model.

Then we could begin to train the same model with the 2018 election data that we have so far also by county, as long as we build the dataframe in a consistent and scalable enough way, this should not be a problem.

In this way we will be able to use labeled machine learning to predict a binary output, goes for GOP vs goes for democrat. The possibility that the state goes to some extra third party is marginal, based on real data, this almost never seems to happen, however a county in theory that goes to a third party should just be mapped to [0][0] by our machine, that is not GOP and not democrat. This simplifies the network structure.

Details of first neural network structure (basic architecture).

Hidden layers: >= 2.

Type of neural network: input [1, …. N] input fields each in the range [0, 1], output: [0, 1] x [0, 1], gradient descent optimization, sigmoid activation function; labeled machine learning, simple neural network model (not convolutional net, or RNN, just simple NN. )

We will flatten the input for each county into a 1D array with N fields, initialize random weights for all our layers, multiply through the next layer, add a bias and then activate for our 2 or more layers. The final layer will shape the output into just 2 output classes, GOP [1][0] or democrats [0][1]. (or neither [0][0]). In reality the output elements will be some probability from [0, 1] not 0 or 1.

At each step we feed the county data into the network, calculate the error and backpropagate it. We can accomplish this with keras and tensorflow to do the heavy lifting for us, and we should be able to have a beautiful simple neural network model running by late october.

During this time, darwin will be working to implement bayesian reasoning to do a predictor along a parallel pathway. He will lay out his project architecture below. Our cumulative result this semester will be the results of all algorithms we apply the problem to.