Analytics Storytelling:

Assignment II

Austin Koenig

Due 20 September, 2019

Submission Idea 1

Working Title The Effect of Noise in Training Data for Predictive Models, A Brief Noise Experiment

Plot, Twists, and Endings This blog will explore the effect of noise in data on different models. We will test three types of models – random forest, polynomial regression, and neural network – using the error with respect to the actual data value. We expect to see that as the amount of noise increases, the error of all models increase. Which model will be able to withstand noise the most?

Data Description and Research Plan The training data and testing data will be generated from a sine wave on the domain $[0, 2\pi]$. The training input will be uniform random samples from $[0, 2\pi]$ and the training output will use the sine of the inputs as a mean for the normal distribution with variable standard deviation. The testing data will be generated similarly, beside the fact that we will instead use the exact sine value as the output. The three mentioned models above will be trained upon several datasets with varying degrees of noise (at least five, maybe more), for a total of at least 15 models to be tested against one another.

Submission Idea 2

Working Title A Study on Global Energy Production

Plot, Twists, and Endings This blog will examine the global production of renewable energy sources against that of nonrenewable energy sources between different countries. The object is to outline which countries in the world exemplifies the

1. overproduction of nonrenewable energy and underproduction of renewable energy, as well as
2. an appropriate use of both renewable and nonrenewable energy.

In the end, we will have an idea of who to look towards for the most energy production reform as well as an idea of who to look towards as an archetype for energy production.

Data Description and Research Plan The data in question will come from UNdata, a trustworthy source of information. We will pull several datasets which encompass a variety of energy production indicators.

I've included two submission ideas because I cannot decide which is more interesting. Frankly, this was a hard assignment for me because I'm so used to receiving specific instructions as to what kind of results we want in the end. Thus, I don't know which idea is more interesting because I can't pinpoint the audience I'm writing for.

In the dreadful case that neither of these ideas are interesting enough, I will continue to think of other ideas. Have you any input for whether I should go one way or another? Thanks for taking the extra time to read, as I know this was probably much longer than you were hoping for.