Springboard Capstone Project – Hunter Browning-Smith

As the world transitions from fossil fuels to more sustainable sources of energy, the energy grid must adapt to these changes. The grid is impacted because these sustainable sources of energy – such as wind and solar – do not produce a steady stream of electrical power. They are intermittent. The power output of a solar panel depends on the amount of sunlight impending on itself, and wind turbines depend on the kinetic energy of the wind blowing.

Solar power in particular may more strongly impact the electrical grid. As more homes and commercial facilities install solar panels, this "decentralized" model of energy production adds stress to the existing grid through its variable energy output and the ability to push electricity back to the energy providers.

One way to mitigate the negative impacts of intermittent sources of energy on grid stability is to implement a decentralized smart grid control system. Such a system would be able to assign accurate pricing to electricity to better meet supply and demand, thereby reducing costs.

The dataset analyzed is one which contains 11 predictive attributes – tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, and g4. There are two dependent variables to represent grid stability – one is the numerical stab' variable, whose value is positive for an unstable grid and negative for a stable one, and the other is stabf, represented as text as either "unstable" or "stable." The data set can be found here: https://archive.ics.uci.edu/ml/datasets/Electrical+Grid+Stability+Simulated+Data+

I used logistic and linear regression techniques to build the most accurate model of the system. The logistic regression model performed better than the linear regression model.

The logistic regression model resulted in accurate grid stability predictions based on the variable stabf and the independent variables tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, and g4. The model was iterated 100 times to determine the most accurate score using the sets of independent variables which produced the top 10 highest accuracy scores.

Similarly, I iterated through all of the permutations of linear models possible given the independent variables tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, and g4, and the dependent variable stab. The model which yields the highest Rsquared value and lowest AIC value is the one which includes all of the mentioned independent variables. The highest Rsquared value achieved is 0.46396, indicating that the linear model only accounts for 46.396% of the variation of the predicted stab value from the actual stab value.

Based on the linear and logistic regression results, I recommend using a logistic regression model using the independent variables tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, and g4, to determine stabf, the binary stability value.

Energy providers, energy policy makers, and homeowners with solar panels would be interested in this data model because they can more accurately quantify the impacts of variable energy types when choosing an energy production system.