Applying Machine Learning Techniques to Evaluate Electrical Grid Stability in Decentralized Models

A Springboard Capstone Project

By: Hunter Browning-Smith

Introduction

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.

Motivation

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.

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.

Data Description

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+. Table 1 describes the variables used in the data set.

Table 1: Independent Variables of Decentralized Smart Grid Control

P_x	Mechanical power produced/consumed
τ_x	Reaction time, the delay between a price change and adaptation to it
γ_x	Coefficient, proportional to price elasticity
See the following paper for more detail about the variables used: Arzamasov, Vadim, Klemens Böhm, and Patrick Jochem. 'Towards Concise Models of Grid Stability.' Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2018 IEEE International Conference on. IEEE, 2018 (Section V-A)	

Modelling Details

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. First, the data was fitted to a logistic regression model with each possible combination of independent variables. The resulting dataframe contained 2047 combinations of independent variables and their corresponding accuracy scores. Of those models, the ten combinations of independent variables with the highest accuracy scores were placed into a new dataframe. A logistic regression was then again applied to the top 10 combinations 100 times. The average accuracy score of the regression modelling was found, and the combination of independent variables which produced the highest average accuracy score after 100 iterations was deemed to be the best model. Table 2 shows the average accuracy scores of different models after 100 iterations.

Table 2: Top 10 Logistic Regression Models and Accuracy Scores

Independent Variables	Accuracy Score
[tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, g4]	0.815915
[tau1, tau2, tau3, tau4, p3, p4, g1, g2, g3, g4]	0.81583
[tau1, tau2, tau3, tau4, p2, p3, g1, g2, g3, g4]	0.815275
[tau1, tau2, tau3, tau4, p4, g1, g2, g3, g4]	0.814865
[tau1, tau2, tau3, tau4, p2, p4, g1, g2, g3, g4]	0.81472
[tau1, tau2, tau3, tau4, p2, g1, g2, g3, g4]	0.81429
[tau1, tau2, tau3, tau4, g1, g2, g3, g4]	0.81289
[tau1, tau2, tau3, tau4, p2, p4, g2, g3, g4]	0.798405
[tau1, tau2, tau3, tau4, p2, p3, g1, g2, g3]	0.794655
[tau1, tau3, tau4, g1, g2, g3, g4]	0.79453

For linear regression modelling, I performed several EDA steps to identify any obvious trends between the independent variables and the numerical dependent variable 'stab,' which is a value that is either positive or negative. Figure 1 shows the numerical relationships between each independent variable and the outcome variable stab.

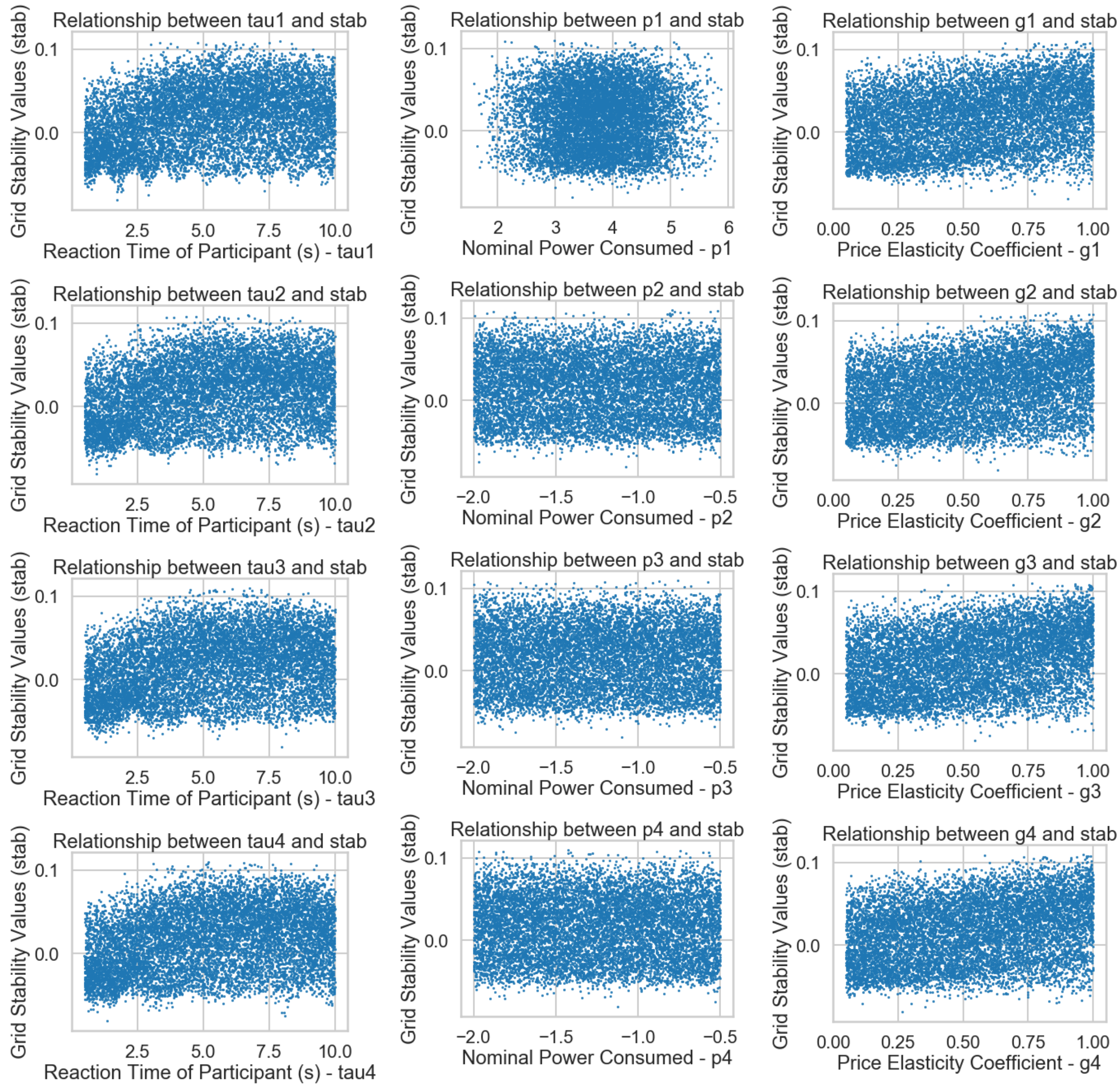

Figure 1: Linear relationship between all independent variables and numerical representation of grid stability (stab)

Similar to the methodology used for the logistic regression, I utilized the same iteration technique to establish which 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 R-squared value and lowest AIC value is the one which includes all of the mentioned independent variables. The highest R-squared value achieved is 0.46396, indicating that the best linear model only accounts for 46.396% of the variation of the predicted stab value from the actual stab value.

Table 3: Top 5 Linear Regression Models Based on OLS R-Squared, OLS F-Statistic, and OLS AIC

Independent Variables	OLS R-Squared Value	OLS F-statistic	OLS AIC
[tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, g4]	0.46396	785.981	-42147.8
[tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3]	0.44653	805.974	-41829.8
[tau1, tau2, tau3, p2, p3, p4, g1, g2, g3, g4]	0.446041	804.383	-41821
[tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, g4]	0.445351	802.141	-41808.5
[tau1, tau3, tau4, p2, p3, p4, g1, g2, g3, g4]	0.445242	801.786	-41806.6

Results

Based on the linear and logistic regression results, logistic regression produced the most accurate model using the independent variables tau1, tau2, tau3, tau4, p2, p3, p4, g1, g2, g3, and g4, to determine stabf, the binary stability value. Linear regression provided a model which could account for less than half of the overall variability.

Conclusion

I recommend using logistic regression for predicting grid stability using these types of data inputs. Logistic regression produced the best model, which predicted grid stability with an accuracy of greater than 80%. Providing additional data inputs may enhance the models ability to accurate predict grid stability. There appears to be no significant correlation between the provided independent variables and the numerical value for grid stability. If a linear regression technique is preferred, I recommend exploring how the numerical grid stability variable ‘stab’ is created.