James Taylor

|  |  |
| --- | --- |
| **Data Analytics Project/Project Scope** | **Key Performance Indicators (how success will be measured)** |
| Predict the amount of power generated by solar panels given current conditions. This would help power grid officials coordinate power sources to meet demands of the grid. | * Residuals (Actual vs Predicted) * Accuracy on binned Actuals vs Predicted * Models tested on non-training data |
| **Project Importance** | **Planned Milestones** |
| Power grids need to plan how to meet power demand. Accurately predicting solar’s input to the grid would make energy production more efficient/environmental friendly since things like coal and natural gas will be needed less. Knowing how much power can be produced in a certain region can plan for transmission loss. | * Determine useful applications for models * Clean/partition data * Determine what model type(s) works best * Train and test on unseen data |
| **Background** | **Proposed Data Sets** |
| Data publicly available from the energy transmission company. I have not seen analysis done on it. | Dataset from Belgian electrical transmission company. Solar production from solar plants in a region. Have another weather data source. |
| **Project Domain/Business Objectives** | **Proposed Tools (We understand these may change.)** |
| The area of interest for this is the energy sector. Some business objectives involved are optimization, prediction and efficiency. | All the work will be done in Python 3. The software will be open source like numpy, pandas, tensorflow, matplotlib and others. |
| **Success Factors** | **Data Analysis** |
| The quality of the data will determine accuracy in both training and test set. How far in the future the model is meant to predict will have a large influence on the accuracy and purpose of the model. | I expect cloud cover to be the largest determinant of energy generation. Solar intensity will likely be important to, were morning and evenings won’t be as producing as mid-day. |