Seattle Power Outage Predictor

Yuanyuan Shi [a], Yanbo Qi [b], Ryan Kastilani [b], and Jerry Chen [b]

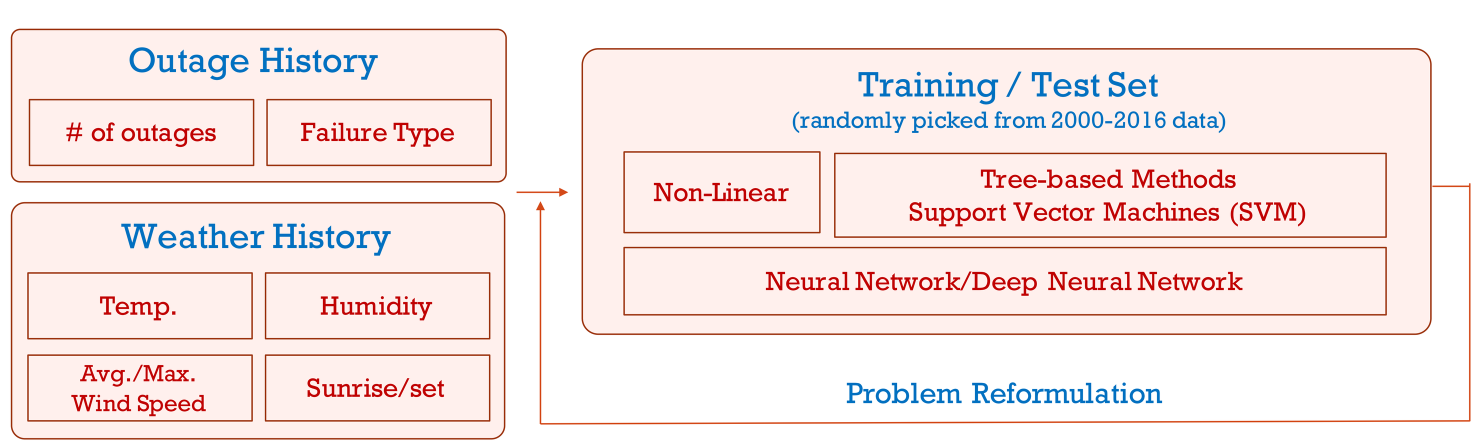
[a] Department of Electrical Engineering , [b] Department of Chemical Engineering
University of Washington , Seattle, Washington

Background

Power outage happens almost every day in Seattle. Large-scale blackouts may affect thousands of people, resulting in millions of dollars economic loss. Due to the limited prediction ability, Seattle City Light, a major utility company in the Seattle area, cannot estimate the number of power outages accurately. As a result, they do not know how many specialists should they have on call for a certain day to minimize the outage damage. In this project, we attempt to use machine learning techniques including Tree-based Methods, Support Vector Machines, and Neural Network methods to predict the possibility of major power outage based on the weather forecast. We also developed a graphical user interface to visualize our prediction as well as history data. This project can help the Seattle City Light to understand the pattern of power outages better and prepare proactive maintenance plan accordingly.

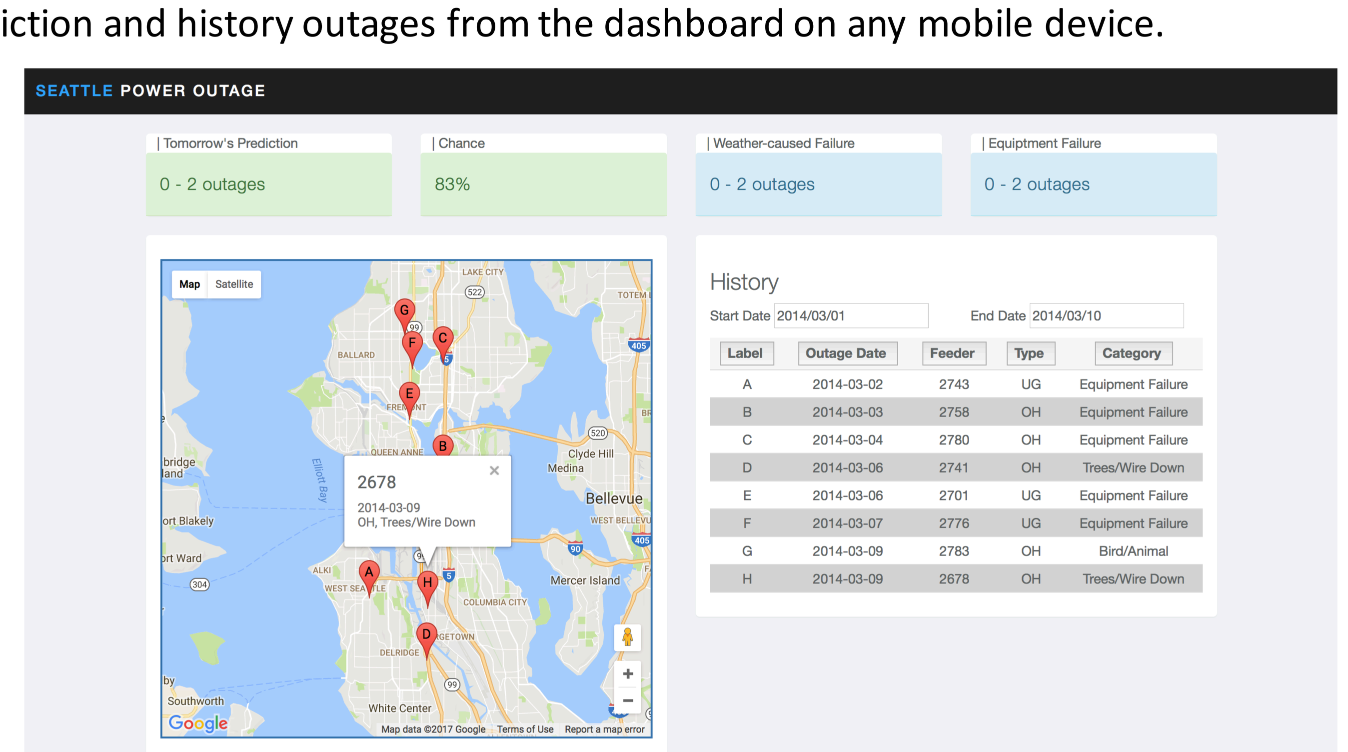
Data Processing

We studied the Seattle secondary network outage events caused by equipment, treefall, animal and lighting-caused failures from 2000/09/11 to 2016/03/14. We incorporated weather data including temperature, wind speed, precipitation, visibility, humidity, weather event (rain, fog, snow, lightening etc.), and day length. We consider each outage happens independently, so it is only correlated with the weather data itself. We considered it as classification problem: a **normal** (0 - 2 outages), **bad** (3 - 7 outages), and **extreme** (8+ outages) cases are studied.



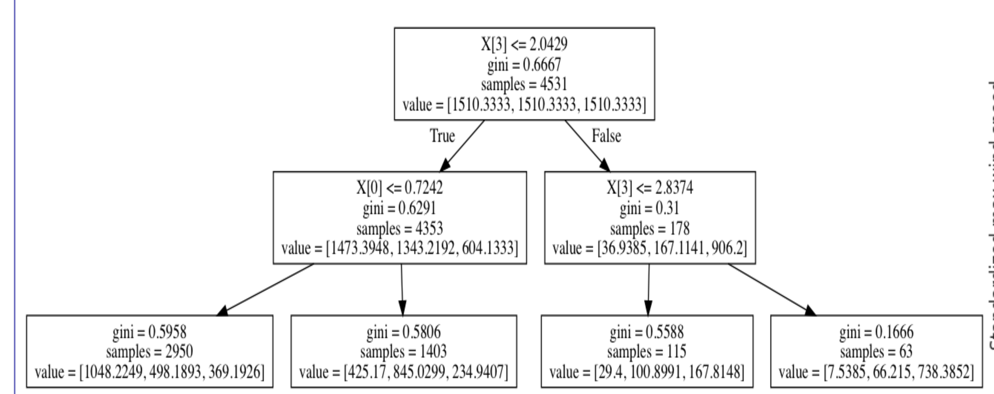
Graphic User Interface

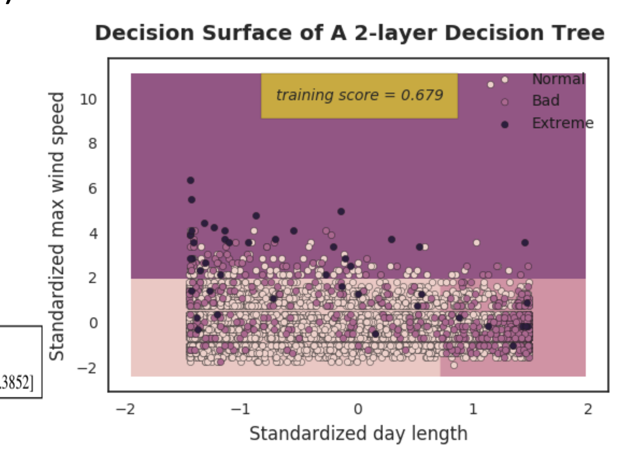
A web-based user interactive interface is built using Javascript. Users could read the prediction and history outages from the dashboard on any mobile device.



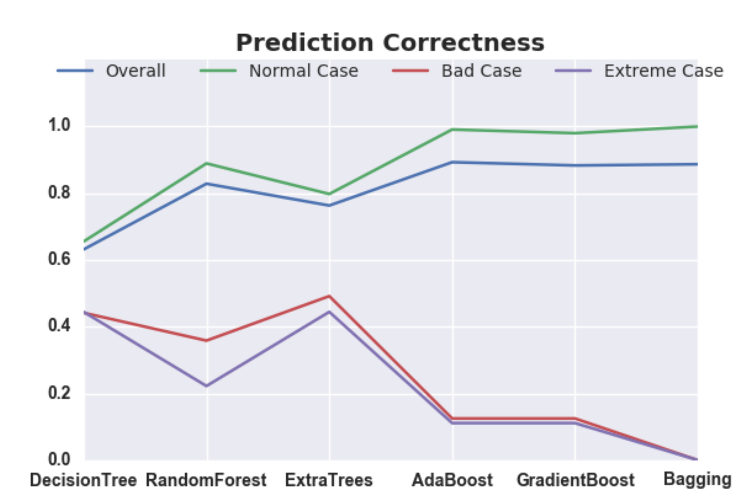
Tree-based Methods

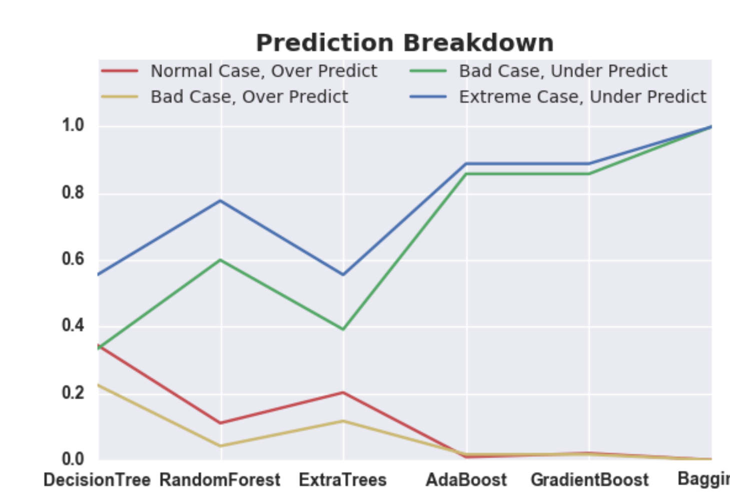
- **Tree-based classification methods**: hierarchical way of partitioning the space
- **3 general elements**: the selection of the splits; declare a node terminal; assign class
- **Features**: day length (hr), average temperature (F), average humidity (%), max wind speed (mph), precipitation (in)





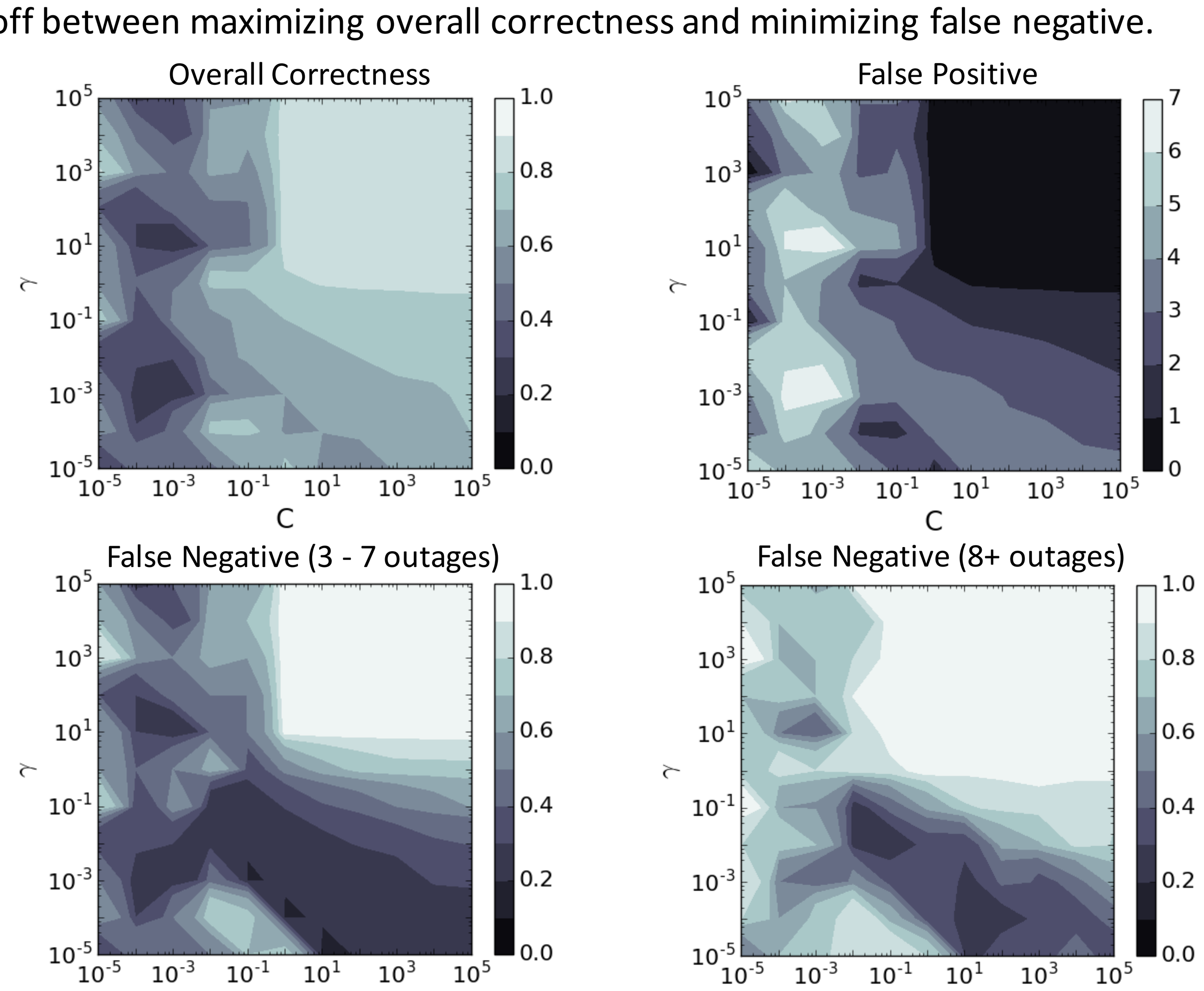
- **Ensemble Tree Methods**: grow multiple trees to yield a consensus prediction bagging, boosting, random forest





Support Vector Machines

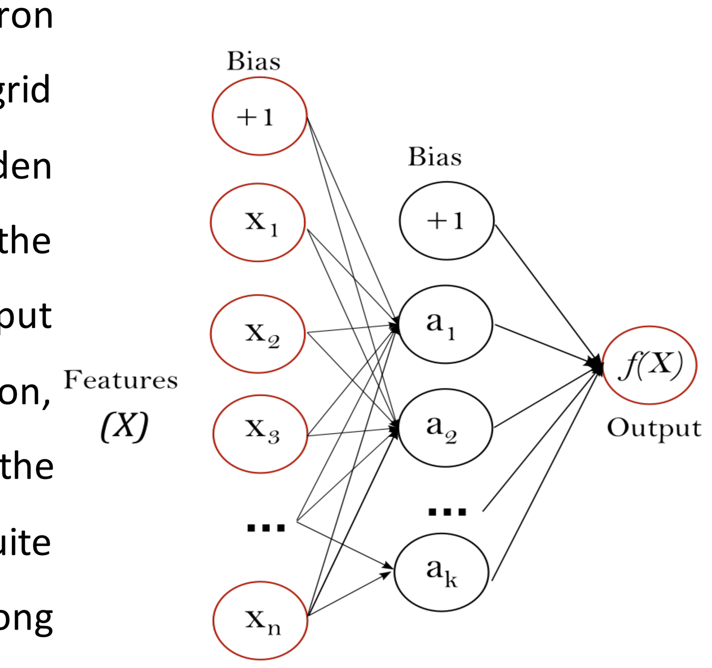
Support vector classifier was trained using radial basis function (RBF) kernel with two parameters C and γ, and balanced sample weighing. Grid search indicated that there is a tradeoff between maximizing overall correctness and minimizing false negative.

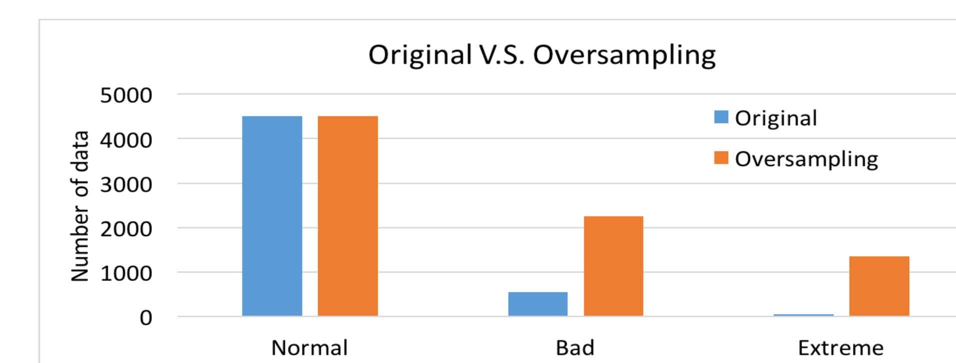


Overall correctness is normalized to all testing data, false negatives are normalized to their corresponding class frequency, and false positive is normalized against the sum of bad (3 - 7 outages) and extreme cases (8+ outages).

Neural Network

We implement a Multiple-layer Perceptron algorithm for outage number classification. By grid search, we set the network structure as 2 hidden layers, with the first hidden layer 5 nodes and the second layer 3 nodes (with a softmax output activation function). Using 10-fold cross-validation, we find though the testing 0/1 error is low, the prediction for "bad" and "extreme" case is quite poor. This is because our data have strong between-class imbalance.





We tackle the between-class imbalance problem by oversampling from two under-represented classes.

Method	Overall correctness	Right predicting normal	Right predicting bad	Right predicting extreme
Original	0.883	0.99	0.033	0
Oversampling	0.844	0.92	0.203	0.6

Conclusions

- Non-linear machine learning methods were successfully implemented to predict the number of power outages by classifying them into three categories: **normal** (0 - 2 outages), **bad** (3 - 7 outages), and **extreme** (8+ outages).
- Difficulties were encountered in predicting bad and extreme outages despite sample weighing. We hypothesize that this is because there too few cases of these, especially the extreme outages, and there are other important features that are not incorporated into the models.
- No single method is decisively better than the rest. Until a better model is implemented, Seattle City Light must make an economics-driven choice to balance the trade-off between false positives and false negatives.
- GUI allows easy access to predictions. Python packages are also available for implementation in other media.

Reference:

[1] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.
[2] Classification and Regression Trees by L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Chapman & Hall, 1984.
[3] Scikit-learn Documentation http://scikit-learn.org/stable/