Boston Fire Incident Prediction: The Weather Factor

Zijian Lin, Weixi Li

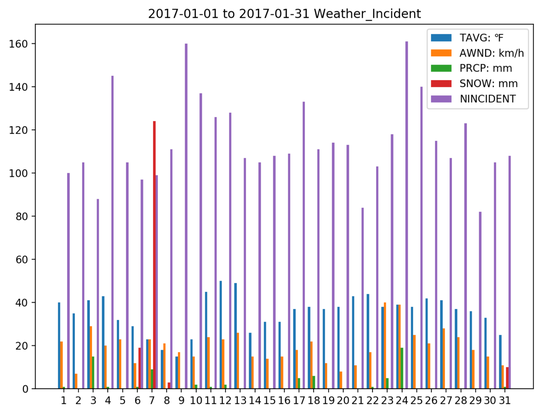
Introduction

There are about one hundred fire incident happens in Boston every day. Some of them cause little losses while some cause a lot. If we could identify the dangerous fire incident signal, then we can enable the City to better direct its preventive outreach efforts (including safety inspections, smoke and carbon monoxide detector installation, and fire safety education) to address these hazards before they turn into tragedies. In this project, we collected the data of Boston daily weather information, Boston fire facilities and Boston fire incident report. We tried several machine learning methods to training models that use the weather factor to predict the fire incident risk.

Dataset

In this project, we collected following dataset:

- Boston Fire Incident Report
- Boston Building and Property Violations
- Boston Fire Facilities (Including fire department, fire hydrants and fire alarm boxes)
- Boston Daily Weather Information (Including daily temperature, wind speed, precipitation and snow fall.



Algorithm

In our project, since we want to focus on daily prediction, we only use the daily weather information to predict the fire incident. We use four features as the input: Daily average temperature, wind speed, precipitation and snow. The predict label is the risk of the fire incident, where we label each as high/medium/low with number of fire incidents happens.

For the machine learning algorithms, we tried several classification models. The result shown below:

Model Name	Training Accuracy	Testing Accuracy
Stochastic Gradient Descent	0.497	0.455
Logistic Regression	0.508	0.440
Support Vector Machine	0.520	0.425
Ada Boosting	0.604	**0.477**
Random Forest	0.972	0.432
Gradient Boosting Decision Tree	0.790	**0.477**

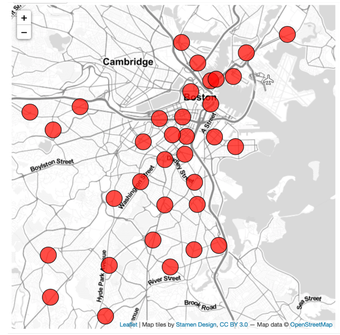
According to above result, we notice that the Support Vector Machine model and Gradient Boosting Decision Tree Model achieve best result, which is 0.477 with 3 classes (0.333 for random guess). The accuracy is 50% above random gauss. We showed at least weather information can be used to identify the fire risk.

Below is the correlation and p-value between weather features and true label.

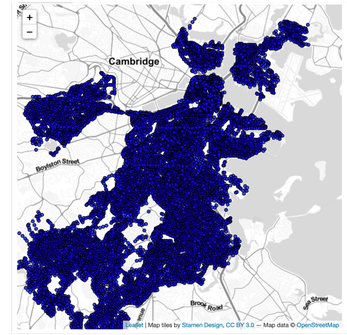
Feature	TMAX	TAVG	TMIN	AWND	PRCP	SNOW	PRED
Correlation	0.157	0.146	0.143	0.153	0.260	0.08	0.187
P-value	4.36E-5	1.47e-04	2.06e-04	6.97e-05	8.44e-12	3.77e-02	3.07e-02

Visualization

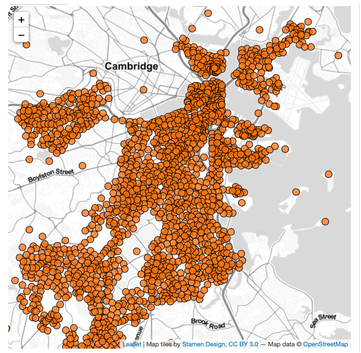
In data visualization part, we create a web service that plot every Boston facilities (fire department, fire hydrants and fire alarm boxes) on the map. We also generate the heat map that counts the number of incident in each area.



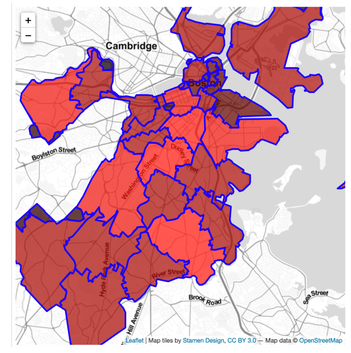
Fire Department



Fire Hydrants



Fire Alarm Boxes



Fire Count Hot Map

Conclusion and Future Work

In this project, we proved that weather factor can be used to identify the fire risks. However, there are still lots of features that could potentially be used to get better prediction result. Factors such as number of property violations in each area could also contribute to our model.