CS504 Final Report

Boston Fire Incident Prediction: The Weather Factor

Zijian Lin
Weixi Li

Abstract

Fire incidents happen every day in Boston. Severe fire incidents not only causes property loss but also threats people's life. In this project, we collected fire incidents related dataset from multiple sources and built up a machine learning model that tried to use weather factor to predict the fire risk of the day. We also implemented the visualization of fire facilities and fire hotmap in Boston area which may help people find the potential fire risks.

1. 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 did a transformation to map the three dataset into one dataset which can reflect the overall fire facilities condition of Boston. For Boston weather condition, we did a transformation to aggregate the daily weather and number of incidents to see if they have certain relations. Finally, we collected the data of building and property violations in Boston to see if such violations can largely improve the risks of fire incidents.

For the optimization problem, we tried several machine learning methods to training models that use the weather factor to predict the fire incident risk. Then we analyzed the correlation and p-value between each weather factor and the true label

Finally, we implemented a web service with Flask that can visualize the fire facilities and fire incidents of Boston area. We also visualized the histogram of daily weather and number of fire incidents in the day, which allow users to explore more relations.

2. Datasets

In this project, we collecting following dataset:

- Boston Fire Incident:
  https://data.boston.gov/dataset/fire-incident-reporting
- Boston Code Enforcement - Building and Property Violation:
  https://data.boston.gov/dataset/code-enforcement-building-and-property-violations
- Boston Fire Alarm Boxes:
  http://bostonopendata-boston.opendata.arcgis.com/datasets/fire-alarm-boxes
- Boston Fire Department:
  http://bostonopendata-boston.opendata.arcgis.com/datasets/fire-departments
- Boston Fire Hydrants:
  http://bostonopendata-boston.opendata.arcgis.com/datasets/fire-hydrants
- Boston Weather:
  https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00014739/detail

3. Models and Analysis

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. We generated the label with equal number methods that the boundary of each label class is decided so that number of samples in each class is same. Since different season may influence the training process, we just normalize all features and then shuffle the data.

For the machine learning algorithms, we tried six different classification models. The results are 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**

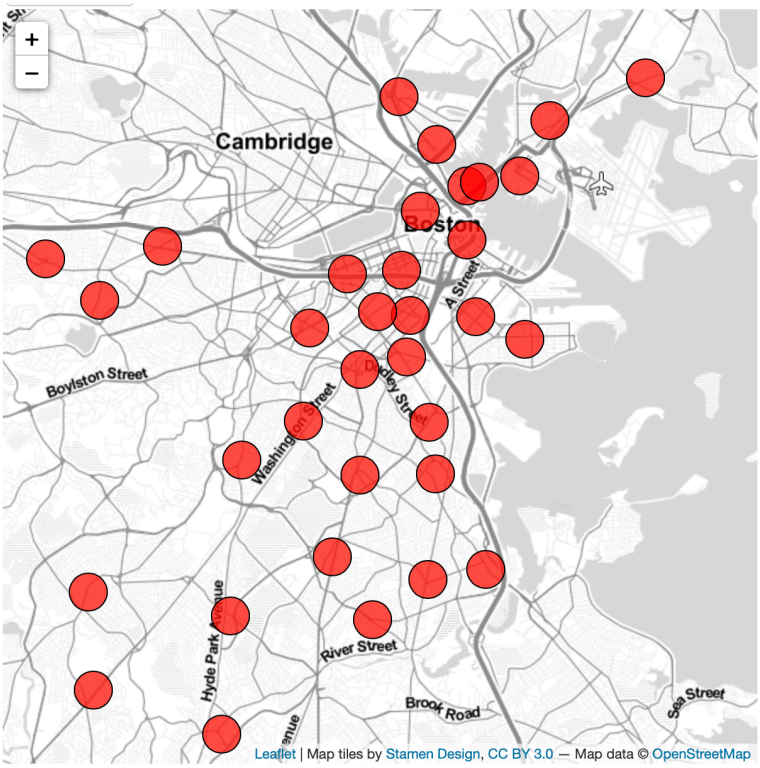
According to above result, we notice that the Support Vector Machine model and Gradient Boosting Decision Tree Model achieve best result, which is 0.477 with 3 classes (0.333 for random guess). The accuracy is 50% above random gauss. We showed at least weather information can be used to identify the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

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

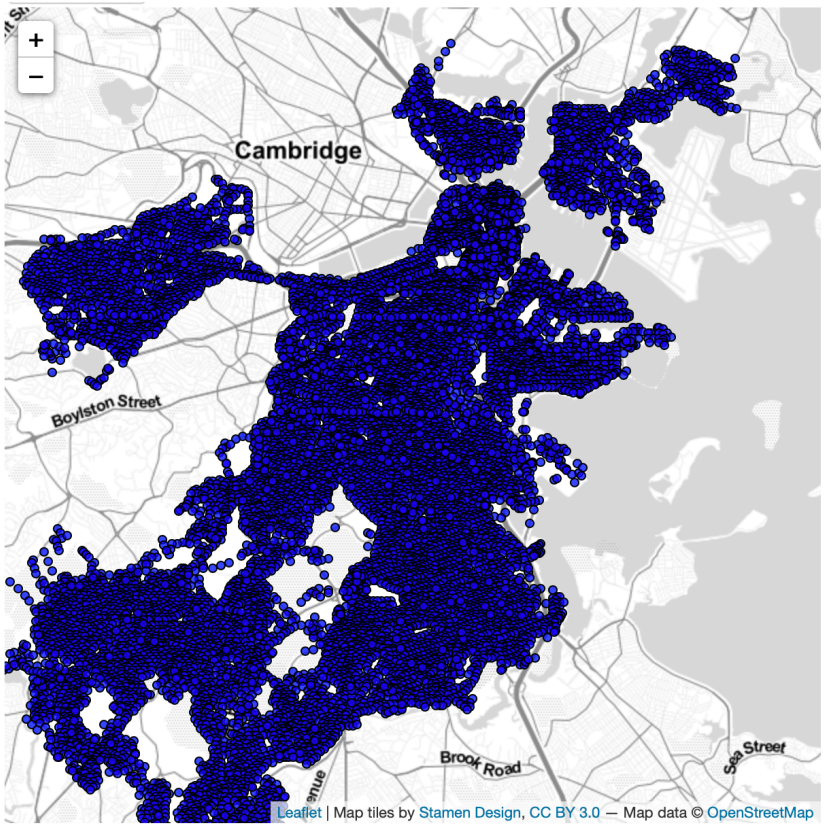
Feature	TMAX	TAVG	TMIN	AWND	PRCP	SNOW	PRED
Correlation	0.157	0.146	0.143	0.153	0.260	0.08	0.187
P-value	4.36e-5	1.47e-04	2.06e-04	6.97e-05	8.44e-12	3.77e-02	3.07e-02

4. Visualization

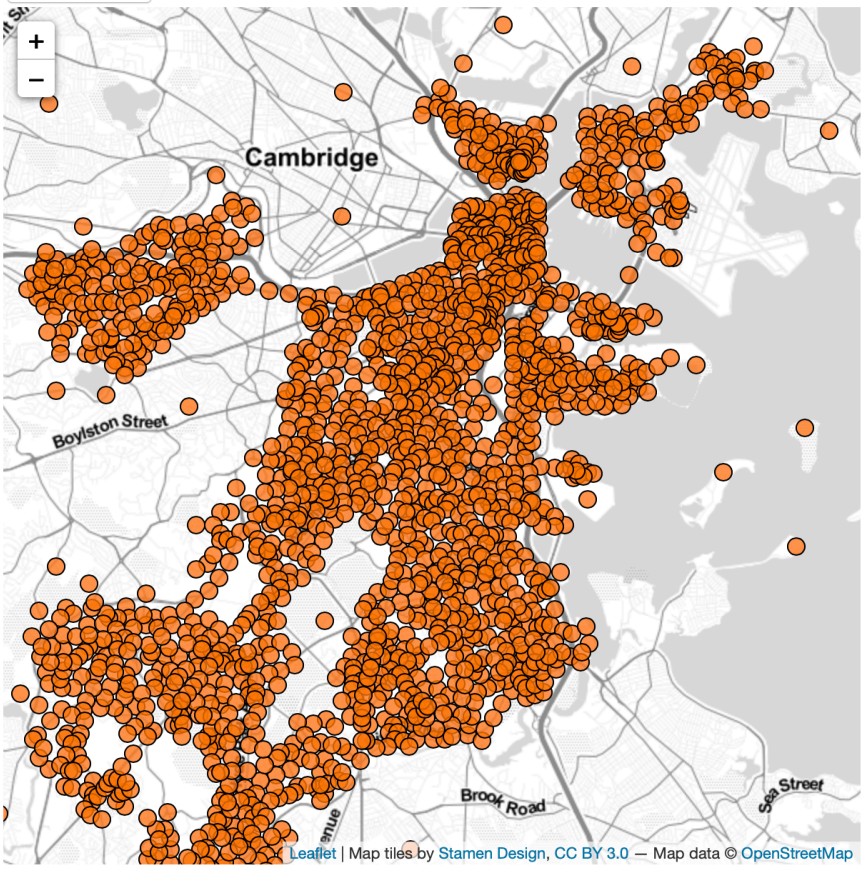
In data visualization part, we create a web service that plot every Boston facilities (fire department, fire hydrants and fire alarm boxes) on the map.



Boston Fire Department

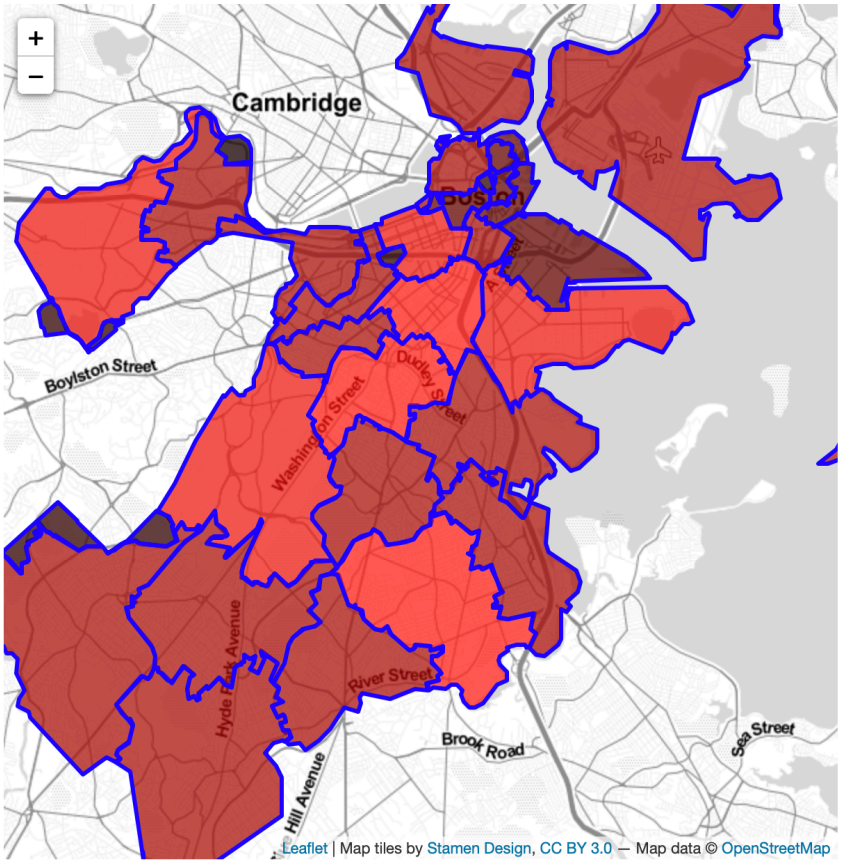


Boston Hydrants

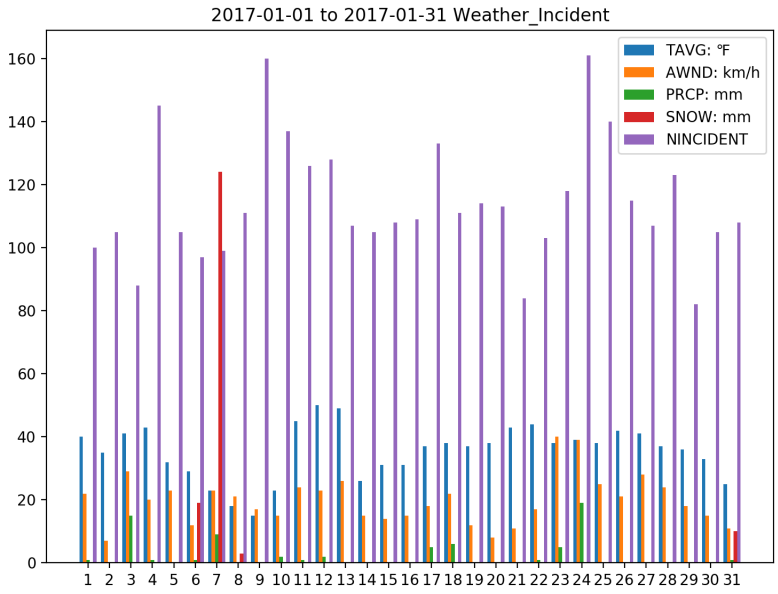


Boston Fire Alarm Boxes

We generated the heat map that counts the number of incidents in each area. (Brighter represents more fire incidents)



We also implement the daily weather and number of fire incidents histogram for two years:



5. 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. For future work, we may try to add more features into the training model and see if there is significantly improvement.