Predicting Traffic Accident Likelihoods in Seattle

Ming Huang

Why?

- Approximately 10,000 collision related traffic accidents in Seattle streets according to the 2013 Traffic Report
- Being able to estimate traffic accident likelihoods will allow for better resource allocation of 911 response teams
- Will also allow for better awareness of risky situations

Goals of this Project

- Use publically available and relevant data
- Identify features and models predictive of traffic accidents
- Create a data pipeline suitable of automation and live app support
- Build a live app reporting real time traffic accident likelihoods

Overall Project Summary

Workflow:

- Build scrapers, API wrappers, and data processing modules to collect data.
- Feature engineer through the use of python and sql.
- Evaluate multiple models select the most suitable.
- Build data pipeline models capable of automation and flexible model selection.
- Build a live app to report current conditions using the above model.

Data Sources:

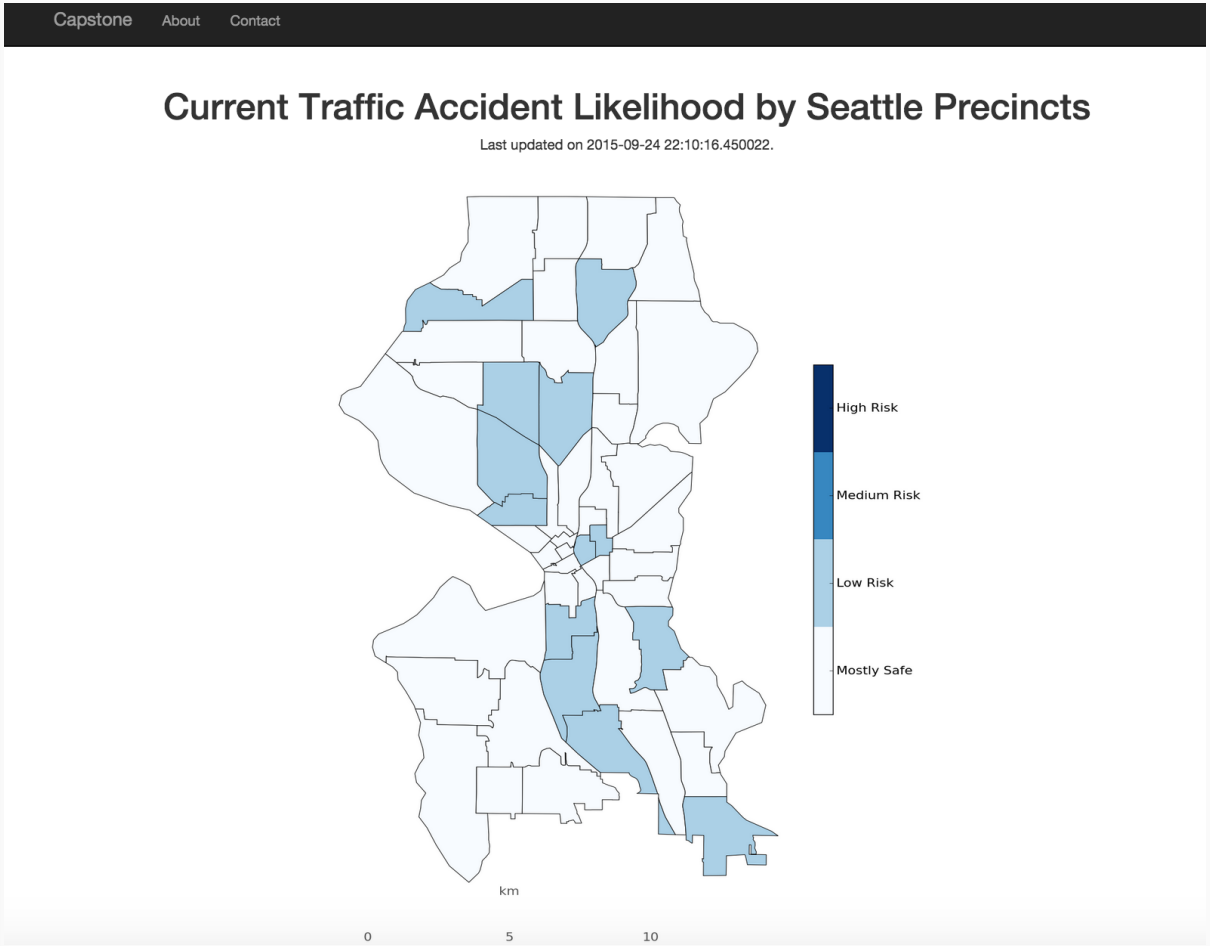
- Socrata 911 Incident Response records
- Wunderground historical weather data
- Wunderground real time weather API
- Mariners game schedule
- Seahawks game schedule

CURRENT Findings

- There are a lot of randomness not accountable by the current set of features and data.
- There is a huge imbalance of classifiers, making prediction difficult.
- Most models performed poorly, but random forest seems to be consistently better.
- Need more time and feature experimentation.

My App

- Choropleth of Seattle precincts shaded by the rate of accident risks.
- Updates automatically every 60 minutes by using the most recent Wunderground weather report..



Next Steps

- Add and improve on existing features
- Implement non-conventional models
- Improve App functionality and design

Thanks!

Ming Huang:

mnghuang@gmail.com