Introduction

- Machine learning competition hosted by Kaggle.com
- 12 years of San Francisco arrests data given
- Goal is to predict the category of crime, given time and location
- Assist in building a safer society
- Work in a highly competitive leaderboard to evolve our skills
- Gain valuable experience with Data Mining and Machine Learning techniques

Approach

- Utilize multiple Machine Learning libraries for more classifiers.
- Create a modular design for easy to maintain code
- Use an internal debugging module to improve our logloss score
- Derive additional attributes from the dataset such as hotspots and holidays

System Architecture

- Predict and Categorize Crime based off of past crimes
- Combines multiple algorithms from multiple languages to make the best predictions.
- Uses hotspots visualization to analyze crimes

Features

- Kaggle Scores
  - Initial Submission: 13.61204
  - Prediction Offset: 2.83145
  - Random Forest Tree Classifier

Figure 1: Submission Timeline

Figure 2: Hotspots Overlay

Figure 3: System Design