Data Mining: Project 2
Stop-Question-and-Frisk

Tasks: Data Preparation, Classification

Assigned: September 14th,2017

Due: October 5th,2017

Points: 90

For this project we will look again at stop-and-frisk data for the year 2016. The cleaned dataset obtained from project 1 will be used as an input dataset for this project.

The test set: The data objects located at rows {10, 20, 30, …, 12,400} (1,240 objects)

The training set: The remaining data objects (11,164 objects)

Write a report covering in detail all steps of the project. The results (rule set, decision tree, probabilities, confusion matrix, etc.) have to be reproducible using your report. Carefully describe every assumption and every step in your report. Here are classification tasks you have to predict:

1. Predict if a person is armed?
2. Predict if an arrest is made?
3. Predict what physical force used by officer?

Follow the CRISP-DM framework

1. Data Preparation [30 points]

> 1.1 Define and prepare your class attributes for each classification tasks if needed. You may have to combine different columns. [20]
>
> 1.2 Remove variables that are not needed/useful for the analysis. [5]
>
> 1.3 Describe the final dataset that is used for classification (include the scale/range for the new combined variables) [5]

2. Modeling [50 points]

    2.1 Create at least 3 different classification models (different techniques) for each of the classification tasks. [30]

    2.2 Implement models obtained from 2.1 on the test set for each classification task. [15]

    2.3 Compare classification results obtained from 2.2 for each classification task. [5]

3. Evaluation and Deployment [10 points]

    3.1 How useful is your model for the police? [5]

    3.2 How would your implement your model to improve policing? What other data should be collected? [5]