|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Target** | **Features** | **Model** | **Score** | **Findings** |
| Crime Code | Time Occurred, Area ID, Reporting District, VictimAge, Premise Code, weather features | RandomForestRegressor | 83.3% | Crime codes are represented as numbers, but they’re really just classifications. This score might not be correct because all of the classifiers should be one-hot encoded. |
| Crime Code | Month, day of week | Linear Regression |  | Weight coefficients: [[-0.11888112]]  y-axis intercept: [612.77272727] |
| Crime Code | Year, month, day, dayofweek, quarter, DRNumber, Time Occurred, AreaID, Reporting Disrict, CrimeCode, MoCodes, VictimAge, Premise Code, WeaponUsedCode | RandomForestClassifier | 98.92% | Use RandomForest to classify which column is more influence to y. The highest importance number is the most relative effector. Use classifier in each individual crime code. If combine together, it will lower the score. |
| Crime Code | Premise Code | Grid Search | 70% | It is the only successful model for this specific crime code 510. I tried for top 10 crime code and they all failed. The score is below 30% |
| Crime Code | Premise | *k*NN | 70% | The only Premise code has the score is 101-street, which is the predict location happened for vehicle stolen. The other crime type also failed, the score is below 30% |
| Victim Age | Population, Crime Classification, Weather | DecisionTreeClassifier | 16.5% | This was a bust. |
| Crime Classification | Time Occurred, AreaID, Reporting District, Crime Code, Victim Age, Premise Code, Weather features | LogisticRegression | 43% | Not a very good result. Possibly too many features, or the features are just not predictive. |
| Crime Classification | Month, Time of Day, Premise, Weather Type, Location | RandomForestClassifier | 51% | Not a very good result. The weather is not predictive, nor is the location in the city. |
|  |  | LogisticRegression |  |  |
| VictimAge | AreaName, CrimeCodeDescription | Linear Regression | MSE – 1.01  R2 - .0056 | Data was binary encoded for AreaName and Crime Code Description. The MSE score needs to be as close to 0 and the R2 score needs to be close to 1. Thus this model is not good for the Crime Data set. |
| VictimAge | AreaName, CrimeCodeDescription | LASSO Model | MSE – 1.012  R2 - .00539 | Data was binary encoded for AreaName and Crime Code Description. The MSE score needs to be as close to 0 and the R2 score needs to be close to 1. Thus this model is not good for the Crime Data set. |
| VictimAge | AreaName, CrimeCodeDescription | Ridge Model | MSE – 1.0001  R2 - .005597 | Data was binary encoded for AreaName and Crime Code Description. The MSE score needs to be as close to 0 and the R2 score needs to be close to 1. Thus this model is not good for the Crime Data set. |
| VictimAge | AreaName, CrimeCodeDescription | ElasticNet | MSE – 1.0010  R2 - .005536 | Data was binary encoded for AreaName and Crime Code Description. The MSE score needs to be as close to 0 and the R2 score needs to be close to 1. Thus this model is not good for the Crime Data set. |
| VictimAge | AreaName, CrimeCodeDescription | Random Forest Regressor | Average baseline error: 30.95  Mean Absolute Error:  11.48 degrees  Accuracy: 63.13% | This model seems to the best to use for analyzing model. Data is not one hot encoded. It’s possible the CrimeCode, AreaName and VictimAge was treated as integer instead of a string. |