May 5th, 2024

Team VariUNTce

Young Yu, Tao Xiong, Sonali Sabnam, Awais Khan, Sonam Pohuja

EXPANDING BASTION SECURITY SYSTEMS LLC’S FOOTPRINT IN THE DFW AREA:

A DATA-DRIVEN APPROACH TO A GOVERNMENT CONTRACT ACQUISITION

**Problem Question**

Bastion Security Systems LLC, a leader in advanced security solutions, aims to expand its market presence in the Dallas-Fort Worth (DFW) area by securing a strategic government contract. The primary challenge is to substantiate the need for Bastion's cutting-edge technologies to enhance local security measures, ensuring a competitive edge in the bidding process.

**Data Processing**

Our team, VariUNTce Inc., has been tasked with analyzing over 1.2 million crime incident reports from the Dallas Police Department spanning from 2014 to the present. The dataset includes sensitive information, with certain limitations due to privacy concerns and ongoing investigations, impacting the project's results. Acknowledging these challenges is crucial as they influence the reliability and predictive power of our analysis.

**Research & Limitations**

The data provided by the Dallas Police Department comes with the caveat that crime classifications may change following further investigations, highlighting the preliminary nature of the dataset. This affects the accuracy, volume, and completeness of the data, which are essential for robust predictive modeling.

**Data Cleaning & Preparation**

Data cleaning was an iterative process requiring substantial effort to evaluate the relevance of each of the 86 features in the dataset. We employed various strategies to handle large datasets effectively, such as breaking down the data by year and using tools like Tableau Data Prep for local processing. This approach helped in managing the project efficiently, allowing parallel tasks while the final dataset was being prepared.

**Exploratory Data Analysis (EDA)**

We focused on analyzing variables like crime severity, division, and days of the week with the highest crime rates using visual tools like heat maps, histograms, and pair plots. Our EDA revealed that motor vehicle theft was the most recorded crime, with the Northwest Division of Dallas showing the highest crime rates. Seasonal trends indicated higher crime rates from April to October, with a peak in July. The challenges we faced during EDA were primarily due to the lack of numeric data and the need to synthesize "Severity" and “is\_high\_risk” columns to group high-risk crimes.

**Implications of EDA Results**

The Northwest division, identified as having the highest crime rates and severity, underscores the need for enhanced security solutions that Bastion Security Systems can provide. This area’s vulnerability to high-severity crimes positions Bastion ideally for addressing these challenges through its advanced security technologies.

**Data Modeling**

Our data modeling involved using KNN, Logistic Regression, and Random Forest. These models are well-suited for classification problems and provide a good balance between interpretability and predictive power. We used techniques like cross-validation and hyperparameter optimization to enhance model performance and prevent overfitting. The models achieved an accuracy of approximately 94%, with Logistic Regression performing slightly better, making it the preferred model for identifying high-risk areas.

**Predictions**

We have implemented the below models and calculated the output for our models.

* kNN:

A screenshot of a graph

Description automatically generated

* Logistic Regression

A screenshot of a computer

Description automatically generated

* Random Forest

A screenshot of a computer

Description automatically generated

Based on the above models we have the below results consolidated.

|  |  |
| --- | --- |
| **Model** | **Accuracy** |
| kNN | 0.939831 |
| Logistic Regression | 0.940669 |
| Random Forest | 0.939638 |

Our models predict that high-severity crimes are likely to occur in the Northwest Division of Dallas, emphasizing the critical need for deploying Bastion’s security systems in this area to mitigate such risks.

This executive summary outlines the rigorous data processing, EDA, and modeling efforts undertaken by VariUNTce Inc. to support Bastion Security Systems LLC’s bid for a government contract in the general Dallas area. Our findings demonstrate a clear need for enhanced security measures that Bastion can provide, particularly in high-risk areas identified through our analysis.

**Reference:**

*Police incidents | Dallas OpenData*. (2024, April 27). <https://www.dallasopendata.com/Public-Safety/Police-Incidents/qv6i-rri7/about_data>