⑂ main ▾   Go to file

randellmwania Uploaded after analysis and modelling   ...   13 minutes ago   ⟳ 1

View code

☰ README.md

Analysis of Terry Stops in Seattle

Prepared and Presented by: Randell Mwania

Business Problem

In the United States, a Terry stop allows police to briefly detain a person based on reasonable suspicion of criminal activity. This project explores the demographic variables influencing the arrest outcome of Terry stops. The goal is inference, as predicting with the model could introduce human bias.

Data

This dataset records police-reported stops under Terry v. Ohio (1968), provided by Seattle, WA. With 58,167 rows and 23 variables, the classification target is 'Arrest Flag' (N - 42,585, Y - 2,732).

Modeling Process

Data cleaning addressed placeholder and missing values, preserving data integrity.

Model Performance

The Decision Tree, KNN, and Logistic Regression models perform remarkably well in predicting the outcomes of Terry stops, showcasing their potential for aiding law enforcement decision-making processes. Further exploration, including fine-tuning hyperparameters and evaluating additional metrics, can contribute to a comprehensive understanding of their effectiveness in real-world scenarios.

1. **Decision Tree:**

   - **Accuracy Score: 85.85%**
   - The Decision Tree model exhibits the highest accuracy among the three models. This suggests that, based on the provided features, the Decision Tree algorithm is effective in accurately classifying whether a Terry stop results in an arrest or not.

2. **K-Nearest Neighbors (KNN):**

   - **Accuracy Score: 85.74%**
   - The KNN model closely trails the Decision Tree in accuracy, showcasing its competitive performance. KNN relies on the similarity of instances, and its ability to achieve a high accuracy score suggests its suitability for this classification task.

3. **Logistic Regression:**

   - **Accuracy Score: 85.74%**
   - Similar to KNN, Logistic Regression also achieves an accuracy score of 85.74%. Logistic Regression is commonly used for binary classification tasks, and its performance in this context aligns with that of the KNN model.

Key Observations:

- All three models demonstrate strong predictive capabilities, surpassing an 85% accuracy threshold. This implies that the chosen features and model configurations effectively capture patterns in the data related to Terry stops and arrest outcomes.

Considerations for Further Analysis:

- While accuracy is a valuable metric, further analysis could involve examining other metrics such as precision, recall, and F1-score to gain insights into the models' performance across different aspects of classification.

Conclusions

- Most arrests are made by male officers, reflecting the overall gender distribution among officers.
- Most arrests show no involvement of weapons.
- Notably, individuals aged 26-35 are more frequently involved in arrests.
- This analysis uncovers that, on average, less than 10% of stops result in immediate arrests.

Recommendations:

1. Conduct training programs for law enforcement officers to enhance their judgment on deciding when it is suitable to make an arrest during a Terry stop. Providing clear guidelines on differentiating situations requiring immediate action from those that can be addressed later can contribute significantly to reducing unnecessary arrests.

2. Emphasize the importance of recording the officer's precinct in all Terry stops. This additional data point can enhance the predictive capabilities of the model, allowing for a more nuanced analysis of factors contributing to potential arrests.

3. Implement training modules for officers to recognize optimal situations for conducting a 'frisk' during Terry stops. Understanding the appropriate circumstances for frisking individuals can serve as a crucial indicator in predicting arrests accurately.

Future Work

1. Further analyze unknown or missing values.
2. Update 'Arrest Flag' with arrest values from 'Stop Resolution'.
3. Experiment with no SMOTE.
4. Tune Support Vector Classification.

Releases

No releases published

Packages

No packages published

Languages

- **Jupyter Notebook** 100.0%