

Solvability of murder

Kyle Kingston

DSC630

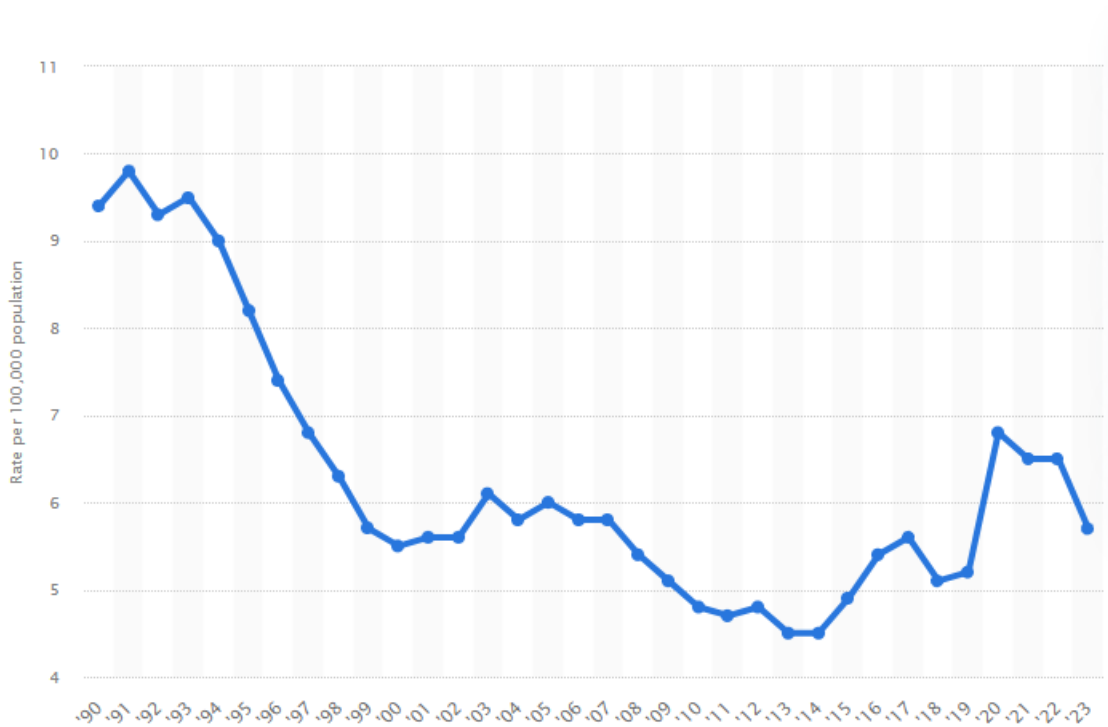
Professor Andrew Hua

Murder stats and catch prediction

Introduction

Crime in the us has been a hot button issue at both a national, and local level.

Overall murder rates appear to have been dropping until recent years when we have seen a spike.



I would like to see if we can accurately predict what crime is being committed based on various aspects of the victim. I am wanting to find out if a model can accurately predict whether a murder will be solved or unsolved based on the victim and circumstance information.

Data Selection

This data set was gathered by the Us Government, Crime department in collaboration with local law enforcement agencies. This data was gathered and uploaded to Kaggle.com. The dataset consists of fields that describe the law enforcement agency who filed the report, location data for the incident, information about the incident, information about the perpetrator and information about the victim(s).

Model Selection

When selecting a model, I first evaluated what kind of target variable and what problem we are looking at. This is a binary classification problem and due to that consideration, I narrowed down to three models: logical regression, k nearest neighbors, and random forest. Using hyperparameter tuning we will let the grid search model decide which of these models fits the data the best and gives the best result. After running this model, it shows that the best model is a random forest classifier.

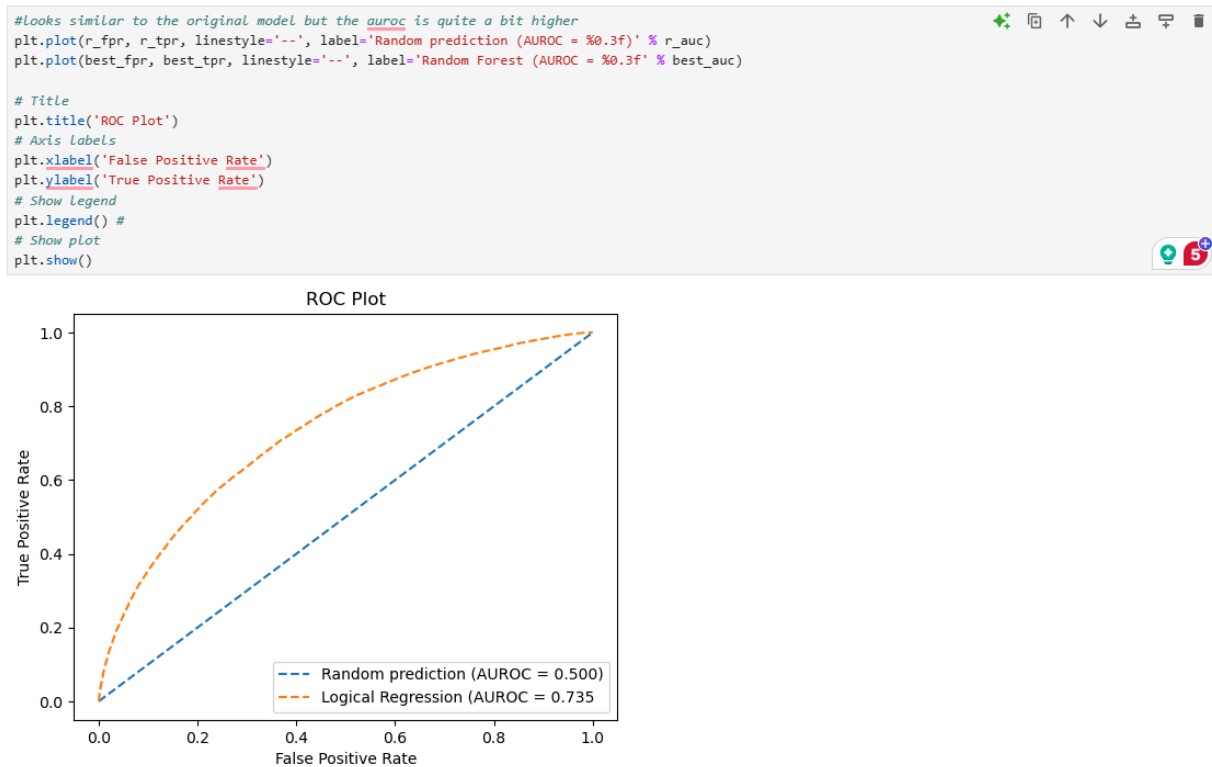
```
: #show the best model based on the grid search
grid.best_params_

: {'classifier': RandomForestClassifier(random_state=42),
  'classifier__max_depth': None,
  'classifier__max_features': 'sqrt',
  'classifier__min_samples_split': 5,
  'classifier__n_estimators': 200}

best params {'classifier': RandomForestClassifier(random_state=42), 'classifier__max_depth': 20,
'classifier__max_features': 'sqrt', 'classifier__min_samples_split': 5, 'classifier__n_estimators': 200}
```

Interpretations of results

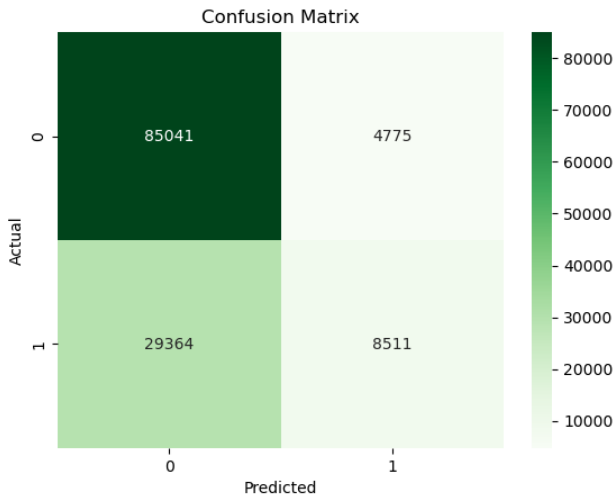
This model appears to be functioning at about 66 percent accurate, I feel like that is a decent result. After the previous submission, I fixed the imbalance in the target variable and that changed how the model preformed. The hyperparameter trained model performed a bit better than the original model showing that random forest may have been a better suited for this dataset then the logistical regression model or knn.



```
#create and plot the confusion matrix

cm = confusion_matrix(y_test, y_pred, labels=['Yes', 'No'])

# Visualize the confusion matrix using a heatmap
sns.heatmap(cm, annot=True, fmt="d", cmap="Greens")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix")
plt.show()
```



```
#the precision shows 74 percent accurate
#the recall shows we caught more positives than negatives, this lends to our false negatives being high in our confusion matrix
#the f1 shows the model is much better at identifying positives than negatives
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
No	0.64	0.22	0.33	37875
Yes	0.74	0.95	0.83	89816
accuracy			0.73	127691
macro avg	0.69	0.59	0.58	127691
weighted avg	0.71	0.73	0.68	127691

Conclusion

Murder has and will continue to be an issue in this country and around the world. I would like to see models like this to help determine if a murder is solvable based on the circumstances of the murder and hopefully isolate why those features seem to aid in the solving of those particular incidents. If we can isolate the causes that lend to the solving of

these cases hopefully we can leave less crime unsolved and result in less re-occurrence of events.

References

Agrawal, A. (2023, October 2). *US crime dataset*. Kaggle.

<https://www.kaggle.com/datasets/mrayushagrawal/us-crime-dataset>

Jatmiko, N. A. (2024, November 20). *Depression professional dataset*. Kaggle.

<https://www.kaggle.com/datasets/ikynahidwin/depression-professional-dataset>

Korhonen, V. (2024, November 12). *USA: Reported murder and Nonnegligent Manslaughter*

Rate 2023. Statista. <https://www.statista.com/statistics/191223/reported-murder>

[-and-nonnegligent-manslaughter-rate-in-the-us-since-1990/](https://www.statista.com/statistics/191223/reported-murder-and-nonnegligent-manslaughter-rate-in-the-us-since-1990/)