Catching Hackers in Call of Duty: Modern Warfare

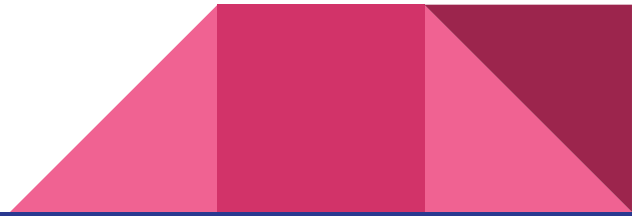
By Jenny Reschke
Data Analyst

Summary

Goal: Determine if a player has hacked the game by building a predictive model to classify outliers in the players' gaming statistics.

Steps:

1. Data Wrangling
2. Exploratory Data Analysis
3. Feature Selection
4. Model Selection
5. Conclusions and Future Research



Problem Statement

Call of Duty: Modern Warfare

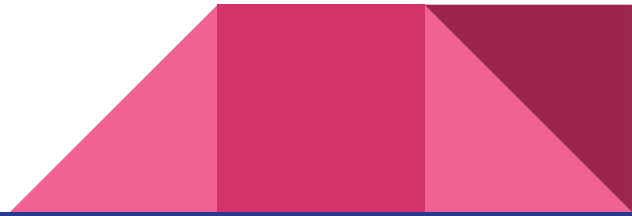
- Over $1 Billion in sales in the first 50 days
- Continuously updates with weapons and game-play
- Has a problem with players hacking the game to improve their weapons and abilities.

Problem Statement:

Identify the outliers in the player's data set and build a classification model to prediction outliers. These outliers may determine if the player is a hacker.

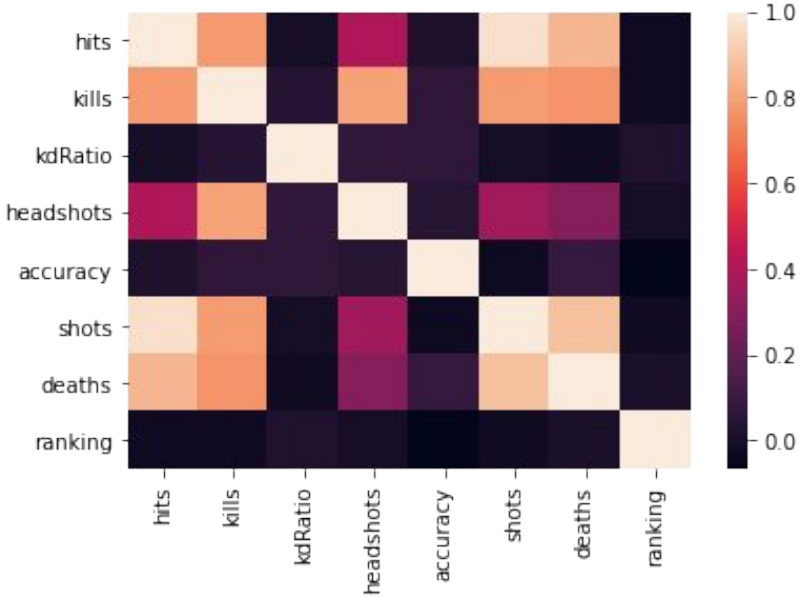
Data Wrangling

- Scrap player’s game tags from a rating website
- Ran game tags though a CoD API
- Cleaned data set to the required statistics
- Final data set contained 3,676 rows and 11 columns



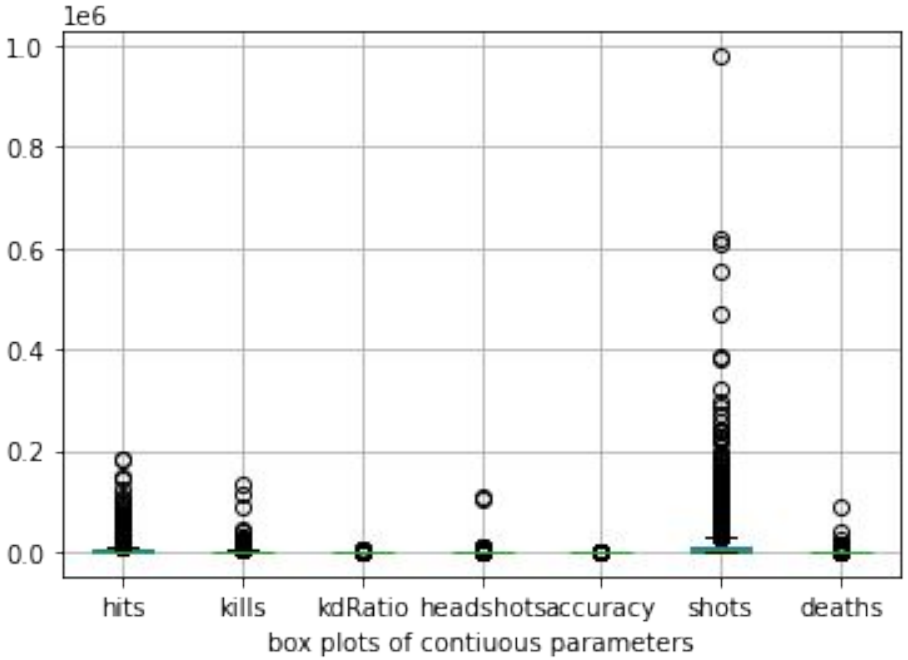
Exploratory Data Analysis

Ranking, KD Ratio and Accuracy are the three statistics that measure the skills and do not have significant correlation to another parameter.



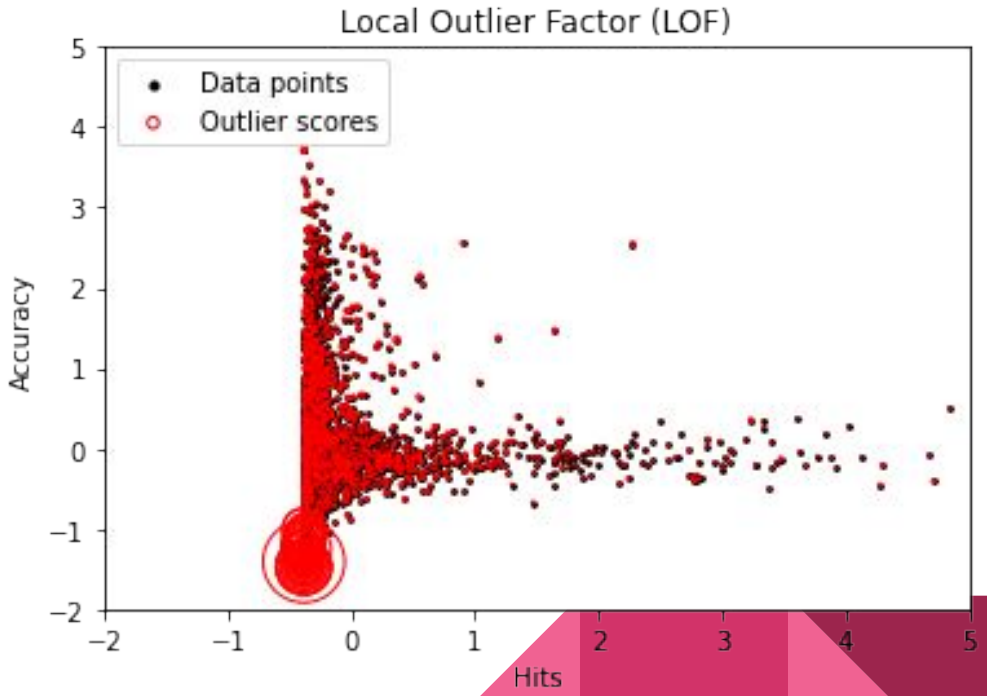
Exploratory Data Analysis

- Looking at the bar plot, there are many outliers in the data set.
- A statistical method will be required to determine points are outliers that could be caused by hacking



Feature Selection and Outlier Detection

- All continuous variables are normalized from 0 to 1 The discrete variables are set from either 0 or 1.
- The data sets are divided into training and testing.
- To determine outliers, LOF is performed where the bigger red circle's radius means an increased chance of being an outlier.

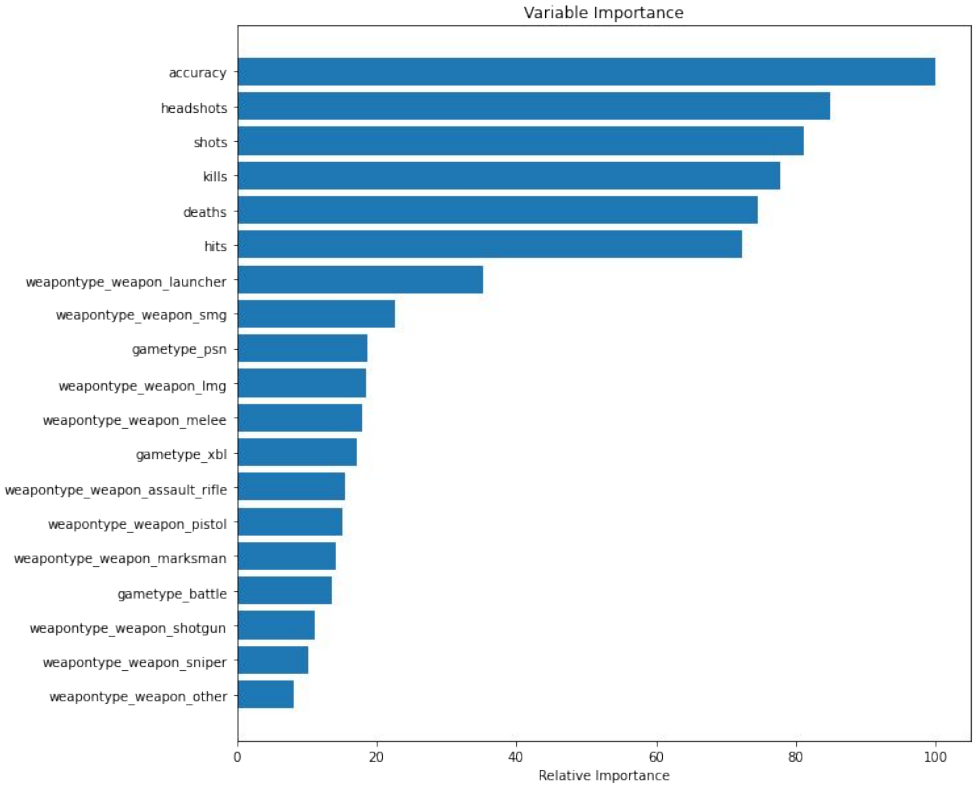


Model Selection

	KMeans = 3	KMeans + Bagging Best Option	Random Forest
Accuracy	0.75741	0.78704	0.70032
Balanced Accuracy	0.47983	0.50433	0.49329
Precision score for non outlier points	0.79401	0.79069	0.78663
Precision score for outlier points		0.37500	0.19548
Recall score for not outlier points	0.95966	0.99304	0.85118
Recall score for outlier points		0.01562	0.13541

Precision is the most important as we are more concerned about false positive or labelling someone a hacker that isn't than false negatives.

Results



The Random Forest model's results include ranking the parameters from most to least important in the model's understanding of the relationship.

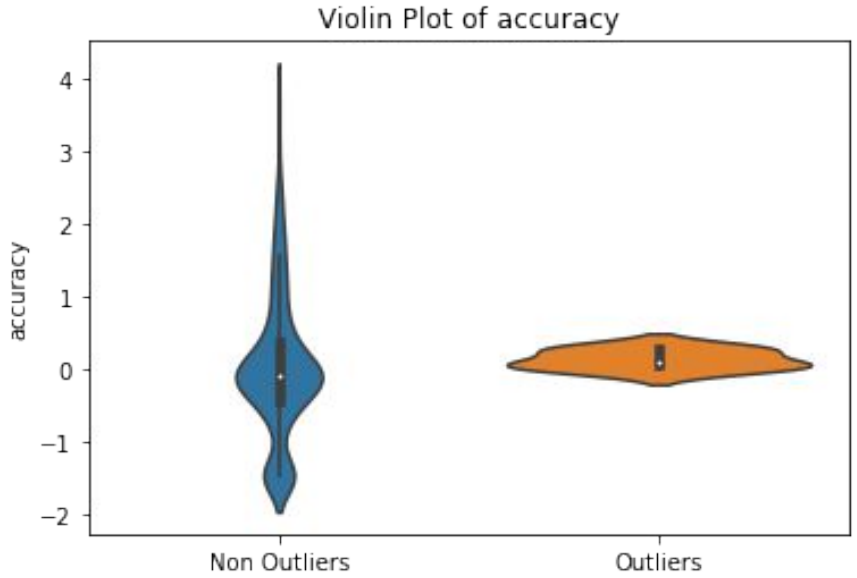
Accuracy with headshots close behind (an accurate shot) is the most important.

This makes sense as a player making every shot could be a sign of hacking.

Results - Accuracy's Impact

The plot shows the distribution of the accuracy parameter for the kmeans+bagging model.

The outliers are focused around 0 whereas the non outliers fit the whole distribution.



Conclusion and Future Research

- This project showed the initial analysis of outliers in the CoD Modern Warfare video game. The results showed a relationship between the accuracy metric and the outliers in the data set.
- Improvement for future research include:
  - Explore other input parameter that may better capture skill vs.hacking
  - Develop a data set with known hackers so ensure the LOF method works for determining outliers
  - Explore additional methods of unsupervised learning