Dheric Seney

Professor Mehmet Ergezer

COMP-3800

Data Science Final Project

UFC 236 Analysis

Overview:

Comp-3800 Data Science taught me how to efficiently use python and its libraries such as NumPy, pandas, matplot, and many more. It allowed me to represent data using arrays, organize data, munge missing data, and work with time-series data. Not only that, I was able to learn how to visualize data using different types of graphs, learn the principles of machine learning, and become aware of ethical consequences of automated decisions-making. With the information I got from the semester, the final project of the course is to provide a data-driven solution to a problem that excites me in using the tools discussed.

Problem:

In UFC, Data Analytics is not used or utilized as much as it should be. What it can be used for is to help an MMA fighter learn about their opponent. They could learn what their opponent mostly does and does not do. With this information, the fighter can train their skills to go up against their opponent rather.

Selection of Data:

For the project I wanted to get Data for the fighters who are fighting in the Main Card of UFC 236. The fighters are, Max Holloway, Dustin Poirier Kelvin Gastelum, Israel Adesanya, Khalil Rountree Jr., Eryk Anders, Dwight Grant, Alan Jouban, Nikita Krylov, and Ovince Saint Preux. The data retrieved for this project was found on ufcstats.com. The data consists of stats of rounds 1 to 5. If the match did not go to 5 rounds either because the match ended before round 5 or the match limit was 3 rounds, those stats were given 0’s. The stats that were taken down included Significant Strikes that were towards the head, body, legs, distance, clinch, or on the ground. The data that was taken for the stats were manually recorded into an excel file that was later saved as a .csv file. There was data from Kaggle that could have been used; however, the data did not consist of all the fighters that were fighting on the main card, and it did not contain stats that were recent.

Methods & Tools Discussion:

I used a few libraries for my Data Science Project. I used NumPy, Pandas, and Matplotlib. NumPy was mainly used for the methods calculations. In the project I have two methods that were mainly used. The first one is getRoundStats(). The method is used to get the specific stats for one fighter per round. Within the dataset each person has more than one fight of stats. Therefore, the method gets all the fights for that one fighter, and with NumPy, I take the mean of all the stats and return it. The next method that uses NumPy is the category() method. This method uses the data retrieved from getRoundStats() and gets the specific significant strike stat per category. The different significant strike categories of stats you can get includes normal, distance, clinch, and ground. Within this category, it sums up all the columns pertaining to that category. For example, there are multiple normal significant strike columns like head, body, and leg. Using NumPy, I add those three columns together to get the total landed normal strikes.

Pandas was used to help create a dataframe for all the data so it would be easier to visualize it. I also used Pandas for slicing the data so I can return certain values of the dataframe. Lastly, I used Matplotlib. I used Matplotlib.pyplot to create visuals for the data I took from the file. I used Matplotlib.pyplot to create pie charts to show the percentages of each category stat (normal, distance, clinch, and ground). By visualizing the stats using a piechart, fighters can see what their opponent would most likely do per round. With the information, the fighter can train based on knowing what their opponent does and get the upper hand against them.

Methods & Tools Questionnaire:

Overall, the tools I used helped me with analyzing the fighter’s stats. They were able to help calculate the means and sums of the stats. They also helped me slice the data and separate them from the rest of the data. Lastly, they helped me visualize the data in a pie chart representation to help compare stats per round.

Summary:

After using the methods I used to predict what a fighter would most likely do, I used the results of UFC 236 to find whether my findings were accurate. Based on the Totals data, I predicted that Max Holoway would throw 50% Normal Significant Strikes, 3.0% ground, 42.7% distance, and 4.3% clinch. Based on the actual results, Max Holloway threw 50% Normal Significant strikes, 0% ground, 46.5% distance, and 3.5% clinch. Comparing the total and the results, my findings were close to the results with a few minor differences between the stats. Below is a chart comparing Totals and Results with every fighter who fought in UFC 236.

|  |  |
| --- | --- |
| **Totals** | **Results (UFC 236)** |
| **Max Holloway** | |
| Normal - 50.0% | Normal – 50.0% |
| Ground – 3.0% | Ground – 0.0% |
| Distance – 42.7% | Distance – 46.5% |
| Clinch 4.3% | Clinch – 3.5% |
| **Dustin Poirier** | |
| Normal – 50.0% | Normal – 50.0% |
| Ground - 3.9% | Ground – 0.0% |
| Distance – 41.0% | Distance – 45.8% |
| Clinch – 5.1% | Clinch – 4.2% |
| **Kelvin Gastelum** | |
| Normal – 50.0% | Normal – 50.0% |
| Ground – 8.6% | Ground – 0.0% |
| Distance – 36.7% | Distance – 50.0% |
| Clinch - 4.7% | Clinch – 0.0% |
| **Israel Adesanya** | |
| Normal - 49.5% | Normal – 50.0% |
| Ground – 0.3% | Ground – 0.9% |
| Distance – 45.2% | Distance – 49.1% |
| Clinch – 5.1% | Clinch – 0.0% |
| **Eryk Anders** | |
| Normal - 49.9% | Normal – 50.0% |
| Ground – 9.5% | Ground – 0.0% |
| Distance - 28.7% | Distance – 50.0% |
| Clinch - 11.8% | Clinch – 0.0% |
| **Khalil Rountree Jr.** | |
| Normal - 50.0% | Normal – 50.0% |
| Ground - 4.9% | Ground – 4.1% |
| Distance - 39.4% | Distance – 45.9% |
| Clinch – 5.7% | Clinch – 0.0% |
| **Alan Jouban** | |
| Normal - 50.0% | Normal – 50.0% |
| Ground - 5.5% | Ground – 8.3% |
| Distance – 40.8% | Distance – 39.6% |
| Clinch 3.7% | Clinch – 2.1% |
| **Dwight Grant** | |
| Normal – 50.0% | Normal – 50.0% |
| Ground – 7.8% | Ground – 0.0% |
| Distance – 34.4% | Distance – 48.7% |
| Clinch – 7.8% | Clinch – 1.3% |
| **Ovince Saint Preux** | |
| Normal - 50.0% | Normal – 50.0% |
| Ground - 2.0% | Ground – 2.5% |
| Distance - 43.3% | Distance – 25.0% |
| Clinch – 4.7% | Clinch – 12.5% |
| **Nikita Krylov** | |
| Normal - 50.0% | Normal – 50.0% |
| Ground – 7.9% | Ground – 7.4% |
| Distance – 26.1% | Distance – 31.5% |
| Clinch - 16.0% | Clinch – 11.1% |