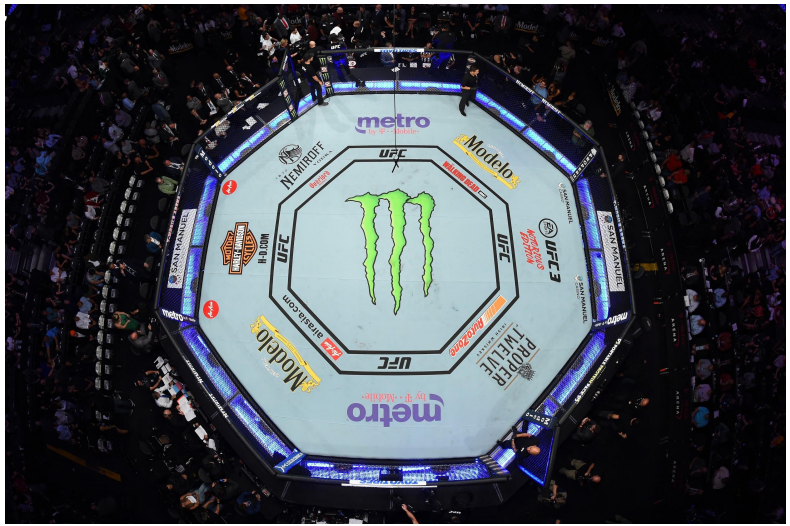
Fighter Analysis

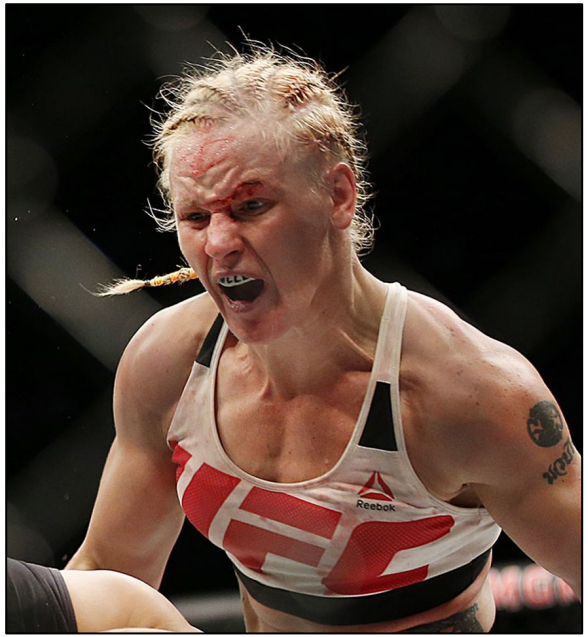
Predicting Match Winners

Overview of Project

Our group members are UFC and MMA fans and would like to use the skills we learned in the course to examine fighting techniques to determine which have the most effect the win in a match.



Mixed martial arts (MMA) is a full-contact combat sport based on striking, grappling and ground fighting, incorporating techniques from various combat sports and martial arts from around the world.



Ultimate Fighting Championship (UFC) is a Las Vegas based promotion company that has revolutionized the fighting business since 1993. UFC features some of the highest-level fighters in the sport on its roster and produces events worldwide that showcase twelve weight divisions (eight men's divisions and four women's divisions). As of 2020, the UFC has held over 500 events and grown into a globally popular multi-billion-dollar enterprise.

Purpose of Analysis

Using a Kaggle dataset containing various attributes of UFC fighter stats, fighting techniques and body metrics, we will predict winning fighters with machine learning.

Our CSV file is small (23 columns and 8.990 rows) but complete as it contains roughly every match under the UFC umbrella.

- A Git Hub repository was created for the analysis so everyone in the group can contribute and review information.
- The group will meet twice a week during our scheduled class sessions on Zoom to work on the project and use our team Slack channel to communicate during the week.

Role Distribution

After establishing the communication structure, we created the foundation for our UFC fighter analysis project by defining roles that play to our individual strengths.

	Segment 1	**Segment 2**	**Segment 3**	**Segment 4**
Square	Mohammed	Alexandra	Mohammed	
Triangle	Alexandra	Mohammed	Felicia	
Circle	Oybek	Oybek	Oybek	
X	Felicia	Felicia	Alexandra	

Results

Segment 1 Pre-Processing the Data

- Our CSV file is small (23 columns and 8,990 rows) but complete
- To clean the data and make it appropriate for machine learning we:
  - Renamed the Win Column to Win or Lose to better predict wins and losses with the model and for easier interpretation.
  - Dropped unnecessary columns and NaN rows to further clean our data.
- We then used OneHotEncoder to encode and read the data into the model.

Segment 2 Training the Model and Integrating the Database

- To make sure we used the best entities for our SQL database, we mapped it with an Entity Relationship Diagram.
- We used StandardScalar to train and test our data and selected Random Forest to model our data

Results

Segment 3 Refining and Visualizing the Data

- We performed an exploratory analysis and established a baseline accuracy score of 64%
- We decided to use feature selection to find the best attributes to explain the relationship between a fighter's characteristics and winning matches
- A linear regression model helped us identify which variables were most significant.
    - This removed the noise in our model but the accuracy didn't improve.
    - We were able to improve the false negatives and positives - making our model more precise
    - However, we didn't have enough data to explain the variance.

Summary

- We created the foundation for our UFC fighter analysis project by defining roles that play to our individual strengths and establishing the communication structure. We preprocessed our data for easier encoding and modeling.
- We performed an exploratory analysis to establish a baseline accuracy score, created a database and used a linear regression to fine tune our model's accuracy.
- While we were able to remove the noise and improve the model's precision, we didn't have enough data to explain the variance so we used our dashboard to visualize interesting observations from the dataset such as countries with the most winning fighters, winning stances and winning finishes.

Dashboard/Viz