

UFC 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.

	<u>Segment 1</u>	<u>Segment 2</u>	<u>Segment 3</u>	<u>Segment 4</u>
Square	Mohammed	Alexandra		
Triangle	Alexandra	Mohammed		
Circle	Oybek	Oybek		
X	Felicia	Felicia		

Progress and Deliverables

Segments 1 and 2

- We used Python and Scikit-Learn to build and evaluate several algorithms to predict match winners.
 - We also created a database in pgAdmin using SQL to store the CSV data files and started working on a dashboard in Tableau for our presentation.
- Random Forest was our preferred modeling tool because it:
 - Runs efficiently on large data sets
 - Works against overfitting
 - Can be used to rank input variables

Results

Segment 1 Pre-Processing the Data

- 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.

Results

Segment 2 Training the Model and Integrating the Database

- We performed an exploratory analysis and established a baseline accuracy score.
- Next Steps: Decide how we'd like to improve our model's accuracy.
 - Options
 - Feature Selection - Visualization, like our Tableau Dashboard, will help.
 - Algorithm Tuning - We will need to decide if speed or performance is more important
 - Ensemble Methods - The most common solution

Summary