**UFC Fighter Analysis**

**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.

*Role Distribution: Week 1*

Square: Mohammed

Triangle: Alexandra

Circle: Oybek

X: Felicia

*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. In Week 1 of our project, we will use Python and Scikit-Learn (a Python Machine Learning Library) 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.

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. Our final results will be presented to the class using Google Slides.

**Results**

While our CSV file is small (237 columns and 4,495 rows), it is complete as it contains roughly every match under the UFC umbrella. Alexandra doubled our data by listing each fighter individually instead of working with the Red and Blue Fighter designations in the original dataset. She also renamed the Win Column to Win or Lose to better predict wins and losses with the model and for easier interpretation. Allie dropped unnecessary columns and NaN rows to further clean our data.

Images: Updated Data Shape and Types

Updated Dataframe List

We selected Random Forest to model our data and found the initial accuracy is 64%. Random Forest is our preferred modeling tool because it:

* + Runs efficiently on large data sets
  + Works against overfitting
  + Can be used to rank input variables

Image: 12.20 Accuracy Score

Allie is sure we can improve the model’s accuracy as we work. A few ways to improve our accuracy are:

* Feature Selection – Finding the best subset of attributes which better explain the relationship between a fighter’s characteristics (independent variables) and winning matches (target variable). Visualization, like Mohammed’s work with our Tableau Dashboard, will help.
* Algorithm Tuning- Finding the optimum value for each parameter. We will need to decide which is more important- speed or performance. Max\_features will improve the performance of the model by having more options to consider but we will need to maintain the right balance and select the best features for our model.
* Ensemble Methods- Combining multiple weak models to produce better results; The most common solution

**Summary**

This week 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 performed an exploratory analysis and established a baseline accuracy score. Next, we’ll need to decide how we’d like to improve our model’s accuracy.

After the break, we will continue to refine our model, make edits to our database and update our dashboard and presentation slides.