**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.

*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 Segment 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.

*Role Distribution: Segment 1*

Square: Mohammed

Triangle: Alexandra

Circle: Oybek

X: Felicia

*Role Distribution: Segment 2*

Square: Alexandra

Triangle: Mohammed

Circle: Felicia

X: Oybek

**Results**

*Segment 1*

While our CSV file is small (23 columns and 8,990 rows), it is complete as it contains roughly every match under the UFC umbrella. To clean the data and make it appropriate for machine learning. Alexandra dropped unnecessary columns such as Fighter\_total, Date\_total and Location\_total and NaN rows.

Image: Updated Data Shape and Types

Image: Updated Dataframe List

*Note: Dropping the NaN rows reduced our data by 754 rows.*

She also renamed the Win Column to Win or Lose to better predict wins (1) and losses (0) with the model and for easier interpretation, converted columns with data type object to a list so the model can read the encoded information and used OneHotEncoder to encode and read the data into the model.

Image: Revised Data Shape 1.10

*Segment 2*

After preprocessing the data, we decided to use feature selection to find the best attributes to explain the relationship between a fighter’s characteristics (independent variables) such as gender, class, weight, reach, height or odds and winning matches (target variable). After using StandardScalar to train and test our data, 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).
* 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

This week, Mohammed created a database using SQL to store static data for use during the project. Our database includes two tables, UFC Dataset and Mastertable\_Text, and is joined using Match ID. To make sure we used the best entities, Mo mapped our database with an Entity Relationship Diagram.

Image: 1.10 ERD Diagram

Lastly, we began to develop our dashboard. Felicia generated images in Tableau to add to the data story being told. She created images that showed winning match finishes and winning stances and then reached out to the team on Slack for feedback on the visuals.

Image 1.10 Slack Convo A

Image 1.10. Slack Convo B

Image 1.10 Slack Convo C

**Summary**

*Segment 1*

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 preprocessed our data for easier encoding and modeling.

*Segment 2*

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. If given more time, visualization, like Felicia’s work with our Tableau Dashboard, will help us identify areas for fine tuning to improve our model’s accuracy. An interactive dashboard is a powerful tool that can lend great strength to analysis.