A close up of a sign

Description automatically generated IST707

PREDICTING OF UFC FIGHT OUTCOMES USING MACHINE LEARNING

|  |  |  |
| --- | --- | --- |
| Laxman Kumar  [lakumar@syr.edu](mailto:lakumar@syr.edu)  School of Information Studies  Syracuse University |  |  |

**Introduction**

Sports betting is a $155 billion industry. Fighting ranks among the top in the industry, and the Ultimate Fighting Championship (UFC) is currently taking steps to push it even further. Mixed Martial Arts (MMA) fighter statistics involve everything from skill centric values such as wins, and significant strikes landed to physiological measurements such as height and reach. There are over one hundred different features up to analyze before any given fight, and machine learning can be used to best understand which are most relevant, and to indent trends and predict the outcomes (win/draw/loss) of each fight.

**Problem Statement**

The goal of this study is to explore our ability to predict the outcome of UFC fights based on each match’s pre-fight statistics using machine learning models. An accurate prediction model could both inform the best placed bets (and potential risk associated) for each fight, but also could provide insight to coaches when accepting fights to begin with, simply by looking at the opponent’s statistics relative to their fighter. It could also be used to help to identify which features are most significant in this prediction.

**Dataset and Features**

There are over one hundred different fighter statistics on UFC Stats for each of the fights in UFC record from 1993 to 2019, which include information such as fighters’ height, weight, reach, and stance, as well as statistics such as win streaks, strike percentage, guard passes, and strikes landed by location. The dataset used in this study was scraped from UFCStats and statistics pertaining to the circumstances of the fight (i.e. location, number of allocated rounds, etc.) were removed to as they were deemed irrelevant and did not align with the fighter-centric intention of the study.

Dataset link : <https://www.kaggle.com/rajeevw/ufcdata>

| Column name | Description |
| --- | --- |
| R*and B* | prefix signifies red and blue corner fighter stats respectively |
| *opp* | containing columns is the average of damage done by the opponent on the fighter |
| KD | is number of knockdowns |
| SIG\_STR | is no. of significant strikes 'landed of attempted' |
| SIG\_STR\_pct | is significant strikes percentage |
| TOTAL\_STR | is total strikes 'landed of attempted' |
| TD | is no. of takedowns |
| TD\_pct | is takedown percentages |
| SUB\_ATT | is no. of submission attempts |
| PASS | is no. times the guard was passed? |
| REV | *Probably reversels* |
| HEAD | is no. of significant strinks to the head 'landed of attempted' |
| BODY | is no. of significant strikes to the body 'landed of attempted' |
| CLINCH | is no. of significant strikes in the clinch 'landed of attempted' |
| GROUND | is no. of significant strikes on the ground 'landed of attempted' |
| win\_by | is method of win |
| last\_round | is last round of the fight (ex. if it was a KO in 1st, then this will be 1) |
| last\_round\_time | is when the fight ended in the last round |
| Format | is the format of the fight (3 rounds, 5 rounds etc.) |
| Referee | is the name of the Ref |
| date | is the date of the fight |
| location | is the location in which the event took place |
| Fight\_type | is which weight class and whether it's a title bout or not |
| Winner | is the winner of the fight |
| Stance | is the stance of the fighter (orthodox, southpaw, etc.) |
| Height\_cms | is the height in centimeter |
| Reach\_cms | is the reach of the fighter (arm span) in centimeter |
| Weight\_lbs | is the weight of the fighter in pounds (lbs) |
| age | is the age of the fighter |
| title\_bout | Boolean value of whether it is title fight or not |
| weight\_class | is which weight class the fight is in (Bantamweight, heavyweight, Women's flyweight, etc.) |
| no\_of\_rounds | is the number of rounds the fight was scheduled for |
| current\_lose\_streak | is the count of current concurrent losses of the fighter |
| current\_win\_streak | is the count of current concurrent wins of the fighter |
| draw | is the number of draws in the fighter's ufc career |
| wins | is the number of wins in the fighter's ufc career |
| losses | is the number of losses in the fighter's ufc career |
| total\_rounds\_fought | is the average of total rounds fought by the fighter |
| total\_time\_fought(seconds) | is the count of total time spent fighting in seconds |
| total\_title\_bouts | is the total number of title bouts taken part in by the fighter |
| win\_by\_Decision\_Majority | is the number of wins by majority judges decision in the fighter's ufc career |
| win\_by\_Decision\_Split | is the number of wins by split judges decision in the fighter's ufc career |
| win\_by\_Decision\_Unanimous | is the number of wins by unanimous judges decision in the fighter's ufc career |
| win\_by\_KO/TKO | is the number of wins by knockout in the fighter's ufc career |
| win\_by\_Submission | is the number of wins by submission in the fighter's ufc career |
| win\_by\_TKO\_Doctor\_Stoppage | is the number of wins by doctor stoppage in the fighter's ufc career |
| avg | average over number of rounds in the fight |

**Approach**

We have followed the Crisp DM – Cross industry standard procedure for data mining methodology. This is the most widely used analytics model. This model has five major phases –

• Business Understanding

Diagram

Description automatically generated• Data Understanding

• Data Preparation

• Modeling

• Evaluation

**Evaluation**

We are planning to use Regression, Random Forest and XGBoost.

**Reference**

1. https://www.kaggle.com/rajeevw/ufcdata