Exploring UFC Statistics:

A Comprehensive Analysis of Fights and Fighters

Members:
Meng-Wei (Vivian) Wu
Kexin (Shera) Huang
Zonghai (Liam) Wan

Table of Contents

Executive Summary

Competitive Sports have always been prevalent in making professional or amateur predictions and share moderate economic impacts. Yet, relying solely on industry experts for predicting match results might lead to inaccuracies in presumed outcomes. This report demonstrates the process of data scraping and discusses the development of a structured database for potential analysis and prediction of UFC (Ultimate Fighting Championship) combat outcomes. We employed web-scraping routines to gather comprehensive data using Python and the BeautifulSoup package, aiming for a table-based presentation. We target completed weekly UFC events, including match outcomes, fighter statistics and other pertinent information. Document-oriented model was used to store the collected data in a NoSQL database, MongoDB, allowing for fighter data stores in JSON format documents. The organized datasets might be used for betting or by potential fighter coaches in taking advantage of the information for future UFC results prediction using machine learning techniques. In order to handle the complexity of web-scraped data, the article emphasizes the need for a flexible and scalable data storage solution. Because of this, NoSQL MongoDB database is the best option for storing data with a variety of structures and complicated data types.

Background and Introduction

Sports prediction has always been top of the topics. However, there isn't quite a prediction model that can achieve high accuracy. Among all the sports competition, UFC (Ultimate Fighting Championship) is the one with the most complex and thorough statistics for each match. As a jiu-jitsu practitioner, personally speaking, it's one of the most entertaining sports to watch, bet, and analyze. After researching through the web, we found that among all the attempts to predict UFC results, the data sources people used are generally quite limited and not comprehensive. Hence, we'd like to build up a more structured database to touch on as many features as possible for the future UFC fight prediction.

Data Description

UFCStats.com[1] is a website that provides comprehensive statistics on all the completed UFC events. It is a valuable resource for anyone interested in the sport of mixed martial arts and offers detailed information on individual fighters, events, and championships. In order to gather this data, we would employ weekly or monthly web-scraping routines to ensure the database is a reliable and up-to-date source of information that could be used by fans, journalists, and industry professionals.

We have compiled data on the weekly fights that take place, as well as information on all active fighters currently participating in these events. This data includes relevant details such as match outcomes, fighter statistics, and other pertinent information that may be useful for analysis or record-keeping purposes. The web-scraping routine involved using Python and the BeautifulSoup package to extract the necessary information from the UFC website. Specifically, we targeted a table-based format on the website, which stored the information of each fighter row by row with a header on the top.

For the weekly fights, we began by scraping the entire page of all nights that hosted UFC events. Typically, these events occur every sunday. We want to obtain information of all completed fights. However, the first event shown on the web page is actually an upcoming fight for the following week. We differentiated between fights by selector class. Within each night, there were different numbers of fights scheduled, ranging from ten to fourteen per night. The information being extracted includes event name, hosting date and location, fighters' name and results, as well as combat status.

To create a simple and clear database schema, we inserted each fight into the same row as its event. The resulting table (see table 1) includes event name, date, location, web URL and a list of fights arranged in sequence. Additionally, there are fourteen subsequent tables associated with the main table.

By clicking on each match, for example match 1, one will find all matches of each UFC event, with the desired match highlighted.

title	date	location	link	match1	match2	match3	match4	match5
UFC Fight Night: Yan vs. Dvalishvili 3	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC 285: Jones	March 04, 2023	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	February 25, 20	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	February 18, 20	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC 284: Makha	February 11, 20	Perth, Western A	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	February 04, 20	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC 283: Teixei	January 21, 202	Rio de Janeiro, F	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	January 14, 202	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	December 17, 2	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC 282: Blacho	December 10, 2	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	December 03, 2	Orlando, Florida,	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	November 19, 2	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC 281: Adesa	November 12, 2	New York City, N	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	November 05, 2	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }
UFC Fight Night:	October 29, 202	Las Vegas, Neva	http://www.ufcst	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }	{ 7 fields }

Table.1 The overview of the table of nightly fights

To extract the relevant data for each fighter, we first identified all the rows associated with each fighter and then scraped the values for each row. The collected data included the fighter's FirstName, LastName, Nickname, Height, Weight, Reach, Stance, Win, Lose, Draw, and Belt. It should be noted that the data on the website was arranged by LastName and contained several null values. To prepare the data for analysis, we removed the units of Weight and Reach and converted the units of Height from foot and inches to inches only. This conversion facilitated straightforward comparison of the data. We then stored the information of each fighter in a list.

To store the collected data, we opted for a NoSQL database management system - MongoDB. We chose MongoDB due to its flexible data model, dynamic schema, and scalability. These features enabled the storage of unstructured data with varying data types and volumes, ideal for storing data scraped from the web.

In designing the database schema, we employed a document-oriented model that accommodated storing fighter information in JSON-like documents. Each document contained the extracted data for a single active fighter, and the field names were FirstName, LastName, Nickname, Height (in inches), Weight (in pounds), Reach (in inches), Stance, Win, Lose, Draw, and Belt. We also included an auto-generated ObjectID field to facilitate efficient retrieval of individual records (see table 2).

_id	First	Last	Nickname	Ht.	Wt.	Reach	Stance	W
641b10053572a	Tom	Aaron		null	155	null		5
641b10053572a	Danny	Abbadi	The Assassin	71	155	null	Orthodox	4
641b10053572a	Nariman	Abbasov	Bayraktar	68	155	66	Orthodox	28
641b10053572a	David	Abbott	Tank	72	265	null	Switch	10
641b10053572a	Hamdy	Abdelwahab	The Hammer	74	264	72	Southpaw	5
641b10053572a	Shamil	Abdurakhimov	Abrek	75	235	76	Orthodox	20
641b10053572a	Hiroyuki	Abe	Abe Ani	66	145	null	Orthodox	8
641b10053572a	Daichi	Abe		71	170	71	Orthodox	6
641b10053572a	Papy	Abedi	Makambo	71	185	null	Southpaw	10
641b10053572a	Ricardo	Abreu	Demente	71	185	null	Orthodox	5
641b10053572a	Klidson	Abreu	White Bear	72	205	74	Orthodox	15
641b10053572a	Daniel	Acacio		68	180	null	Orthodox	30

Table.2 The overview of the table of fighters

Overall, the NoSQL MongoDB database design choice was informed by the requirement for a flexible and scalable data storage solution capable of handling the web scraped data's complexity. The document-oriented database schema design was an ideal choice for storing data with varying structures and complex data types.

Business Insights

UFC stands for Ultimate Fighting Championship. In the early stages of UFC, it was set for identifying the most effective martial arts with no weight class differentiation and minimal rules.. Later it developed more fully into different weight classes, specified techniques and more rigorous rules used in the matches. In 2007, UFC caught the attention of the betting community and now UFC betting is officially listed in the online betting websites. UFC in turn also officially lists the "betting odds" of each match. Just like any old-fashioned fighting games where the "fighter statistics" are displayed when you choose your character, UFC has an official website to store all fights and fighters' statistics including but not limited to: (each match) number of knockdown (KD), number of significant strikes (STR), number of takedowns (TD), number of submission (SUB), and the winning method of that match (e.g. if winning by submission, it will list the submission types such as real naked choke, guillotine choke, arm triangle; or if winning by decision, and number of rounds and the total time last. Within each UFC fighter, UFC also lists their body measure and historical fights. Therefore, if we can scrape the comprehensive data from UFC statistics, the data can be used to analyze and predict using machine learning models for the future

UFC results. The prediction can be used for betting, and for potential fighters coaches who want to take a big picture of the opponents statistics and prediction to improve their fighters' skills to another level.

Summary and Conclusions

In conclusion, sports prediction has always been a hot topic, and UFC has the most comprehensive statistics among all sports competitions. Unfortunately, existing prediction models have not achieved high accuracy due to limited and incomplete data sources. In order to address this, we have built a structured database based on data scraped from UFCStats.com, a website that provides detailed information on individual fighters, events, and championships. This data includes information such as match outcomes, fighter statistics, and other pertinent details. For the collected data, we have decided to use a NoSQL MongoDB database for data storage, owing to its flexible data model, dynamic schema, and scalability. This database schema design was ideal for storing data with varying structures and complex types. In data utilization, machine learning models can be used to analyze and predict future UFC results, which can be useful for betting and potential fighter coaches who want a bigger picture of their opponents.