



Clustering UFC Fighters

What is MMA?

Mixed Martial Arts (MMA) is a combat sport that incorporates techniques from multiple martial art styles.

Broadly speaking, fighters employ:

- Boxing
- Brazilian Jiu-Jitsu (BJJ)
- Kickboxing,
- Wrestling
- Muay Thai,
- Taekwondo
- Karate



BOXING
GREECE | 688 BCE



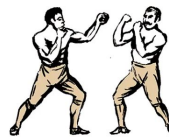
PANKRATION
GREECE | 648 BCE



KUNG FU
CHINA | 500s AD



SUMO
JAPAN | 20 BCE



PRIZEFIGHTING
ENGLAND | 1500s



GRECO-ROMAN
FRANCE | EARLY 1800s



KARATE
OKINAWA | 1800s



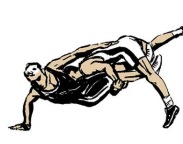
JUJUTSU
JAPAN | 1600s



MUAY THAI
THAILAND | 1600s



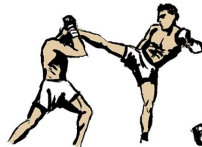
COLLEGIATE
USA | LATE 1800s



FREESTYLE
USA/UK | LATE 1800s



JUDO
JAPAN | 1882



KICKBOXING
USA/NETH/JAPAN | 1980s



SAMBO
SOVIET UNION | 1930



BRAZILIAN JIU JITSU
BRAZIL | 1909



TAE KWON DO
KOREA | 1955

Source: <https://blackbeltmag.com/>

MMA x Data Mining



Understanding

MMA, as a sport, is still in its nascent stages and is growing rapidly. We can use Data Mining to understanding the sport better through analysis.



Gaining a competitive advantage

Use past data to predict opponent's strategies and counter that effectively



Evaluation

Help quantify performances and in turn help in improving skill set

Related Work

A paper talks about clustering Basketball players on the basis of playing position using Self Organizing Maps.

Most work focus on predicting the winner of a fight.

Clustering on UFC datasets using K-means has been done before. It did not include round-wise statistics

Another article discusses the idea of analysing a fighter's approach. But it is based on just 3 aspects—Standing time, control time and controlled time

Dataset

The dataset statistics for all the fights that happened between Mar 11, 1994 and Nov 12, 2022.

49

Attributes

6000

Fights

16K

Fight Rounds

3000

Fighters

L

Israel Adesanya

"THE LAST STYLEBENDER"

W

Alex Pereira

"POATAN"

UFC

PERF

UFC MIDDLEWEIGHT TITLE BOUT

METHOD:

KO/TKO

ROUND:

5

TIME:

2:01

TIME FORMAT:

5 Rnd (5-5-5-5-5)

REFEREE:

Marc Goddard

DETAILS:

Punches to Head At Distance

TOTALS

FIGHTER	KD	SIG. STR.	SIG. STR. %	TOTAL STR.	TD	TD %	SUB. ATT	REV.	CTRL
Israel Adesanya	0	86 of 162	53%	119 of 209	1 of 4	25%	0	0	6:34
Alex Pereira	0	91 of 157	57%	140 of 214	1 of 1	100%	0	0	0:31

PER ROUND ▶

SIGNIFICANT STRIKES

FIGHTER	SIG. STR	SIG. STR. %	HEAD	BODY	LEG	DISTANCE	CLINCH	GROUND
Israel Adesanya	86 of 162	53%	41 of 103	21 of 29	24 of 30	77 of 148	3 of 4	6 of 10
Alex Pereira	91 of 157	57%	42 of 89	27 of 33	22 of 35	76 of 139	14 of 17	1 of 1

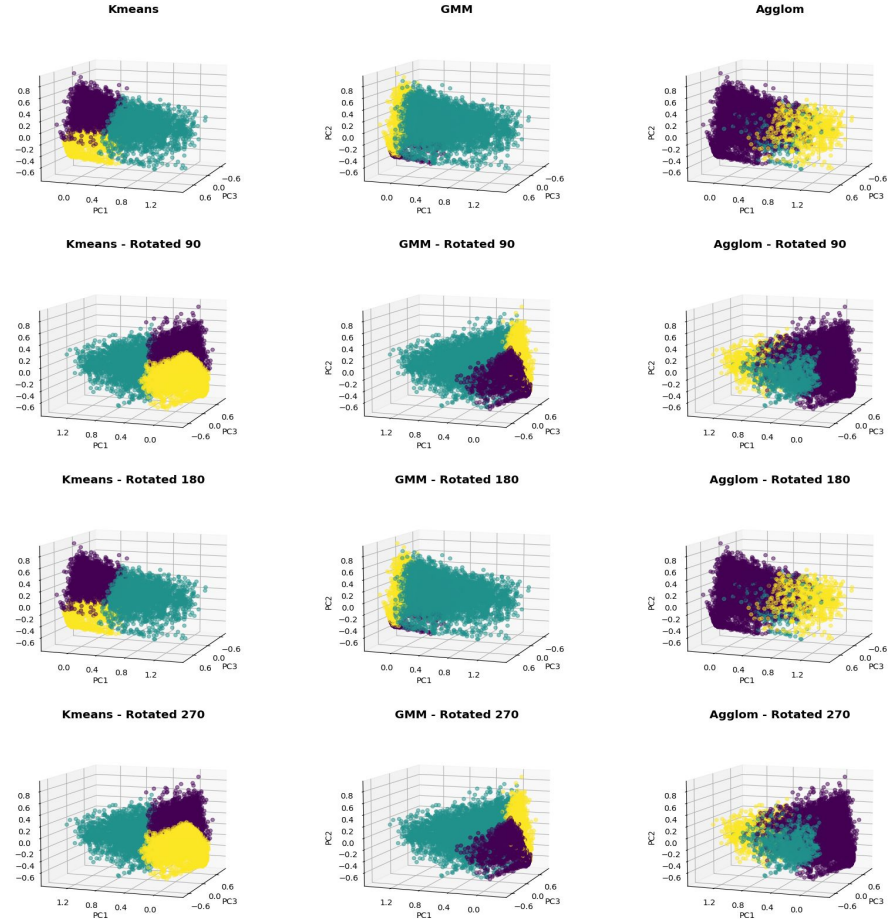
PER ROUND ▶

Source: <http://www.ufcstats.com/>

Pipeline

3D Visualization of results from 3 clustering algorithms

1. Data Preparation
 - a. Restructuring
 - b. Cleansing
 - c. Fill Missing Values
 - d. Standardization
 - e. Feature Reduction
 - f. Dimensionality Reduction
2. Clustering
 - a. K-means
 - b. Gaussian Mixture
 - c. Agglomerative Clustering
3. Evaluation



How did they do?

Handpicked a few fighters from each category, assigned labels and compared with results from each algorithm

Gaussian Mixture Model (GMM) produced results close to actual

Silhouette scores

- K-means 0.225286
- GMM 0.062508
- AGNES 0.265400

Cluster Labels for Hand-Picked Fighters

Fighter	Labels			
	K-Means	GMM	AGNES	Actual
Demetrious Johnson	Tactician	Grappler	Tactician	Grappler
Georges St-Pierre	Grappler	Grappler	Tactician	Grappler
Khabib Nurmagomedov	Grappler	Grappler	Grappler	Grappler
Anderson Silva	Power Puncher	Tactician	Tactician	Tactician
Conor McGregor	Power Puncher	Power Puncher	Tactician	Tactician
Israel Adesanya	Tactician	Tactician	Tactician	Tactician
Francis Ngannou	Power Puncher	Power Puncher	Tactician	Power Puncher
Derrick Lewis	Power Puncher	Power Puncher	Tactician	Power Puncher
Justin Gaethje	Tactician	Power Puncher	Tactician	Power Puncher

Conclusion & Future Work

Given that MMA fighters use skills from multiple fighting disciplines, both K-means and Gaussian Mixture Model performed well to cluster the fighters meaningfully.

Some of these fighters could belong to multiple clusters. In such cases, applying non-exclusive clustering algorithms could also produce compelling results. This can be good direction to take in the future.

Also, I plan on testing out more clustering algorithms like Mean Shift, Self Organizing Maps on this dataset.



Thank You!

Dall-E Prompt: Panda fighting
another panda in the UFC octagon