



FIFA WORLD CUP
Qatar2022

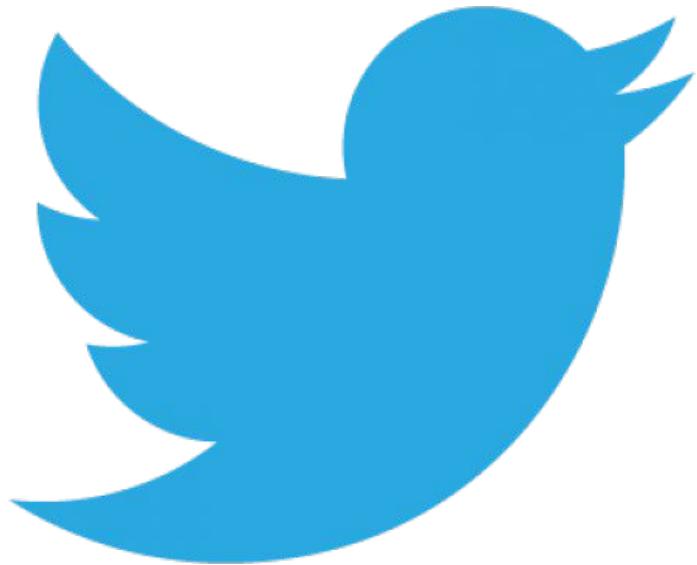
SENTIMENT ANALYSIS OPENING FIFA WORLD CUP 2022 QATAR

By : Alir Adn Haqqu

BACKGROUND



WHY TWITTER ???





#fifaworldcup2022

Harsha Purasinghe @harshap 16m
#Portugal 🇵🇹 #Argentina & #Messi 🦸 vs #Ronaldinho 🦸 still alive 😊 #FIFAWorldCup 🏆 #FIFAWorldCup2022

umar @jaemy_joker12 16m
👉 I found THE best way to bet on #FIFAWorldCup2022!

I'll sign up for @SixSigmaSports' very own #sportsbetting testnet to discover their crypto betting experience and enjoy native token rewards 🎉🎉

YB Digital @yb.digital 3m
09/12 #fifaworldcup 🏆 Quarter final results: [wcifly.com/en/worldcup-fi...](#) Netherlands 2 (3) - (4) 2 #Argentina qualified #Croatia qualified 1 (4) - (2) 1 Brazil 10/12 Quarter finals:
#Morocco - #Portugal 🇲🇦
#France - #England 🇬🇧 #FIFAWorldCup2022 #FIFAWorldCupQatar2022 #FIFA2022

Mugdho Sapiens @MugdhoSapiens 8m
A Match Full of Energy, Sweat and Bruises. It was extreme in all level; Good, Bad and Ugly. Finally, the better team won. #FIFAWorldCup2022 #ARGNED

bellaanovita 🌸 @bellanovitaaa 1m
Malam beb,aku ready untuk bo and vcs yaa ❤️❤️ wa.me/message/VUM2MY... #Balikpapan

<https://twitter.com/search?q=%23FIFAWorldCup>



#FIFAWorldCup2022

Bhavik Sharma @bhavik_sharma10 17s
Aaj Morocco जीती छाये #siu #FIFAWorldCup2022 #FIFAWorldCup 🇲🇶

Abdullah @Abdullahiwhab0 27s
i'm still selling tickets for the (Morocco 🇲🇶 vs Portugal 🇵🇹) match if you need them dm me if you're interested i'm selling at good prices #FIFAWorldCup2022 #FIFAWorldCup 🇲🇶 #WorldCup2022 #QatarWorldCup2022

Naseer Isah Ameen @nasey30 56s
Dear Morocco,
You are Africa's last hope.Give us thisday.Please Crush Portugal. #FIFAWorldCup 🇲🇶 #FIFAWorldCup2022

bellaanovita 🌸 @bellanovitaaa 1m
Malam beb,aku ready untuk bo and vcs yaa ❤️❤️ wa.me/message/VUM2MY... #Balikpapan



FIFA WORLD CUP
Qatar2022

Project Overview

Goal :

To Classify whether the opening of the world cup in Qatar 2022 has a positive or negative Sentiment on consumers who enjoy football

Dataset :

Subset include tweets in English containing hashtag #WorldCup2022 on Opening Ceremony Day.

Source: Kaggle



FIFA WORLD CUP
Qatar2022

DATASET

- Tweet → Contain Hashtag #FifaWorldCup2022
- Tweet for 24 hours in Opening Day Ceremony at 20 November 2022 GMT
- This Project contain 22524 tweet & 4 Features

	Date Created	Number of Likes	Source of Tweet	Tweet
0	2022-11-20 23:59:21+00:00	4	Twitter Web App	What are we drinking today @TucanTribe \n@MadB...
1	2022-11-20 23:59:01+00:00	3	Twitter for iPhone	Amazing @CanadaSoccerEN #WorldCup2022 launch ...
2	2022-11-20 23:58:41+00:00	1	Twitter for iPhone	Worth reading while watching #WorldCup2022 ht...
3	2022-11-20 23:58:33+00:00	1	Twitter Web App	Golden Maknae shinning bright\n\nhttps://t.co/...
4	2022-11-20 23:58:28+00:00	0	Twitter for Android	If the BBC cares so much about human rights, h...
...
22519	2022-11-20 00:00:21+00:00	1	Twitter Web App	Here We go World cup 2022 #WorldCup2022
22520	2022-11-20 00:00:03+00:00	0	DenetPro	Anderlecht confirms former Viborg FF's Jesper ...
22521	2022-11-20 00:00:01+00:00	2	Twitter for iPhone	Great thread to read before the start of #Wor...
22522	2022-11-20 00:00:00+00:00	11	Twitter Web App	Raphinha wants Brazil to be united at the #Wor...
22523	2022-11-20 00:00:00+00:00	24	Twitter Web App	How to buy \$SOT on PinkSale? 😊\n\nHave you been...

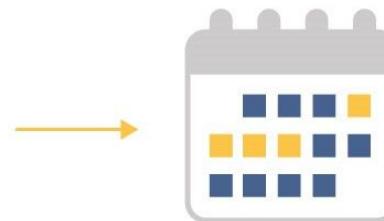
Workflow



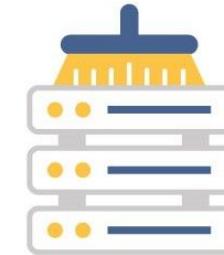
FIFA WORLD CUP
Qatar2022



1. TEXT INFORMATION



2. SEGMENTATION AND TOKENIZATION



3. TEXT CLEANING



7. INTERPRETATION
OF THE RESULT



6. MACHINE LEARNING
ALGORITHMS



5. TEXT LEMMATIZATION
AND STEMMING



4. VECTORIZATION AND
FEATURE ENGINEERING

DATA PRE PROCESSING

Cleaning Tweet



- Step A : Converting html entities
- Step B : Removing "@user" from all the tweets
- Step C : Changing all the tweets into lowercase
- Step D : Apostrophe Lookup
- Step E : Short Word Lookup
- Step F : Emoticon Lookup
- Step G : Replacing Special Characters with space
- Step H : Replacing Numbers (integers) with space
- Step I : Removing words whom length is 1
- Step J : remove Additional stopwords
- Step K : Stemming
- Step L : Retainment of Relevant Parts of Speech



FIFA WORLD CUP
Qatar 2022

Linguistic Data Reprocessing

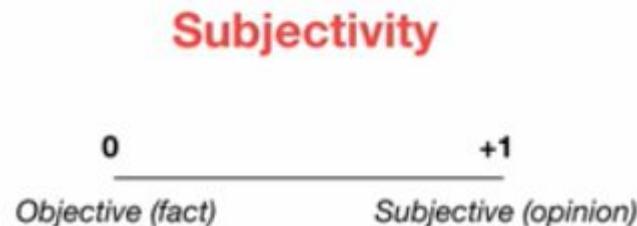
- TextBlob is a python library and offers a simple API to access its methods and perform basic NLP

Step :

- Collection of tweet
- Sentimental Classification
- Identifying polarity and subjectivity using TextBlob API.



TextBlob



Result

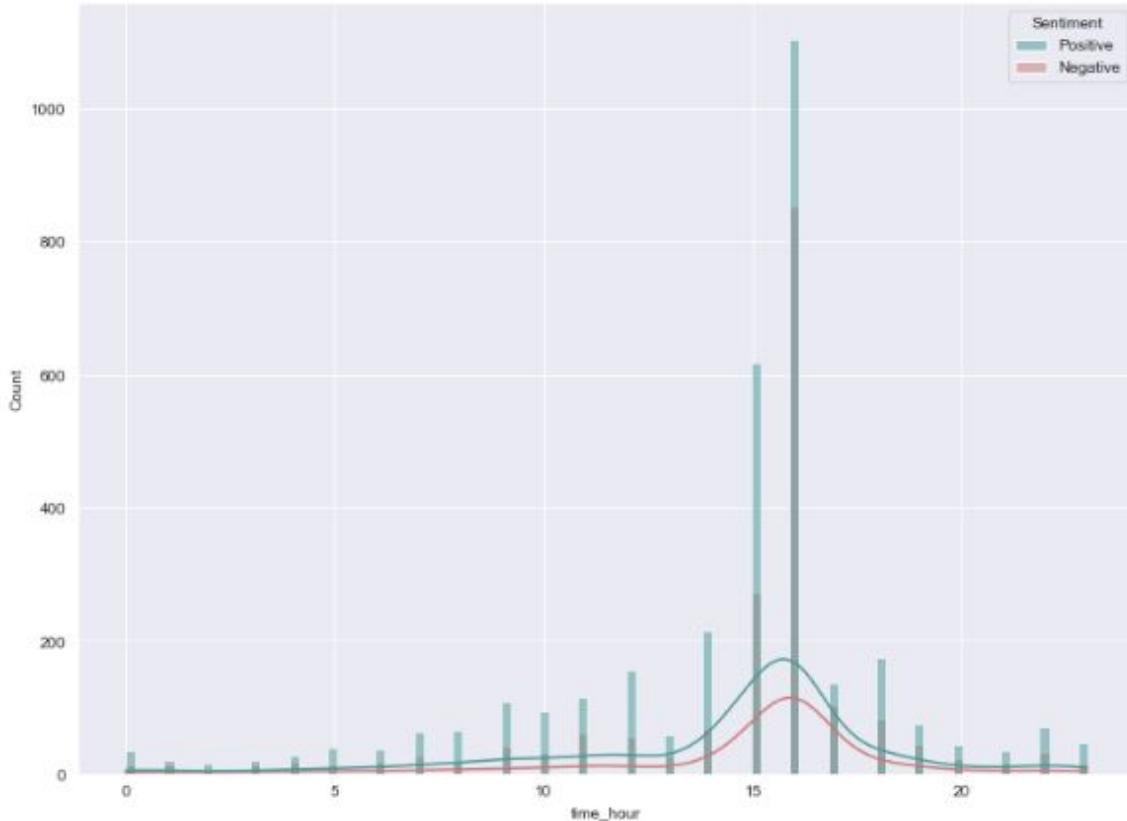


FIFA WORLD CUP
Qatar2022

Tweet_clean	Subjectivity	Polarity	Sentiment
someon free kick taken shown replay	0.800000	0.400000	Positive
less minut ecuador score	0.066667	-0.166667	Negative
incred centuri score maradona first commentari...	0.333333	0.250000	Positive
fun support england let us keep	0.200000	0.300000	Positive
even bafana bafana better host stokvel disguis...	0.500000	0.500000	Positive
wi connect slim found game he play canmnt	0.400000	-0.400000	Negative
need sort rule host countri fight spot type to...	1.000000	-0.500000	Negative
ayo wasnt fuck offsid	0.600000	-0.400000	Negative
help littl comfort safeti beauti pup confus lo...	0.600000	0.500000	Positive
delay deni first ecuador	0.333333	0.250000	Positive



Traffic Tweet



Observation :

1. Based on the time, consumer traffic tweeted at 14.49 GMT during the closing ceremony of the opening ceremony of the FIFA world Cup

Subjectivity vs Polarity



Observation :

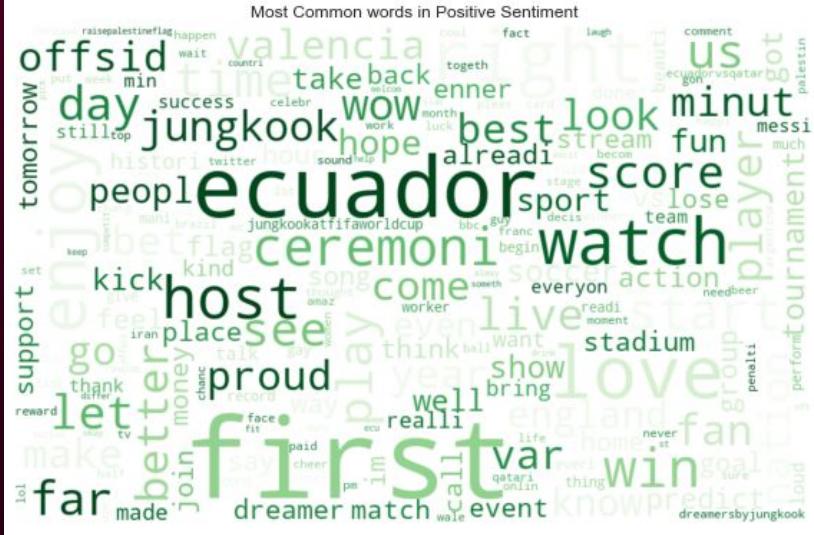
1. Based on subjectivity, the average score is 0.53, which means that consumer sentiment tends to be subjective
2. Based on Polarity, the average score is 0.1, which means that consumer sentiment tends to be slightly positive

EDA WordCloud



FIFA WORLD CUP
Qatar2022

Most Common words in Positive Sentiment



Positive Tweet : Jungkook, Win, Ceremony

Most Common words in Negative Sentiment

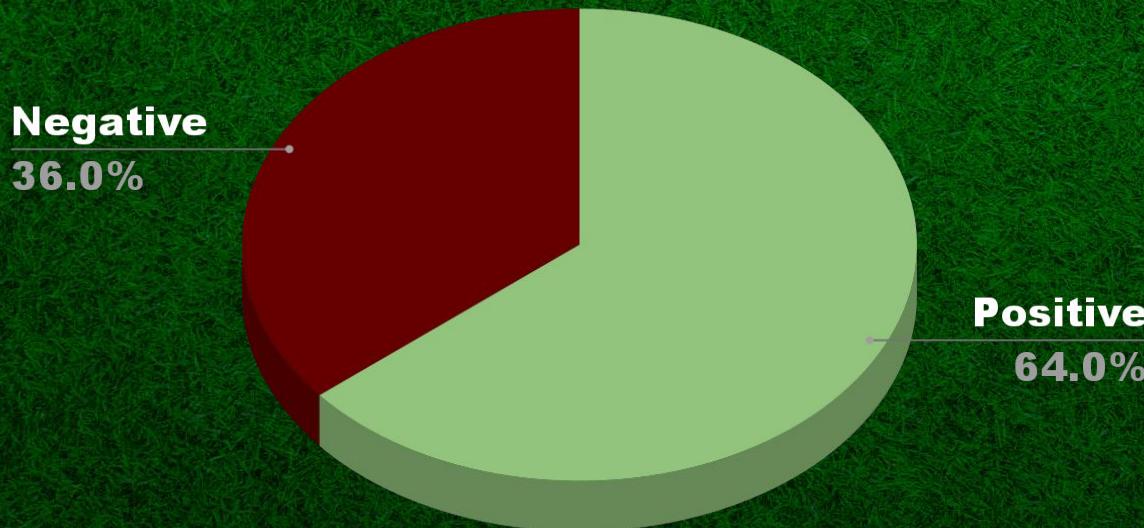


Negative Tweet : Corrupt, Money, Spent



FIFA WORLD CUP
Qatar 2022

Validation Model Using Supervised Learning Target Variable



Observation :

Target variable is Imbalanced. So we use AUPRC Score to evaluate models than Accuracy.
(Rosenberg,2022)



FIFA WORLD CUP
Qatar2022

Model Building

Using Word Embedding:

1. **BOW (Bag-of-Word)**
2. **TF-IDF (Term Frequency-Inverse Document Frequency)**

Model:

1. **Logistic Regression**
2. **Naive Bayes**
3. **SVM**
4. **Random Forest**
5. **XGB**

BASELINE MODELL - BOW (Bag-of-Word)



FIFA WORLD CUP
Qatar2022

MODEL	ACCURACY	AUPRC	RECALL	PRECISION	F1 SCORE
Logistic Reg	0.93	0.97	0.97	0.92	0.94
Naive Bayes	0.89	0.94	0.92	0.92	0.92
SVM	0.93	0.97	0.96	0.94	0.95
Random Forest	0.93	0.95	0.96	0.94	0.95
XGB Classifier	0.93	0.97	0.97	0.93	0.95

BASELINE MODELL - TF-IDF



FIFA WORLD CUP
Qatar2022

MODEL	ACCURACY	AUPRC	RECALL	PRECISION	F1 SCORE
Logistic Reg	0.86	0.98	0.99	0.83	0.90
Naive Bayes	0.75	0.90	0.98	0.73	0.84
SVM	0.95	0.99	0.99	0.93	0.96
Random Forest	0.92	0.97	0.98	0.90	0.94
XGB Classifier	0.94	0.99	0.99	0.93	0.96

HYPERPARAMETER TUNING & EVALUATION



Model	Accuracy	Recall	Precision	F1 Score
Baseline	0.95	0.99	0.93	0.96
Tuning	0.95	0.94	0.96	0.95
Tuning in Test Data	0.96	0.99	0.95	0.97



FIFA WORLD CUP
Qatar2022

Machine Learning Conclusion :

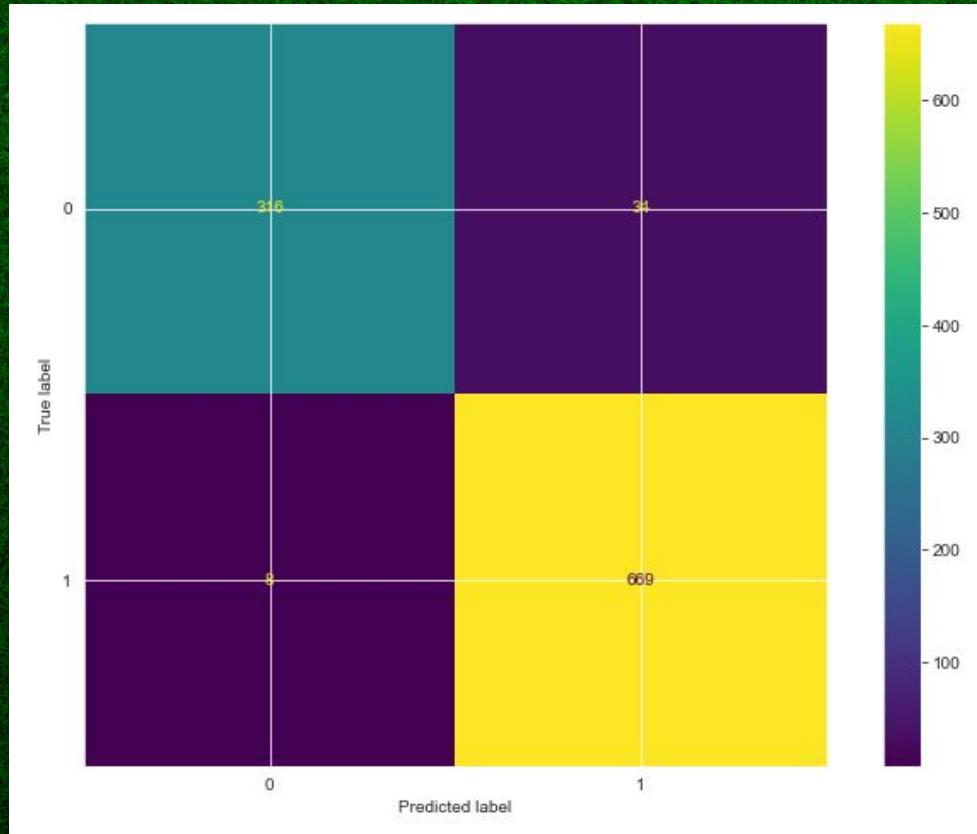
- Best Model = Tuned SVM + TF-IDF.
- Hyperparameter:

Hyperparameter Tuning	
C	1000
Gamma	0.001
Kernel	rbf



FIFA WORLD CUP
Qatar2022

Model Evaluation



Summary :

- True Positive = 669
- True Negative = 316
- False Positive = 34
- False Negative = 8
- Accuracy = 0.96
- F1 score = 0.97

Baseline Model without tuning
has accuracy 95%



Conclusion

- ❖ In dataset, most sentiment are positive with 64% with most frequent word Jungkook, Win, Ceremony. Which reasonable because this event has K-Pop Star who singing the official Anthem FIFA World Cup Qatar. The traffic boost up when the singer show up at 14.39 Pm GMT.
- ❖ Based on subjectivity, the average score is 0.53, which means that consumer sentiment tends to be subjective
- ❖ Best model is Tuned SVM + TF-IDF with 96% accuracy. Accuracy increase with 1% using tuning.

Business Insight

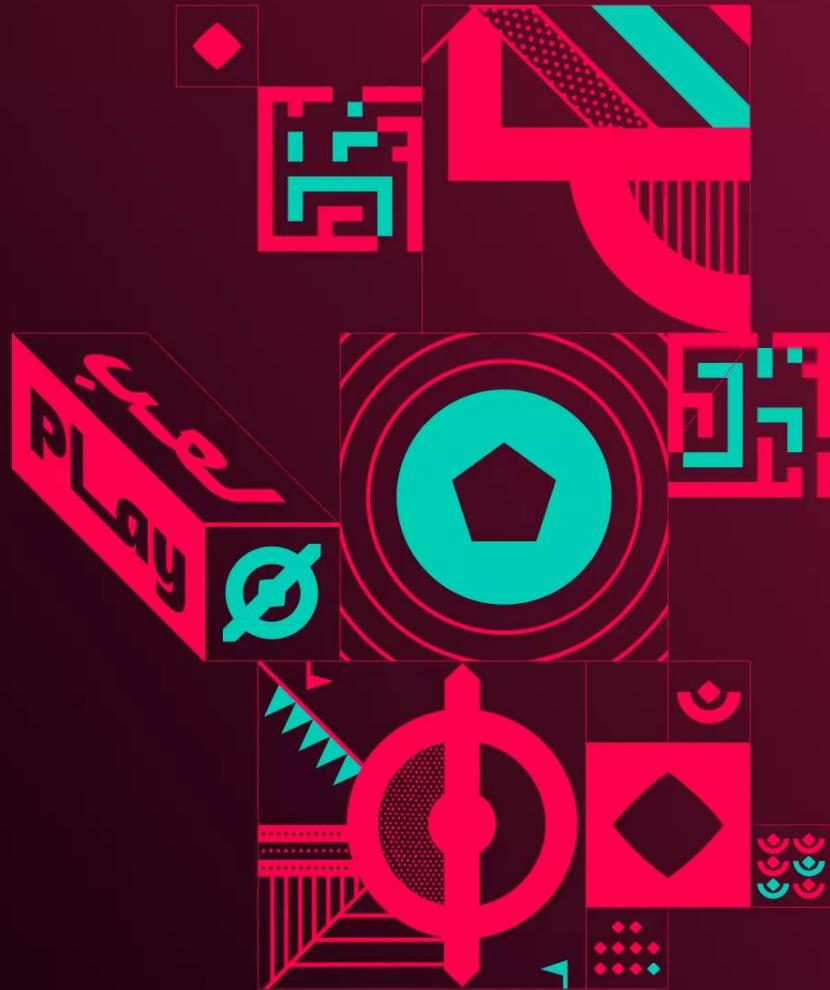
- Has South Korean pop singer Jung Kook of K-pop boyband BTS as singing the official Anthem FIFA World Cup Qatar is very well impact to the customer Sentiment also using him as a headlined the event quite success.
- With this Model Machine Learning has capability to classify sentiment consummate very good, with 96% accuracy



FIFA WORLD CUP
Qatar2022

Reference

- Machine Learning | [Tuning Hyperparameters Logistic Regression Menggunakan Grid Search #UcupStory](#) by Adipta Martulandi
- Machine Learning | [BOW+ TF-IDF in Python For Unsupervised Learning](#) by Eleonora Fontana
- Machine Learning | [Twitter Sentiment Analysis](#) by Anand Kumar
- Machine Learning | [Twitter Sentiment Analysis using a Classification Algoithm](#) by Pallavi Gupta
- Machine Learning | [Sentiment Analysis – Let TextBlob do all the Work!](#) by Abdul Hafeez Fahad
- Machine Learning | [Twitter Sentiment Analysis with full code and explanation \(Naive Bayes\)](#) by Koshu Takatsui



THANK YOU



FIFA WORLD CUP
Qatar2022

