



**FIFA WORLD CUP**  
**Qatar 2022**

**15.095 Machine Learning**  
Yen Hann Yoo | Shannan Liu



# Introduction & Overview

## 2018 FIFA World Cup

- Over 3.5 billion spectators
- Betting valuations: \$155 billion

Betting strategy often unclear due to risk-reward trade-off

- Do we bet on Team X to win because the potential returns are high even though the odds are stacked against them?

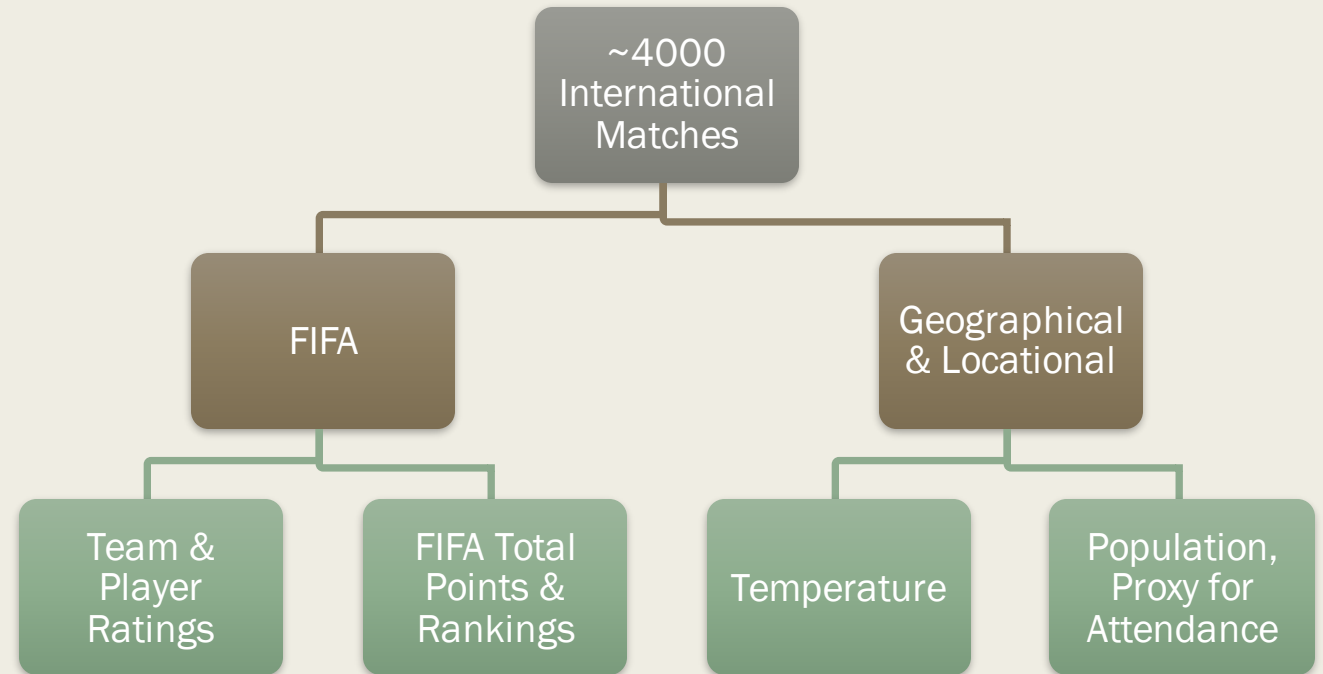
We sought to build a betting framework for the Group Stages

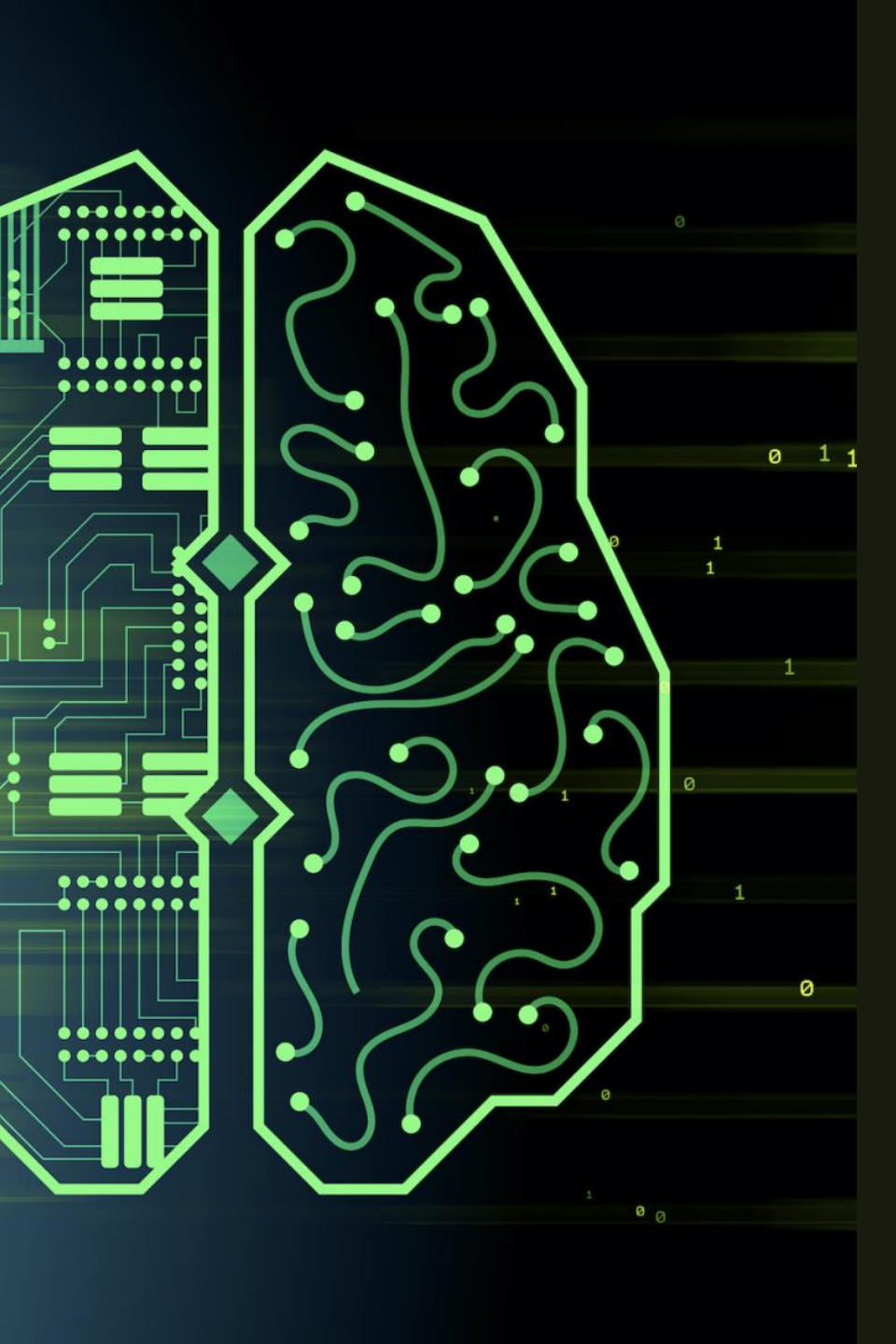
- How to allocate bets to maximize expected returns?





# Data Pre-Processing





# Modelling

	XGBoost	Random Forest	LGBM	Logistic Regression
AUC	0.914	0.859	0.903	0.73
Accuracy	0.79	0.67	0.78	0.59
Precision	0.79	0.76	0.77	0.55
Recall	0.76	0.59	0.76	0.52
F1	0.77	0.58	0.76	0.52

- For each match  $i$ , and match outcome  $j$  (win, draw, loss), XGBoost was used to generate probabilities,  $p_{ij}$

# Betting Framework

$$\max \sum_{i=1}^{48} \sum_{j=1}^3 p_{ij} R_{ij} x_{ij} - \sum_{i=1}^{48} \sum_{j=1}^3 x_{ij}$$

## Decision Variables

- $x_{ij}$ : Betting amount on match  $i$ , outcome  $j$
- $z_{ij} = \begin{cases} 1, & \text{if a bet is made on match } i, \text{ outcome } j \\ 0, & \text{otherwise} \end{cases}$

## Constraints

- Budget constraints
- Linking constraints between  $x_{ij}$  and  $z_{ij}$
- Place bets on multiple matches





# Results

$$\text{Profit from Match } i = \begin{cases} R_{ij}x_{ij} - x_{ij}, & \text{if predicted} = \text{actual} \\ -x_{ij}, & \text{if predicted} \neq \text{actual} \end{cases}$$

BetAmt. : \$5000	Model (XGB)	God's Eye
Total Profit	\$3,150	\$19,937.50

Model (XGB)		God's Eye	
Match	Profit	Match	Profit
CAM v. BRA	\$1732.5	ARG v. KSA	\$5880.0
JPN v. CRC	\$1367.5	CAM v. BRA	\$1732.5
AUS v. DEN	\$1345.0	JPN v. CRC	\$1367.5
JPN v. ESP	\$1260.0	AUS v. DEN	\$1345.0
KOR v. POR	\$565.0	GER v. JPN	\$1320.0

# 2022 FIFA WORLD CUP

Predictions for R016, Quarter Finals, Semi-Finals, and Finals

# Knockout Stage Predictions: R016





# Knockout Stage Predictions: QF



# Knockout Stage Predictions: SF



# Knockout Stage Predictions: FINALS







# Conclusions

- Difficult to correctly predict the outcomes of every match due to "upsets"
- Team's FIFA points most important predictor; midfield is the most important part of a team's success

## What can we say for certain?

- England, football is not coming home
- If our model got anything right, the probability of Saudi Arabia winning the World Cup is more than Italy's





**Thank You**  
(France, please don't win)

**FIFA WORLD CUP**  
**Qatar2022**