Football data Challenge

By Ibrahima Ndiaye

Competition & Goal

- All the "Serie A" football matches of the last 8 years are provided.
- Predict the final result, having at your disposal the odds of various bookmakers.

Datasets - Train & Test

- ID = An anonymous ID unique to a given match
- Date = Date of the match
- FTR = Full Time Result
- HomeTeam = Home Team
- AwayTeam = Away Team
- B365H = Bet365 home win odds
- B365D = Bet365 draw odds
- B365A = Bet365 away win odds
- BWH = Bet&Win home win odds
- BWD = Bet&Win draw odds
- BWA = Bet&Win away win odds
- IWH = Interwetten home win odds
- IWD = Interwetten draw odds
- IWA = Interwetten away win odds
- LBH = Ladbrokes home win odds
- LBD = Ladbrokes draw odds
- LBA = Ladbrokes away win odds
- VCH = VC Bet home win odds
- VCD = VC Bet draw odds
- VCA = VC Bet away win odds
- WHH = William Hill home win odds
- WHD = William Hill draw odds
- WHA = William Hill away win odds

Benefits to the client

- Leverage the opportunity provided by bookmakers to make easy money
- To transform sports betting into a lucrative investment (land, stocks etc...)
- This will give the client a immediate return of investment which is good for both short term and long term.

What do we want to achieve

- Predict the outcome of football games as accurately as possible
- Kaggle competition winner got a score of 0.66885 , runners up got a score of 0.53442
- A score prediction of at least 0.5 and above will be considered a success

How we expect to achieve it?

- Clean and enrich the data
- Add additional variables :
  - Tier A, B. C
  - Weather conditions
  - Current form
  - Head to Head
  - Referee of the game

How we expect to achieve it?

Multiclass classification problem which can be solved a number of ways, but we will choose one of the two :

- Change it to a binary classification and use Logistic Regression
  - Home team wins or not
  - Then do the same for Away team (Away team wins or not?)
  - If result is the same, we can conclude its a draw
- Multiclass classification and use Neural Networks or decision tree
  - More difficult but with less steps

Ethical Implications

It could affect sports and its value to consumers:

- Ruin the fun of sports and its unpredictability
- Players salaries, lack of interest in sports from the customers

If any ethical implications:

- Gambling

THANK YOU