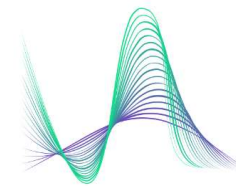




MACHINE LEARNING APPLICATIONS

IN ONLINE GAMBLING

Máté Sinkó



Prospect AI

2022. MÁJUS 03-04.
BUDAPEST ML FORUM
MACHINE LEARNING | DATA SCIENCE | ARTIFICIAL INTELLIGENCE

ONLINE GAMBLING – a brief history

1994

First ever online bet placed

2018

121 licensed operators across Europe

2020

234 licensed operators - 94% growth in 2 yrs

2022

market value of € 93 bn
28 million players – 19% YoY growth



Transaction volume



Source: <https://entaingroup.com/wp-content/uploads/2021/08/Entain-Inv-day-slides-combined-FINAL2-12-Aug-21.pdf>

Introduction



1999 - 2013

Data roles, Data Mining, BI, DWH

accenture

T · Systems



2014

Moved to Gibraltar – Advanced Analytics



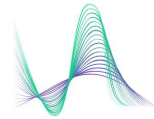
2016

Head of Advanced Analytics

LADBROKES CORAL
GROUP

2019

Founder



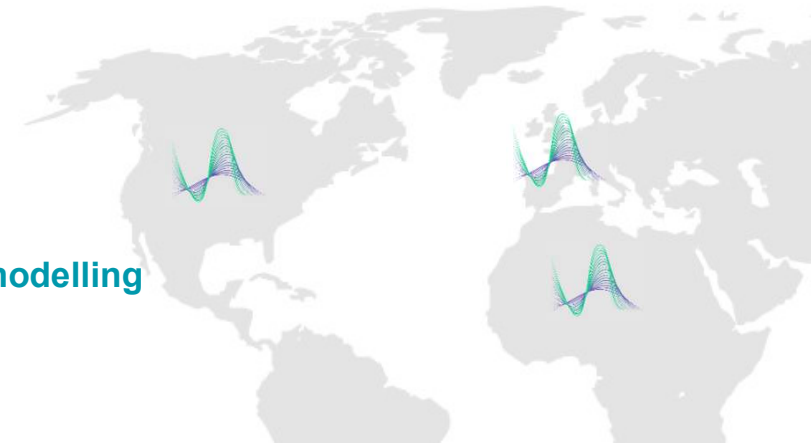
Prospect AI

Working with leading gambling brands internationally

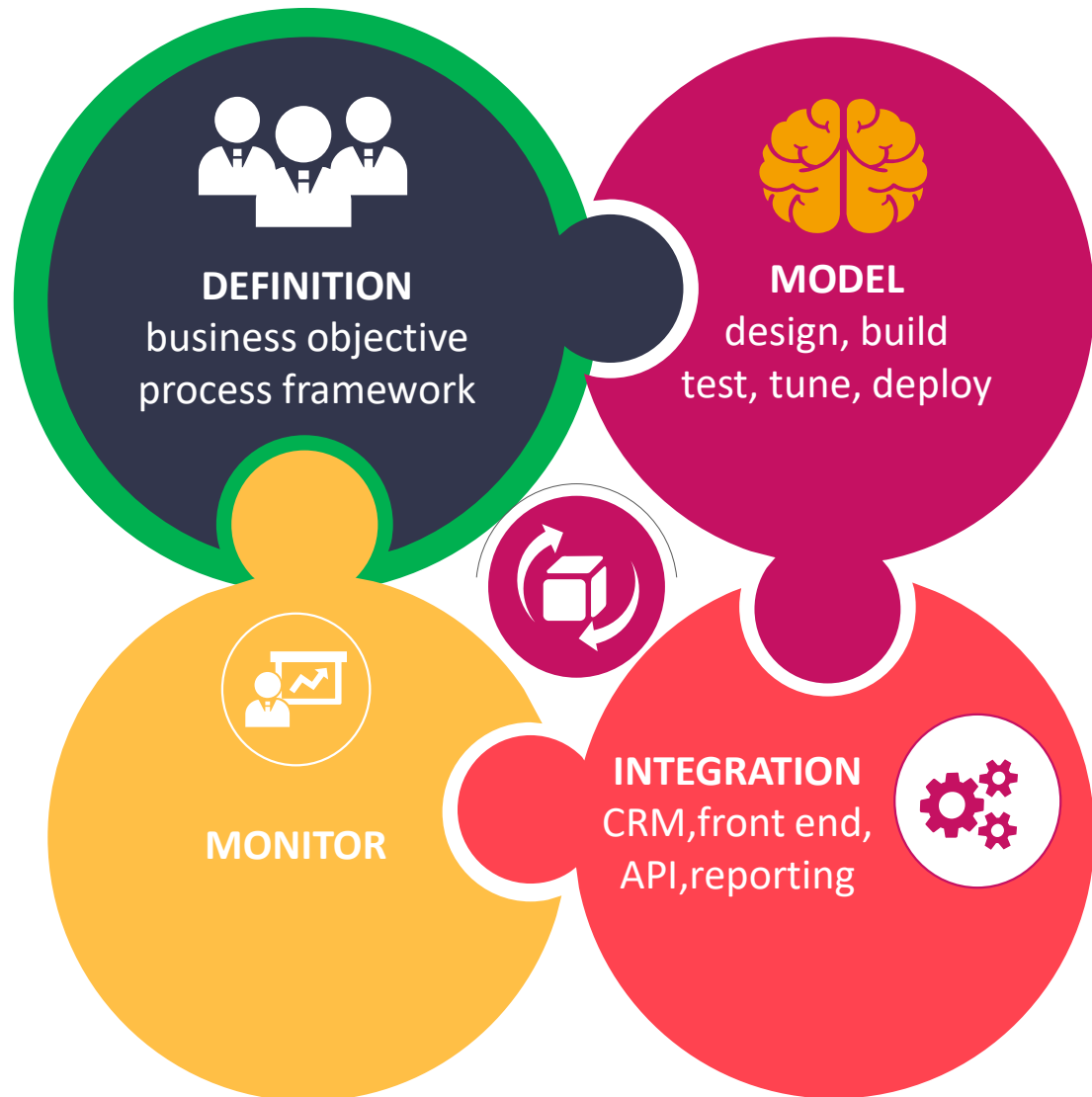
Consulting, analytics, player behaviour modelling, forecasting, predictive modelling

Interim Head of Analytics

Projects with C level sponsorship and involvement



Machine Learning project cycle







The sportsbook business model

Sports outcome probability models

Mallorca to win	Draw	Granada to win	True probability
50%	26%	24%	True odds - neutral payout level 1/TrueProb (0% margin)
2.00	3.82	4.14	Reduced odds (~ 6.2 % margin works for the bookmaker)
1.95	3.50	3.80	

bet365

Sáb 07 may	1	 Mallorca	X	2	 Granada
14:00		Mallorca			Granada
	1.95		3.50		3.80

Bookmaker Payout ODDS

Implied probabilities (1/Bookmaker Odds)			Overround
51%	29%	26%	106.2%

Sports outcome probability models

Large set of features

Recent games history

Wins / loses

Home / Away stats

Number of goals

Passes

Attacks

Dangerous attacks

Shots on goal

Corners

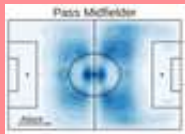
Player performance,

injuries,

VAR, XG

Weather

Visual game data,
pitch heat map, etc.



Markov
Chains

Monte
Carlo

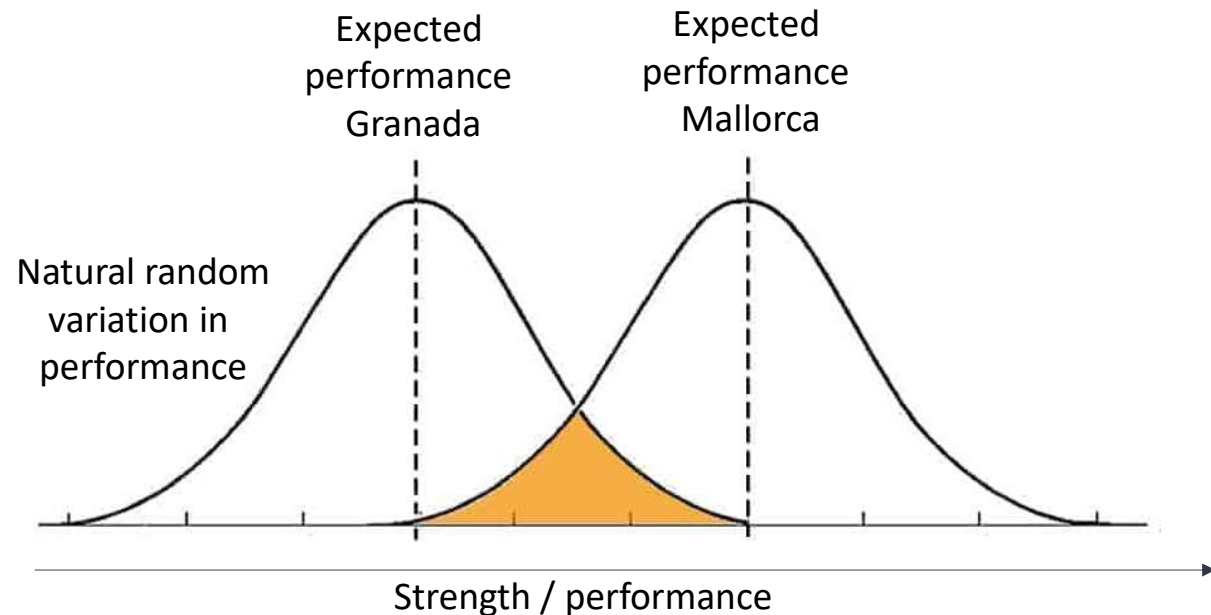
Neural
Networks

Linear
Models

Mallorca to WIN

50.0 %

True probability



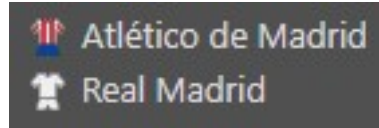
Event popularity pyramid

Top events

Wealth of data

Super accurate True Prob

Large betting volume, bookmaker can afford low margin



Non-top teams

Rich data

Accurate True Prob

Less volume, higher margin

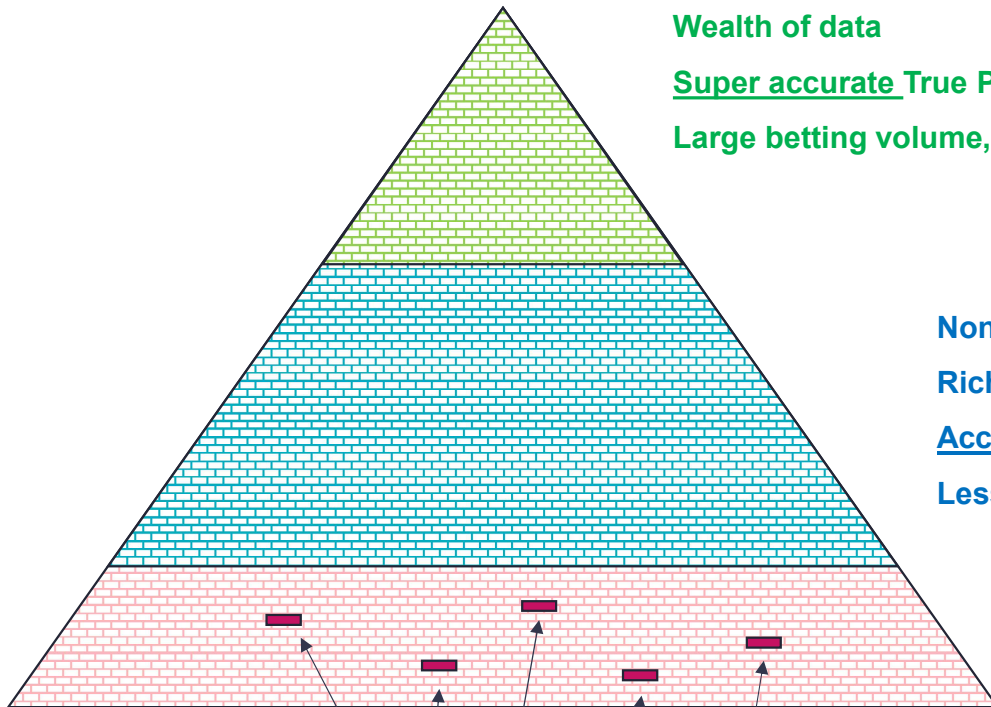
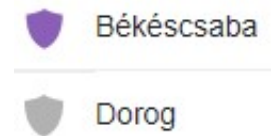


Least popular

Narrow set of data

Limited accuracy of TrueProb

Bookmakers defend themselves with high margin



“SHARP” players (Wiseguys) step in to exploit cracks in the structure

X referenced Odds, Tipsters, People building their home-made models

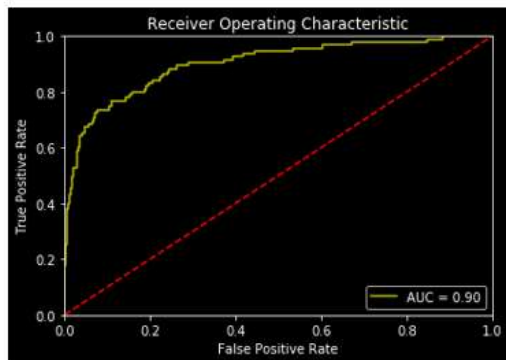
Exploit profiling – WiseGuy modeling

- Binary classifier
- Estimates probability of player being a WiseGuy
(likelihood of long term negative revenue expectancy)
- ~ 50+ original features, 15 used in the model
- Py XGBoost
- Imbalanced data, 0.5% positive target baseline
- AUC ceiling at 0.9

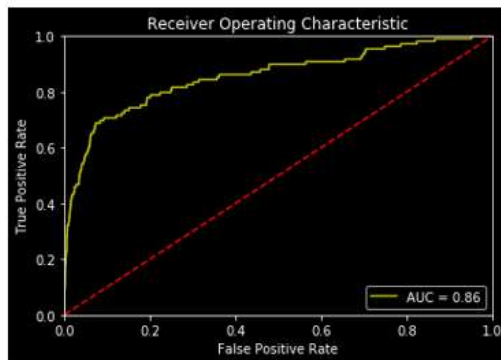
TREATMENT

- Bookmaker will limit maximum stakes on top risk players
- Players need to accept T & C when signing up

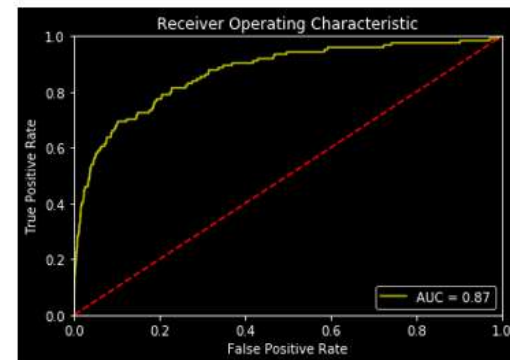
After 50 bets



After 20 bets



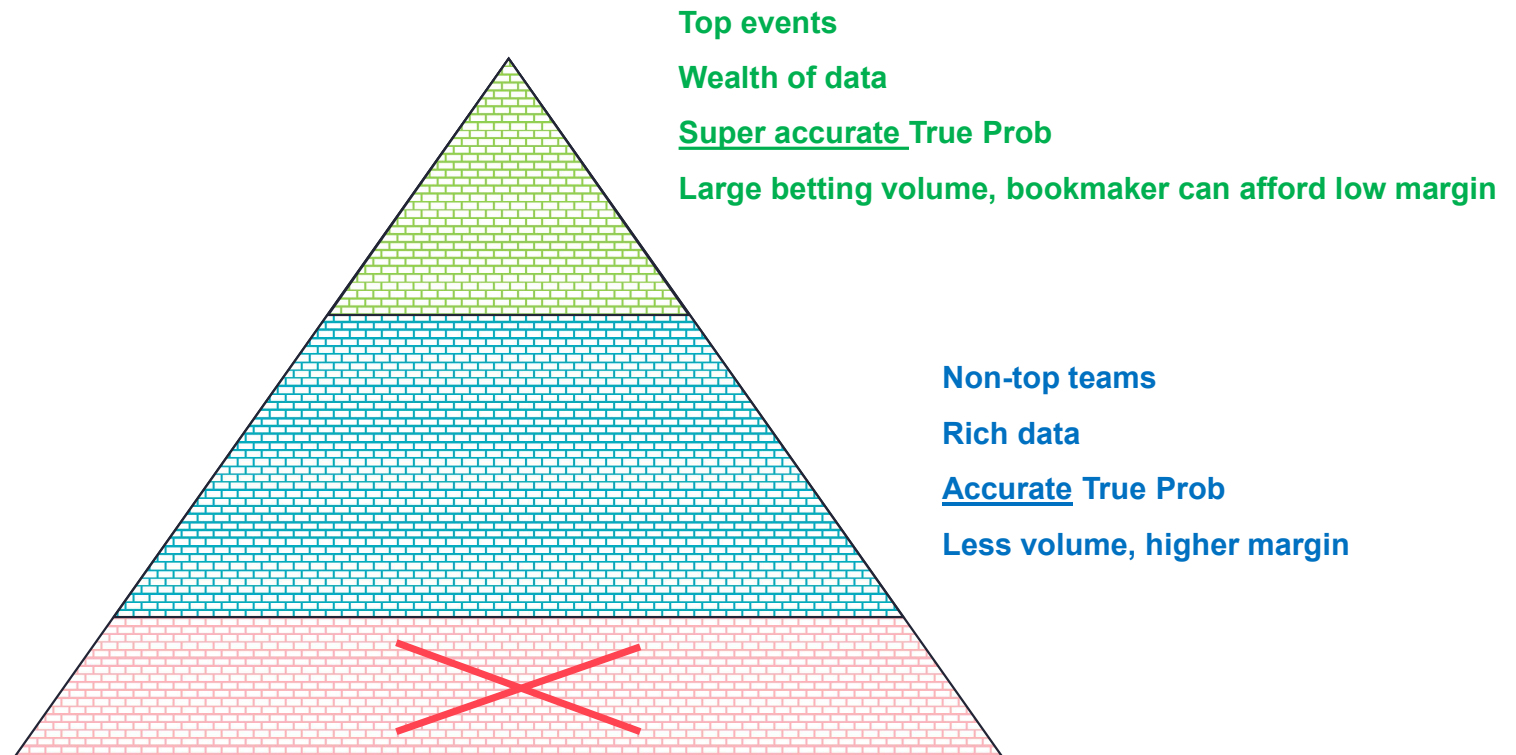
After 10 bets



Some “sharp” bookmakers do not limit players



- Special model: ODDS self correct as players place bets (like stock exchange price movements)
- Lowest popularity games are less available



Player protection - gambling harm minimization

- Evaluates risk of excessive play & gambling harm
 - Identify patterns of self control issues
 - Binary classifier
 - Python XGBoost
 - Key Markers of Harm
 - Late night play
 - Loss chasing
 - Frequency of play increase
 - Declined deposits
 - Cancelled withdrawals
 - Multiple payment methods
 - ~ 100+ original features, 10 used in the model
 - Highly imbalanced data, 0.25% positive target baseline
 - AUC ceiling around 0.85
-

TREATMENT of top risk players

- Direct notifications
- Promoting responsible gambling tools
(e.g. self exclusion)
- Deposit limits, ban of deposit
- Request source of funds proofing
(affordability verification)

BeGambleAware.org[®]



Other predictive models

- Player activity prediction
- Product recommendation systems (collaborative filtering)
- Media Mix Optimisation (multiple regression) for budgeting
- Player value
- Risk of churn
- Cross-sell propensity
- Promotion response prediction, bonus optimization
- Withdrawal risk profiling
- Payment fraud profiling



Thank you

Q / A

