

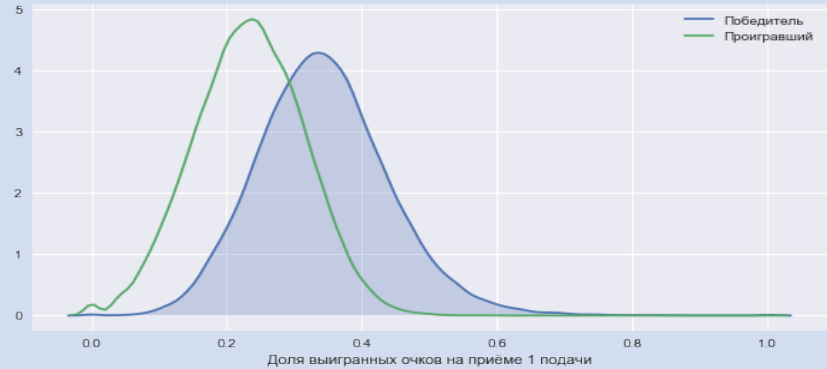
Анализ формирования ставок на теннисные матчи на букмекерском рынке

Анисимов Максим Андреевич

Датафрейм: статистика матчей

tourney	date	winner name	loser name	score
20000501		Antony Dupuis	Andrew Ilie	3 – 6 7 – 6(6) 7 – 6(4)
20000501		Fernando Gonzalez	Cecil Mamiit	6 – 2 7 – 5
20000501		Paradorn Srichaphan	Sebastien Lareau	6 – 1 6 – 3
20000501		Jan Siemerink	Justin Gimelstob	4 – 6 6 – 2 7 – 5
20000501		Jason Stoltenberg	Alex Lopez Moron	6 – 1 6 – 4

График плотности: доля выигранных очков на приёме 1 подачи
Заметна разница между распределением показателя победителя и проигравшего



Доля эйсов в общем числе очков на подаче - рост
Тенденция: чем выше игрок, тем больше он подаёт эйсов в процентном соотношении



Качество предсказания вероятностей

- RandomForestClassifier MAE = 0.5
- XGBoost ROC-AUC = 0.5

Примечание: предсказания делались для игрока 1

Цели работы:

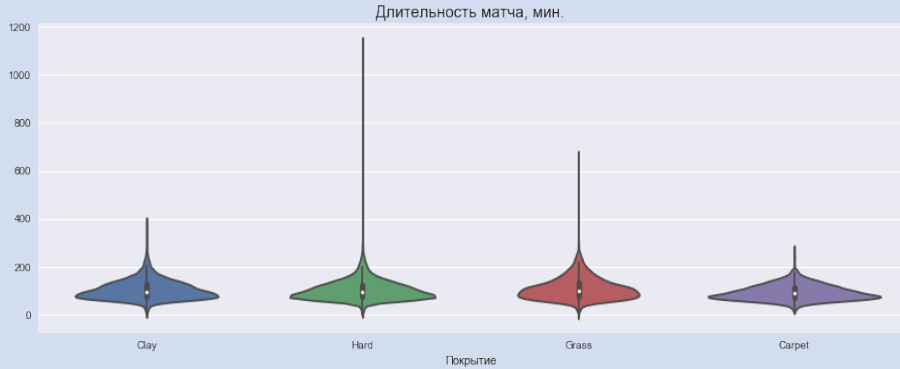
- Работа с «сырыми» данными
- Визуализация и анализ данных

Датафрейм: ставки

Date	Winner	Loser	AvgWBet	AvgLBet
2001 – 01 – 01	Clement A.	Gaudenzi A.	NaN	NaN
2001 – 01 – 01	Goldstein P.	Jones A.	NaN	NaN
2001 – 01 – 01	Haas T.	Smith L.	NaN	NaN
2001 – 01 – 01	Henman T.	Rusedski G.	NaN	NaN
2001 – 01 – 01	Hewitt L.	Arthurs W.	NaN	NaN

AbgWBet - средняя ставка на победителя, AvgLBet - средняя ставка на проигравшего

Длительность матча - покрытие
Наибольший разброс значений наблюдается на харде



Динамика средней длительности матча за год
Средняя длительность матча выросла за 16 лет



Самые важные признаки для предсказания: RFC

- Возраст игрока 1
- Среднее число сыгранных очков на приёме 2-ой подачи за матч игрока 2
- Средний процент двойных ошибок за матч игрока 2
- Средний процент выигранных очков на приёме 1-ой подачи за матч игрока 1
- Среднее число двойных ошибок за матч игрока 2

Основные трудности:

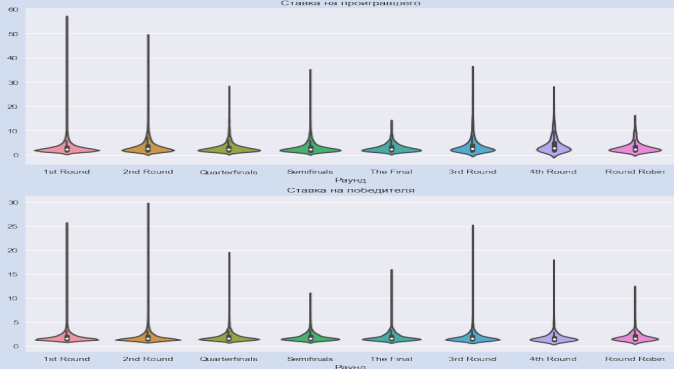
- Соединение датафреймов
- Оптимизация качества предсказаний

Новый датафрейм

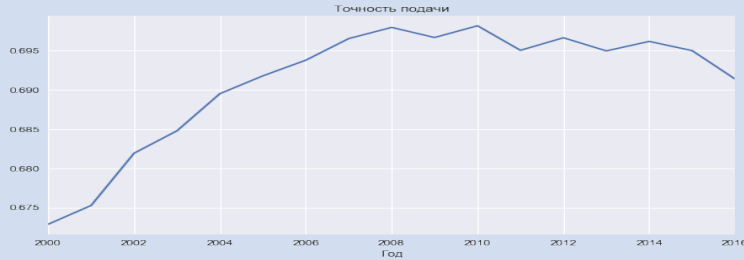
P1 rank	P1 age	P2 rank	P2 age	Year	P1 MPS aces	P2 MPS serve accuracy
67	27	2	29	2016	4.2	0.69
123	28	13	27	2016	5.8	0.67
29	26	37	25	2016	9.2	0.7
9	30	28.	19	2016	7.7	0.68
21	34	18	21	2016	10.2	0.71

MPS - Mean Previous Stats (стат. показатель игрока в среднем за карьеру)

Ставки - раунд
Наибольший разброс - в первых трёх раундах



Динамика средней точности подачи в матче за год
Доля точных подач в общем числе подач за игру падает последние 3 года



Самые важные признаки для предсказания: XGBoost

- Точность 2-ой подачи игрока 2
- Среднее число сыгранных очков на приёме 2-ой подачи за матч игрока 1
- Среднее число сыгранных очков на приёме 1-ой подачи за матч игрока 2
- Точность подачи игрока 2
- Среднее число выигранных очков за матч игрока 2