

В этом проекте предлагается марковская модель прогнозирования, которая предсказывает вероятности победы в теннисном матче для обоих игроков.

Для этого проекта я использую файлы:

```
atp_matches_qual_chall_2022.csv
atp_matches_qual_chall_2023.csv
atp_matches_qual_chall_2024.csv
atp_matches_2022.csv
atp_matches_2023.csv
atp_matches_2024.csv
```

Все датасеты были взяты из https://github.com/JeffSackmann/tennis_atp. Эти датасеты содержат данные за матчи 2022, 2023, 2024 годов из турниров АТР, челленджеров, квалификаций (одиночный разряд). Файл https://github.com/JeffSackmann/tennis_atp/blob/master/matches_data_dictionary.txt содержит объяснение параметров датасетов. Также я использую файл odds.xlsx, он имеет ту же структуру, но 2 дополнительных столбца в конце, которые содержат коэффициенты букмекерской конторы (bet365) на линии открытия. Я использую этот файл для оценки моделей.

Объяснение кода:

- Эти функции рассчитывают статистику игроков и вероятность выигрыша очка:

```
combined_prob_calc
prob_calc
players_statistics
```

Идея описана в разделе 2.3.3 файла.madura.pdf.

- Эти функции вычисляют вероятности выигрыша гейма, тай-брейка, сета, матча:

```
game_prob_calc
tiebreaker_prob_calc
set_prob_calc
match_prob_calc
```

Идеи описаны в разделе 2.3.1 файла a.madurska.pdf.

Данная функция используется для метода общих соперников:

```
match_odds
```

Идеи описаны в разделе 3.1 файла a.madurska.pdf.

После этого, используя прогнозируемые коэффициенты, мы моделируем ситуацию, в которой мы ставим 1 доллар на каждый матч, где какой-либо прогнозируемый коэффициент ниже соответствующего коэффициента, предлагаемого bet365.

Наилучший результат дает базовая марковская модель, которая учитывает тип поверхности (#method2 в коде) и имеет ROI= 0,004

Базовая модель, которая не учитывает тип покрытия (#method1), имеет точность = 0,597 и ROI = -0,08. Точность букмекерской конторы для этих данных составляет 0,592.