For this project I use files:

```
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
```

All datasets have been taken from https://github.com/JeffSackmann/tennis_atp repository. These datasets have data for 2022, 2023, 2024 matches from ATP, challengers, qualifications (solo).
File https://github.com/JeffSackmann/tennis_atp/blob/master/matches_data_dictionary.txt has explanation of features of datasets.
Also I use file odds.xlsx, it has the same structure, but 2 additional columns in the end which contain bookmaker's odds (bet365) on opening line for . I use this file for evaluating the model.
All approaches are described in a.madurska.pdf, u can find it in current repository.

Code explanation:

- These functions calculate players statistics and probability of winnin a point:

```
combined_prob_calc
prob_calc
players_statistics
```

Idea is described in section 2.3.3 of a.madurska.pdf file.

- These functions calculate probabilities to win game, tiebreak, set, match:

```
game_prob_calc
tiebreaker_prob_calc
set_prob_calc
match_prob_calc
```

Ideas are described in sections 2.3.1 of a.madurska.pdf file.

This function is used for common opponents method:

```
match_odds
```

This method is described in section 3.1. of a.madurska.pdf file.

Best result is given from basic markov model which take into account type of surface(#method2 in code) and it has ROI= 0.004

Basic model which doesn't take into account type of surface(#method1) has accuracy = 0.597 and ROI = -0.08. Bookmakers accuracy for this data is 0.592.