**Features** **engineering**

**Features based on team’s detailed performance in previous matches:**

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Description** | **Notes** |
| Points | sum of all points scored in a last game |  |
| Serve\_number | number of serves performed in a last game |  |
| Serve\_errors | number of service errors in a last game |  |
| Serve\_aces | number of service aces in a last game |  |
| Serve\_%\_effic | service efficiency in a last game | (number of service aces - number of service errors)/number of all serves performed |
| Rec\_number | number of serve receptions in a last game performed by a given team |  |
| Rec\_errors | number of receiving errors in a last game |  |
| Rec\_%\_pos | percentage of positive receptions in a last game | (number of positive receptions + number of perfect receptions)/number of all receptions |
| Rec\_%\_perf | percentage of perfect receptions in a last game | number of perfect receptions/number of all receptions in a last game |
| Atack\_number | number of attacks performed in a last game |  |
| Atack\_error | number of attack errors performed in a last game |  |
| Atack\_blocked | number of attacks that were blocked in a last game |  |
| Atack\_%\_eff | attack efficiency in a last game | number of successful attacks /number of all attacks performed |
| Block\_points | number of blocks performed in a last game |  |
| Points\_net | difference in number of points scored by two teams in a last game | for example, value of 40 for home team means, that home team in the last match scored 40 points more than their opponent at the time |

**Features based on players characteristics:**

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Description** | **Notes** |
| Height\_avg | average height of all players in a given team |  |
| Weight\_avg | average weight of all players in a given team |  |
| Range\_avg | average range of all players in a given team |  |
| Opposite\_Height | average height of players playing as an opposite hitter in a given team |  |
| Libero\_Height | average height of players playing as a libero in a given team |  |
| Outside\_Height | average height of players playing as an outside hitter in a given team |  |
| Setter\_Height | average height of players playing as a setter in a given team |  |
| Middle\_Height | average height of players playing as a middle blocker in a given team |  |
| Opposite\_Weight | average weight of players playing as an opposite hitter in a given team |  |
| Libero\_Weight | average weight of players playing as a libero in a given team |  |
| Outside\_Weight | average weight of players playing as an outside hitter in a given team |  |
| Setter\_Weight | average weight of players playing as a setter in a given team |  |
| Middle\_Weight | average weight of players playing as a middle blocker in a given team |  |
| Opposite\_Range | average range of players playing as an opposite hitter in a given team |  |
| Libero\_Range | average range of players playing as a libero in a given team |  |
| Outside\_Range | average range of players playing as an outside hitter in a given team |  |
| Setter\_Range | average range of players playing as a setter in a given team |  |
| Middle\_Range | average range of players playing as a middle blocker in a given team |  |

**Features based on team’s overall performance in previous matches:**

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Description** | **Notes** |
| Points\_all | points scored in all matches preceding the current game. | in PlusLiga team might obtain: 3 points for winning 3:0 or 3:1, 2 points for winning 3:2, 1 point for loosing 2:3 and 0 points for losing 1:3 or 0:3 |
| Matches\_won | number of matches won before the current game |  |
| Matches\_lost | number of matches lost before the current game |  |
| Sets\_won | number of sets won before the current game |  |
| Sets\_lost | number of sets lost before the current game |  |
| Points\_won | number of in-game points won before the current game | these are points scored during each game, not the ‘big points’ used for ranking teams |
| Points\_lost | number of in-game points lost before the current game | these are points lost during each game, not the ‘big points’ used for ranking teams |

**Additional variable:**

|  |  |  |
| --- | --- | --- |
| %\_points\_poss | percentage of all league points possible to obtain in all matches preceding the current game scored by given team | for example, if home team has 1 point after 3 rounds, these variable is set to 0.11 as home team obtained 1 out of 9 points possible |

**Outcome variable:**

Results\_home – 1 if home team won, 0 otherwise

**Basic info:**

Variables from the first category were calculated for three time periods, depending on number of matches taken into account: only last match, average from three last matches, and average from last 5 matches.

Variables from the second category were calculated for each team depending on the roster in particular season.

Variables from the third category were, similarly to first category, calculated for three time periods, depending on number of matches taken into account: only last match, average from three last matches, and average from last 5 matches.

Additional variable was not averaged across any periods, and reflects the share of all points possible to obtain in all matches up to the current game scored by analysed team.

All of these variables, whether averaged or not, were calculated for both home and away team, as well as difference between home and away team was considered.

In the end, the final dataset contains 1809 matches for seasons between 2010/2011 to 2021/2022, with 276 features each.