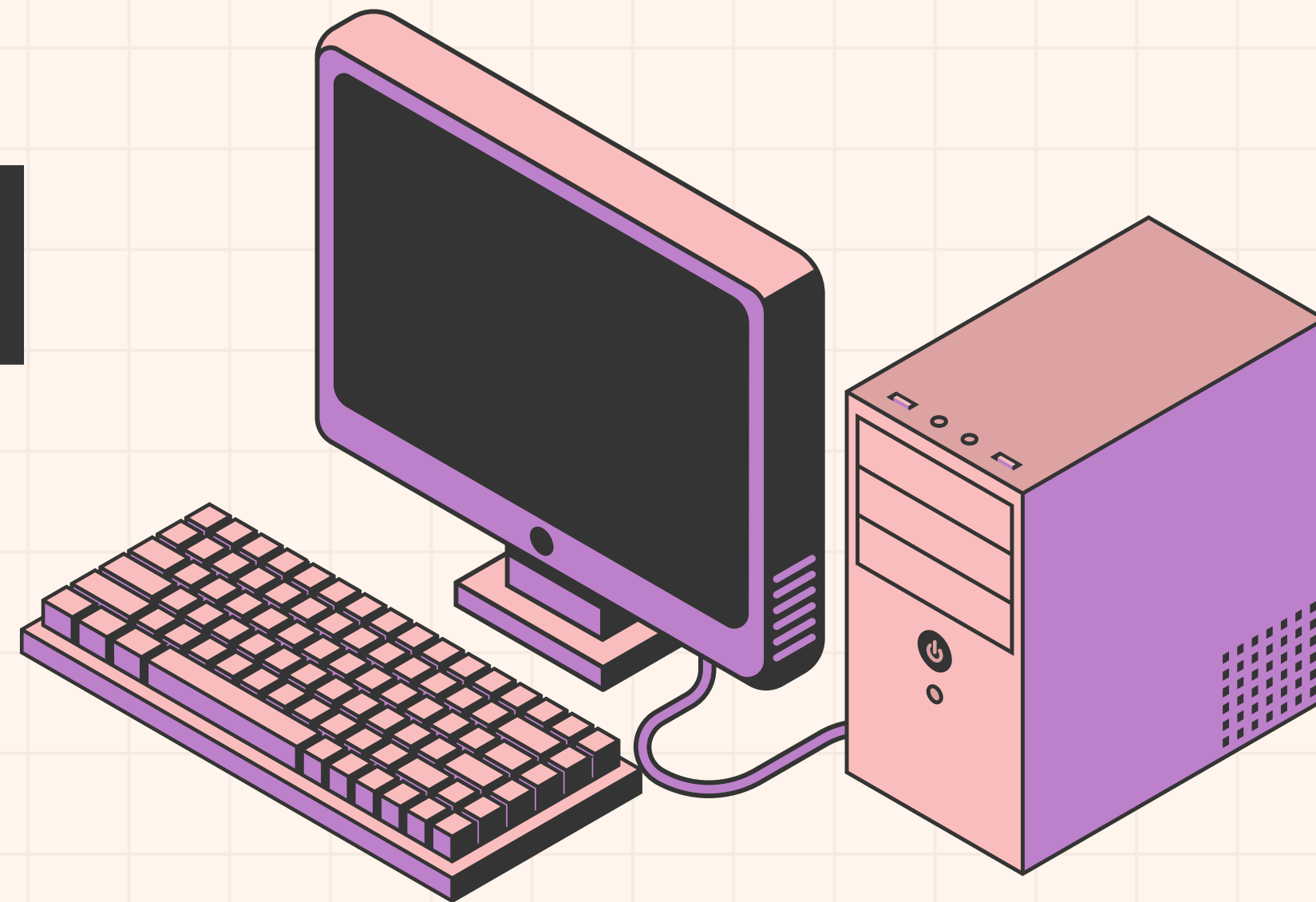


PREDICTION IN SPORTS

Can we predict sports results? Let's explore two methods and check if predictions can be trusted. In this presentation, I'll show two different approaches to predicting football results and discuss their reliability.



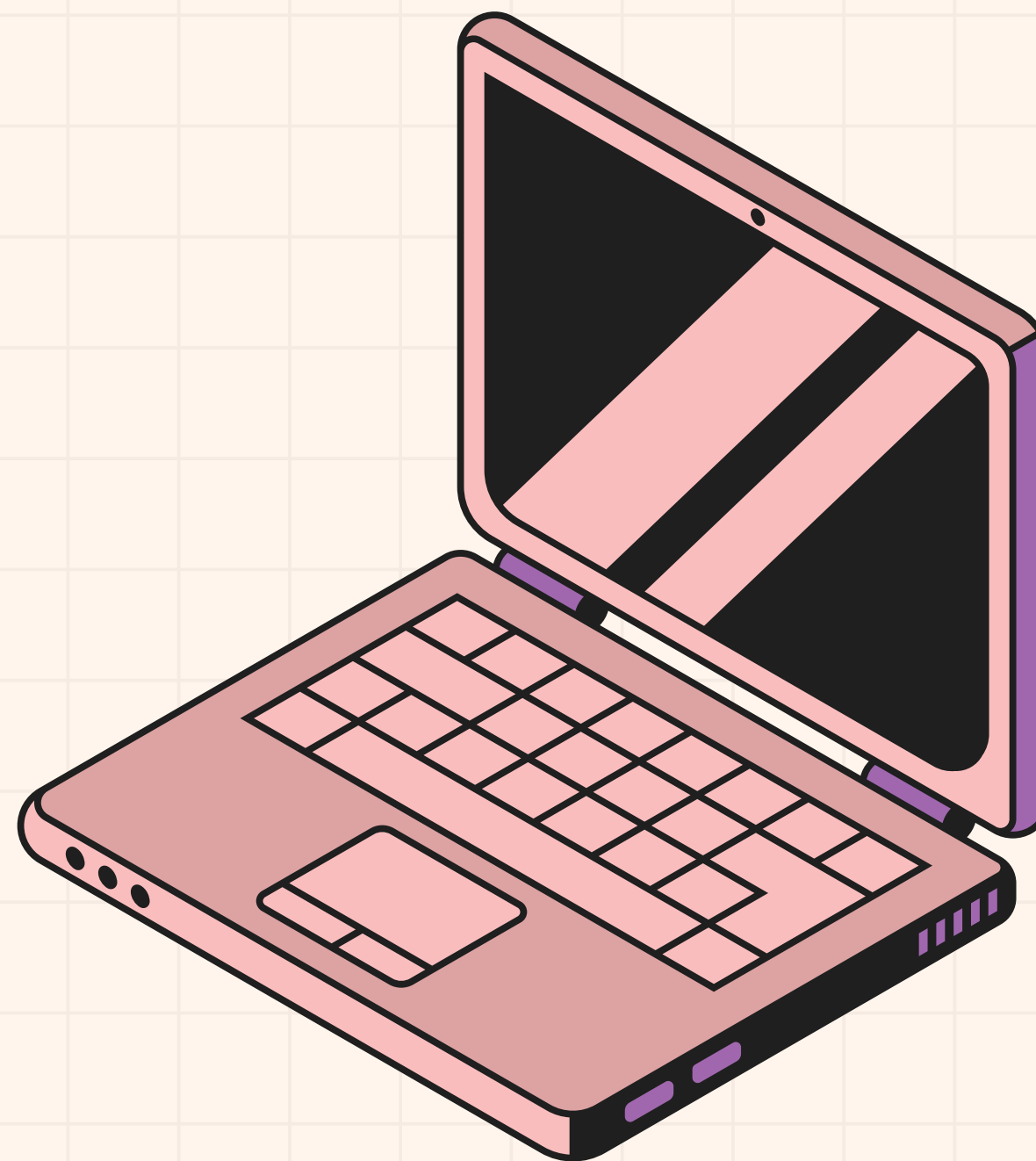
INTRODUCTION

Sports are full of surprises. Predicting results is difficult because of the human factor, unexpected events, and the nature of the game itself.

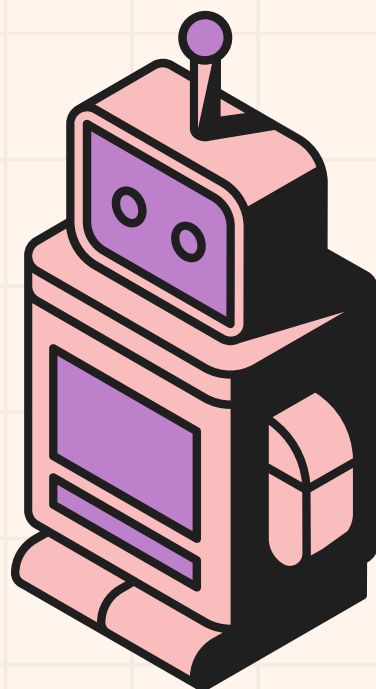
In this presentation, I will test two prediction methods — one based on statistics and another using machine learning — to check if it is possible to predict the final results of a football season.

The main goal is to answer:

Are sports predictable, or are predictions always uncertain?

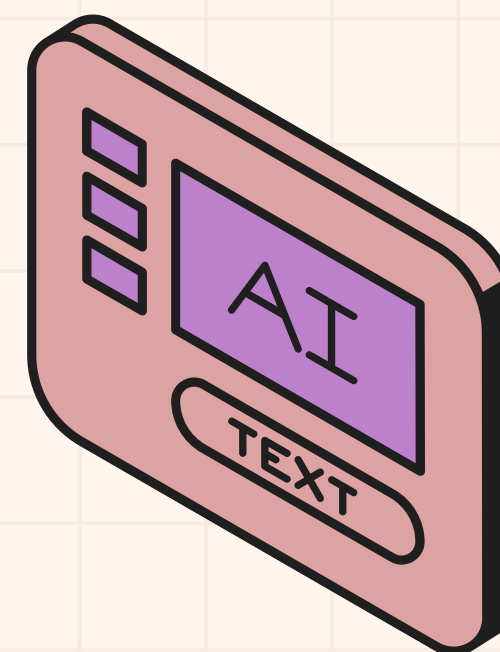


PREDUCTION METHODS OVERVIEW



STATISTICAL METHOD (ELO + PROBABILITY)

Based on historical team performance and ELO ratings, this method calculates the probability of each match outcome and simulates the entire season.



MACHINE LEARNING METHOD

A model trained on past seasons' data predicts the average number of points each team will get in the upcoming season.

STATISTICAL METHOD ELO & PROBABILITY

This method uses historical ELO ratings to calculate the probability of each match result.

Based on these probabilities, two types of results are generated:

- Expected Points — the average points a team should get, calculated from probabilities.
- Simulated Points — random points generated from the same probabilities, representing one possible scenario of the season.*

This approach shows how likely each outcome is, but not the exact final result.

```
# Funkcja do symulacji meczów z losowaniem wyniku
def simulate_season(df): 3 usages
    Premier_League_Season = generate_matches(df)
    expected_points = {club: 0 for club in df["Club"].values} # Słownik z początkowymi punktami 0 dla w
    actual_points = {club: 0 for club in df["Club"].values} # Rzeczywiste punkty po losowaniu wyników

    # Symulacja meczów
    for match in Premier_League_Season:
        home_team = match[0]
        away_team = match[1]
        home_game = match[2] == 'home'

        # Pobieranie rankingów ELO dla drużyn
        elo_home = df[df["Club"] == home_team]["Elo"].values[0]
        elo_away = df[df["Club"] == away_team]["Elo"].values[0]

        # Obliczanie oczekiwanych punktów dla drużyn
        p_win_home, p_draw, p_win_away, p_exp_home, p_exp_away = final_score_probability_logistic(elo_home, elo_away)

        # Losowanie wyniku na podstawie prawdopodobieństw
        outcome = random.choices(
            population=['home_win', 'draw', 'away_win'],
            weights=[p_win_home, p_draw, p_win_away],
            k=1
        )[0]

        # Przyznawanie punktów na podstawie wyniku meczu
        if outcome == 'home_win':
            actual_points[home_team] += 3
        elif outcome == 'away_win':
            actual_points[away_team] += 3
```

STATISTICAL METHOD RESULTS

-----Premier League-----					
	Club	Actual Points	Rank	Expected Points	Difference
0	Man City	78	1	76.355615	1.644385
1	Liverpool	73	2	67.645085	5.354915
2	Man United	68	3	63.558072	4.441928
3	Tottenham	67	4	59.386378	7.613622
4	Arsenal	64	5	69.685536	-5.685536
5	Chelsea	62	6	61.065446	0.934554
6	Crystal Palace	58	7	44.997469	13.002531
7	Everton	58	8	39.300318	18.699682
8	Brentford	53	9	53.163620	-0.163620
9	Newcastle	49	10	58.137762	-9.137762
10	Aston Villa	49	11	51.057780	-2.057780
11	West Ham	49	12	49.015515	-0.015515
12	Brighton	48	13	56.210599	-8.210599
13	Leeds	48	14	42.218570	5.781430
14	Wolves	47	15	41.117324	5.882676
15	Burnley	44	16	46.240575	-2.240575
16	Fulham	42	17	44.010574	-2.010574
17	Bournemouth	41	18	35.780241	5.219759
18	Southampton	38	19	36.959734	1.040266
	Leicester	24	20	47.905534	-23.905534

Klub	RM	W	R	P	BZ	BS	RB	Pkt	Ostatnie 5
1 Man City	38	28	5	5	94	33	61	89	✓✓✓✓✗
2 Arsenal	38	26	6	6	88	43	45	84	✓✓✗✗✓
3 Man Utd	38	23	6	9	58	43	15	75	✗✓✓✓✓
4 Newcastle	38	19	14	5	68	33	35	71	✗-✓- -
5 Liverpool	38	19	10	9	75	47	28	67	✓✓✓- -
6 Brighton	38	18	8	12	72	53	19	62	✓✗✓-✗
7 Aston Villa	38	18	7	13	51	46	5	61	✗✗✓-✓
8 Tottenham	38	18	6	14	70	63	7	60	✗✓✗✗✓
9 Brentford	38	15	14	9	58	46	12	59	✓✗✓✓✓
10 Fulham	38	15	7	16	55	53	2	52	✗✓✓-✗
11 Crystal Palace	38	11	12	15	40	49	-9	45	✓✗✓- -
12 Chelsea	38	11	11	16	38	47	-9	44	✓-✗✗-
13 Wolverhampton	38	11	8	19	31	58	-27	41	✗✓✗-✗
14 West Ham	38	11	7	20	42	55	-13	40	✗✓✗✓✗
15 Bournemouth	38	11	6	21	37	71	-34	39	✓✗✗✗✗
16 Nottingham F...	38	9	11	18	38	68	-30	38	✗✓-✓-
17 Everton	38	8	12	18	34	57	-23	36	-✓✗-✓
18 Leicester	38	9	7	22	51	68	-17	34	-✗✗-✓
19 Leeds	38	7	10	21	48	78	-30	31	✗✗-✗✗
20 Southampton	38	6	7	25	36	73	-37	25	✗✗✗✗-

MACHINE LEARNING METHOD PREDICTING POINTS

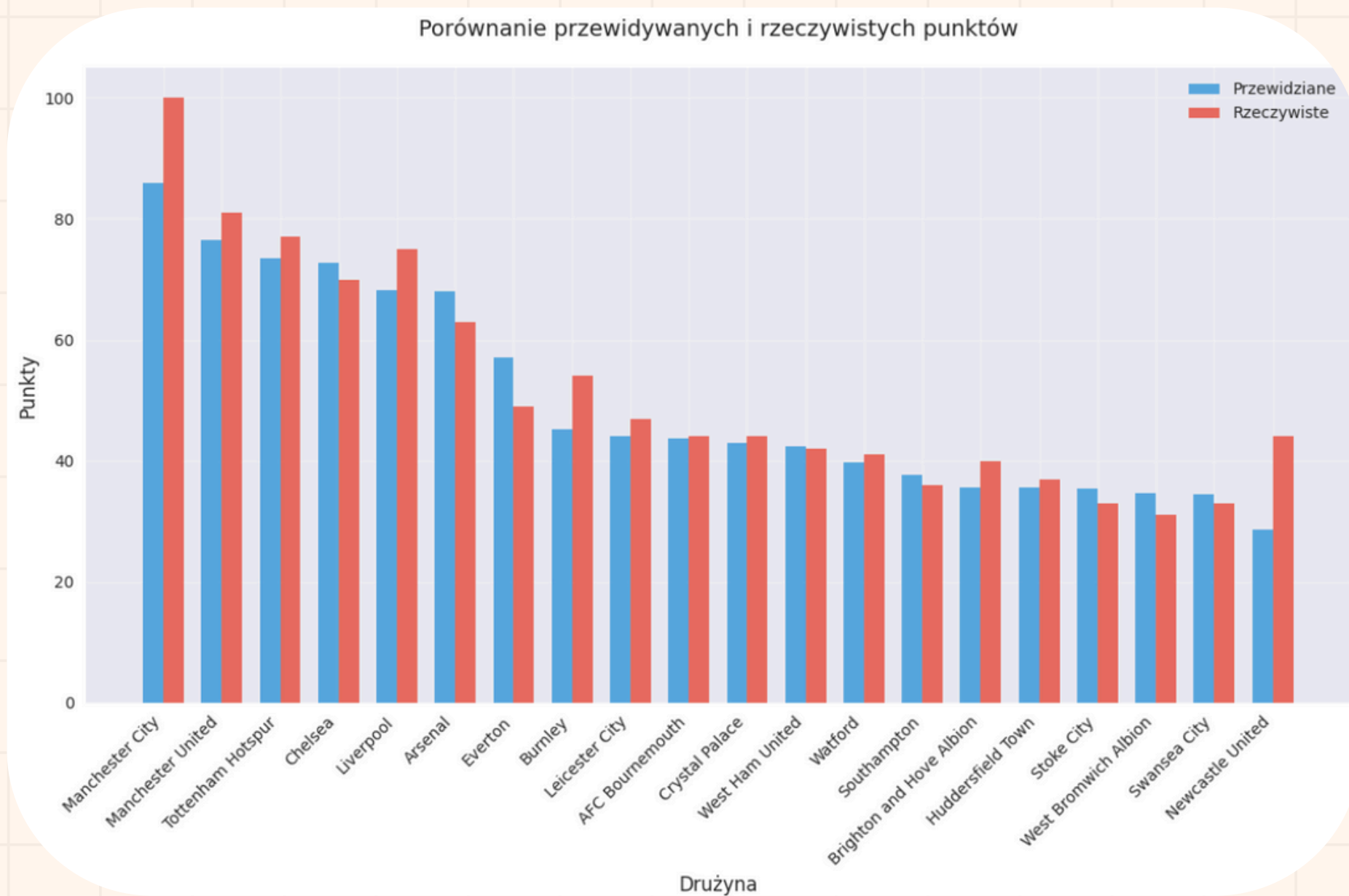
This method uses historical season data to train a machine learning model.

The model predicts the expected average number of points a team will collect in the upcoming season, based on patterns from past results.

Unlike the statistical method, here the prediction is not based on probabilities, but on learned trends in the data.

```
15 # Załaduj dane
16 df = pd.read_csv("results.csv")
17
18 # Kopia danych
19 df = df.copy()
20
21 # Przekształcenie wyniku meczu na punkty
22 df['home_points'] = df['result'].map({'H': 3, 'D': 1, 'A': 0})
23 df['away_points'] = df['result'].map({'H': 0, 'D': 1, 'A': 3})
24
25 # Sumowanie punktów dla gospodarzy i gości
26 home_points = df.groupby(['season', 'home_team'])['home_points'].sum().reset_index()
27 home_points.rename(columns={'home_team': 'team', 'home_points': 'points'}, inplace=True)
28
29 away_points = df.groupby(['season', 'away_team'])['away_points'].sum().reset_index()
30 away_points.rename(columns={'away_team': 'team', 'away_points': 'points'}, inplace=True)
31
32 # Łączymy punkty
33 all_points = pd.concat([home_points, away_points])
34 season_team_points = all_points.groupby(['season', 'team'])['points'].sum().reset_index()
35
36 # Lista wszystkich drużyn
37 teams = sorted(set(df['home_team'].unique().tolist() + df['away_team'].unique().tolist()))
38
39 # Lista wszystkich sezonów
40 seasons = sorted(df['season'].unique())
41
42 # Tworzymy wszystkie kombinacje drużyn i sezonów
43 full_index = pd.MultiIndex.from_product([seasons, teams], names=['season', 'team'])
44 full_table = pd.DataFrame(index=full_index).reset_index()
```

MACHINE LEARNING RESULTS



		RM	W	R	P	BZ	BS	RB	Pkt	Ostatnie 5
1	Man City	38	32	4	2	106	27	79	100	✓✓✓✓✓
2	Man Utd	38	25	6	7	68	28	40	81	✓✓✓✓✓
3	Tottenham	38	23	8	7	74	36	38	77	✓✓✓✓✓
4	Liverpool	38	21	12	5	84	38	46	75	✓✓✓✓✓
5	Chelsea	38	21	7	10	62	38	24	70	✓✓✓✓✓
6	Arsenal	38	19	6	13	74	51	23	63	✓✓✓✓✓
7	Burnley	38	14	12	12	36	39	-3	54	✓✓✓✓✓
8	Everton	38	13	10	15	44	58	-14	49	✓✓✓✓✓
9	Leicester	38	12	11	15	56	60	-4	47	✓✓✓✓✓
10	Newcastle	38	12	8	18	39	47	-8	44	✓✓✓✓✓
11	Crystal Palace	38	11	11	16	45	55	-10	44	✓✓✓✓✓
12	Bournemouth	38	11	11	16	45	61	-16	44	✓✓✓✓✓
13	West Ham	38	10	12	16	48	68	-20	42	✓✓✓✓✓
14	Watford	38	11	8	19	44	64	-20	41	✓✓✓✓✓
15	Brighton	38	9	13	16	34	54	-20	40	✓✓✓✓✓
16	Huddersfield	38	9	10	19	28	58	-30	37	✓✓✓✓✓
17	Southampton	38	7	15	16	37	56	-19	36	✓✓✓✓✓
18	Swansea City	38	8	9	21	28	56	-28	33	✓✓✓✓✓
19	Stoke City	38	7	12	19	35	68	-33	33	✓✓✓✓✓
	West Brom	38	6	13	19	31	56	-25	31	✓✓✓✓✓

CONCLUSIONS PREDICTION LIMITS



MIDDLE OF THE TABLE IS PREDICTABLE

Predictions for mid-table teams are usually close to reality, but extreme cases — top and bottom teams — are much harder to predict.



TOO MANY HUMAN FACTORS

Sports results depend on injuries, motivation, mistakes, and emotions. Numbers alone can't capture all of this.



PREDICTIONS ARE NOT REALITY

Even advanced models can only show likely scenarios, not exact outcomes. The real world is always more random.

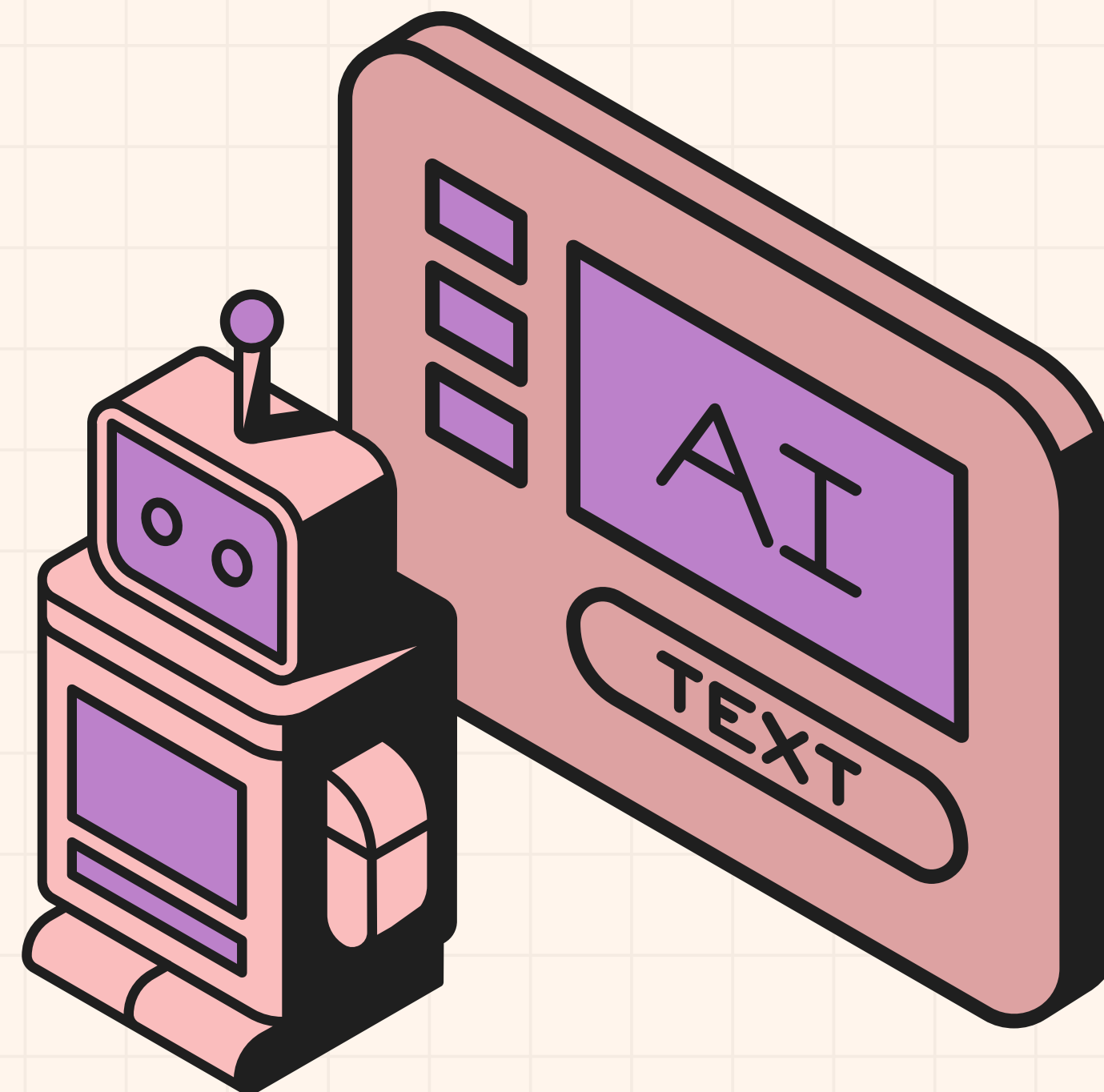
FINAL CONCLUSION

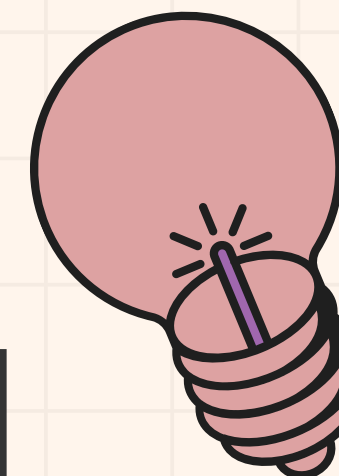
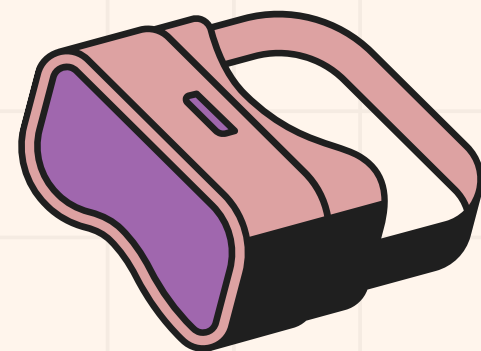
After testing both prediction methods, the hypothesis has been confirmed.

It is not possible to predict sports results with full accuracy. While mathematical models and machine learning can show us tendencies and give a general view of possible outcomes, the real world is far more complex.

Human factors, emotions, unexpected events and pure luck make football — and sports in general — too unpredictable for any model to be fully reliable.

Predictions are useful for understanding the game, but they will always remain only an approximation, not a guarantee.





THANK YOU

FEEL FREE TO ASK QUESTIONS

