Football Predictions Program

Data Science Final Project

Ido Ashkenazi
Daniel Bugdari

The stages of ML

Problem Definition

Defining and understanding the problem we are looking to solve

Data Collection

Web crawling for collecting and obtaining the data

Data Preparation

Data validation, cleansing, filtering and formatting

Data Visualization

Visualizing and understanding the data

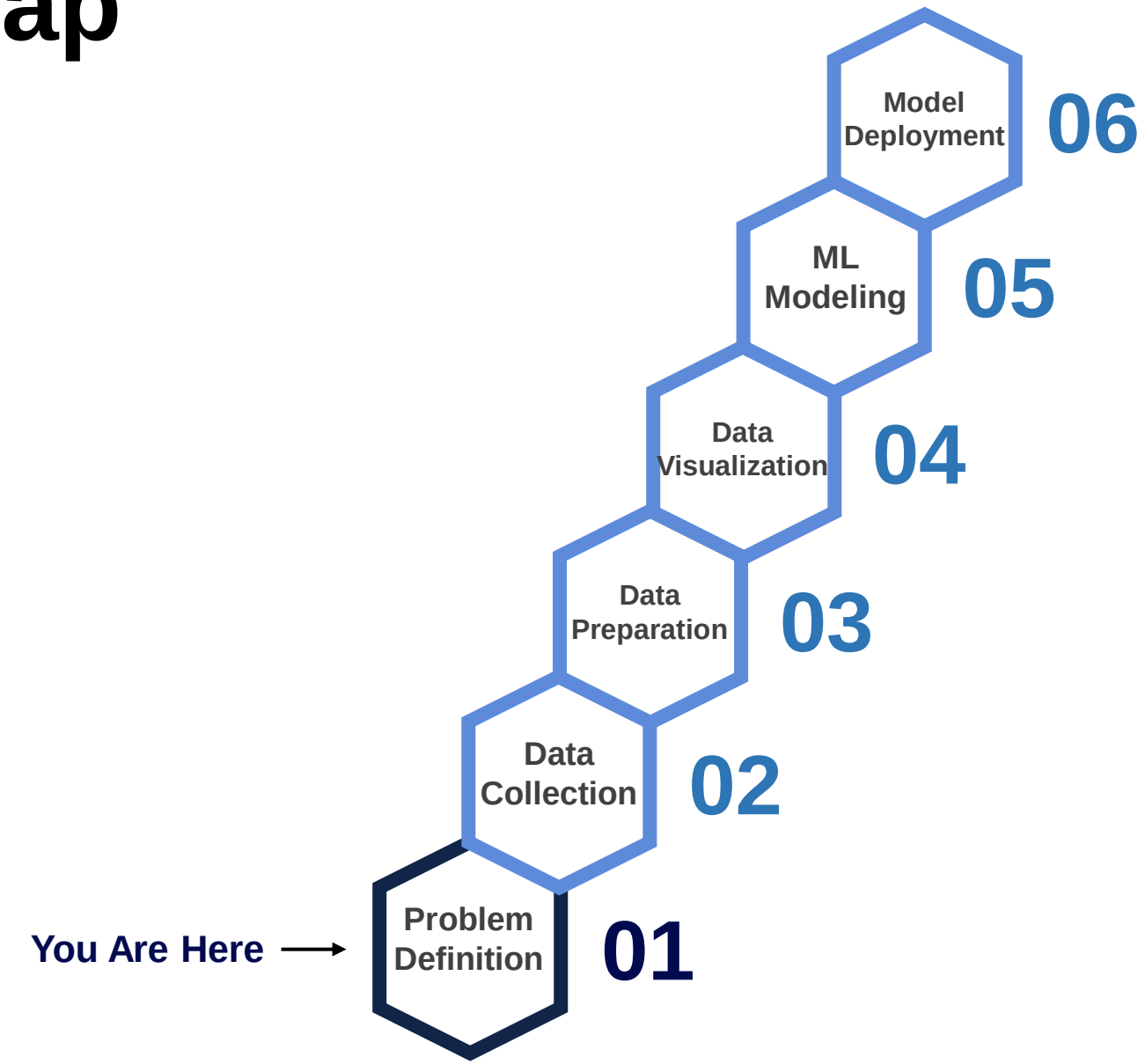
ML Modeling & Feature Engineering

Finding patterns in data, ML algorithms for data analysis and predictive modeling based on the provided data

Model Deployment

Final predict model, making predictions using the model based on the relevant data

Stages Map


Model Deployment
06
ML Modeling
05
Data Visualization
04
Data Preparation
03
Data Collection
02
Problem Definition
01
You Are Here

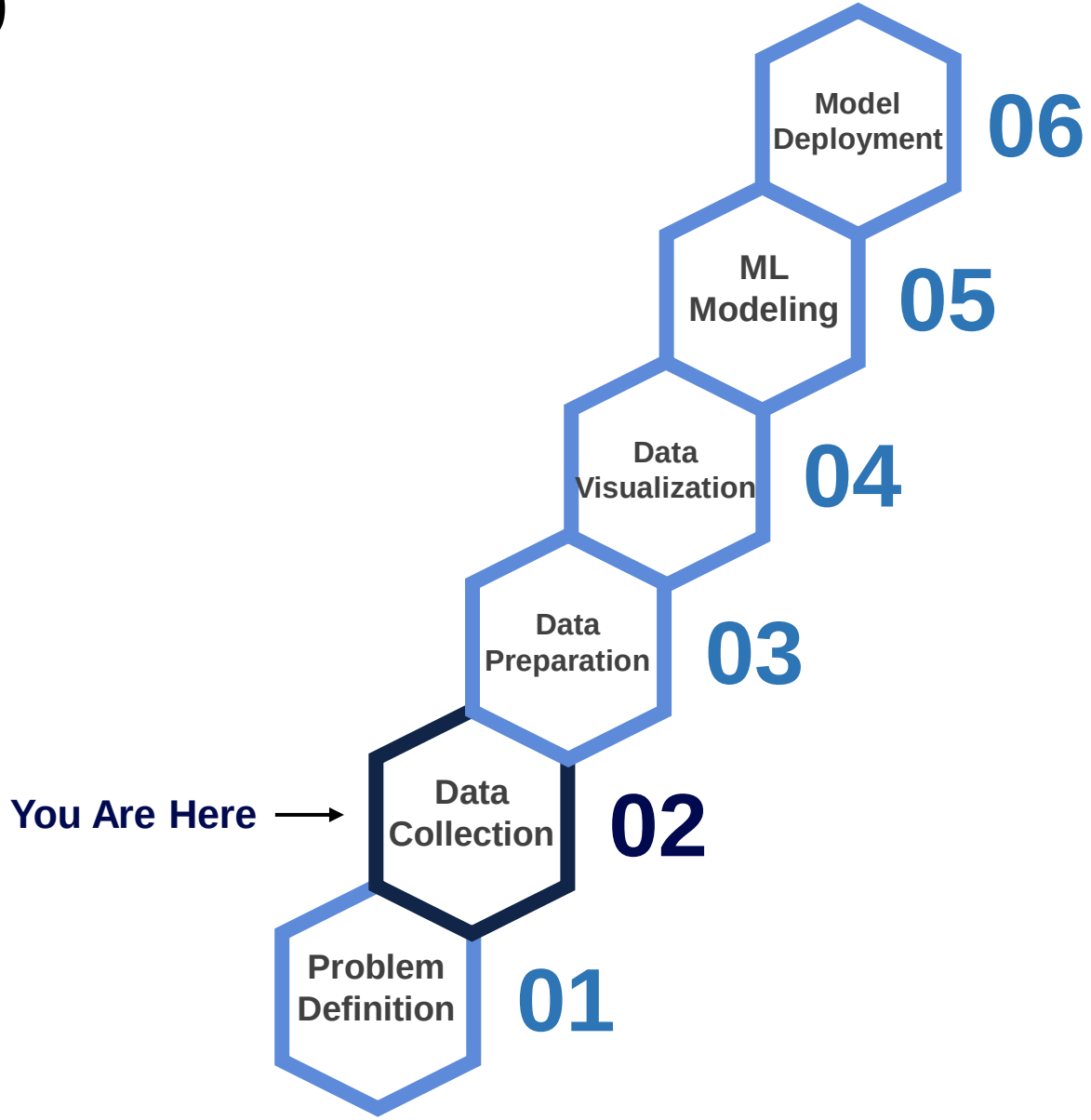
Problem Definition

Let’s start with a little practice.
Every month, your workplace chooses a bonding social game to play and the chosen game for the current month is a football matches predictions.
In this game, each week, one football match will be chosen and the employees will be asked to predict how it will end.
Employees that correctly predict will receive a bonus of 500ILS.
This week's match is between Maccabi Haifa and Hapoel Hadera.

Now I'd like to ask you: what is your prediction?
Almost certainly you will bet on Maccabi Haifa to win.

But why? What motivates you to bet on Maccabi Haifa's victory?
Is it possible to build an automatic machine that predicts match results based on the same features you used?

Stages Map


Model Deployment
06
ML Modeling
05
Data Visualization
04
Data Preparation
03
You Are Here
Data Collection
02
Problem Definition
01

Data Collection – Web Crawling

Data Source:

www.transfermarkt.com

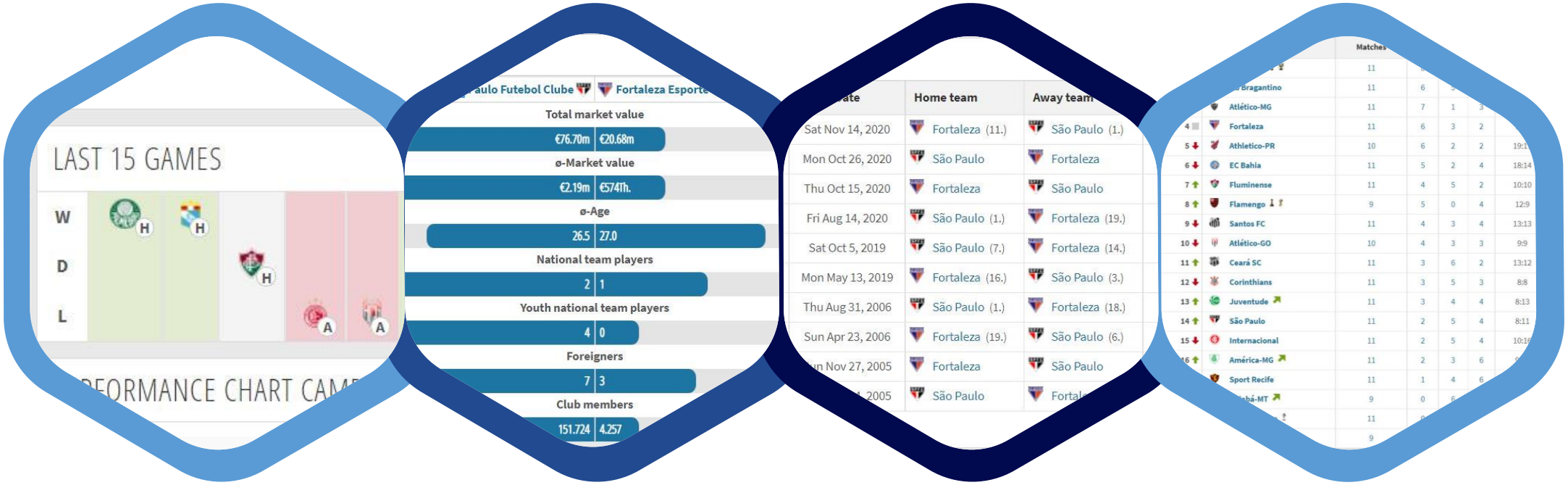
	Competition	Status	Home Team Name	Away Team Name	link
0	Campeonato Brasileiro Série A	matchday 12	São Paulo	Fortaleza	https://www.transfermarkt.com/spielbericht/index/spielbericht/3583831
1	Campeonato Brasileiro Série A	matchday 12	Ceará SC	Athletico-PR	https://www.transfermarkt.com/spielbericht/index/spielbericht/3583834
2	Campeonato Brasileiro Série B	matchday 12	Confiança-SE	Guarani	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584117
3	Campeonato Brasileiro Série B	matchday 12	Sampaio Corrêa	Coritiba FC	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584118
4	Campeonato Brasileiro Série B	matchday 12	Brasil-RS	EC Vitória	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584113
5	Campeonato Brasileiro Série B	matchday 12	Goiás EC	Londrina-PR	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584110
6	Campeonato Brasileiro Série B	matchday 12	Cruzeiro	Avaí FC	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584112
7	Campeonato Brasileiro Série B	matchday 12	Ponte Preta	Remo	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584111

We used ‘BeautifulSoup’ Python library for pulling data out of the HTML pages and ‘Pandas’ library for data analysis and manipulation.

```
def get_url_content(url):
    url = "https://www.transfermarkt.com" + url
    response = requests.get(url, headers={'User-Agent': 'Custom'})
    return BeautifulSoup(response.content, 'html.parser')
```

```
games_df = pd.DataFrame(games, columns = ['Competition', 'Status', 'Home Team Name', 'Away Team Name', 'link'])
games_df_styler = games_df.style.set_properties(**{'text-align': 'left'})
games_df_styler = games_df_styler.set_table_styles([dict(selector = 'th', props=[('text-align', 'left')])])

games_df.to_csv(os.path.join(r'Matchdays' , str(chosen_date)+'.csv'))

display(games_df_styler)
```

Data Collection – Web Crawling



Recent matches results

Getting the results of each team's recent matches

Players Strengths

Getting the market values of the teams

Bigger market value = better players

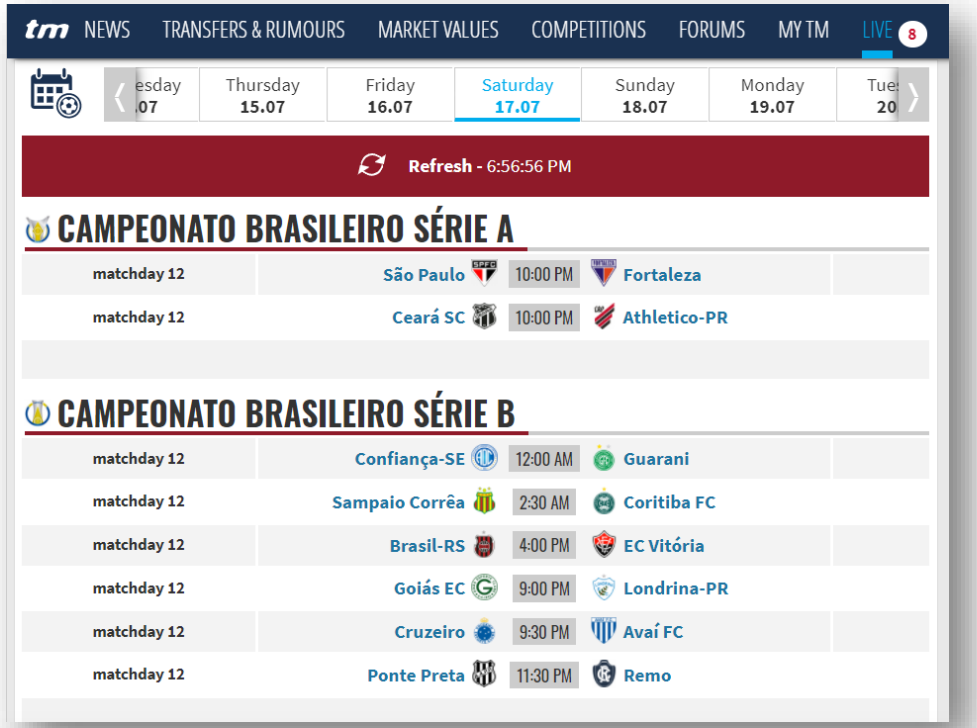
Past meetings results

Getting the results of their previous meetings

Positions

Getting the teams positions from the league table

Data Collection – Web Crawling



Getting the chosen date games from the website and exporting them to a panda's data frame. Exporting the data frame to a CSV file.

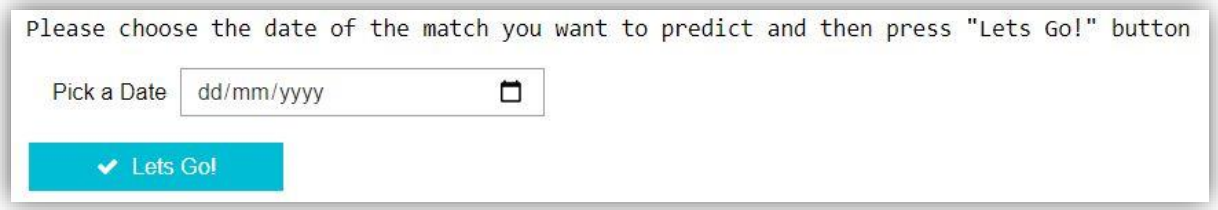
	F	E	D	C	B	A	
	link	Away Team Name	Home Team Name	Status	Competition		1
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3583831	Fortaleza	SÃ£o Paulo	matchday 12	Campeonato Brasileiro SÃ©rie A	0	2
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3583834	Athletico-PR	CearÃ¡ SC	matchday 12	Campeonato Brasileiro SÃ©rie A	1	3
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584117	Guarani	ConfianÃ§a-SE	matchday 12	Campeonato Brasileiro SÃ©rie B	2	4
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584118	Coritiba FC	Sampaio CorrÃªa	matchday 12	Campeonato Brasileiro SÃ©rie B	3	5
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584113	EC VitÃ³ria	Brasil-RS	matchday 12	Campeonato Brasileiro SÃ©rie B	4	6
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584110	Londrina-PR	GoiÁ¡s EC	matchday 12	Campeonato Brasileiro SÃ©rie B	5	7
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584112	AvaÃ- FC	Cruzeiro	matchday 12	Campeonato Brasileiro SÃ©rie B	6	8
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584111	Remo	Ponte Preta	matchday 12	Campeonato Brasileiro SÃ©rie B	7	9
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602794	Oeste FC-SP	EC SÃ£o JosÃ©	Group Stage	Campeonato Brasileiro SÃ©rie C	8	10
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602795	CriciÃºma EC	Mirassol-SP	Group Stage	Campeonato Brasileiro SÃ©rie C	9	11
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602789	Botafogo	Jacuipense-BA	Group Stage	Campeonato Brasileiro SÃ©rie C	10	12
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602790	Tombense	Santa Cruz	Group Stage	Campeonato Brasileiro SÃ©rie C	11	13
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602796	Botafogo-SP	Figueirense FC	Group Stage	Campeonato Brasileiro SÃ©rie C	12	14
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3589789	El Ahly	Kaizer Chiefs	Final	CAF-Champions League	13	15
	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602425	CS Mioveni	Gaz Metan	matchday 1	Liga 1	14	16

	Competition	Status	Home Team Name	Away Team Name	link
0	Campeonato Brasileiro Série A	matchday 12	São Paulo	Fortaleza	https://www.transfermarkt.com/spielbericht/index/spielbericht/3583831
1	Campeonato Brasileiro Série A	matchday 12	Ceará SC	Athletico-PR	https://www.transfermarkt.com/spielbericht/index/spielbericht/3583834
2	Campeonato Brasileiro Série B	matchday 12	Confiança-SE	Guarani	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584117
3	Campeonato Brasileiro Série B	matchday 12	Sampaio Corrêa	Coritiba FC	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584118
4	Campeonato Brasileiro Série B	matchday 12	Brasil-RS	EC Vitória	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584113
5	Campeonato Brasileiro Série B	matchday 12	Goiás EC	Londrina-PR	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584110
6	Campeonato Brasileiro Série B	matchday 12	Cruzeiro	Avaí FC	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584112
7	Campeonato Brasileiro Série B	matchday 12	Ponte Preta	Remo	https://www.transfermarkt.com/spielbericht/index/spielbericht/3584111
8	Campeonato Brasileiro Série C	Group Stage	EC São José	Oeste FC-SP	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602794
9	Campeonato Brasileiro Série C	Group Stage	Mirassol-SP	Criciúma EC	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602795
10	Campeonato Brasileiro Série C	Group Stage	Jacuipense-BA	Botafogo	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602789
11	Campeonato Brasileiro Série C	Group Stage	Santa Cruz	Tombense	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602790
12	Campeonato Brasileiro Série C	Group Stage	Figueirense FC	Botafogo-SP	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602796
13	CAF-Champions League	Final	Kaizer Chiefs	El Ahly	https://www.transfermarkt.com/spielbericht/index/spielbericht/3589789
14	Liga 1	matchday 1	Gaz Metan	CS Mioveni	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602425
15	Liga 1	matchday 1	Univ. Craiova	FC Arges	https://www.transfermarkt.com/spielbericht/index/spielbericht/3602426
16	Belgian Supercup	Final	Club Brugge	KRC Genk	https://www.transfermarkt.com/spielbericht/index/spielbericht/3604605
17	Super liga Srbije	matchday 1	FK TSC	Novi Pazar	https://www.transfermarkt.com/spielbericht/index/spielbericht/3598593
18	Super liga Srbije	matchday 1	Proleter	Partizan	https://www.transfermarkt.com/spielbericht/index/spielbericht/3598595
19	1.HNL	matchday 1	NK Lokomotiva	Hajduk Split	https://www.transfermarkt.com/spielbericht/index/spielbericht/3591218
20	1.HNL	matchday 1	HNK Rijeka	HNK Gorica	https://www.transfermarkt.com/spielbericht/index/spielbericht/3591215
21	Ykkönen	matchday 12	RoPS	JIPPO	https://www.transfermarkt.com/spielbericht/index/spielbericht/3512727
22	Ykkönen	matchday 12	Ekenäs IF	Mikkelin	https://www.transfermarkt.com/spielbericht/index/spielbericht/3512728
23	Ykkönen	matchday 12	Klubi 04	FF Jaro	https://www.transfermarkt.com/spielbericht/index/spielbericht/3512729
24	Veikkausliiga	matchday 12	AC Oulu	HJK Helsinki	https://www.transfermarkt.com/spielbericht/index/spielbericht/3511683
25	Prva Liga	matchday 1	NK Bravo	NK Radomlje	https://www.transfermarkt.com/spielbericht/index/spielbericht/3593718
26	Prva Liga	matchday 1	NK Aluminij	FC Koper	https://www.transfermarkt.com/spielbericht/index/spielbericht/3593719
27	Prva Liga	matchday 1	NS Mura	Tabor Sezana	https://www.transfermarkt.com/spielbericht/index/spielbericht/3593720

Data Collection – Date Picker

We used a “Date Picker” widget from ‘ipywidgets’ library for picking match date.

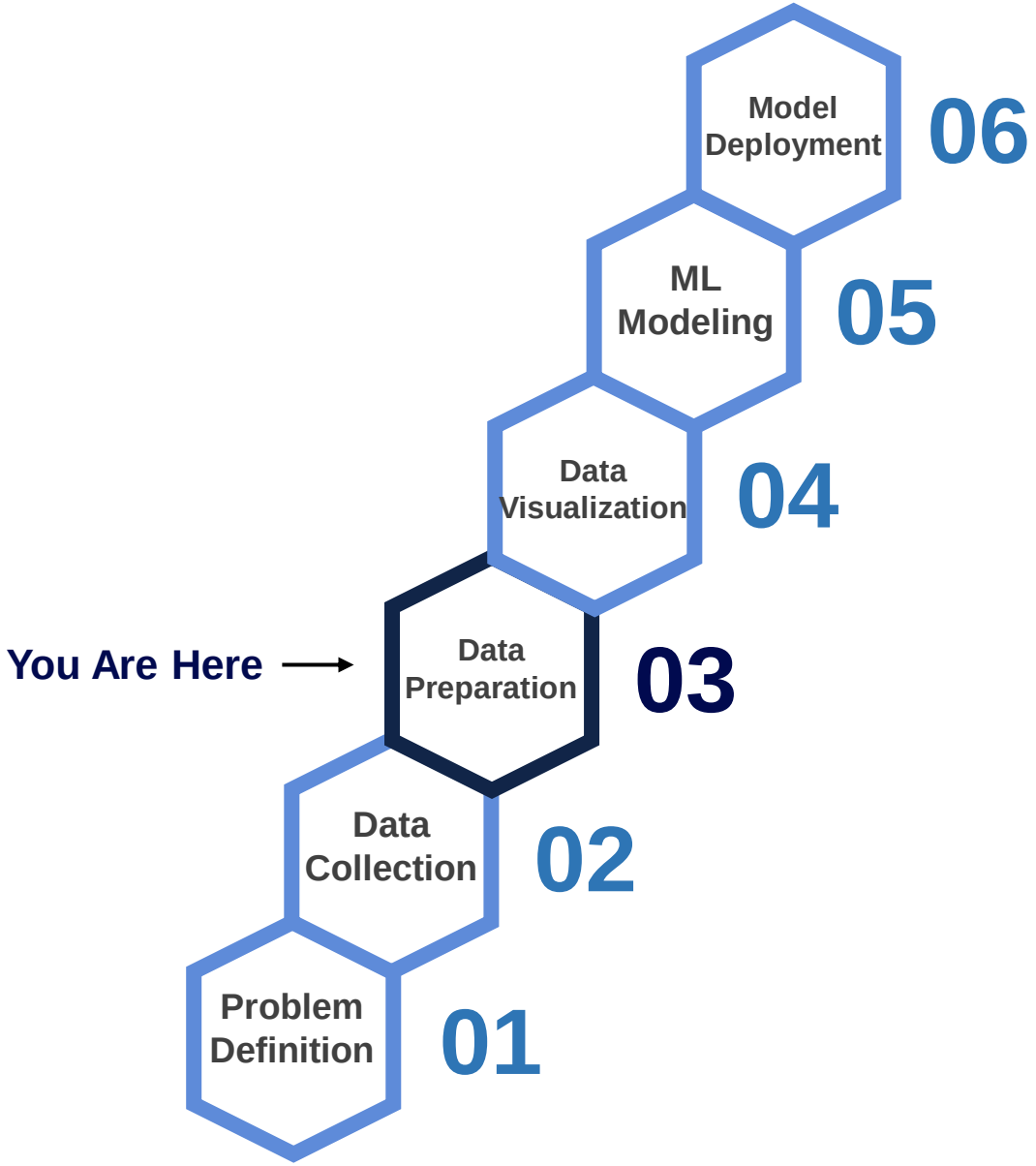
We used ‘ipython_blocking’ library to be able to use the data we received from the user's input into the date picker widget.




```
def games_date(a):
    global chosen_date
    chosen_date = str(date_picker.value)
    if chosen_date == 'None':
        print("Please pick a valid date")
    else:
        print('The chosen date is: ', chosen_date)
        button.disabled=True
        %store chosen_date
        %store


date_picker = widgets.DatePicker(
    description='Pick a Date',
    disabled=False
)

button = widgets.Button(
    description=' Lets Go!',
    disabled=False,
    button_style='info', # 'success', 'info', 'warning', 'danger' or ''
    tooltip='Click me',
    icon='check' # (FontAwesome names without the `fa-` prefix)
)

display(date_picker)
display(button)
button.on_click(games_date)
```

```
%blockrun button
```

Stages Map


Model Deployment
06
ML Modeling
05
Data Visualization
04
You Are Here
Data Preparation
03
Data Collection
02
Problem Definition
01

Data Preparation – formatting & cleansing

Formatting and cleansing the collected data so it can be used.
We used ‘re’ library for that mission.

For example,

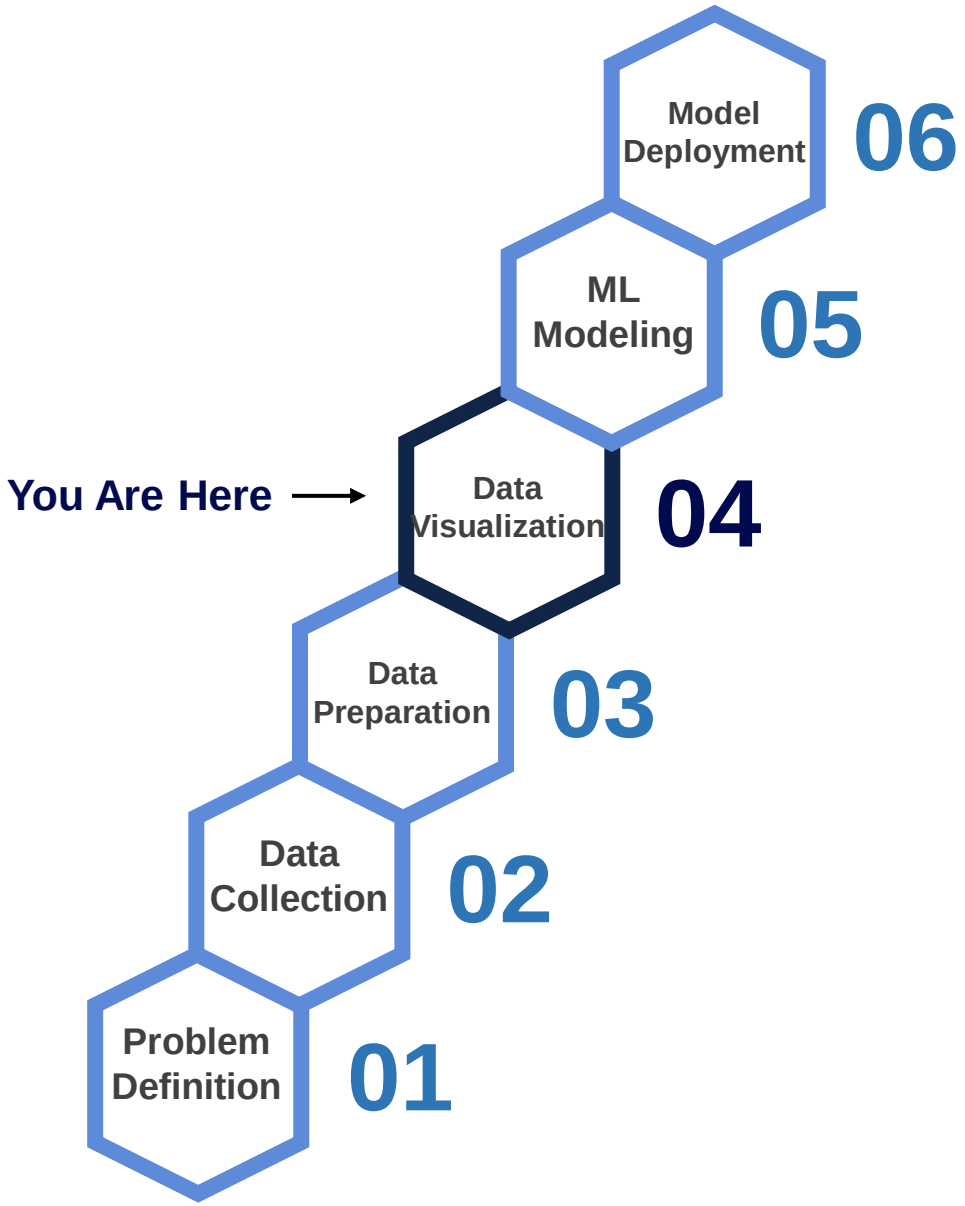
Converting the market values from 200m to 200000000

Converting the game result from 3:2 to 3 home team goals and 2 away team goals, what means that the home team won the match.

```
teams_market_value = []

temp = match_url_content.find("div", class_="box daten-und-fakten")
home_team_market_value = temp.find("td", class_= "daten-und-fakten-linker_balken").get_text()
away_team_market_value = temp.find("td", class_= "daten-und-fakten-rechter_balken").get_text()

print(home_team + ' Market Value: ' + home_team_market_value)
print(away_team + ' Market Value: ' + away_team_market_value)

final_home_value = float(re.search(r'\d+\.\d+', home_team_market_value).group())
final_away_value = float(re.search(r'\d+\.\d+', away_team_market_value).group())

if 'm' in home_team_market_value.lower():
    final_home_value *= 1000000
elif 'b' in home_team_market_value.lower():
    final_home_value *= 1000000000
final_home_value = int(final_home_value)

if 'm' in away_team_market_value.lower():
    final_away_value *= 1000000
    final_away_value = int(final_away_value)
elif 'b' in away_team_market_value.lower():
    final_away_value *= 1000000000
```

```
for game_result in each_row.findAll("td", class_= "zentriert hauptlink"):
    res = game_result.find("a", class_= "ergebnis-link").get_text(strip=True)
    res = res.partition(':')
    home_team_goals = int(re.search(r'\d+', res[0]).group())
    away_team_goals = int(re.search(r'\d+', res[2]).group())
    if home_team_goals > away_team_goals:
        home_team_points = 2
        df_target_rows.append(2)
    elif home_team_goals < away_team_goals:
        away_team_points = 2
        df_target_rows.append(0)
    else:
        home_team_points = 1
        away_team_points = 1
        df_target_rows.append(1)
    last_matches_between_teams_url += game_result.find("a", class_= "ergebnis-link").get('href')
```

Stages Map


Model Deployment
06
ML Modeling
05
You Are Here
Data Visualization
04
Data Preparation
03
Data Collection
02
Problem Definition
01

Data Visualization

São Paulo Last 15 Matches (Home Team)
Position: 14
Market Value:
€76.70m

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
W	H	H				H								A	H
D			H					H		H	A	A			
L				A	A		A		A				H		

Home W : 4
Home D : 3
Home L : 1
Total Home Games : 8

Fortaleza Last 15 Matches (Away Team)
Position: 4
Market Value:
€20.68m

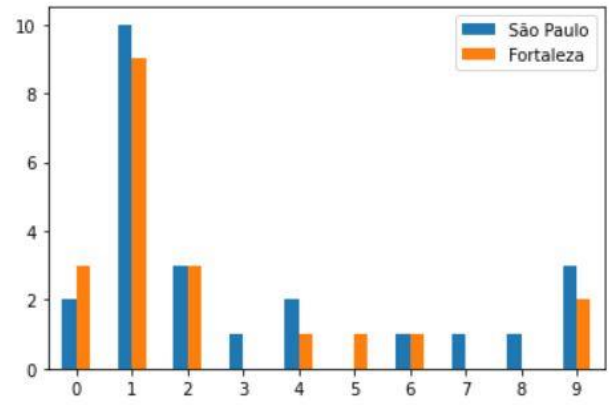
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
W	H		A		H	A	H					H		H	H
D		H		H				A	H		A				
L										A			A		

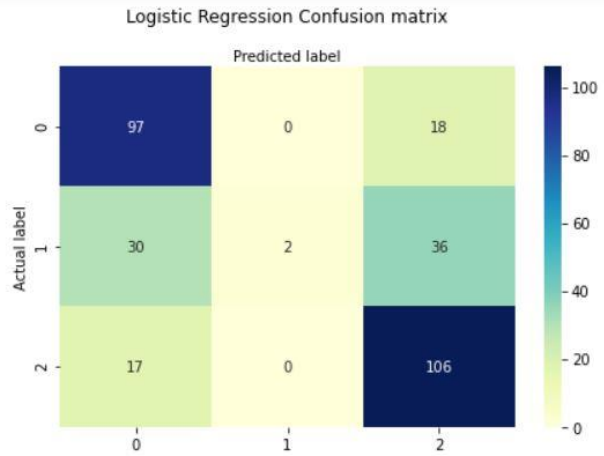
Away W : 2
Away D : 2
Away L : 2
Total Away Games : 6

Last matches between teams results:

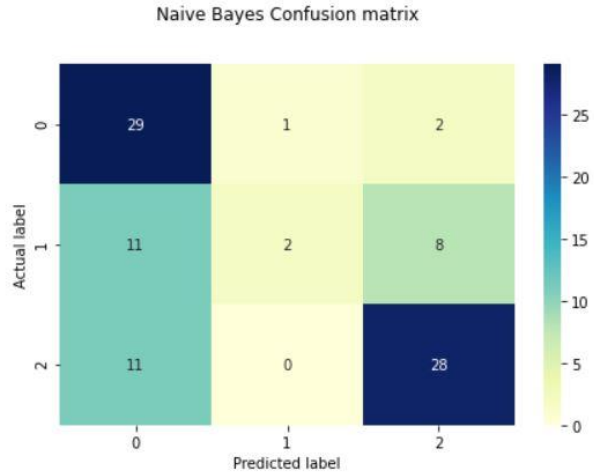
	São Paulo	Fortaleza
0	2	3
1	10	9on pens
2	3	3
3	1	0
4	2	1
5	0	1
6	1	1
7	1	0
8	1	0
9	3	2

<AxesSubplot:>

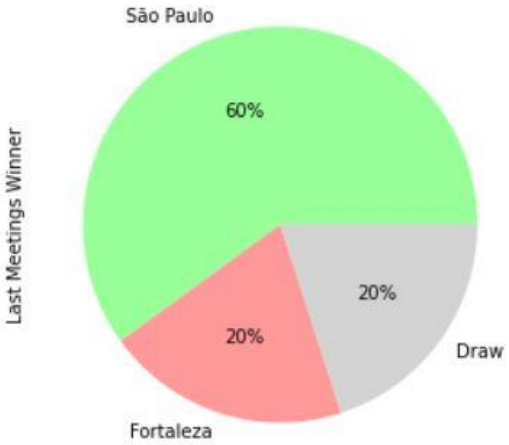




Logistic Regression - Accuracy on test data= 0.66993

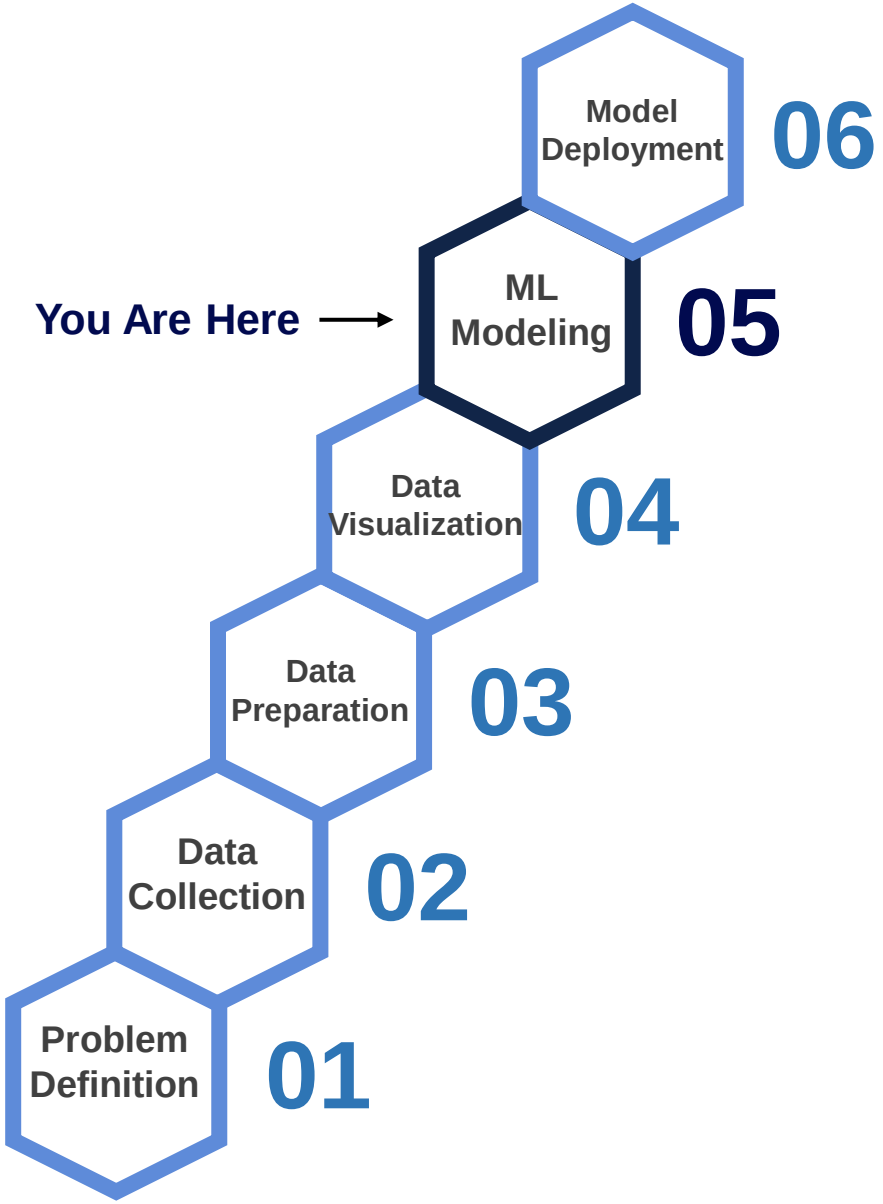


Naive Bayes - Accuracy on test data= 0.6413043478260869



We used ‘matplotlib’ library for that mission.

Stages Map


Model Deployment
06
You Are Here
ML Modeling
05
Data Visualization
04
Data Preparation
03
Data Collection
02
Problem Definition
01

ML Modeling & Feature Engineering

In our project, the user can choose which league and season he wants to learn from.

The advantage of this method is that the user can choose to learn from the same league as his match, making the prediction even more accurate.

	Competition
0	Bundesliga
1	Primier League
2	La Liga
3	Seria A
4	Ligue 1

```
Please choose the competition's number you want to learn:
```

	Seasons
0	19/20
1	18/19
2	17/18
3	16/17
4	15/16

```
Please choose the season's number you want to learn:
```

ML Modeling & Feature Engineering

Choosing the ML model

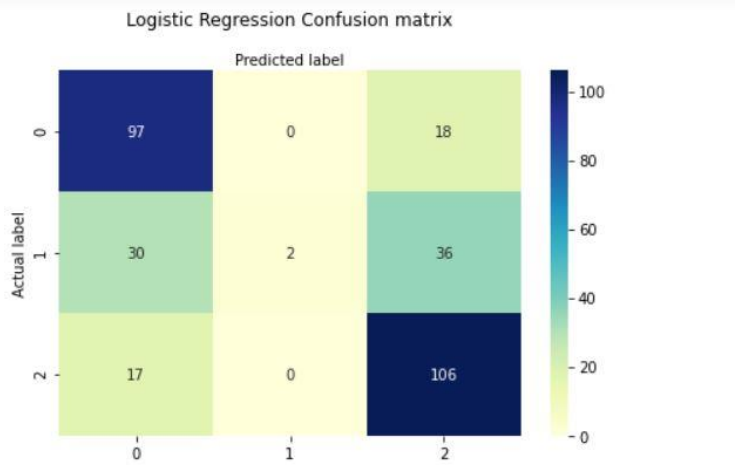
After collecting the necessary data about the chosen league and season, we used `sklearn` library to be able to train a model that predicts the probability of each results of the selected league's matches.
Our predictions will be the results with the highest probability.

The user can choose which algorithms he wants to use:

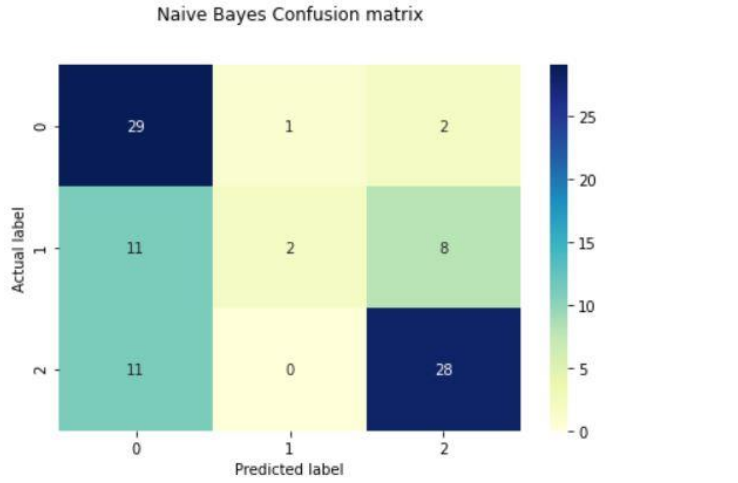
1. **Best – the most accurate algorithm.**
2. **Logistic Regression**
3. **Naive Bayes**

```
Please choose the ML algorithm's number you want to use:
```

	ML_Algorithms
0	Best
1	Logistic Regression
2	Naive Bayes



Logistic Regression - Accuracy on test data= 0.6699346405228758



Naive Bayes - Accuracy on test data= 0.6413043478260869

ML Modeling & Feature Engineering

01

Positions

The team ranked higher in the league table has a better chance of winning.

02

Recent matches results

The team that did better results in her last matches has a better chance of winning.

03

Past meetings results

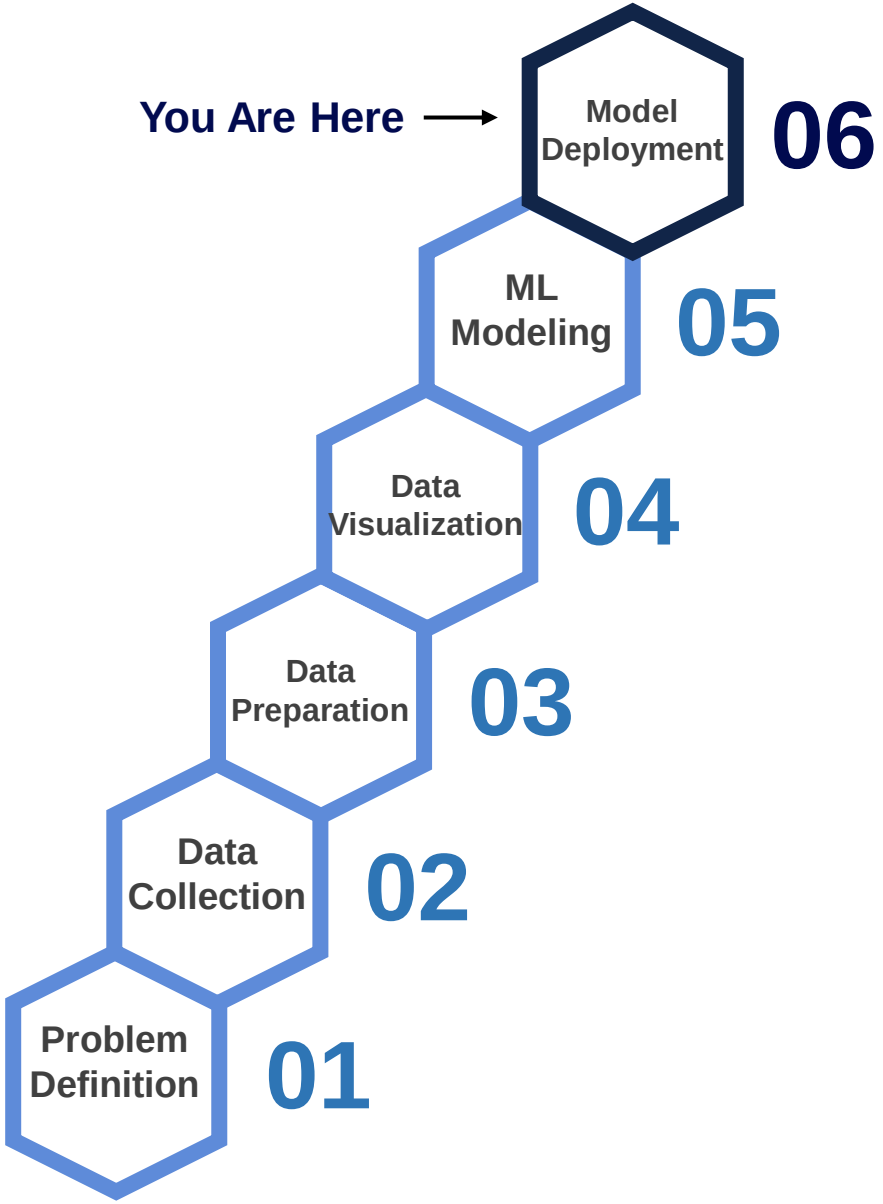
The team that did better results in their previous meetings has a better chance of winning.

04

Players Strengths

The team with the better players has a better chance of winning. Bigger market value indicates on a better players.

Stages Map


You Are Here
Model Deployment
06
ML Modeling
05
Data Visualization
04
Data Preparation
03
Data Collection
02
Problem Definition
01

Model Deployment

The final prediction

This is the last step of the machine learning process.

The model is now ready to make predictions on his own, based on the relevant match data.

All we have is to choose the match we want to predict and our model will do the rest.

Model Deployment

Home Team	0	SOCCER	0	Away Team

Home Team		Away Team
Market Value		Market Value
Position		Position
Past home meetings between teams results	Vs	Past away meetings between teams results
Recent home matches results		Recent away matches result
Match points		Match points

Thank You.