

```
In [1]: '''
* @FileName : iplPointPrediction.py
* @Author : Pradyumn Joshi
* @Brief : Given details of ipl teams and players predict the total points they scored on the basis of runs, balls,wickets
* @Date : 24 Jan 2021
*
* Copyright (C) 2021
'''
```

```
Out[1]: '\n* @FileName : iplPointPrediction.py\n* @Author : Pradyumn Joshi\n* @Brief : Given details of ipl teams and players predict
the total points they scored on the basis of runs, balls,wickets\n* @Date : 24 Jan 2021\n*\n* Copyright (C) 2021\n'
```

```
In [2]: '''
1. Extract Features
    a. Features and target should not have null value
    b. Features should be numeric in nature
    c. Features should be of the type array/dataframe
    d. Features should have some rows and columns
2. Split the dataset into training and testing datasets.
    e. Features should be on same scale
3. Train the model on training dataset
4. Test the model on testing dataset
'''
```

```
Out[2]: '\n1. Extract Features\n    a. Features and target should not have null value\n    b. Features should be numeric in nature\n
c. Features should be of the type array/dataframe\n    d. Features should have some rows and columns \n2. Split the dataset
into training and testing datasets.\n    e. Features should be on same scale\n3. Train the model on training dataset\n4. Test
the model on testing dataset\n'
```

```
In [3]: import pandas as pd
ipl = pd.read_csv("ipl2017.csv")
```

```
In [4]: ipl
```

```
Out[4]:
```

	mid	date	venue	bat_team	bowl_team	batsman	bowler	runs	wickets	overs	runs_last_5	wickets_last_5	striker	non-striker	total
0	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	SC Ganguly	P Kumar	1	0	0.1	1	0	0	0	222
1	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	1	0	0.2	1	0	0	0	222
2	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.2	2	0	0	0	222
3	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.3	2	0	0	0	222
4	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.4	2	0	0	0	222
...
76009	617	2017-05-21	Rajiv Gandhi International Stadium, Uppal	Mumbai Indians	Rising Pune Supergiant	KH Pandya	DT Christian	121	7	19.2	40	0	40	12	129
76010	617	2017-05-21	Rajiv Gandhi International Stadium, Uppal	Mumbai Indians	Rising Pune Supergiant	KH Pandya	DT Christian	127	7	19.3	46	0	46	12	129
76011	617	2017-05-21	Rajiv Gandhi International Stadium, Uppal	Mumbai Indians	Rising Pune Supergiant	KH Pandya	DT Christian	128	7	19.4	47	0	47	12	129
76012	617	2017-05-21	Rajiv Gandhi International Stadium, Uppal	Mumbai Indians	Rising Pune Supergiant	MG Johnson	DT Christian	129	7	19.5	48	0	47	13	129
76013	617	2017-05-21	Rajiv Gandhi International Stadium, Uppal	Mumbai Indians	Rising Pune Supergiant	KH Pandya	DT Christian	129	8	19.6	47	1	47	13	129

76014 rows × 15 columns

```
In [5]: y=ipl["total"]
```

```
In [6]: X=ipl.drop(["total"],axis=1)
```

```
In [7]: #checking null values
X.isna().sum()
```

```
Out[7]: mid          0
date            0
venue           0
bat_team        0
bowl_team       0
batsman         0
```

```

bowler      0
runs        0
wickets     0
overs       0
runs_last_5 0
wickets_last_5 0
striker     0
non-striker 0
dtype: int64

```

```
In [8]: y.isna().sum()
```

```
Out[8]: 0
```

```
In [9]: X.shape
```

```
Out[9]: (76014, 14)
```

```
In [10]: X.dtypes
```

```

Out[10]: mid          int64
date          object
venue         object
bat_team      object
bowl_team     object
batsman       object
bowler        object
runs          int64
wickets       int64
overs         float64
runs_last_5   int64
wickets_last_5 int64
striker       int64
non-striker   int64
dtype: object

```

```
In [11]: X
```

```

Out[11]:
   mid  date      venue  bat_team  bowl_team  batsman  bowler  runs  wickets  overs  runs_last_5  wickets_last_5  striker  non-striker
0    1  2008-04-18  M Chinnaswamy Stadium  Kolkata Knight Riders  Royal Challengers Bangalore  SC Ganguly  P Kumar    1     0    0.1         1         0         0         0
1    1  2008-04-18  M Chinnaswamy Stadium  Kolkata Knight Riders  Royal Challengers Bangalore  BB McCullum  P Kumar    1     0    0.2         1         0         0         0
2    1  2008-04-18  M Chinnaswamy Stadium  Kolkata Knight Riders  Royal Challengers Bangalore  BB McCullum  P Kumar    2     0    0.2         2         0         0         0
3    1  2008-04-18  M Chinnaswamy Stadium  Kolkata Knight Riders  Royal Challengers Bangalore  BB McCullum  P Kumar    2     0    0.3         2         0         0         0
4    1  2008-04-18  M Chinnaswamy Stadium  Kolkata Knight Riders  Royal Challengers Bangalore  BB McCullum  P Kumar    2     0    0.4         2         0         0         0
...  ...  ...      ...          ...          ...          ...      ...      ...      ...      ...      ...      ...      ...      ...
76009  617  2017-05-21  Rajiv Gandhi International Stadium, Uppal  Mumbai Indians  Rising Pune Supergiant  KH Pandya  DT Christian  121     7   19.2         40         0        40        12
76010  617  2017-05-21  Rajiv Gandhi International Stadium, Uppal  Mumbai Indians  Rising Pune Supergiant  KH Pandya  DT Christian  127     7   19.3         46         0        46        12
76011  617  2017-05-21  Rajiv Gandhi International Stadium, Uppal  Mumbai Indians  Rising Pune Supergiant  KH Pandya  DT Christian  128     7   19.4         47         0        47        12
76012  617  2017-05-21  Rajiv Gandhi International Stadium, Uppal  Mumbai Indians  Rising Pune Supergiant  MG Johnson  DT Christian  129     7   19.5         48         0        47        13
76013  617  2017-05-21  Rajiv Gandhi International Stadium, Uppal  Mumbai Indians  Rising Pune Supergiant  KH Pandya  DT Christian  129     8   19.6         47         1        47        13

```

76014 rows × 14 columns

```
In [12]: X=X.drop(['mid','date'],axis=1)
```

```

In [13]: #encoding the data (object-> numeric)
from sklearn.preprocessing import LabelEncoder
l_encoder = LabelEncoder()

```

```

In [14]: l_venue = X['venue']
l_bat = X['bat_team']
l_bowl = X['bowl_team']
l_batsman = X['batsman']
l_bowler = X['bowler']

```

```
In [15]: l_encoder.fit(X['venue'])
X['venue']=l_encoder.transform(X['venue'])
```

```
In [16]: #decoded value with encoded value
lVen = dict(zip(l_venue,X['venue']))
```

```
In [17]: l_encoder.fit( X['bat_team'] )
X['bat_team']=l_encoder.transform(X['bat_team'])
```

```
In [18]: lBat = dict(zip(l_bat,X['bat_team']))
```

```
In [19]: l_encoder.fit( X['bowl_team'] )
X['bowl_team']=l_encoder.transform(X['bowl_team'])
```

```
In [20]: lBowl = dict(zip(l_bowl,X['bowl_team']))
```

```
In [21]: l_encoder.fit( X['batsman'] )
X['batsman']=l_encoder.transform(X['batsman'])
```

```
In [22]: lBatsman = dict(zip(l_batsman, X['batsman']))
```

```
In [23]: l_encoder.fit( X['bowler'] )
X['bowler']=l_encoder.transform(X['bowler'])
```

```
In [24]: lBowler = dict(zip(l_bowler,X['bowler']))
```

```
In [25]: X
```

```
Out[25]:
```

	venue	bat_team	bowl_team	batsman	bowler	runs	wickets	overs	runs_last_5	wickets_last_5	striker	non-striker
0	14	6	12	328	201	1	0	0.1	1	0	0	0
1	14	6	12	61	201	1	0	0.2	1	0	0	0
2	14	6	12	61	201	2	0	0.2	2	0	0	0
3	14	6	12	61	201	2	0	0.3	2	0	0	0
4	14	6	12	61	201	2	0	0.4	2	0	0	0
...
76009	23	7	10	172	96	121	7	19.2	40	0	40	12
76010	23	7	10	172	96	127	7	19.3	46	0	46	12
76011	23	7	10	172	96	128	7	19.4	47	0	47	12
76012	23	7	10	215	96	129	7	19.5	48	0	47	13
76013	23	7	10	172	96	129	8	19.6	47	1	47	13

76014 rows × 12 columns

```
In [26]: from sklearn.model_selection import train_test_split as tts
X_train, X_test, y_train, y_test = tts(X,y, test_size=0.25, random_state=42)
```

```
In [27]: #scaling the data
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
```

```
In [28]: from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()
```

```
In [29]: model.fit(X_train,y_train)
```

```
Out[29]: RandomForestRegressor()
```

```
In [30]: model.score(X_test,y_test)
```

```
Out[30]: 0.9349628143324903
```

```
In [31]: pred_data = pd.DataFrame({'venue':[14], 'bat_team':[7], 'bowl_team':[12], 'batsmen':[224], 'bowler':[192], 'runs' : [100], 'wicke
```

```
In [32]: pred_data
```

```
Out[32]:
```

	venue	bat_team	bowl_team	batsmen	bowler	runs	wickets	overs	runs_last_5	wickets_last_5	striker	non-striker
0	14	7	12	224	192	100	0	2.0	44	0	0	2

```
In [33]: model.predict(pred_data)
```

```
Out[33]: array([187.43])
```