

```
import pandas as pd
```

submitted by : joswin v jaison email:[joswinvjaison@karunya.edu.in](mailto:joswinvjaison@karunya.edu.in),[joswin2222@gmail.com](mailto:joswin2222@gmail.com)

```
import numpy as np
```

```
df=pd.read_csv('/content/ipl2017.csv')
```

```
df.head()
```



|   | mid | date       | venue                 | bat_team              | bowl_team                   | batsman     | bowler  | runs | wickets | overs |
|---|-----|------------|-----------------------|-----------------------|-----------------------------|-------------|---------|------|---------|-------|
| 0 | 1   | 2008-04-18 | M Chinnaswamy Stadium | Kolkata Knight Riders | Royal Challengers Bangalore | SC Ganguly  | P Kumar | 1    | 0       | 0.1   |
| 1 | 1   | 2008-04-18 | M Chinnaswamy Stadium | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar | 1    | 0       | 0.2   |
| 2 | 1   | 2008-04-18 | M Chinnaswamy Stadium | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar | 2    | 0       | 0.2   |
| 3 | 1   | 2008-04-18 | M Chinnaswamy Stadium | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar | 2    | 0       | 0.3   |

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 76014 entries, 0 to 76013
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   mid                    76014 non-null  int64  
1   date                   76014 non-null  object  
2   venue                  76014 non-null  object  
3   bat_team               76014 non-null  object  
4   bowl_team              76014 non-null  object  
5   batsman                76014 non-null  object  
6   bowler                 76014 non-null  object  
7   runs                   76014 non-null  int64  
8   wickets                76014 non-null  int64  
9   overs                  76014 non-null  float64

```

```

10 runs_last_5      76014 non-null int64
11 wickets_last_5   76014 non-null int64
12 striker          76014 non-null int64
13 non-striker      76014 non-null int64
14 total            76014 non-null int64
dtypes: float64(1), int64(8), object(6)
memory usage: 8.7+ MB

```

```
df.drop(labels=['mid','date'],axis=1,inplace=True)
```

```
print(df.head())
```

```

      venue      bat_team  ... non-striker total
0  M Chinnaswamy Stadium  Kolkata Knight Riders  ...      0    222
1  M Chinnaswamy Stadium  Kolkata Knight Riders  ...      0    222
2  M Chinnaswamy Stadium  Kolkata Knight Riders  ...      0    222
3  M Chinnaswamy Stadium  Kolkata Knight Riders  ...      0    222
4  M Chinnaswamy Stadium  Kolkata Knight Riders  ...      0    222

```

```
[5 rows x 13 columns]
```

```
df.dtypes
```

```

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
total       int64
dtype: object

```

```

y=df['total']
x=df.drop('total',axis=1)

```

```
x=pd.get_dummies(data=x,columns=['venue','bat_team','bowl_team','batsman','bowler'],drop_firs
```

```
x.head()
```

|   | runs | wickets | overs | runs_last_5 | wickets_last_5 | striker | non-striker | venue_Brabourne Stadium |
|---|------|---------|-------|-------------|----------------|---------|-------------|-------------------------|
| 0 | 1    | 0       | 0.1   | 1           | 0              | 0       | 0           | 0                       |
| 1 | 1    | 0       | 0.2   | 1           | 0              | 0       | 0           | 0                       |
| 2 | 2    | 0       | 0.2   | 2           | 0              | 0       | 0           | 0                       |

```
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.3,random_state=38)
```

5 rows x 805 columns

```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
```

```
xtrain=scaler.fit_transform(xtrain)
xtest=scaler.transform(xtest)
```

```
from sklearn.ensemble import RandomForestRegressor
forest=RandomForestRegressor()
```

```
df.head()
```

|   | venue                 | bat_team              | bowl_team                   | batsman     | bowler  | runs | wickets | overs | runs_last_5 |
|---|-----------------------|-----------------------|-----------------------------|-------------|---------|------|---------|-------|-------------|
| 0 | M Chinnaswamy Stadium | Kolkata Knight Riders | Royal Challengers Bangalore | SC Ganguly  | P Kumar | 1    | 0       | 0.1   |             |
| 1 | M Chinnaswamy Stadium | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar | 1    | 0       | 0.2   |             |

```
df.fillna(df.mean())
```

|       | venue                                     | bat_team              | bowl_team                   | batsman     | bowler       | runs | wickets | overs | runs. |
|-------|---|-----------------------|-----------------------------|-------------|--------------|------|---------|-------|-------|
| 0     | M Chinnaswamy Stadium                     | Kolkata Knight Riders | Royal Challengers Bangalore | SC Ganguly  | P Kumar      | 1    | 0       | 0.1   |       |
| 1     | M Chinnaswamy Stadium                     | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar      | 1    | 0       | 0.2   |       |
| 2     | M Chinnaswamy Stadium                     | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar      | 2    | 0       | 0.2   |       |
| 3     | M Chinnaswamy Stadium                     | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar      | 2    | 0       | 0.3   |       |
| 4     | M Chinnaswamy Stadium                     | Kolkata Knight Riders | Royal Challengers Bangalore | BB McCullum | P Kumar      | 2    | 0       | 0.4   |       |
| ...   | ...                                       | ...                   | ...                         | ...         | ...          | ...  | ...     | ...   | ...   |
| 76009 | Rajiv Gandhi International Stadium, Uppal | Mumbai Indians        | Rising Pune Supergiant      | KH Pandya   | DT Christian | 121  | 7       | 19.2  |       |
| 76010 | Rajiv Gandhi International Stadium, Uppal | Mumbai Indians        | Rising Pune Supergiant      | KH Pandya   | DT Christian | 127  | 7       | 19.3  |       |
| ...   | ...                                       | ...                   | ...                         | ...         | ...          | ...  | ...     | ...   | ...   |

```
df.isnull().sum()
```

```
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
total          0
dtype: int64
```

```
forest.fit(xtrain,ytrain)
```

```
RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
                        max_depth=None, max_features='auto', max_leaf_nodes=None,
                        max_samples=None, min_impurity_decrease=0.0,
                        min_impurity_split=None, min_samples_leaf=1,
                        min_samples_split=2, min_weight_fraction_leaf=0.0,
                        n_estimators=100, n_jobs=None, oob_score=False,
                        random_state=None, verbose=0, warm_start=False)
```

```
foresttesting=forest.score(xtrain,ytrain)
```

```
print(foresttesting)
```

```
0.9914383467525102
```

```
foresttraining=forest.score(xtrain,ytrain)
```

```
print(foresttraining)
```

```
0.9914383467525102
```

```
data=x.iloc[[0,150]]
```

```
print(data.head())
```

```
      runs  wickets  overs  ...  bowler_YS Chahal  bowler_Yuvraj Singh  bowler_Z Khan
0         1         0   0.1  ...              0              0              0
150      33         1   4.1  ...              0              0              0
```

```
[2 rows x 805 columns]
```

```
rfrre=forest.predict(data)
```

```
rfrre
```

```
array([198.73, 247.48])
```

```
original_score=y.iloc[[0,150]]
```

```
print(original_score)
```

```
0      222
```

```
150    240
```

```
Name: total, dtype: int64
```