

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_squared_error, r2_score
```

```
test = pd.read_csv('/content/Test.csv')
Train= pd.read_csv('/content/Train.csv')
```

```
test.head()
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	
0	FDW58	20.750	Low Fat	0.007565	Snack Foods	107.8622	OUT049	1999	
1	FDW14	8.300	reg	0.038428	Dairy	87.3198	OUT017	2007	
2	NCN55	14.600	Low Fat	0.099575	Others	241.7538	OUT010	1998	
3	FDQ58	7.315	Low Fat	0.015388	Snack Foods	155.0340	OUT017	2007	
4	FDY38	NaN	Regular	0.118599	Dairy	234.2300	OUT027	1985	

```
Train.head()
```

t_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type
Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket
Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket
Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	Tier 1	Supermarket
Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	Tier 3	Grocery store
Low Fat	0.000000	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket

```
test.isnull().sum()/len(test)
```

```
Item_Identifier      0.000000
Item_Weight          0.171801
Item_Fat_Content     0.000000
Item_Visibility      0.000000
Item_Type            0.000000
Item_MRP             0.000000
Outlet_Identifier    0.000000
Outlet_Establishment_Year 0.000000
Outlet_Size         0.282697
Outlet_Location_Type 0.000000
Outlet_Type          0.000000
dtype: float64
```

```
test.Item_Weight .fillna(test.Item_Weight.mean(),inplace=True)

test.Outlet_Size.fillna(test.Outlet_Size.mode(),inplace=True)
```

```
Train.isnull().sum()

Item_Identifier      0
Item_Weight         1463
Item_Fat_Content      0
Item_Visibility      0
Item_Type            0
Item_MRP            0
Outlet_Identifier    0
Outlet_Establishment_Year  0
Outlet_Size         2410
Outlet_Location_Type  0
Outlet_Type          0
Item_Outlet_Sales    0
dtype: int64
```

Train.head()

Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Sales
FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	101.2796
DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	101.2796
FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999	101.2796
FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	101.2796
NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013	1987	101.2796

```
Train.Item_Weight .fillna(Train.Item_Weight.mean(),inplace=True)
Train.Outlet_Size.fillna(Train.Outlet_Size.mode(),inplace=True)
```

```
Train=pd.get_dummies(Train)
Train
```

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Item_Outlet_Sales	Item_Identifier_DRA12	Item_Identifier_DRA24
0	9.300	0.016047	249.8092	1999	3735.1380	0	0
1	5.920	0.019278	48.2692	2009	443.4228	0	0
2	17.500	0.016760	141.6180	1999	2097.2700	0	0
3	19.200	0.000000	182.0950	1998	732.3800	0	0
4	8.930	0.000000	53.8614	1987	994.7052	0	0
...
8518	6.865	0.056783	214.5218	1987	2778.3834	0	0
8519	8.380	0.046982	108.1570	2002	549.2850	0	0
8520	10.600	0.035186	85.1224	2004	1193.1136	0	0
8521	7.210	0.145221	103.1332	2009	1845.5976	0	0
8522	14.800	0.044878	75.4670	1997	765.6700	0	0

8523 rows × 1605 columns

```
test=pd.get_dummies(test)
test
```

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Item_Identifier_DRA12	Item_Identifier_DRA24	Item_Identifier_I
0	20.750000	0.007565	107.8622	1999	0	0	
1	8.300000	0.038428	87.3198	2007	0	0	
2	14.600000	0.099575	241.7538	1998	0	0	
3	7.315000	0.015388	155.0340	2007	0	0	
4	12.695633	0.118599	234.2300	1985	0	0	
...	
5676	10.500000	0.013496	141.3154	1997	0	0	
5677	7.600000	0.142991	169.1448	2009	0	0	
5678	10.000000	0.073529	118.7440	2002	0	0	
5679	15.300000	0.000000	214.6218	2007	0	0	
5680	9.500000	0.104720	79.7960	2002	0	0	

5681 rows × 1588 columns

```
y = Train['Item_Outlet_Sales']
x = Train.drop(['Item_Outlet_Sales'], axis = 1)
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=True,test_size=0.2)
```

```
print(x_train)
```

```

Item_Weight  Item_Visibility  Item_MRP  Outlet_Establishment_Year  \
1945  18.350000      0.089345  191.9504      2009
1720  17.350000      0.168065  176.2712      2009
1954  10.100000      0.053887  225.6088      2007
1919  10.850000      0.162904  104.9622      2009
2461   7.170000      0.059717  130.9968      2004
...      ...      ...      ...      ...
2895  12.857645      0.277459  156.3946      1985
7813  12.857645      0.178193   54.9956      1985
905   17.850000      0.065810  149.1050      1997
5192   9.600000      0.049618   44.2086      1998
235   12.857645      0.043574  192.8846      1985

Item_Identifier_DRA12  Item_Identifier_DRA24  Item_Identifier_DRA59  \
1945      0      0      0
1720      0      0      0
1954      0      0      0
1919      0      0      0
2461      0      0      0
...      ...      ...      ...
2895      0      0      0
7813      0      0      0
905      0      0      0
5192      0      0      0
235      0      0      0

Item_Identifier_DRB01  Item_Identifier_DRB13  Item_Identifier_DRB24  \
1945      0      0      0
1720      0      0      0
1954      0      0      0
1919      0      0      0
2461      0      0      0
...      ...      ...      ...
2895      0      0      0
7813      0      0      0
905      0      0      0
5192      0      0      0
235      0      0      0

...  Outlet_Size_High  Outlet_Size_Medium  Outlet_Size_Small  \
1945  ...      0      1      0
1720  ...      0      1      0
1954  ...      0      0      0
1919  ...      0      1      0
2461  ...      0      0      1
```

```

...      ...      ...      ...      ...
2895 ...      0      0      1
7813 ...      0      0      1
905 ...      0      0      1
5192 ...      0      0      0
235 ...      0      1      0

```

```

Outlet_Location_Type_Tier 1 Outlet_Location_Type_Tier 2 \
1945      0      0
1720      0      0
1954      0      1
1919      0      0
2461      0      1

```

```
print(x_test)
```

```

Item_Weight Item_Visibility Item_MRP Outlet_Establishment_Year \
1070 13.500000 0.055102 37.0874 2002
6305 12.500000 0.074035 87.9198 2009
8504 12.857645 0.124111 111.7544 1985
5562 12.500000 0.073735 87.1198 1997
1410 15.850000 0.007140 40.8480 1987
...      ...      ...      ...
376 7.575000 0.055390 195.2768 1997
7708 12.600000 0.074222 255.9356 2009
3812 9.600000 0.006693 164.9184 1987
3928 12.857645 0.097411 52.7324 1985
7654 15.500000 0.103423 144.8470 2004

```

```

Item_Identifier_DRA12 Item_Identifier_DRA24 Item_Identifier_DRA59 \
1070      0      0      0
6305      0      0      0
8504      0      0      0
5562      0      0      0
1410      0      0      0
...      ...      ...
376      0      0      0
7708      0      0      0
3812      0      0      0
3928      0      0      0
7654      0      0      0

```

```

Item_Identifier_DRB01 Item_Identifier_DRB13 Item_Identifier_DRB24 \
1070      0      0      0
6305      0      0      0
8504      0      0      0
5562      0      0      0
1410      0      0      0
...      ...      ...
376      0      0      0
7708      0      0      0
3812      0      0      0
3928      0      0      0
7654      0      0      0

```

```

... Outlet_Size_High Outlet_Size_Medium Outlet_Size_Small \
1070 ...      0      0      0
6305 ...      0      1      0
8504 ...      0      1      0
5562 ...      0      0      1
1410 ...      1      0      0
...      ...      ...
376 ...      0      0      1
7708 ...      0      1      0
3812 ...      1      0      0
3928 ...      0      1      0
7654 ...      0      0      1

```

```

Outlet_Location_Type_Tier 1 Outlet_Location_Type_Tier 2 \
1070      0      1
6305      0      0
8504      0      0
5562      1      0
1410      0      0

```

```
print(y_train)
```

```

1945 5369.0112
1720 1230.3984
1954 4250.4672
1919 1482.0708
2461 2348.9424
...
2895 473.3838

```

```
7813    109.1912
905     2247.0750
5192     44.6086
235     2293.0152
Name: Item_Outlet_Sales, Length: 6818, dtype: float64
```

```
print(y_test)
```

```
1070     952.7598
6305    1133.8574
8504    4138.6128
5562    1657.1762
1410     679.1160
...
376     5715.2272
7708    4832.3764
3812    2972.1312
3928    2492.7552
7654    1717.7640
Name: Item_Outlet_Sales, Length: 1705, dtype: float64
```

```
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
x_train=sc.fit_transform(x_train)
x_test=sc.transform(x_test)
```

```
#before train and test specify model u r using
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

```
from sklearn.neighbors import KNeighborsClassifier
model=KNeighborsClassifier()
```

```
from sklearn.ensemble import RandomForestClassifier
#class      #function
clf = RandomForestClassifier()
```