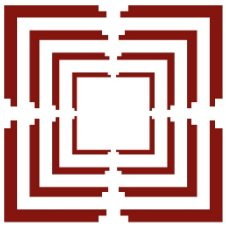

Ahmedabad University

CSE523 Machine Learning

Weekly Project Report

Date: 18-02-2023

Project title: Big Mart Sales Prediction

Group 10

Name	Enrolment no.
Meet Patel	AU2040010
Dev Patel	AU2040056
Kush Patel	AU2040137
Vatsal Shah	AU2040019

1. **Task performed this week**

- After reviewing the literature and researching, we started the programming part this week.
- We studied and analysed the dataset.
- We first checked for the null or missing values in the data.
- Replaced the missing values with mean for numerical data and mode for categorical data.

2. **Outcomes of task performed**

- The basic information of the columns:

```
[18] # Infomation about the data
     df_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8523 entries, 0 to 8522
Data columns (total 12 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Item_Identifier            8523 non-null   object
 1   Item_Weight                7060 non-null   float64
 2   Item_Fat_Content           8523 non-null   object
 3   Item_Visibility            8523 non-null   float64
 4   Item_Type                  8523 non-null   object
 5   Item_MRP                   8523 non-null   float64
 6   Outlet_Identifier          8523 non-null   object
 7   Outlet_Establishment_Year  8523 non-null   int64
 8   Outlet_Size                6113 non-null   object
 9   Outlet_Location_Type       8523 non-null   object
 10  Outlet_Type                8523 non-null   object
 11  Item_Outlet_Sales          8523 non-null   float64
dtypes: float64(4), int64(1), object(7)
```

- Checking for missing values:

```
# Checking for missing values
df_data.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
```

- After replacing the missing values:

```
# filling the missing values in "Item_weight" column with Mean
df_data['Item_Weight'].fillna(df_data['Item_Weight'].mean(), inplace=True)
```

```
df_data.isnull().sum()
```

```
Item_Identifier               0
Item_Weight                   0
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
```

```
# filling the missing values in "Outlet_Size" column with Mode
df_data['Outlet_Size'].fillna(df_data['Outlet_Size'].mode()[0],inplace=True)
```

```
df_data.isnull().sum()
```

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

3. Tasks to be performed in the upcoming week

- We will perform more data analysis on the dataset.
- We will perform the required data cleaning and dimensionality reduction on the data.
- We try to perform exploratory data analysis on the dataset to know more about the dataset.