

```
In [ ]: ▶ # Importing the Libraries
import numpy as np
import pandas as pd
from sklearn.preprocessing import OneHotEncoder, LabelEncoder, MinMaxScaler, S
from sklearn.model_selection import train_test_split, ParameterGrid
from sklearn.metrics import accuracy_score, confusion_matrix, mean_squared_er
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer

In [ ]: ▶ # Importing the Data
Phones = pd.read_csv("C:\\Users\\Darp\\Documents\\GitHub\\CS-513-Project\\Fir

In [ ]: ▶ Phones.info()

In [ ]: ▶ # Checking the null Values
Phones.isnull().sum()

In [ ]: ▶ # Saperating the Numerical Valuse and Categorical
numeric_columns = ['RAM', 'primary_camera', 'secondary_camera', 'weight_g', 'i
categorical_columns = [c for c in Phones.columns if c not in numeric_columns]
print(categorical_columns)
print(numeric_columns)

In [ ]: ▶ numeric = Phones[numeric_columns]
category = Phones[categorical_columns]

In [ ]: ▶ # Replcaing the Numeric Values by the Mean of the Column
num_imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
numerical = pd.DataFrame(num_imputer.fit_transform(numeric))
numerical.columns = numeric.columns
#numerical = numerical.iloc[:,0:].fillna(numerical.iloc[:,0:].mean())

In [ ]: ▶ # Normalyzing the Numeric Columns
scaler = MinMaxScaler()
scaler.fit(numerical)
Nor_numerical = pd.DataFrame(scaler.transform(numerical))
Nor_numerical.columns = numerical.columns

In [ ]: ▶ # Replcaing the Original Price
Nor_numerical['Price'] = Phones['Price']
#Nor_numerical['Price'][5003] = Phones['Price'].mean()

In [ ]: ▶ Nor_numerical.head(20)
```

```
In [ ]: ▶ # Replacing the Categorical Values by the Most Frequent value
cat_imputer = SimpleImputer(missing_values=np.nan, strategy='most_frequent')
categorical = pd.DataFrame(cat_imputer.fit_transform(category))
categorical.columns = category.columns
#categorical = category.iloc[:,0:].fillna(category.iloc[:,0:].mode())
```

```
In [ ]: ▶ # LabelEncoding
le = LabelEncoder()
Cat = categorical.apply(le.fit_transform)
Cat.head()
```

```
In [ ]: ▶ #Concatinating the Numerical and Cateforical Dataframes
Phones_rna = pd.concat([Cat, Nor_numerical], axis=1, sort=False)
```

```
In [ ]: ▶ Phones_rna.head(20)
```

```
In [ ]: ▶ # Again Checking the Null Values
Phones_rna.isnull().sum()
```

```
In [ ]: ▶ # Replcaing the Price by the categories
bins = [0, 100, 200, 300]
#bins = [0, 50, 150, 250, 500, 1000]
names = [1, 2, 3, 4]

label_dict = dict(enumerate(names, 1))
price = pd.Series(np.vectorize(label_dict.get)(np.digitize(Phones_rna['Price'],
price
```

```
In [ ]: ▶ Phones_rna['Price'] = price
```

```
In [ ]: ▶ # Exporting to CSV
Phones_rna.to_csv("Final_Refined_Encoded.csv", index=False)
```

```
In [ ]: ▶
```