Wireframe

Restaurant Rating Prediction

Problem statement: The main goal of this project is to perform extensive Explor atory Data Analysis(EDA) on the Zomato Dataset and build an appropriate Machine Lea rning Model that will help various Zomato Restaurants to predict their respective Ratings based on certain features.

Introduction: Zomato is the most reputed company in the field of food reviews. Founded in 2008, this company started in India and now is in 24 different countries. Its is so big that the people now use it as a verb. “Did you know about this restaurant? Zomato it”. The rating is the most important feature of any restaurant as it is the first parameter that people look into while searching for a place to eat. It portrays the quality, hygiene and the environment of the place. Higher ratings lead to higher profit margins. Notations of the ratings usually are stars or numbers scaling between 1 and 5. Zomato has changed the way people browse through restaurants. It has helped customers find good places with respect to their dining budget. Different machine learning algorithms like SVM, Linear regression, RandomForestRegressor, ExtraTreesRegressor

Approach: The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing. Try out different machine learning algorithms that’s best fit for the above case.

Tools used:

1)Jupyter notebook

2)Python language

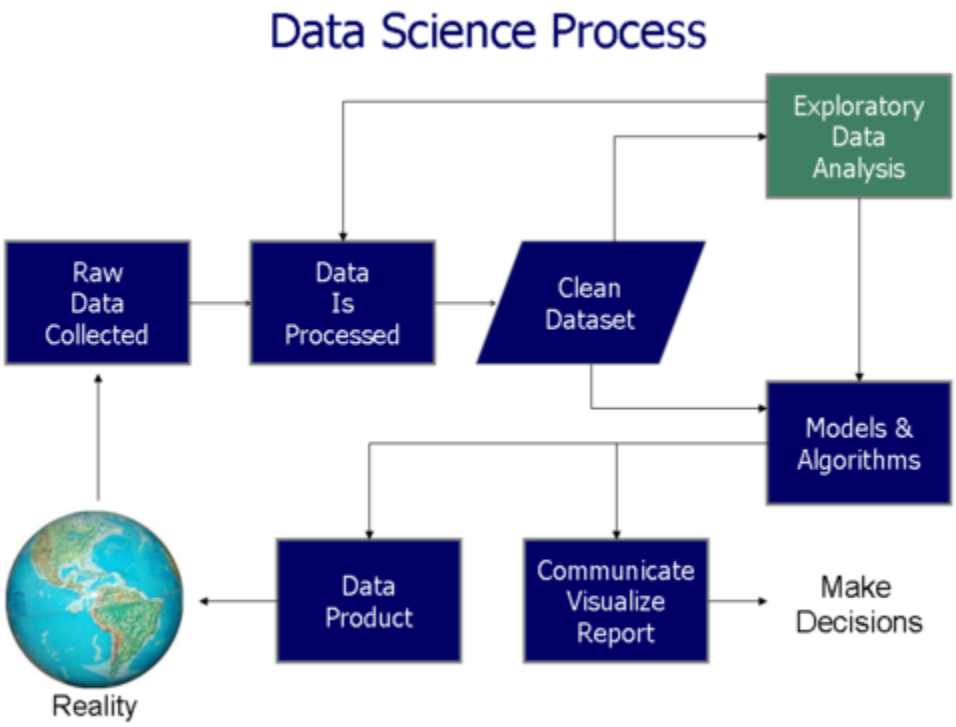
Libraries used:

Wireframe

```
 import numpy as np
import pandas as pd
pd.set_option('max_columns', None)

import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import  ExtraTreesRegressor
from sklearn.model_selection import train_test_split
```

Design process

Wireframe



Performance:

Wireframe

```
from sklearn.ensemble import RandomForestRegressor
RF_Model=RandomForestRegressor(n_estimators=650,random_state=245,min_samples_leaf=.0001)
RF_Model.fit(x_train,y_train)
y_predict=RF_Model.predict(x_test)
r2_score(y_test,y_predict)
```

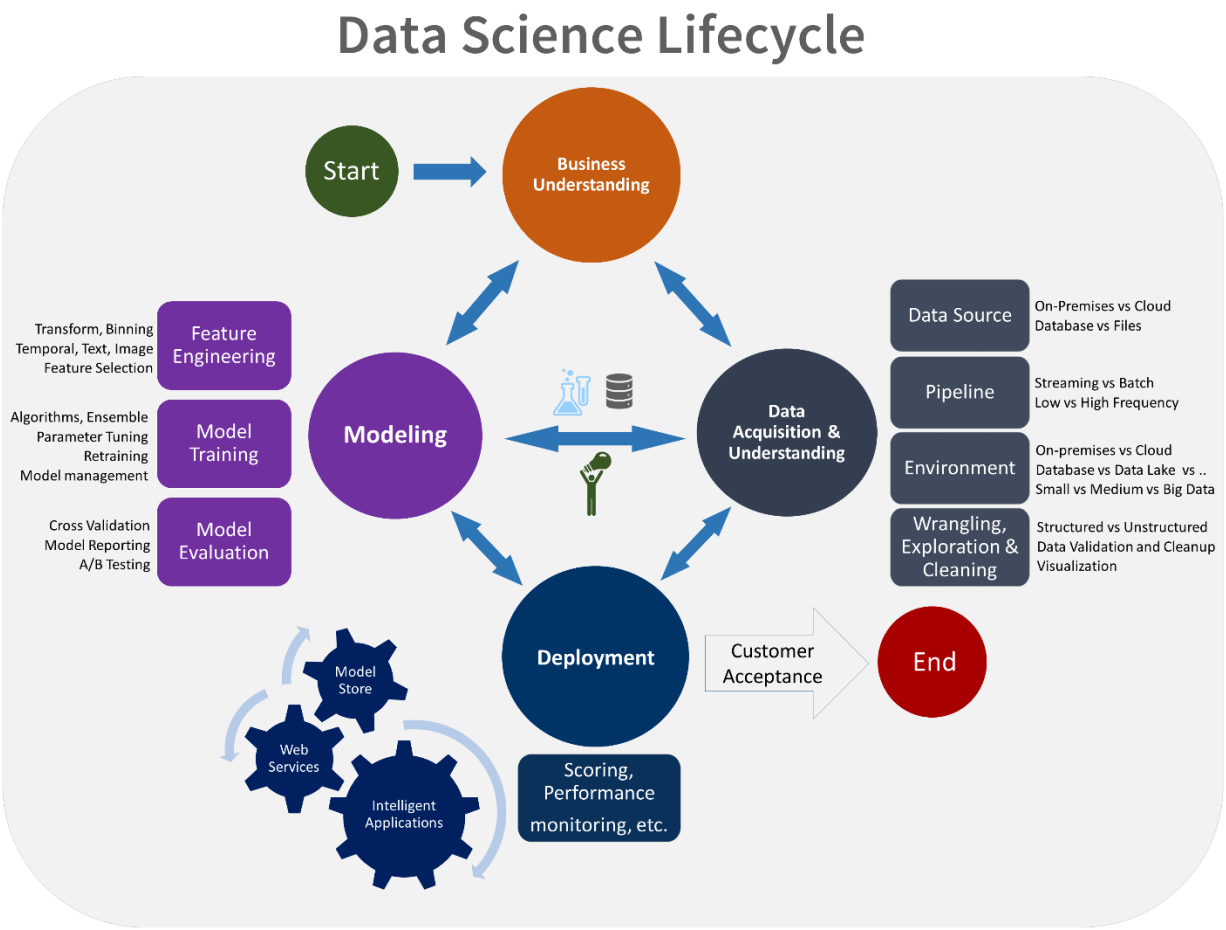
```
0.8809706960047533
```

```
#Preparing Extra Tree Regression
from sklearn.ensemble import  ExtraTreesRegressor
ET_Model=ExtraTreesRegressor(n_estimators = 120)
ET_Model.fit(x_train,y_train)
y_predict=ET_Model.predict(x_test)


from sklearn.metrics import r2_score
r2_score(y_test,y_predict)
```

```
0.9323240658158621
```

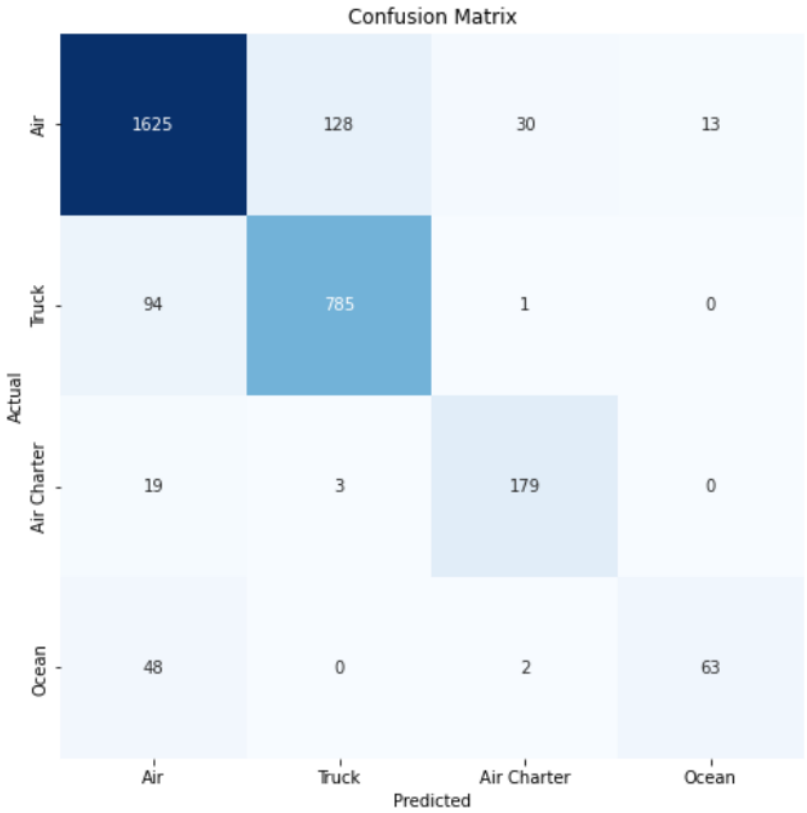
Deployment:

Wireframe



Confusion matrix:

Wireframe



A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix com pares the actual target values with those predicted by the machine learning mo del. This gives us a holistic view of how well our classification model is perform ing and what kinds of errors it is making.

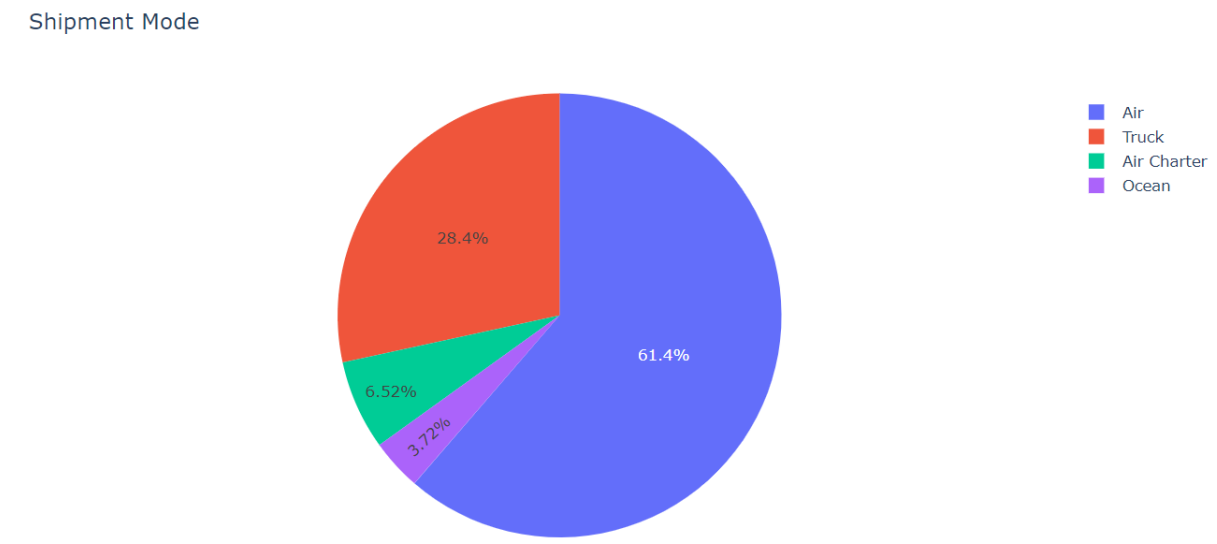
Model training :

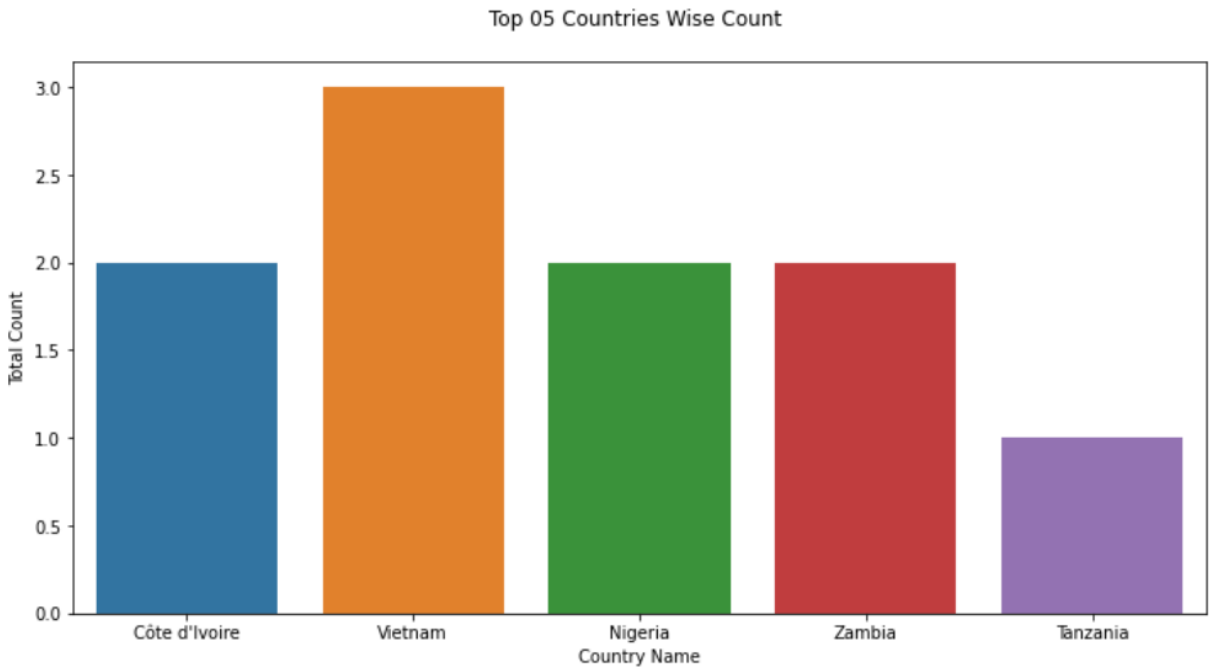
```
# Split df into X and y
    y = df['Shipment Mode']
    X = df.drop('Shipment Mode', axis=1)

    # Encode the labels
    y = y.replace(label_mapping)

    # Train-test split
    X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, shuffle=True, ra
ndom_state=1)

    # Scale X
    scaler = StandardScaler()
    scaler.fit(X_train)
```

Wireframe

```
    X_train = pd.DataFrame(scaler.transform(X_train), index=X_train.index, columns=X_t
rain.columns)
    X_test = pd.DataFrame(scaler.transform(X_test), index=X_test.index, columns=X_tes
t.columns)

    return X_train, X_test, y_train, y_test




inputs = tf.keras.Input(shape=(771,))
x = tf.keras.layers.Dense(128, activation='relu')(inputs)
x = tf.keras.layers.Dense(128, activation='relu')(x)
outputs = tf.keras.layers.Dense(4, activation='softmax')(x)

model = tf.keras.Model(inputs=inputs, outputs=outputs)

model.compile(
    optimizer='adam',
    loss='sparse_categorical_crossentropy',
    metrics=['accuracy']
)

history = model.fit(
    X_train,
    y_train,
    validation_split=0.2,
    batch_size=32,
    epochs=100,
    callbacks=[
        tf.keras.callbacks.EarlyStopping(
            monitor='val_loss',
            patience=3,
            restore_best_weights=True
        )
    ]
)
```

Wireframe





Conclusion: The designed model can easily predict the data and has good accuracy and precison.Which can help the logistics to take good business decisions

#Top Country for Pack Price : Nigeria - 25,620.72

#Top Shipping Mode : Air

#The Max Air Shipment Mode is : 1000

#The Min Air Shipment is : 1

Wireframe

#The Mean Air Shipment is : 82.35
#Top Manufacturing Site : Aurobindo Unit III, India - 3172
#Top Air Manufacturing Site : Aurobindo Unit III, India - 1694