Artificial Intelligence with Python – LAB TASK 3

Kartik Srinivas 16BIS0085

Question:

Lab Task 3

Download the dataset from the following link:

https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview

The Ames Housing dataset was compiled by Dean De Cock for use in data science education. It's an incredible alternative for data scientists looking for a modernized and expanded version of the often cited Boston Housing dataset.

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, you have to predict the final price of each home.

Goal

Predict the sales price for each house. For each Id in the test set, you have to predict the value of the SalePrice variable.

Answer:

```
from typing import List
import numpy as np
import pandas as pd
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_log_error
import matplotlib.pyplot as plt

#READ CSV FILE
train_file = r'C:\Users\karti\Desktop\Fall Semester 2019-20\Artificial
Intelligence with Python\Lab\task3\train.csv'
data = pd.read_csv(train_file)
summary = data.describe()

#SELECT THE FEATURES, DEFINE MODEL
y = data.SalePrice
data_features = [x for x in data.columns if str(data[x][0]).isdigit()][:-1]
X = data[data_features]
describe = X.describe()
head = X.head()
```

```
#FIT MODEL
data_model = DecisionTreeRegressor(random_state=1)
data_model.fit(X, y)
X = np.array(X)

#PREDICTION
test_file = r'C:\Users\karti\Desktop\Fall Semester 2019-20\Artificial Intelligence
with Python\Lab\task3\test.csv'
test_data = pd.read_csv(test_file)
X_test = test_data[data_features]
X_test = pd.DataFrame(X_test).fillna(X_test.mean())
print(X_test)
result = data_model.predict(X_test)
print(result)

# %%
submission_file = r'C:\Users\karti\Desktop\Fall Semester 2019-20\Artificial
Intelligence with Python\Lab\task3\submission.csv'
submission_data = pd.read_csv(submission_file)

# ROOT MEAN SQUARE ERROR & ROOT MEAN SQUARE LOG ERROR
print('RMSE is:  ', (mean_absolute_error(submission_data['SalePrice'], result)) **
0.5)
print('RMSLE is: ', (mean_squared_log_error(submission_data['SalePrice'], result))
** 0.5)
```