Name : Dion D Rodrigues Assignment.no: 03

Roll.no : 53 Class : FYMCA B Subject : ML Lab Batch : B3

import numpy as np

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

# Define column names for the California Housing dataset

column\_names = ['longitude', 'latitude', 'housing\_median\_age', 'total\_rooms', 'total\_bedrooms', 'population', 'households', 'median\_income', 'ocean\_proximity', 'median\_house\_value']

# Load the dataset

bos1 = pd.read\_csv('C:/Users/Admin/Desktop/ML/Assignment\_3/california\_housing\_train.csv')

# Display the first 10 rows

print(bos1.head(10))

# Preprocessing the data: Removing NaN values

bos1 = bos1.dropna()

# One-hot encoding for categorical data

bos1 = pd.get\_dummies(bos1, columns=['ocean\_proximity'])

# Splitting model data with 70% for training

X = bos1.drop('median\_house\_value', axis=1)

Y = bos1['median\_house\_value']

# Split the data into training and testing sets

x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.30, random\_state=5)

# Using Linear Regression Model

lr = LinearRegression()

# Train the model on training data

lr.fit(x\_train, y\_train)

# Predict the testing data so that we can later evaluate the model

pred\_lr = lr.predict(x\_test)

# Model Evaluation: error for linear regression

mse\_lr = mean\_squared\_error(y\_test, pred\_lr, squared=False)

print("Error for Linear Regression = {}".format(mse\_lr))

