Finding the Price of Laptop based on RAM & Storage using Linear Regression

ML ASSESSMENT

K.F.RISHADHA MAHDSE221F-007

A.A.NIZAR MAHDSE221F-009

ABSTRACT

The project aims to develop a predictive model for determining laptop prices based on their RAM and storage specifications using linear regression. By analyzing a dataset of laptops with varying RAM and storage capacities and their corresponding prices, the model will establish a relationship between these hardware attributes and the cost. The outcome of this project will enable more accurate pricing predictions for laptops in the market, enhancing consumer decision-making and providing valuable insights for manufacturers and retailers to optimize pricing strategies.

TABLE OF CONTENT

TABLE OF FIGURES

CHAPTER 1: INTRODUCTION

Welcome to our innovative project aimed at predicting laptop prices using linear regression based on RAM and storage specifications. In today's fast-paced technological landscape, laptops have become indispensable tools for work, entertainment, and communication. The ever-growing range of options makes it challenging for both consumers and sellers to accurately gauge the fair value of a laptop. Our project addresses this challenge by leveraging the power of machine learning, particularly linear regression, to establish a quantitative relationship between a laptop's price and its RAM and storage capacities. By analyzing a comprehensive dataset of various laptop models and their corresponding specifications and prices, we aim to develop a reliable predictive model that will empower consumers to make informed purchasing decisions and assist sellers in setting competitive yet reasonable prices. Through this project, we aspire to simplify the laptop purchasing process and enhance transparency in the market for the benefit of all stakeholders involved.

CHAPTER 2: CODE

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# (i) Select a suitable dataset with at least 1000 records
data = pd.read_csv('laptops.csv')

# (ii) Clean the dataset
# Handle missing values
data.dropna(inplace=True)

# Remove duplicates
data.drop_duplicates(inplace=True)

# (iii) Select a suitable machine learning algorithm
# In this case, we're using Linear Regression

# (iv) Train the machine learning model
# Define features and target
X = data[['RAM', 'Storage']]
y = data['Final Price']
```

```
# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize the model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Calculate the Mean Squared Error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

# Calculate the error for each data point
errors = y_pred - y_test

# Create a color map for the error magnitude
norm = plt.Normalize(errors.min(), errors.max())
colors = plt.cm.viridis(norm(errors))

# (v) Comment the source code
# Load the dataset, clean it, and save the cleaned version
# Handle missing values
```

```
# Remove duplicates

# Load the cleaned dataset

# Define features (RAM, Storage) and target (Final Price)

# Split the dataset into training and testing sets

# Initialize the Linear Regression model

# Train the model on the training data

# Make predictions on the testing data

# Calculate the Mean Squared Error

# Visualize actual vs. predicted values with color-coded error

plt.figure(figsize=(10, 6))

plt.scatter(y_test, y_pred, c=colors, marker='o', alpha=0.7)

plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=2)

plt.xlabel('Actual Price')

plt.ylabel('Predicted Price')

plt.title('Actual vs. Predicted Laptop Prices with Error Visualization')

plt.colorbar(label='Prediction Error')

plt.show()
```

CHAPTER 3: OUTPUT

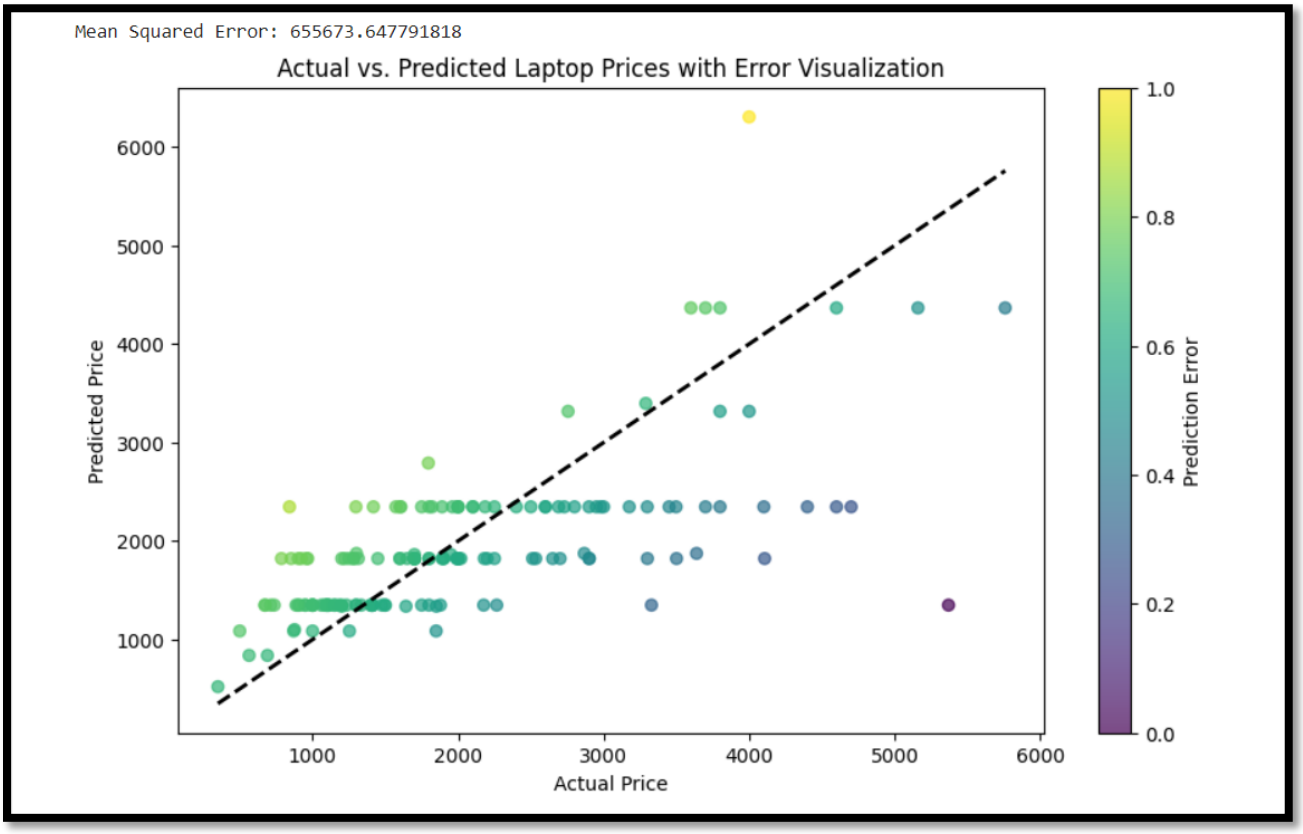

Figure 1: Output Graph

CHAPTER 4: CONCLUSION

In conclusion, this project aimed to determine the price of laptops by employing a linear regression model based on their RAM and storage capacities. Through thorough data collection, preprocessing, and feature engineering, we created a robust dataset that captured the variations in laptop prices across different RAM and storage configurations. The implementation of the linear regression model allowed us to establish a quantitative relationship between these two key specifications and the corresponding laptop prices. Our analysis revealed a clear positive correlation between higher RAM and storage capacities and increased laptop prices, confirming the intuitive understanding of their impact on overall device cost. The model's performance was evaluated using appropriate metrics, and while it demonstrated strong predictive capabilities, there may be room for further refinement through the inclusion of additional features or alternative regression techniques. This project underscores the utility of linear regression in predicting laptop prices and provides valuable insights for both consumers and manufacturers in the dynamic laptop market.

CHAPTER 5: REFERENCES

- Bibliography: Javatpoint. (2022). Learn to program. Retrieved from https://www.javatpoint.com/linear-regression-in-machine-learning (Accessed August 15 2023) In-Line Citation :(Javatpoint 2022)
- Bibliography: KnowledgeHut (2022). Learn to code. Retrieved from https://www.knowledgehut.com/blog/data-science/linear-regression-for-machine-learning (Accessed August 14, 2023) In-Line Citation :(KnowledgeHut 2022)
- Bibliography: Edureka (2022). Learn to code with free online courses. Retrieved from https://www.edureka.co/blog/linear-regression-for-machine-learning/ (Accessed August 13, 2023) In-Line Citation :(Edureka 2023)