LAPTOP PRICE PREDICTION PROJECT REPORT

GROUP H ANALYSIS

EXECUTIVE SUMMARY:

This report presents the findings and outcomes of a group project undertaken by Group H to develop a machine learning model capable of predicting laptop prices based on their hardware specifications. The project followed the complete data science workflow: data cleaning, exploratory analysis, feature engineering, model training, evaluation, and finally deployment through a user-friendly web application.

The results show that machine learning can be used to predict laptop prices with reasonable accuracy, offering value to consumers, retailers, and manufacturers alike. Beyond the technical outcomes, this project also represents an exercise in teamwork, planning, and collaboration, with each group member contributing to different stages of the work.

INTRODUCTION:

Laptops today are essential tools in education, business, and entertainment. The global laptop market is highly competitive, with brands offering a wide range of models that differ in price depending on their specifications. Features such as processor type, RAM size, storage, screen quality, and operating system all influence a laptop's price. For customers, this variety can make it difficult to judge whether a laptop is fairly priced, while sellers and manufacturers must constantly evaluate their products against competitors.

This project set out to tackle this problem by building a predictive model that estimates the price of a laptop given its key specifications. Such a system can:

- Help customers compare laptops more effectively.
- Enable retailers to set competitive prices
- Provide manufacturers with insights into how certain features affect price.

In addition to the technical goal, the project also aimed to strengthen the group's teamwork skills. Each member took responsibility for a specific part of the project while also engaging in group discussions to ensure everyone had a shared understanding of the overall process. This balance of individual responsibility and collective collaboration was crucial in delivering a polished final product.

PROJECT OVERVIEW:

Objective:

The main goal was to estimate laptop prices using the following key features:

a. Brand/Company
b. Operating System
c. RAM capacity
d. Weight
e. Screen specifications (resolution, size)
f. Storage capacity
g. Business Value
h. Consumers: Quickly check whether a laptop is overpriced or a good deal.
i. Retailers: Adjust their pricing strategy based on predicted values.
j. Manufacturers: Identify which hardware features justify higher prices.
k. Technical Architecture

The project was organized into a clear structure, ensuring that every dataset, notebook, and model was easy to locate.

WORKFLOW STAGES:

(i) Data Cleaning – fixing missing values, handling inconsistent entries.
(ii) Exploratory Data Analysis (EDA) – discovering patterns and correlations.
(iii) Feature Engineering – creating useful variables like Pixels Per Inch (PPI).
(iv) Model Building – training machine learning models and selecting the best one.
(v) Deployment – creating a web application for real-time predictions.

TECHNICAL IMPLEMENTATION:

Data Processing: The raw dataset was sourced from Kaggle. After cleaning, it was separated into features (X) and target values (y). Processing steps included:

Converting storage to consistent units (GB): Normalizing categorical features such as company and operating system.

Calculating PPI:

$$\text{PPI} = \sqrt{\frac{(screen\ width^2 + screen\ height)}{screen\ size}}$$

This feature combined resolution and size, providing a single powerful metric for screen quality.

Model Deployment: The trained model was deployed using Stream lit. The web application allows users to select laptop specifications via simple input forms and instantly see a price prediction. The app automatically calculates features like PPI in the background, reducing the chance of input errors.

APPLICATION FEATURES:

The application included:

- Company Selection: Major brands such as Dell, Apple, HP, Lenovo, Asus, Acer, MSI, and Toshiba.
- Operating System Options: Windows, MacOS, Linux, or No OS.
- Hardware Inputs: RAM (2–64 GB), weight (0.5–5.0 kg), storage (64–2000 GB), and screen size (10–20 inches).
- Automatic Calculations: Screen PPI computed automatically once dimensions were entered.

As a group, we tested the app with multiple scenarios, ensuring it worked reliably under different inputs.

TECHNOLOGY STACK:

- Data Analysis: pandas, numpy, scipy
- Machine Learning: scikit-learn, category_encoders, joblib
- Visualization: matplotlib, seaborn, missingno
- Deployment: streamlit

By using pinned dependencies, we ensured reproducibility of results across systems.

PROJECT STRENGTHS:

- **Methodology:** A systematic process from raw data to deployment.
- **Feature Engineering:** Creative features like PPI improved model performance.

- **Deployment:** The Streamlit application makes the model usable outside of a coding environment.
- **Teamwork:** Each member focused on different parts but supported integration and testing.

AREAS FOR IMPROVEMENT:

- **Documentation:** More detailed docstrings and inline comments would make the code easier to follow.
- **Model Validation:** Deeper comparisons across algorithms and improved cross-validation would increase confidence in results.
- **Data Quality:** More robust handling of missing values and outliers is needed.
- **Scalability:** Automating retraining and adding monitoring tools would prepare the project for real-world use.

GROUP CONTRIBUTIONS:

- Bassey, Kuyik (22/EG/CO/1775) - Data cleaning and preprocessing.
- John, Daniel Joshua (22/EG/CO/1767) & Oladele, Oluwatemidara David (22/EG/CO/1771) - Exploratory Data Analysis and visualization.
- Edet, Emmanuel Michael (22/EG/CO/1768) - Feature engineering, including PPI calculation.
- Joseph Prince (22/EG/CO/1774) - Model building, training, and evaluation.
- Ekpenyong, Joshua Effiong (22/EG/CO/1764) & Udofia, Aniebietabasi Aniebiet (22/EG/CO/1765) - Model Evaluation.
- All Members: Deployment testing, debugging, and final report writing.

This division of labour allowed each member to specialize while still learning from the entire process through group discussions.

CONCLUSION:

The Laptop Price Prediction Project successfully met its objectives. The group built a complete pipeline from raw data to a deployed web application. Along the way, the team learned how to clean data, engineer features, train predictive models, and design a user interface.

Equally important, this project demonstrated the value of collaboration. Each member contributed different skills, and the group worked together to integrate these pieces into a unified system. The experience improved not only technical skills but also teamwork, planning, and communication.

Key Lessons Learned

- Collaboration makes large projects manageable.
- Careful feature engineering can significantly improve model performance.
- Deployment tools like Streamlit make machine learning accessible to non-technical users.

RECOMMENDATIONS

To further improve the project, the group recommends:

- Expanding the dataset with newer models and additional features like GPU type.
- Performing more rigorous model benchmarking.
- Automating retraining to keep the model updated.
- Incorporating real-world user feedback for continuous improvement.

In conclusion, Group H’s project demonstrates how data science can be applied to solve practical problems in technology markets. Beyond the technical achievement, it stands as an example of effective teamwork, turning raw data into actionable insights through collective effort.