Laptop Price Prediction with Machine Learning

Wijetunga IU
Department of Electrical and Information Engineering
University of Ruhuna
Galle, Sri Lanka
wijetunga_iu_e22@engug.ruh.ac.lk

Senevirathna KMAN
Department of Electrical and Information Engineering
University of Ruhuna
Galle, Sri Lanka
senevirathna_kman_e22@engug.ruh.ac.lk

***Abstract*— This research project addresses the essential of understanding and analyzing laptop prices for enhancing the customer and seller experience. Using machine learning methodologies, this focuses on sentiment analysis to predict the laptop prices. The dataset consists of 1303 samples from Kaggle. It is subjected to a robust pre-processing pipeline, including normalization and vectorization. Implementing machine learning algorithms, particularly Linear Regression and Random forest the analysis yields promising results in accurately predicting prices. Key findings underscore the major factors that influencing the laptop prices and optimizing those factors to have accurate values. Ethical considerations, such as user privacy and potential biases, are discussed. Overall, the project contributes to the society by creating a platform that predict laptop prices and help them to choose most suitable laptop for their own requirements and budget.**

Keywords*— *laptop price prediction, machine learning, linear regression, random forest

I. Introduction

The world of laptops is rapidly evolving with the technology and it creates vast range of choices for consumers to buy one. Therefore, in a world flooded with laptop choices, understanding the factors influencing prices is crucial for consumers and the tech industry. By identifying that our project aims to predict laptop prices based on those key factors.

A. Background

With the current economic situation and import/export limitations in Sri Lanka, it is very hard to decide what is the most suitable laptop to buy in a certain price range. Mostly university students are verdicts of this situation and even for the sellers it is very difficult to predict the prices of the laptops. As a solution we decided to create a platform that can help to predict the laptop prices.

B. Importance

The importance of predicting laptop prices cannot be taken slightly. The project will empower Sri Lankan consumers specially university students by providing them with a tool to make predictions about laptop prices aligned with their requirement and budget. This will not only enhance customer experience but also the business providers experience as well. And as for businesses, this will help them to understand the factors driving laptop prices and it will be crucial for optimize marketing strategies and staying competitive with the price.

C. Project Goals

The primary goals of this research project are developing a predictive model with identification of key factors that decide laptop prices to enhance consumer and business experience. Specially, the project aims to:

Developing a Predictive Model: Create and implement machine learning algorithms, particularly Linear Regression and Random Forest, to develop an accurate predictive model for laptop prices.

User and Business Empowerment: Empower Sri Lankan consumers and businesses, with a user-friendly platform to predict laptop prices aligned with their needs and budget and to keep up with the competition of the fellow laptop businesses.

Enhancing Consumer Experience: Improve the overall purchasing experience by providing a user-centric tool that simplifies laptop choices, aiding users in making informed decisions.

Contributing to Society: Contribute to society by creating a platform that not only predicts laptop prices but also helps users choose the most suitable laptop for their individual needs and budget.

D. Algorithm Choice

In the pursuit of these project goals, the project strategically adopts two robust machine learning algorithms: Linear Regression and Random Forest. Tailored for regression analysis tasks, these algorithms excel in handling regression problems effectively.
Linear Regression: Linear Regression is chosen due to its simplicity and interpretability, making it well-suited for understanding the linear relationships between various features and laptop prices. It serves as a baseline model for the project, providing insights into how individual features contribute to price predictions.
Random Forest: Random Forest is employed for its ability to handle non-linear relationships, capture complex patterns in the data, and resist overfitting. By combining multiple decision trees, Random Forest enhances the predictive power of the model.
The choice of these algorithms aligns with the project's goal of accurately predicting laptop prices while considering the diverse and potentially non-linear relationships among various features. This combination allows for a comprehensive understanding of the factors influencing laptop prices.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. METHODOLOGY

A. Data

Dataset Description: The dataset utilized in this research project was sourced from Github, a prominent platform for data science and datasets. The dataset, titled "Laptop Price Dataset," consists of 1303 samples, each containing unique features essential for price prediction.
Dataset Source: Github
Dataset Link: https://github.com/uvishka/Laptop-Price-Predictor/blob/main/laptop_price.csv
Size: The dataset comprises 1303 samples, providing a substantial volume of data for price prediction.
Content: The dataset includes the 12 key features: Laptop ID, Company, Product, TypeName, Inches, RAM, Weight, Screen Resolution, CPU brand, GPU brand, OpSys

Data Cleaning and Preprocessing:

- Handling Missing Data**:**

There wasn't any missing data or null data in the dataset. Therefore, it doesn't need to handle missing data

- Data Preprocessing (Pre-processing)

There were 12 feature columns in the data set. Therefore, it needed to be preprocessed before building the model.

B. Pre-processing

Category Consolidation:

- Simplify the columns to make the model less complex.
- Created a category called other for company, product, RAM, CPU, GPU and OpSys and items that have less than 10 assigned to that other category.
- Screen Resolution divide into two categories called IPS and TouchScreen and make it as a tick box.
- Removed the Laptop_Id category as it is not a influencing factor for laptop prices
- Change the type of weight and RAM category to float.

Encoding:

- Used one hot encoding to convert all words type column to numerical type columns.

Feature Scaling (Normalization):

- Used MinMax Scaler to normalize the weight and Ram categories as others were already in 0s and 1s.
- For random forest algorithm it doesn't need to do feature scaling but used it for linear regression.

The chosen preprocessing techniques aim to create a normalized and clean dataset with numerical values, reducing complexity and irrelevant information. And collectively contribute to a focused analysis for an accurate model building.

C. Algorithm

The project initially considered two prominent algorithms for price prediction: Linear Regression and Random Forest. Subsequent evaluation led to the selection of Random Forest based on accuracy of RandomForestRegressor.

- Linear Regression – 0.71
- Random Forest – 0.79

The evaluation results played a pivotal role in algorithm selection, with Random Forest demonstrating superior accuracy. The absence of explicit model tuning suggests that the default configurations of both algorithms were sufficient for achieving satisfactory results. Random Forest's complex pattern identification and robustness in regression tasks make it a suitable choice for prediction in the context of laptop price prediction.
GridsearchCV used as the hyperparameter tuner for the Random Forest. Out of absolute error, squared error and poisson absolute error chosen as the best criterion. And the best n_estimator was 100 out of 10,50 and 100. The best model and best parameters were chosen in above and used it in the implementation of the project.

D. Implementation

Programming Languages, Libraries, and Tools:
The implementation phase was executed using the following technologies:
Programming Language:

- *Python:* Selected for its versatile ecosystem of libraries and frameworks tailored for machine learning tasks.

Libraries:

- *NumPy:* Applied for efficient numerical computations and array operations.
- *Pandas:* Employed for seamless data manipulation and analysis.
- *scikit-learn:* Integrated for streamlined implementation of machine learning algorithms and evaluation metrics.
- *Matplotlib:* Utilized to create visually informative plots for data visualization.

Data Preprocessing:

- *Sklearn MinMaxScaler:* It scales features to a specified range (usually between 0 and 1) by linearly transforming them.

Model Training and Evaluation:

- Leveraged the scikit-learn library for the efficient implementation of machine learning algorithms, specifically Linear Regression and Random Forest, for model training and evaluation.

Model Serialization:

- Serialize the trained model into a format suitable for storage and subsequent loading. Commonly, this is achieved using libraries such as *pickle* in Python.

III. RESULTS

CVGridSearch:
Best Model was RandomForestRegressor with the absolute error criterion.
Best parameter was absolute error criterion with n_estimator 100.

```
Best Model: RandomForestRegressor(criterion='absolute_error')
Best Parameters: {'criterion': 'absolute_error', 'n_estimators': 100}
```

Figure 1-Best Model and Best Parameters

Hyperparameter Tuning:

```
#Defining the best model got with hyperparameter tuning
best_model.score(x_test,y_test)
```

```
0.7893810029840271
```

Figure 2 - Accuracy with the hyperparameter tuning

With the best model hyperparameter tuning the accuracy came up for 0.79 for Random Forest algorithm with absolute error criterion with n_estimator 100.

Analysis and Discussion:

```
Index(['Ram', 'Weight', 'Touchscreen', 'IPS', 'Company_Acer', 'Company_Apple',
       'Company_Asus', 'Company_Dell', 'Company_HP', 'Company_Lenovo',
       'Company_MSI', 'Company_Other', 'Company_Toshiba',
       'TypeName_2 in 1 Convertible', 'TypeName_Gaming', 'TypeName_Netbook',
       'TypeName_Notebook', 'TypeName_Ultrabook', 'TypeName_Workstation',
       'OpSys_Linux', 'OpSys_Mac', 'OpSys_Other', 'OpSys_Windows',
       'cpu_name_AMD', 'cpu_name_Intel Core i3', 'cpu_name_Intel Core i5',
       'cpu_name_Intel Core i7', 'cpu_name_Other', 'gpu_name_AMD',
       'gpu_name_Intel', 'gpu_name_Nvidia'],
      dtype='object')
```

```
best_model.predict([[8,1.3,1,1,0,1,0,1,0,0,0,0,0,0,1,0,0,1,0,0,1,0,0,0,0,1,0,1,0,0]])
```

```
/Users/isitha/anaconda3/lib/python3.11/site-packages/sklearn/base.py:464: UserWarning: X d
  warnings.warn(

array([2143.57683333])
```

Figure 3 - Random prediction with the best model

IV. DISCUSSION

A. Discussion

The project's core findings highlight the efficacy of machine learning algorithms, specifically Linear Regression and Random Forest, in predicting laptop prices. The deliberate choice of these algorithms is validated by their accuracy in capturing the intricate relationships between various features and the ultimate price outcomes. Examination of mispredictions offers valuable insights, indicating areas for potential model refinement and optimization.

Implications for E-commerce Businesses:
The outcomes of the laptop price prediction model have substantial implications for both consumers and businesses. Empowering consumers with accurate price predictions aligns purchases with individual needs and budgets. For businesses, understanding the driving factors behind laptop prices optimizes marketing strategies and ensures competitiveness in the market. The model contributes to an enhanced decision-making process for both parties.

Ethical Aspects:
In the realm of predicting laptop prices, ethical considerations are paramount. Upholding user privacy, minimizing biases in the prediction model, and ensuring fairness in pricing are critical aspects. But the privacy of the data is not a major concern as it is laptop prices. But the data integrity should be a major concern as people's choices depend on this laptop prices. Continuous monitoring and ethical oversight are integral to building trust and transparency in providing accurate laptop price predictions.

B. Conclusion

In conclusion, this project illuminates the significant impact of machine learning algorithms, particularly Linear Regression and Random Forest, on predicting laptop prices. The careful selection of algorithms based on their accuracy demonstrates their efficacy in extracting meaningful insights from the dataset. While achieving notable results, the project acknowledges opportunities for further refinement, particularly in addressing mispredictions and upholding ethical considerations.

Future Research Directions:
Future research directions should delve into advanced machine learning techniques, explore domain-specific features, and consider ensemble methods to further refine the accuracy of laptop price predictions. Additionally, investigating the influence of external factors, such as economic trends or technological advancements, on laptop prices could provide valuable insights for a more nuanced predictive model.

This project serves as a foundation for ongoing advancements in predicting laptop prices, paving the way for more sophisticated methodologies. By combining technical rigor with ethical considerations, the laptop price prediction model can continue evolving to meet the dynamic demands of the

market, fostering trust and satisfaction among users and businesses alike.

V. REFERENCES

[1] "Laptop Prices Prediction," *kaggle.com*. https://www.kaggle.com/code/danielbethell/laptop-prices-prediction (accessed Jan. 21, 2024).

[2] M. A. Shaik, M. Varshith, S. SriVyshnavi, N. Sanjana, and R. Sujith, "Laptop Price Prediction using Machine Learning Algorithms," *IEEE Xplore*, Nov. 01, 2022. https://ieeexplore.ieee.org/document/10093357

[3] A. Kumar, "Random Forest for prediction," *Medium*, Jun. 22, 2020. https://towardsdatascience.com/random-forest-ca80e56224c1

[4] Zach, "How to Make Predictions with Linear Regression," *Statology*, Jul. 27, 2021. https://www.statology.org/predictions-regression/

[5] G. Lawton, "Data Preprocessing: Definition, Key Steps and Concepts," *SearchDataManagement*, Jan. 2022. https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing