LAPTOP PRICE PREDICTOR

TIRTH SHAILESH VORA

CHAPTER 1: INTRODUCTION

Selecting the right laptop within one's budget and requirements can be challenging due to wide range of models and prices available. To address this problem, we have developed a laptop price predictor using machine learning.

Our Project aims to assist consumers in selecting the most suitable laptop by providing accurate price predictions based on the laptop's features. By using our predictor, consumers can save time and money by narrowing down their options and making informed decisions.

CHAPTER 2: DATA DESCRIPTON

We have downloaded our data from Kaggle. It consists of 1300 unique combinations of laptop with 12 different features. It contains features like:

Company (Apple/Lenovo etc.)

TypeName (Notebook/Ultrabook/Gaming etc.)

Inches, **Screen Resolution** (IPS, LCD, Touch screen, LED etc.)

CPU (The Processor)

Ram, **Memory** (Storage Capacity), **OS GPU,**

Weight and **Price**.

Our Target column is **Price** and it is our class label which is a continuous value here.

Initially, almost all the features are of type 'object' and unknown to the interpreter, hence we will have to perform lots of preprocessing over it to make ready for numerical operations, descriptive statistics and even feeding the data to the machine learning model.

CHAPTER 3: DATA ANALYSIS

As discussed above, we need our data to be numerically computable. Hence, we started with stripping 'GB' from Ram column and converting it to an integer type, similarly we converted weight to float type.

Then for the Screen Resolution column, we split the column into two, extracted X and Y_res, and combined it with inches to create a better performing feature called 'PPI' (Pixels Per Inch). We dropped the Screen Resolution column thereafter. In a similar way, we selected first 3 words from the CPU column and created 3 most common categories and all the values will lie in these 3 with the help of function NameProcessor.

Moving forward, we checked the correlation between price and different columns and confirmed how strongly does a feature contribute in influencing the price. We found that laptops mostly belonged to Dell, Lenovo and HP. Notebooks are the most common type.

Workstations and Gaming laptops are most expensive. Touchscreen and IPS are more expensive then their counterparts.

Then using df.corr(), after feature engineering as we separated a column into its sub-categories, we analyzed its correlation with price and dropped multiple columns which had a negative or very poor correlation with price.

CHAPTER 4: DATA MODELLING

As we saw in our Notebook, the distplot which contained a histogram and a probability distribution curve was skewed. Hence we fixed it by taking its log and plotting it again, now we separated it as X_Train, X_Test, Y_train, _Y_test.

We implement a data modelling pipeline that performs feature engineering using a ColumnTransformer and uses a LinearRegression model to predict laptop prices. The ColumnTransformer applies a OneHotEncoder to select categorical columns and encodes them into numerical form, while allowing non-categorical columns to pass through unchanged. The pipeline consists of two steps: step1 (ColumnTransformer) and step2 (LinearRegression model). The pipeline is fit to the training data and used to generate predictions. This approach demonstrates the use of data modelling techniques to create an efficient and accurate machine learning pipeline for predicting laptop prices.

The model fitting step is performed using the RandomForestRegressor class from the scikitlearn library. The model is initialized with hyperparameters including the number of estimators, the maximum number of samples, the maximum number of features, and the maximum depth of the tree. The Random Forest Regressor algorithm then fits the model to the data, generating a predictive model that can be used for generating price predictions.

The performance of the model is evaluated using two metrics: the R2 score and the mean absolute error. The R2 score measures the proportion of the variance in the target variable that is explained by the model with a higher score indicating better performance. The MAE measures the average difference between the predicted and the actual prices with a lower score indicating better performance.

CHAPTER 5: CONCLUSION

In conclusion, the laptop recommender system developed in this project demonstrates the effectiveness of machine learning in predicting the price of laptops based on specific features. The project involved several key steps, including data preprocessing, feature engineering, and the application of different machine learning models.

Through the use of OneHotEncoder and ColumnTransformer classes, categorical features were encoded and non-categorical features were transformed, resulting in a more informative dataset for use in model training. The use of multiple models, including Linear Regression, Ridge, Lasso, and Random Forest Regressor, allowed for a thorough exploration of the best fitting model for the given data.

The Random Forest Regressor model ultimately proved to be the most effective in predicting laptop prices, achieving an R2 score of 0.87 on the test set. This indicates that the model has a high level of accuracy and can provide reliable predictions on the price of laptops based on the input features.

Overall, the developed laptop recommender system has the potential to be a valuable tool for both consumers and manufacturers in making informed decisions regarding the pricing and features of laptops. The insights gained from this project can also be applied to other fields requiring machine learning-based prediction, further contributing to the advancement of the field.