**CS 383- Spring 2022**

**Final Project Submission: Result and Actual Steps**

1. **Please specify the objectives of your project:**

|  |
| --- |
| The primary objective of this project is to predict laptop prices while testing with various machine learning models to choose the best one for this application. |

1. **Please list the tools that were used in your project:**

|  |
| --- |
| Python and Jupyter Notebook |

1. **Please state the results achieved by your project:**

|  |
| --- |
| The end result contains laptop predictions performed using models XGB Regressor, Linear Regression, Random Forest Regressor, and Decision Tree Regressor, with their respective model scores. |

1. **Please list at least 5 published papers in the same area (your own final project area). The papers must be published in a peer reviewed journal and/or conferences. For each paper give its main feature, drawback and strength.**

|  |
| --- |
| * **Literature Review: for each published work make sure to add the author names, the title, the publisher , doi, and the publication year (**The two most popular academic writing styles that you may run across are MLA and APA). |
| 1. Prof. Vaishali Surjuse; Sankalp Lohakare; Aayush Barapatre; Abhishek Chapke. Laptop Price Prediction using Machine Learning, KDKCE & RTMNU University, India. 2022 Jan. DOI: 10.47760/ijcsmc.2022.v11i01.021 |
| 1. Raghav Agrawal. Laptop Price Prediction – Practical Understanding of Machine learning project lifecycle. <https://www.analyticsvidhya.com/blog/2021/11/laptop-price-prediction-practical-understanding-of-machine-learning-project-lifecycle/>. 2021 Nov |
| 1. Y R Pratama; Sufa Atin; Irawan Afrianto. Predicting Student Interests Against Laptop Specifications Through Application of Data Mining Using C4.5 Algorithms. 2019 Nov. DOI: [10.1088/1757-899X/662/2/022129](http://dx.doi.org/10.1088/1757-899X/662/2/022129) |
| 1. Ishan Pandey; Mudimela Sathish Kumar Reddy; Mukkamalla Dharani Vamsidhar Reddy; Karumoju Krishna Babu; Harshpreet Kaur. Machine Learning Based Future Priced Prediction of the Products. School of Computer Science and Engineering Lovely Professional University Phagwara, India. 2021 Apr. e-ISSN: 2395-0056. p-ISSN: 2395-0072 |
| 1. Ayesha Ayub Syed; Yaya Heryadi; Lukas; Antoni Wibowo. A Comparison of Machine Learning Classifiers on Laptop Products Classification Task. Proceedings of the International MultiConference of Engineers and Computer Scientists 2021. 2021 Oct |

1. **From the above published papers, select the best one depending on your review.**

I specifically chose to refer to the second paper mentioned above, because data preprocessing is a tedious task as the dataset contains a lot of categorical variables and referring to this paper made the preprocessing a little bit easier by helping me with giving leads about which features that need to be preprocessed and few of them that need to be encoded.

1. **Please list the main tasks of your project, and the steps within each task that you have already taken to achieve your results. Please try to add a screenshot of each step. For example, your steps might look like this:**

**Task 1: Downloading the data**

* [Dataset source](https://www.kaggle.com/datasets/muhammetvarl/laptop-price/download?datasetVersionNumber=1)
* Overview of how the dataset looks like with the columns and respective data,

**Graphical user interface, application, table, Excel

Description automatically generated**

**Task 2: Preprocessing data**

* Data preprocessing involves encoding the dataset with ‘latin-1’, removing null values, converting object data type to the feature’s required data types, and generating columns that are relevant for training like IPS, PPI, Touch screen, CPU brand, CPU speed, Primary memory and its type, Secondary memory and its type, and dropping columns which aren’t required.
* Few screenshots of data preprocessing performed as part of the project,

**Graphical user interface, text

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

**Table

Description automatically generated with low confidence**

**Table

Description automatically generated with medium confidence**

**Task 3: Plotting visualizations to understand features correlation**

* Visualizations were plotted between features like – RAM Vs Price, Primary Memory Type Vs Price, CPU brand Vs Price, RAM Vs Company
* Few screenshots of visualizations generated,

**Chart, bar chart

Description automatically generated**

**Chart

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**Task 4: Using Functions to train and predict**

* I chose to compare four machine learning models – XGB Regressor, Decision Tree Regressor, Linear Regression and Random Forest Regressor
* A function was created to encode the categorical variables to integers
* So, functions were created for fitting the training data into these models to be able to predict them and compare their scores
* Few screenshots of functions created,

**Table

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated**

**Application

Description automatically generated with medium confidence**

1. **Compare your results with the result achieved results by the best solution of step 5.**

|  |  |
| --- | --- |
| **Your own solution** | **The state of art solution** |
| Four machine learning models were compared:  XGB Regressor = 91%  Random Forest Regressor = 88%  Linear Regression = 69%  Decision Tree Regressor = 75%  As we could see XGB Regressor has the highest model score, so, it was declared as the best choice of model for predicting laptop prices. | The paper used Random Forest Regressor but with more intricate changes in parameters present in functions which involves controlling n\_estimators, max\_depth, max\_samples and few more parameters. Author had also applied transformers to columns while encoding the categorical variables to integers. The R2 score of the model was 88.6% |

1. **Please highlight the future works to improve your work:**

|  |
| --- |
| The best paper among the papers listed had developed a web application with all the features relevant to laptops listed out, each of them can be selected with respect to user’s requirements and the predicted price for the configuration will be listed. I would like to work on this in the future and will try to enhance this furthermore by involving much more features like the type of users and filtering specifications with respect to budget. |