Laptop Price Prediction

V. Sarala[1], Padaga Sai Pavan Kumar[2]

[1] **Assistant Professor MCA, DEPT,** Dantuluri Narayana Raju College, **Bhimavaram, Andharapradesh**
Email id:- vedalasarala21@gmail.com

[2]**PG Student of MCA,** Dantuluri Narayana Raju College, **Bhimavaram, Andharapradesh**
Email id :- padagasaipavankumar@gmail.com

ABSTRACT

The laptop price predictor project is a project that aims to predict the price of laptops. The project will be divided into 3 parts, each having its own specific tasks. The first part is to create a model that predicts the price of laptops based on various factors such as the size of screen and CPU speed. The second part is to test this model on real data collected from different websites. Finally, the third part is to present our results and discuss how we built this model in order to make it more accurate. Laptop price predictor is a tool which predicts the price of laptops. it consists of a series of algorithms that predict the price of laptops on the basis of their features and specifications. The results obtained by this project are in close agreement with those obtained using other prediction methods such as neural network and support vector machine (SVM).

1 INTRODUCTION

The laptop price predictor project is a very interesting project that you can use to predict the price of laptops. This will help you in saving money and time, because you don't need to go to different stores and check prices every time you want to buy a new laptop. The project will be divided into three parts: preprocessing, training and testing. The pre- processing process takes care of cleaning up the data, which is done by removing duplicate rows and null values. The training process consists of creating a model based on the data that has been collected and then using this model to predict prices for new laptops.

Finally, we test our model on new data sets to see if it can accurately predict how much a laptop costs. Laptop price prediction especially when the laptop is coming direct from the factory to Electronic Market/ Stores, is both a critical and important task. The mad rush that we saw in 2020 for laptops to support remote work and learning is no longer there. In India, demand of Laptops soared after the Nationwide lockdown, leading to 4.1-Million-unit shipments in the June quarter of 2021, the highest in the five years. Accurate Laptop price prediction involves expert knowledge, because price usually depends on many distinctive features and factors. Typically, most significant ones are brand and model, RAM, ROM, GPU, CPU, etc. In this paper, we applied different methods and techniques in order to achieve higher precision of the used laptop price prediction.

Machine learning is a branch of Artificial intelligence that deals with implementing applications that can make a future prediction based on past data. If you are a data science enthusiast or a practitioner then this article will help build your own end-to-end machine learning project from scratch. There are various steps involved in building a machine learning project but not all the steps are mandatory to use in a single project, and it all depends on the data. In this article, we will build a laptop price prediction project and learn about the machine learning project lifecycle. For the automation of classification tasks, Machine Learning (ML) offers promising methods and algorithms.

Classification in machine learning is a learning problem where a system learns to predict class labels on a set of

data points. As a supervised learning problem, target class labels are also provided as an input to the classification algorithm. Classification can either be binary or multiclass. In binary classification, there are two classes to be predicted while multiclass classification problems involve predicting more than two classes. Some of the machine learning algorithms available for classification include Support Vector Machines (SVM), Decision Tree, Naïve Bayes, K-Nearest Neighbor, Multi-Layer Perceptron (MLP) [3]. The classification algorithms perform differently on different datasets. The performance of the classifier depends on the application, choice of features as well as nature of the dataset This research work focuses on the multiclass classification task for a laptop products dataset. Here, the term 'Laptop Products Classification' refers to the categorization of laptop products in three classes namely, 'Budget', 'Mid-range' and 'Flagship' products.

The inputs to the classifier are features like laptop company, product type, size, weight, RAM, and price. The classifier predicts whether the given product belongs to 'Budget', 'Midrange' or, the 'Flagship' class. Various machine learning classifiers including Support Vector Machines (SVM), Multinomial Logistic Regression, Decision Tree, and Artificial Neural Network (ANN) are used to predict the class of laptop products. The performance of classifiers is compared in terms of classification accuracy, recall, precision, and F1- score the significance of this work is the better management of laptop products on an e-commerce website. From the user's point of view, it helps customers to find the required laptop product easily, efficiently, and according to their financial budget. It eliminates the need to scroll through hundreds of products to find the required one. For the business, it is beneficial because the smooth navigation and shopping experience are likely to bring the customer again to the website in the future, increasing the business sales. It also helps the business to manage the products more effectively resulting in increased productivity. The same research idea is also applicable to the categorization of smartphone products, tablets, and smartwatches, etc.

2.LITERATURE SURVEY AND RELATED WORK

Machine Learning (ML) is a field that is based on concepts and principles from multiple disciplines including Mathematics, Computer Science, Statistics, Cognitive Science, and Optimization Theory [4]. ML tasks are categorized into supervised learning, unsupervised learning, and reinforcement learning. From the supervised learning category, classification and regression are well-known tasks. In classification, the output is discrete e.g., class labels while output in regression takes on continuous values [3]. Researchers in [5] implemented the Multinomial Naïve Bayes algorithm for catalog classification.

The products are categorized into classes like 'Electronics' and then subclasses like 'printer'. The number of products is about 40,000 collected from different databases including Amazon, Flipkart, etc. An overall number of 1000 classes for 40,000 products in the system. Multinomial Naïve Bayes is mostly applied for document classification. The foundation of the Naïve Bayes Classifier is the Bayes theorem based on probability. Naïve Bayes assumption about features is that the features are independent. With X as a feature vector of size n, and y as the class label variable, Naïve Bayes predicts the class label as, the work of [6] is based on the use of the Naïve Bayes classifier and the Decision Tree classifier to predict the classes of mobile phones with given features as 'Economical' or 'Expensive'.

The researchers collected the dataset from GSMArena.com. The features collected in the dataset include display size, weight, thickness, internal memory, camera, video quality, RAM, and battery. Two feature selection algorithms InfoGainEval and WrapperattributEval were applied to select the features that are most important in predicting the output class. The results are compared across the classifiers in terms of accuracy achieved with the selection of minimum features. The Decision Tree algorithm is a popular supervised learning algorithm that works well with classification and regression tasks. The algorithm models a tree-like flowchart that has a root node, decision nodes, branches, and leaf nodes. The algorithm divides the data into small parts to identify the patterns that can be used for making a prediction.

The learning strategy behind decision trees is the divide and conquer strategy. The entire dataset is at the root node. The algorithm chooses the feature that best predicts the target class. The entries are divided into groups of feature values. This decision creates the first set of branches. The divide and conquer process continues on the nodes until a stop criterion is reached [7]. The popular decision tree algorithms include ID3, C4.5, and CART algorithm [8].worked on product categorization on a dataset collected from Amazon distributers, using machine learning classifiers including Naïve Bayes, K-Nearest Neighbor, and Tree Classifier. Features for each item were determined using the bag-of-words model. The features set was processed using the standard pre-processing techniques like stop word removal, punctuation and number removal, lowercasing, and lemmatization. After feature processing, feature importance was determined using a modified MI formula and finally, the features were selected using the forward and backward search strategies. Naïve Bayes finished with 76.9% accuracy, KNN resulted in 69.4% accuracy, and the tree classifier performed the best with 86% accuracy but the execution took a long time (8 hours) to complete as compared to Naïve Bayes (3 seconds) and KNN (4 minutes). Support Vector Machine (SVM) is another powerful machine learning algorithm for solving classification and regression problems. It has been reported to have outperformed other supervised machine learning algorithms and has become quite popular for classification in recent years due to its good generalization ability [10]. SVMs are focused on finding a hyperplane in an n-dimensional feature space that separates/classifies the data points. The algorithm chooses the hyperplane with the maximum margin so that the future data points can be classified more accurately.

The points closest to the hyperplane are the support vectors and these vectors help in maximizing the margin of the hyperplane. To classify data that is not linearly separable, SVM has a technique known as the kernel trick. The kernel function takes a low dimensional input feature space and transforms it to a higher dimensional space. During this process, several complex data transformations take place to classify the data based on the output labels provided. [11] implemented the KNN algorithm using several distance measures like Euclidean distance, Manhattan distance, and Chebyshev distance. The dataset utilized by the researchers is the KDD dataset. It is Knowledge Discovery and Data mining dataset with 41 features and class labels as ‘normal’ or ‘attack’. The dataset is used in Intrusion Detection System (IDS). The results were evaluated using accuracy value, sensitivity and specificity measures, and FPR (False Positive Rate) and FNR (False Negative Rate). The results demonstrated the performance of Manhattan distance to be superior as compared to other distance metrics on the KDD dataset.

The researchers in [12] used Artificial Neural Networks (ANN) for automation of the classification of water quality. The dataset was obtained from the laboratory measurements and included environmental factors like pH, chemical oxygen demand, biological oxygen demand, dissolved oxygen, total suspended solids, and ammonia. The classification accuracy of 80% with an RMSE value of 0.468 was reported. [13] used ANN as well as a hybrid ANN-Bat Optimization Algorithm for the classification of medical diagnosis. The results indicated that the proposed ANN combined with the metaheuristic Bat algorithm performed better in terms of accuracy.

3 PROPOSED WORK AND ALGORITHM

This project introduces the calculation of correlation coefficient on the basis of RNN algorithm, so that the algorithm can consider the weight of each data dimension and eliminate the influence of R value to a certain extent by statistical method. On the basis of the improved RNN algorithm to classify the range of commodity prices, the prediction value of commodity prices is given by using decision tree regression algorithm. The accuracy of this result is better than the direct application of decision tree regression

4.RESULTSANDDISCUSSION SCREENSHOTS

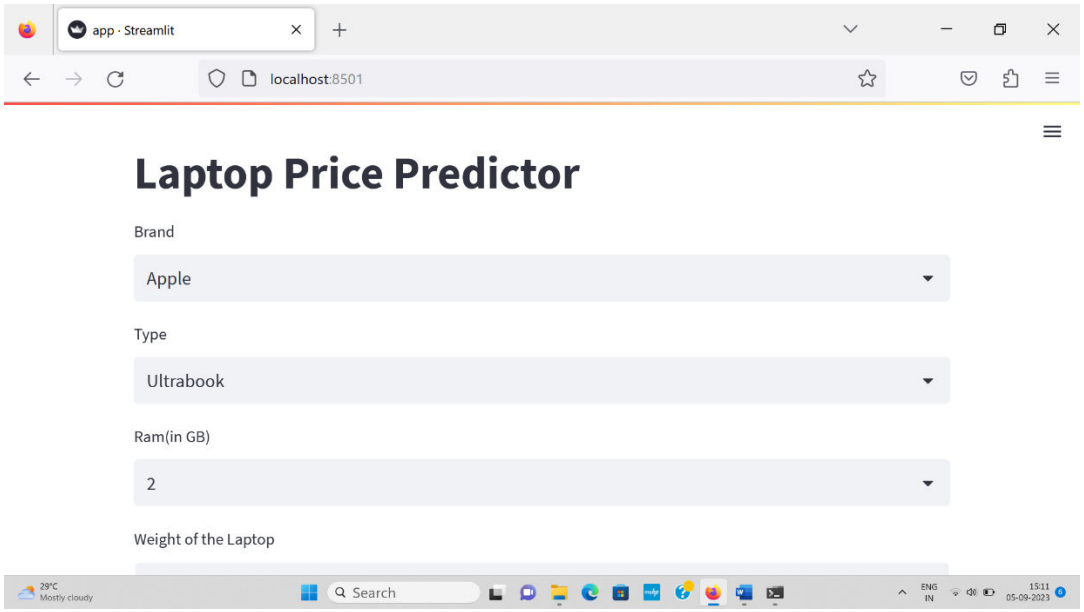

Fig 1:- Interface of laptop price predictor

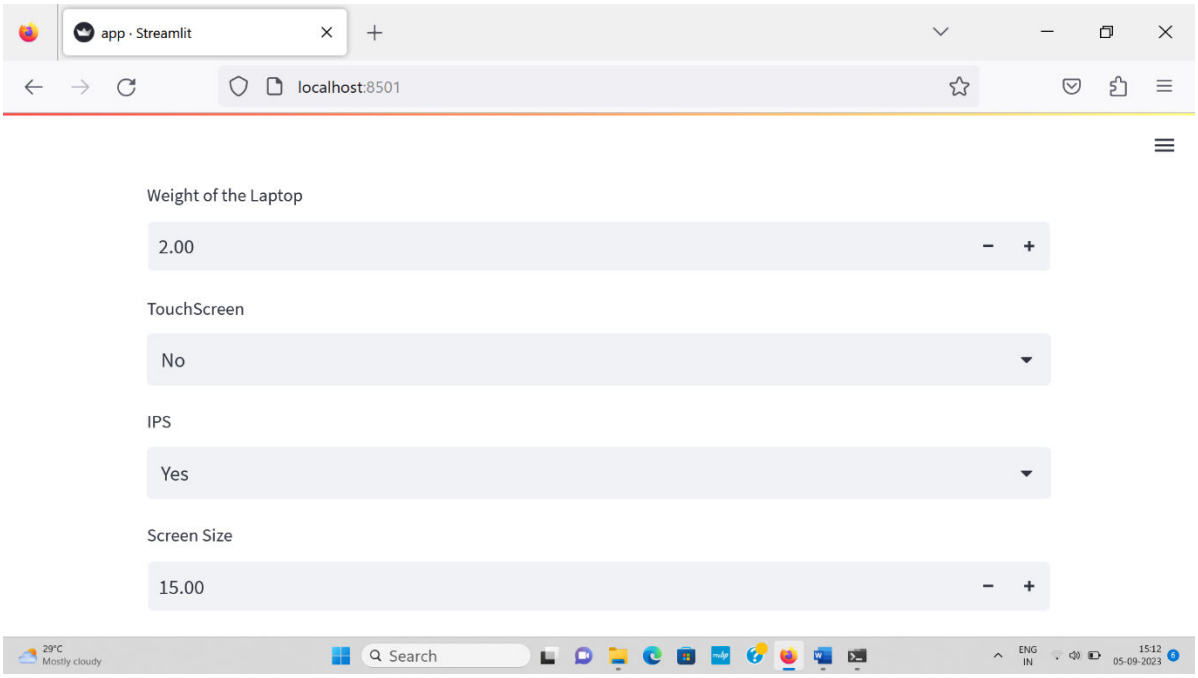

Fig 2:- Filling required fields

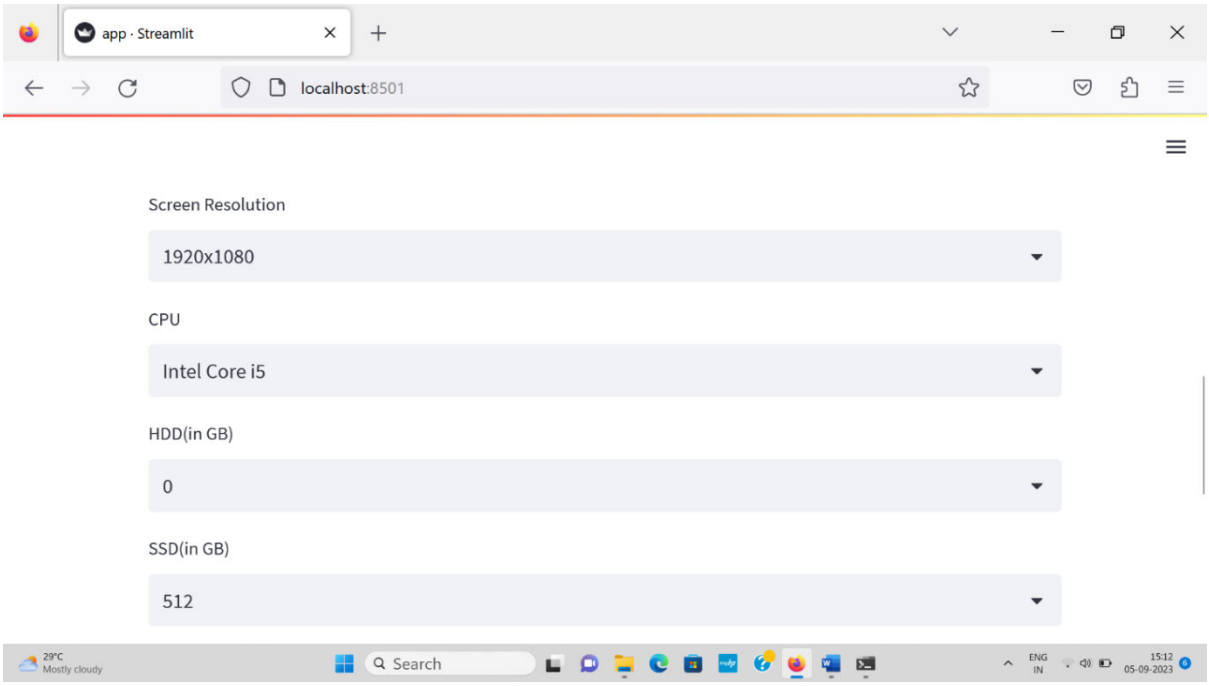

Fig 3 :- filling required fields

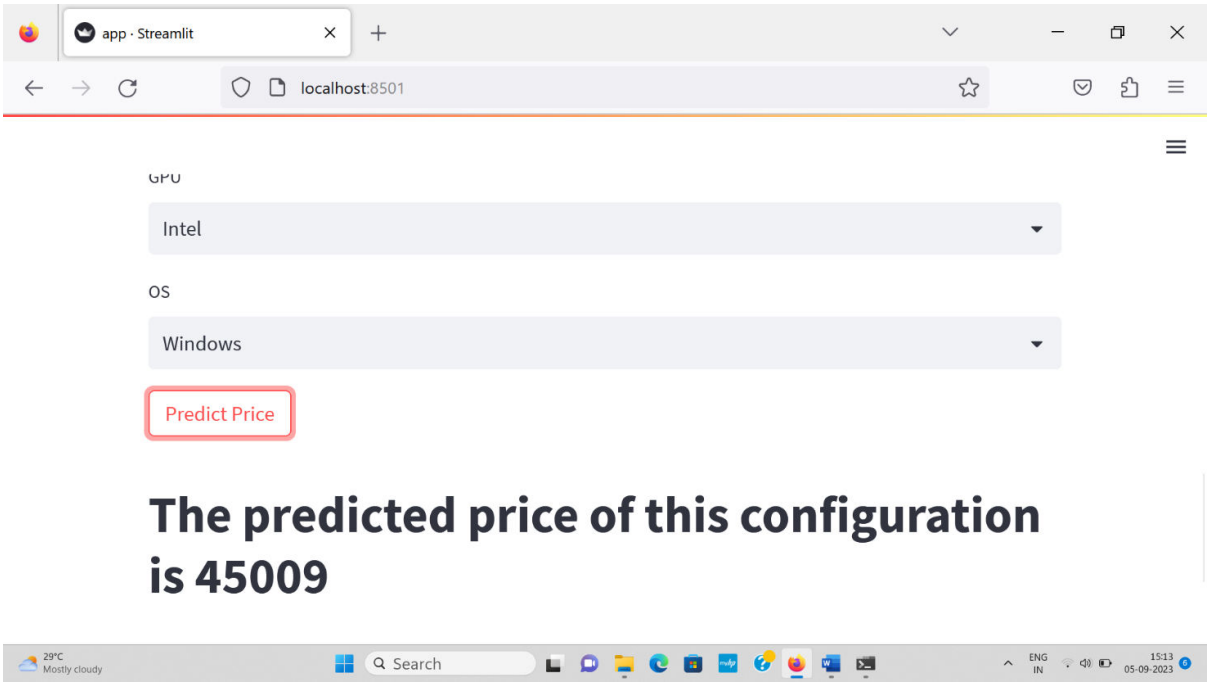

Fig 4 :- Filling all the details and click on enter to get the estimated cost of laptop.

5. CONCLUSION

In this project, the commodity data set collected by crawlers is especially used. After data preprocessing, the improved KNN classification algorithm is firstly wont to classify commodity prices. This classification model will provide merchants with the worth trend of secondhand notebooks within the current period, which is of great significance for merchants to optimize commodity prices reasonably. Then, supported the improved KNN classification, the choice tree regression is employed to work out and predict the worth of commodities. This price model is more accurate than that given directly by the choice tree regression. However, the info set collected during this paper is little, the gathering of commodity categories is incomplete, and therefore the model used still has room for improvement from parameters. within

the next step, we will consider expanding the info set and adopting better parameters.

6. REFERENCES

[1] M. Sinha, L. N. Fukey, and S. Likitha, "Web user Experience and Consumer behavior: The Influence of Color, Usability, and Aesthetics on the Consumer Buying Behavior," Test Engineering and Management, vol. 82, pp. 16592-16600, 2020

[2] A. A. Syed and J. Suroso, "Factors Affecting Consumers' Decision for E-Hotel Booking," CommIT (Communication and Information Technology Journal), vol. 12, no. 2, pp. 111-123, 2018, DOI: 10.21512/commit.v12i2.4917

[3] A. Soofi and A. Awan, "Classification Techniques in Machine Learning: Applications and Issues," Journal of Basic and Applied Sciences, vol. 13, pp. 459–465, 2017, DOI: 10.6000/1927- 5129.2017.13.76

[4] E. Alpaydın, "Introduction to Machine Learning, 2nd ed., The MIT Press, ISBN: 9780262012430, 2009

[5] S. Pandey, M. Supriya, and A. Shrivastava, "Data Classification Using Machine Learning Approach," Intelligent Systems Technologies and Applications, vol. 683, 2018, DOI: https://doi.org/10.1007/978-3-319-68385-0_10

[6] M. Asim and Z. Khan, "Mobile Price Class prediction using Machine Learning Techniques," International Journal of Computer Applications, vol. 179, no. 29, pp. 6–11, 2018, DOI: 10.5120/ijca2018916555

[7] B. Lantz, "Divide and Conquer – Classification Using Decision Trees and Rules," Machine Learning with R, 2nd ed., 2015

[8] H. Sharma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," International Journal of Science and Research, vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954

[9] S. Shankar and I. Lin, "Applying Machine Learning to Product Categorization," Computer Science Project, Stanford University, 2011.

[10] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges, and trends," Neurocomputing, vol. 408, pp. 189-215, 2020, DOI: 10.1016/j.neucom.2019.10.118

[11] P. Mulak and N. Talhar, "Analysis of Distance Measures Using KNearest Neighbor Algorithm on KDD Dataset," International Journal of Science and Research., vol. 4, pp. 2319–7064, 2013, Available: www.ijsr.net

[12] K. Sulaiman, L. Hakim Ismail, M. Adib Mohammad Razi, M. Shalahuddin Adnan, and R. Ghazali, "Water Quality Classification Using an Artificial Neural Network (ANN)," IOP Conference Series: Materials Science and Engineering, vol. 601, no. 1, 2019, doi: 10.1088/1757-899X/601/1/012005

[13] N. Kumar and D. Kumar, "Classification using Artificial Neural Network Optimized with Bat Algorithm," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 3, pp. 696–700, 2020, doi: 10.35940/ijitee.c8378.019320