A

Mini Project Report on

Predictive Analytics for Laptop Pricing by leveraging K-Nearest Neighbors (KNN) and Linear Regression Techniques for Advanced Market Forecasting

Submitted in partial fulfillment of the requirements
for the degree of
BACHELOR OF ENGINEERING
IN
Computer Science & Engineering
Artificial Intelligence & Machine Learning

by

Aditi Gadhave (22106079)
Aabha Bhide (22106093)
Niharika Bandekar (22106136)
Sudhiksha Aradhyula (22106010)

Under the guidance of

Prof. Vijesh Nair

Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)
A. P. Shah Institute of Technology
G. B. Road, Kasarvadavali, Thane (W)-400615
University Of Mumbai
2024-2025

A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled "Predictive Analytics for Laptop Pricing by leveraging K-Nearest Neighbors (KNN) and Linear Regression Techniques for Advanced Market Forecasting**."** is a bonafide work of Aditi Gadhave (22106079), Aabha Bhide (22106093), Niharika Bandekar (22106136) and Sudhiksha Aradhyula (22106010) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

Prof. Vijesh Nair	Dr. Jaya Gupta
Mini Project Guide	Head of Department

A. P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval

This Mini project report entitled ***"Predictive Analytics for Laptop Pricing by leveraging K-Nearest Neighbors (KNN) and Linear Regression Techniques for Advanced Market Forecasting"*** by **Aditi Gadhave (22106079), Aabha Bhide (22106093), Niharika Bandekar (22106136) and Sudhiksha Aradhyula (22106010)** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Science &Engineering***, (AI and ML) ***2024-25***.

External Examiner: _____

Internal Examiner: _____

Place: APSIT, Thane

Date:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Aditi Gadhave (22106079)	Aabha Bhide (22106093)	Niharika Bandekar (22106136)	Sudhiksha Aradhyula (22106010)

ABSTRACT

In the competitive and rapidly evolving laptop market, accurate pricing models are essential for manufacturers and retailers to optimize their strategies and forecast trends effectively. This project explores the use of predictive analytics for laptop pricing, focusing on the application of K-Nearest Neighbors (KNN) and Linear Regression techniques. Linear Regression is employed to model the relationship between laptop prices and various features such as brand, processor speed, RAM, storage, screen size, and weight. K-Nearest Neighbors (KNN), a non-parametric algorithm, is used to capture more complex, non-linear relationships. The models were trained and tested on a dataset of laptop features and prices, with their performance evaluated using metrics like R^2 score and Mean Absolute Error (MAE). Results show that the hybrid approach of Linear Regression and KNN improves the precision of pricing forecasts, providing valuable insights for market strategies. This study demonstrates the potential of predictive analytics to refine pricing models in the laptop market, supporting more data-driven decision-making.

***Keywords**:* Predictive Analytics, Training and Testing, Features, Brand, Processor Speed, RAM, Storage, Screen Size, Weight.

Index

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

In the rapidly evolving technology industry, where new laptop models are frequently introduced and consumer preferences shift quickly, accurately pricing laptops based on their specifications and market trends is a challenging yet essential task. Accurate pricing strategies not only help manufacturers and retailers remain competitive, but they also empower consumers to make informed purchasing decisions. As the number of variables that influence a laptop's price increases, traditional pricing models that rely solely on linear relationships between features and price become less effective. There is a growing need for more sophisticated models that can capture the intricate patterns and relationships inherent in the data.

This project, titled "Predictive Analytics for Laptop Pricing by Leveraging K-Nearest Neighbors (KNN) and Linear Regression Techniques for Advanced Market Forecasting," aims to address this need by developing a hybrid model that combines the strengths of both K-Nearest Neighbors (KNN) and Linear Regression. These two algorithms are widely recognized for their utility in predictive analytics: Linear Regression for its simplicity and interpretability in modeling linear relationships, and KNN for its flexibility in handling non-linear patterns and interactions between features.

By integrating these two approaches, the project seeks to build a robust and versatile predictive model that can accurately forecast laptop prices based on a variety of features such as brand, CPU speed, RAM, storage capacity, and more. The model will be designed to accommodate diverse datasets and varying feature types, ensuring it is applicable across different segments of the laptop market. Additionally, this project emphasizes the importance of data preprocessing, including the handling of categorical data like laptop brands, which must be converted into a numerical format suitable for machine learning algorithms. The careful preparation of input data is crucial to the success of the model, as it ensures that the features fed into the predictive algorithms are in the correct format and represent the underlying market dynamics accurately.

The significance of this project extends beyond just price prediction. By creating a model that can reliably predict laptop prices, the project aims to contribute to more efficient market forecasting, helping stakeholders across the technology ecosystem. For consumers, the mode offers a way to assess whether a laptop's price is fair based on its features, potentially leading to better purchasing decisions. For manufacturers and retailers, it provides insights that can be used to set competitive prices, manage inventory, and plan marketing strategies. Ultimately, the goal is to create a tool that not only addresses current challenges in laptop pricing but also adapts to future changes in the market, ensuring its long-term relevance and utility.

In summary, this project represents a forward-looking approach to laptop price prediction, leveraging the combined strengths of K-Nearest Neighbors and Linear Regression to deliver a model that is both accurate and adaptable. Through this work, the project aims to set a new standard in the field of predictive analytics for consumer electronics, offering a solution that is as innovative as the products it seeks to price.

CHAPTER 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 Literature Review

[1] "Laptop Price Prediction using Machine Learning Algorithms." In *2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, pp. 226-231. IEEE, 2022. Shaik, Mohammed Ali, Medicherla Varshith, Sanka SriVyshnavi, Nagamalla Sanjana, and Rama Sujith.

This paper discusses how machine learning (ML) can improve laptop price predictions by analyzing key features and applying different ML models. It tests models like Decision Trees, Multiple Linear Regression, K-Nearest Neighbors (KNN), and Random Forest to determine which is most accurate in predicting laptop prices, offering a novel approach to assist both manufacturers and consumers.

[2] "Laptop Price Prediction using Machine Learning." *International Journal of Computer Science and Mobile Computing* 11, no. 1 (2022): 164-168.Surjuse, Vaishali, Sankalp Lohakare, Aayush Barapatre, and Abhishek Chapke.

This paper presents a laptop price prediction system by using the supervised machine learning technique. The research uses multiple linear regression as the machine learning prediction method which offered 81% prediction precision. Using multiple linear regression, there are multiple independent variables but one and only one dependent variable whose actual and predicted values are compared to find precision of results. This paper proposes a system where price is dependent variable which is predicted, and this price is derived from factors like Laptop's model, RAM, ROM (HDD/SSD), GPU, CPU,etc.

[3] "Laptop Price Estimation Using Machine Learning." *International Journal of Research in Engineering, Science and Management* 7, no. 3 (2024): 42-44. Pragnatha, Puvvada, Korada Sai Yaswanth Kumar, Vemala SSSM Vikas, Sai Pavan Gorle, Nookala Subrahmanya Sai, and Sangam Hemanth Sai.

This project develops a Laptop Price Prediction Model using the Random Forest algorithm to help students and individuals make informed, budget-conscious laptop purchases. The model, implemented in a user-friendly Streamlit web application, provides real-time price predictions

based on laptop specifications. It aims to simplify the process of finding budget-friendly laptops and enhance decision-making by offering accurate cost estimates.

[4] "Laptop Price Range Prediction with Machine Learning Methods." ***International Journal of Multidisciplinary Studies and Innovative Technologies*** **8, no. 1 (2024): 40-45. Karakuş, Yasin, and Turgay Tugay Bilgin.**

This study addresses the need for predicting laptop price ranges rather than exact prices, which is more practical for real-world applications. Unlike previous research that primarily focuses on regression methods for precise price estimation, this project adapts a dataset for price range prediction and applies preprocessing techniques like data cleaning and feature engineering. This work aims to enhance the practical applicability of laptop price forecasting.

[5] "Laptop Price Prediction Using Real Time Data." In ***2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*****, pp. 1-5. IEEE, 2023. Reddy, Chada Lakshma, K. Bhargav Reddy, G. R. Anil, Sachi Nandan Mohanty, and Abdul Basit.**

This paper addresses the surge in online laptop sales driven by the pandemic, highlighting the increasing need for effective pricing prediction tools. It proposes a feature-based pricing prediction algorithm to assist buyers in making informed purchasing decisions. By leveraging real-time data scraped from an e-commerce website, the study aims to provide accurate price forecasts.

[6] "Human Centric Computing Applications for Laptop Price Prediction." ***American Journal of Advanced Computing*** **2, no. 1 (2023): 62-68. Zahedi, Mehboob, Danish Jamal, and Abhishek Das.**

This paper addresses the challenge of determining laptop prices in a rapidly growing market post-lockdown. With significant demand increases, the proposed model uses Regression machine learning algorithms to predict laptop prices based on features like RAM, ROM, and CPU. This model aims to help consumers make informed decisions and streamline the price evaluation process.

[7] "Research On Laptop Price Predictive Model Based on Linear Regression, Random Forest and Xgboost." ***Highlights in Science, Engineering and Technology*** **85 (2024): 265-271. Tian, Peiru.**

This paper explores predicting laptop prices using a dataset of 1,320 samples and evaluates three models: Linear Regression, Random Forest, and XGBoost. By analyzing 13 features, including brand, screen size, and RAM, the study finds that the XGBoost model outperforms the others, with an RMSE of 294.11 and an R^2 of 0.85. This indicates that XGBoost provides the most accurate and reliable price predictions

[8] "Predicting Laptop Prices Based on Specifications Using Machine Learning Techniques: An Empirical Study." ***IEESE International Journal of Science and Technology*** **13, no. 1 (2024): 14-22. Kafabihi, Ridwan Jauhar.**

This paper explores laptop price prediction using a dataset of 992 laptops and evaluates various features such as RAM, brand, and storage. The study finds that the XGBoost model, with an R^2 value of 0.84, provides the highest predictive accuracy compared to other methods. The model also offers insights into feature importance, benefiting both retailers in pricing strategies and consumers in making informed decisions.

Paper Name	Summary	Limitations	Adaptations
"Laptop Price Prediction using Machine Learning Algorithms" (ICETEMS, 2022)	Evaluates multiple MLmodels (Decision Trees, KNN, Random Forest, etc.) to predict laptop prices, assisting consumers and manufacturers.	Limited focus on real-time data and scalability.	Could integrate real-time web scraping for dynamic predictions
"Laptop Price Prediction using Machine Learning" (IJCSMC, 2022)	Uses multiple linear regression with 81% prediction precision, focusing on features like RAM, ROM, GPU, and CPU.	Restricted to only one model, lacking model comparison.	Testing other algorithms for improved accuracy.
"Laptop Price Estimation Using Machine Learning" (IJRESM, 2024)	Implements Random Forest with a Streamlit app to provide real-time predictions, aiding in budget-conscious purchases.	Limited scope by using only the Random Forest model.	Incorporate additional models like XGBoost for higher precision.
"Laptop Price Range Prediction with Machine Learning Methods" (IJMSIT, 2024)	Predicts price ranges instead of exact prices, focusing on practical real-world applications with data cleaning and feature engineering.	Focuses only on price ranges, missing exact price predictions.	Combine range-based and exact price prediction techniques.

"Laptop Price Prediction Using Real Time Data" (ICAISC, 2023)	Uses real-time e-commerce data for predicting prices and assisting buyers.	May face challenges with data scraping limitations and API restrictions.	Develop strategies to handle missing or inconsistent data
"Human Centric Computing Applications for Laptop Price Prediction" (AJAC, 2023)	Employs regression algorithms to predict prices, considering features like RAM, ROM, and CPU in a post-lockdown market scenario.	Focuses only on regression models without considering ensemble methods.	Explore ensemble techniques to improve prediction robustness.
"Research On Laptop Price Predictive Model Based on Linear Regression, Random Forest and Xgboost" (HSET, 2024)	Evaluates Linear Regression, Random Forest, and XGBoost, finding XGBoost to be the most accurate.	Requires computational resources due to the complexity of XGBoost.	Optimize hyperparameters for performance-efficiency balance.
"Predicting Laptop Prices Based on Specifications Using ML Techniques: An Empirical Study" (IEESE IJST, 2024)	Uses XGBoost to analyze 992 samples, providing insights on feature importance and achieving an R^2 value of 0.84.	Limited generalizability beyond the given dataset.	Use larger datasets to improve model generalizability and performance.

CHAPTER 3

PROBLEM STATEMENT

3. PROBLEM STATEMENT

The primary objective of this project is to develop a predictive model that accurately forecasts laptop prices based on a variety of specifications and market conditions. Traditional pricing models often rely on simple linear relationships, which may not fully capture the complexities of the data, such as non-linear interactions between features like brand, CPU speed, RAM, and storage capacity. To address these challenges, this project proposes a hybrid model that combines the strengths of K-Nearest Neighbors (KNN) and Linear Regression. This approach aims to improve the accuracy of price predictions by leveraging the unique advantages of both algorithms. The model will be designed to handle diverse datasets with varying feature types and will be capable of making reliable predictions that can be used by consumers, manufacturers, and retailers to make informed decisions in the competitive laptop market.

CHAPTER 4

EXPERIMENTAL SETUP

4. EXPERIMENTAL SETUP

4.1 Hardware Setup:

- **Computer System:**

  Processor: The project was executed on a computer equipped with an Intel Core i5 processor (or equivalent). This processor, with at least 4 cores and a clock speed of 2.5 GHz or higher, ensures sufficient computational power for handling data processing and model training tasks. The multi-core architecture facilitates parallel processing, which is beneficial when running algorithms that can take advantage of concurrent execution.

  RAM: A minimum of 8 GB of RAM was utilized to enable smooth multitasking and efficient handling of datasets. This amount of RAM is essential when working with data manipulation libraries like Pandas, as it allows for the efficient loading and processing of large datasets without significant performance degradation.

  Storage: The system was configured with either a 500 GB Hard Disk Drive (HDD) or a 256 GB Solid State Drive (SSD). An SSD is preferable due to its faster read and write speeds, which enhances the overall performance of the system, particularly during data loading and model training. The storage capacity is adequate for holding the dataset and any additional files related to the project.

  Operating System: The project was developed and run on Windows 10/11 or a Linux distribution, such as Ubuntu. These operating systems provide a stable environment for software development and are compatible with all necessary libraries and tools.

- **Additional Peripherals:**

  A monitor with a minimum resolution of 1920x1080 (Full HD) was used to ensure clear visibility of the development environment and results. A larger screen can significantly improve the efficiency of coding and data visualization tasks.

  A standard keyboard and mouse were employed for input and navigation, allowing for a comfortable and efficient coding experience.

4.2 Software Setup

- **Programming Language:**

  Python 3.8 or higher: Python was chosen as the primary programming language due to its simplicity and versatility, as well as the extensive libraries available for data analysis and machine learning.

- **Integrated Development Environment (IDE):**

  Jupyter Notebook: This IDE was used for initial experimentation, data analysis, and model development. Jupyter Notebook allows for an interactive coding environment, making it easier to visualize data and results in real-time.

  Streamlit: Streamlit was used to create the interactive web application interface. This tool enables the rapid development of data-driven web apps with minimal coding, making it ideal for showcasing machine learning models to users.

- **Required Libraries/Packages:**

  Pandas: This powerful data manipulation library was utilized for data cleaning, preprocessing, and analysis. It provides data structures like DataFrames that facilitate efficient handling of structured data.

  NumPy: NumPy was used for numerical operations, enabling efficient computation on large arrays and matrices. This library is fundamental for many scientific computing tasks in Python.

  Scikit-learn: This library encompasses various machine learning algorithms and tools for model training, evaluation, and preprocessing. It was instrumental in implementing both the Linear Regression and K-Nearest Neighbors regression models used in this project.

  Pickle: Pickle was used for saving and loading trained models. This allows the model to be reused without needing to retrain it each time, thus saving time and computational resources.

  Streamlit: As mentioned earlier, Streamlit was used to develop the web application interface that allows users to interact with the model, input their data, and receive predictions in real-time.

- **Additional Tools:**

  CSV Reader: The Pandas library includes built-in functions to read CSV files, which were employed to load the dataset for analysis. This functionality simplifies the process of importing data into the project.

CHAPTER 5

PROPOSED SYSTEM AND IMPLEMENTION

5.PROPOSED SYSTEMS AND IMPLEMENTATION

5.1 Block diagram of proposed system

5.2 Description of block diagram

- **Start:**

  The flowchart begins here, indicating the initiation of the entire process.

- **User Inputs Laptop Specifications**:

  This step consolidates all user inputs into a single process. The user is prompted to provide specifications for a laptop, including:

  Brand: The manufacturer of the laptop.

  Weight: The weight of the laptop in kilograms.

  RAM Size: The amount of RAM (in GB) the laptop has.

  Storage Capacity: The storage capacity (in GB) available on the laptop.

  Screen Size: The size of the laptop's screen (in inches).

  CPU Speed: The speed of the processor (in GHz).

  Condition: The condition of the laptop (e.g., new or used).

- **Preprocess Input Data**:

  In this stage, the provided user inputs are prepared for analysis. The preprocessing includes:

  Clean Data: Ensure that the input data is accurate and free from errors.

  Handle Missing Values: Address any missing or incomplete data by either removing the entries or filling in missing values.

  Encode Categorical Variables: Convert categorical inputs, such as brand and condition, into a numerical format suitable for machine learning models.

- **Split Data into Training and Testing Sets**:

  The processed data is divided into two sets: one for training the models and another for testing their performance. This split is essential for evaluating the models' ability to predict prices on unseen data.

- **Train Models:**

  During this step, the machine learning models are trained using the training data:

Train Linear Regression: A linear regression model is developed to understand the relationship between laptop specifications and their prices.

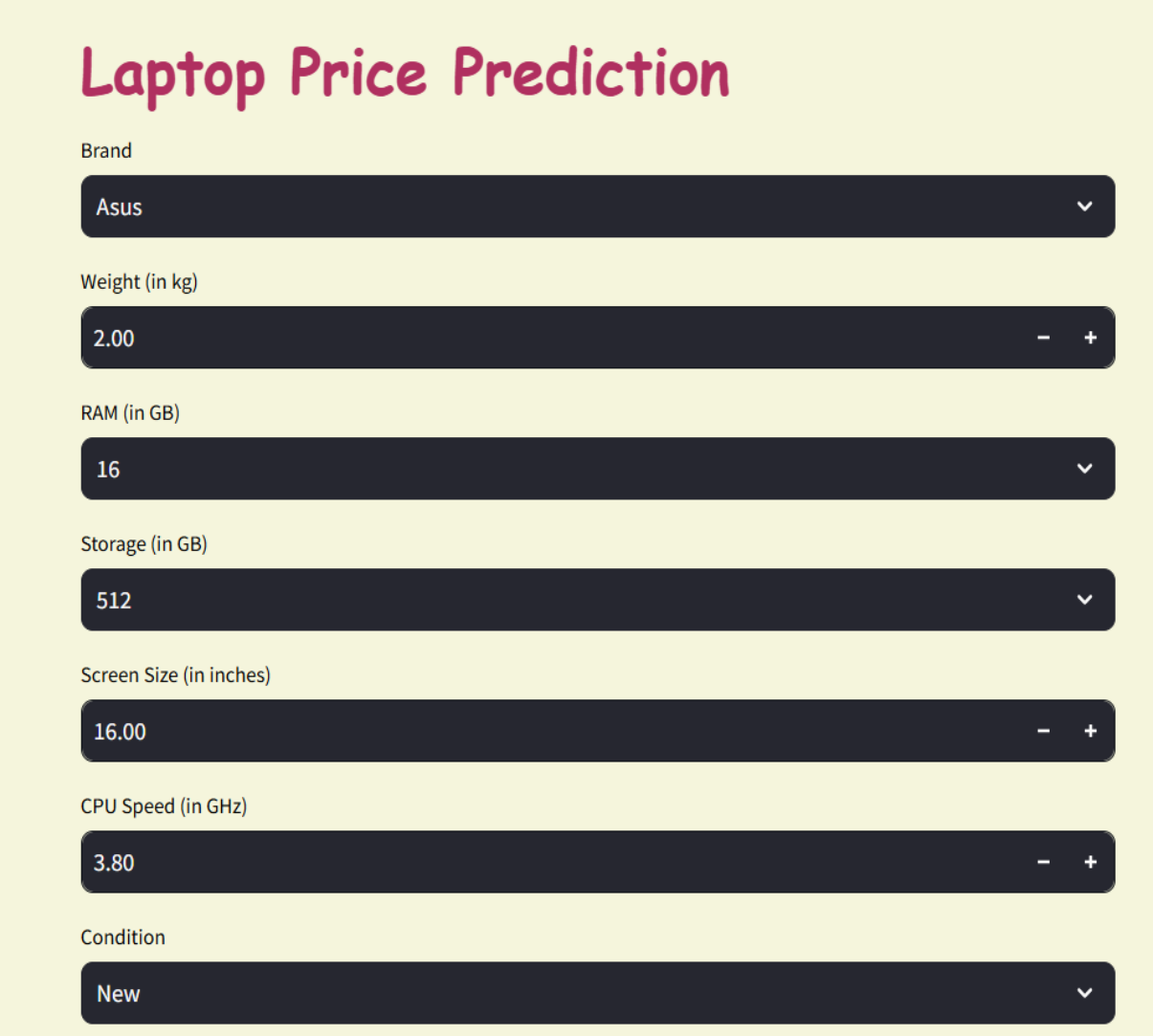
Train KNN: The K-Nearest Neighbors (KNN) algorithm is also trained, allowing it to predict prices based on similarities with existing laptop samples.

- **Predict Price Using Trained Models**: With the models trained, the system predicts the price of the new laptop based on the specifications provided by the user. Both the linear regression and KNN models are utilized for this prediction.
- **Display Predictions to User:**

  The predicted prices from both models are presented to the user. This step delivers the estimated market price for the laptop based on the provided specifications.
- **End:**

  The process concludes here, marking the end of the flowchart. This indicates that the user has received the predictions and the workflow has completed.

5.3 Implementation

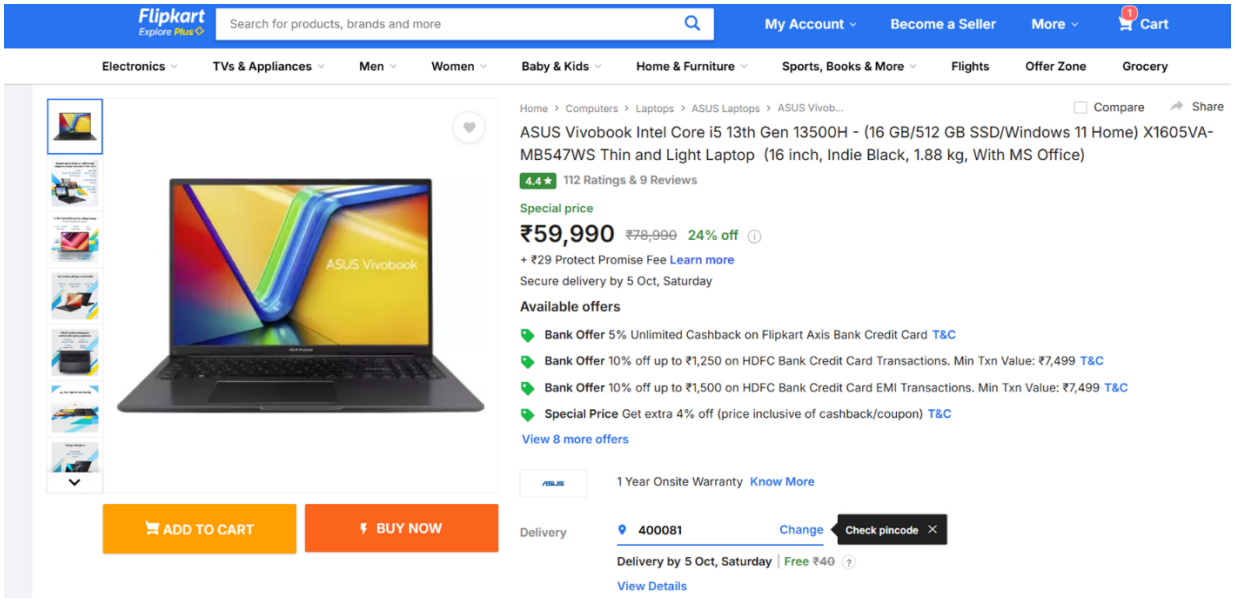


Predict

Prediction for the new sample

Linear Regression Prediction: **59,175.22**

KNN Regression Prediction: **52,600.00**



CHAPTER 6

CONCLUSION

6.CONCLUSION

In this project, we developed a predictive model for estimating laptop prices based on various specifications such as brand, weight, RAM size, storage capacity, screen size, CPU speed, and condition. Through the use of machine learning algorithms, specifically Linear Regression and K-Nearest Neighbors (KNN), we successfully trained our models on a comprehensive dataset. The application offers an intuitive interface for users to input their laptop specifications, providing them with instant price predictions.

The project demonstrates the effectiveness of data preprocessing and feature encoding in enhancing model performance. By cleaning the data and transforming categorical variables, we ensured that our machine learning models could accurately interpret user inputs, leading to reliable price estimations.

Overall, this project highlights the practical application of machine learning in the e-commerce sector, particularly in aiding consumers in making informed purchasing decisions.

Future Scope:

The predictive model established in this project has significant potential for future enhancements, including:

1. **Incorporating More Features**: Future iterations could include additional features such as battery life, warranty, and customer reviews, which could further refine price predictions.
2. **Expanding the Dataset**: Increasing the dataset size and diversity by including laptops from different regions or introducing additional brands may improve the robustness of the model.
3. **Advanced Algorithms**: Exploring other machine learning algorithms such as Random Forests, Gradient Boosting, or Neural Networks could enhance predictive accuracy.
4. **Real-time Price Updates**: Integrating real-time price data from e-commerce platforms could help maintain the model's relevance and accuracy in predicting current market prices.
5. **User Feedback Mechanism**: Implementing a feedback system where users can provide their laptop purchase prices could help in continuously updating and refining the model.
6. **Mobile Application Development**: Creating a mobile version of the application would make it more accessible for users, allowing them to check laptop prices on-the-go.

References

[1] "Laptop Price Prediction using Machine Learning Algorithms." In *2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, pp. 226-231. IEEE, 2022. Shaik, Mohammed Ali, Medicherla Varshith, Sanka SriVyshnavi, Nagamalla Sanjana, and Rama Sujith.

[2] "Laptop Price Prediction using Machine Learning." *International Journal of Computer Science and Mobile Computing* 11, no. 1 (2022): 164-168.Surjuse, Vaishali, Sankalp Lohakare, Aayush Barapatre, and Abhishek Chapke.

[3] "Laptop Price Range Prediction with Machine Learning Methods." *International Journal of Multidisciplinary Studies and Innovative Technologies* 8, no. 1 (2024): 40-45. Karakuş, Yasin, and Turgay Tugay Bilgin.

[4] "Predicting Laptop Prices Based on Specifications Using Machine Learning Techniques: An Empirical Study." *IEESE International Journal of Science and Technology* 13, no. 1 (2024): 14-22. Kafabihi, Ridwan Jauhar**.**

[5] "Laptop Price Prediction Using Real Time Data". Reddy, Chada Lakshma, K. Bhargav Reddy, G. R. Anil, Sachi Nandan Mohanty, and Abdul Basit. In *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*, pp. 1-5. IEEE, 2023.

[6] "Laptop Price Estimation Using Machine Learning". Pragnatha, Puvvada, Korada Sai Yaswanth Kumar, Vemala SSSM Vikas, Sai Pavan Gorle, Nookala Subrahmanya Sai, and Sangam Hemanth Sai. *International Journal of Research in Engineering, Science and Management* 7, no. 3 (2024): 42-44.

[7] "Human Centric Computing Applications for Laptop Price Prediction." *American Journal of Advanced Computing* 2, no. 1 (2023): 62-68. Zahedi, Mehboob, Danish Jamal, and Abhishek Das.

[8] "Predicting Laptop Prices Based on Specifications Using Machine Learning Techniques: An Empirical Study." *IEESE International Journal of Science and Technology* 13, no. 1 (2024): 14-22. Kafabihi, Ridwan Jauhar.