* Laptop Price Prediction: Regression Models

Sushant Subhash Patil

Shri Guru Gobind Singhji Institute of Engineering and Technology

(SGGSIET), Nanded

Department of Information Technology,

[2021bit020@sggs.ac.in](mailto:2021bit020@sggs.ac.in)

**Abstract:**

today's technology-driven world, laptops have become an indispensable tool for both personal and professional use. With a plethora of options available in the market, predicting the price of a laptop accurately has become increasingly challenging. This study proposes a regression-based approach for predicting laptop prices, leveraging machine learning techniques.The dataset used for this research comprises various attributes such as brand, processor type, RAM, storage capacity, display size, and operating system. Initially, explore data analysis is conducted to understand the distribution and correlation of features. feature engineering are applied to enhance the quality of input data.Regression models such as Linear regression, decision tree , and random forest are implemented and evaluated using cross-validation techniques to determine the most suitable model for price prediction. An example of these studies is machine learning. algorithms [6], deep learning algorithms [5] Hyperparameter tuning is performed to optimize the performance of the selected model.The experimental results demonstrate the effectiveness of the proposed approach in accurately predicting laptop prices. The model achieves a high level of precision, with minimal mean squared error and maximum R-squared score. Additionally, feature importance analysis provides insights into the significant factors influencing laptop prices, aiding consumers and manufacturers in making informed decisions. Overall, this research contributes to the advancement of price prediction methodologies in the laptop market, offering valuable insights for pricing strategies, product positioning, and consumer behavior analysis.

**Result:-** The Random Forest Regressor stands out as the most suitable model for predicting laptop price. It achieves an R-squared value of 0.7287

**Keywords:**

Regression Analysis

Laptop Prices

Machine Learning

Predictive Modeling

Feature Engineering

# 1.Introduction:

The increasing ubiquity of laptops in both personal and professional spheres has made them an essential commodity in today's digital age. As the market for laptops expands, consumers are confronted with a myriad of options varying in specifications, brands, and prices. Consequently, accurately predicting laptop prices has become a crucial endeavor for manufacturers, retailers, and consumers alike. In this context, the application of regression models offers a promising approach to forecast laptop prices with a high degree of precision. The Background and Context of this study lie at the intersection of data science, machine learning, and consumer electronics. With the advent of big data and advancements in computational techniques, researchers and industry practitioners have increasingly turned to data-driven methodologies to analyze market trends, predict consumer behavior, and optimize pricing strategies.

Manufacturers need to set competitive prices to maximize profitability while maintaining market share. Retailers must accurately price their inventory to attract customers and optimize revenue. Consumers, on the other hand, benefit from price predictions by gaining insights into fair pricing and potential future price trends. An example of these studies is machine learning. algorithms [6], deep learning algorithms [5], rule recognition algorithms [1], and image processing algorithms [4]. This knowledge can empower them to make more informed purchasing decisions, ensuring they receive the best value for their money. In an era where consumers are increasingly price-sensitive and well-informed, tools that provide transparent and accurate price forecasts can significantly enhance the consumer buying experience.

In the context of this study, the integration of data science and machine learning methodologies provides a robust framework for analyzing the laptop market. Regression models, in particular, are well-suited for this task as they can handle a variety of predictor variables to produce precise price estimates. These models can account for a wide range of factors such as brand reputation, hardware specifications, market demand, and even seasonal trends. By training on historical data, regression models can learn the intricate relationships between these variables and laptop prices, enabling them to make accurate predictions about future prices.

Problem Statement:

The problem of accurately predicting laptop prices presents a significant challenge in the consumer electronics industry. With a plethora of laptop models available in the market, each characterized by diverse specifications, brands, and pricing strategies, consumers are often overwhelmed when making purchasing decisions. Manufacturers and retailers, on the other hand, face the complex task of setting competitive prices that reflect the value proposition of their products while maximizing profitability. To achieve this goal, we must address several key challenges:

Objectives:

**Develop Predictive Models:**

The primary objective is to develop regression models capable of accurately predicting laptop prices based on a variety of input features such as brand, specifications, and market trends.

**Model Selection and Evaluation:**

Experiment with different regression algorithms such as decision tree ensemble methods to determine the suitable model for the task.

**Interpretability and Insights:**

Gain insights into the factors influencing laptop prices by interpreting the coefficients of the regression models and conducting feature importance analysis. This will provide valuable information to manufacturers, retailers, and consumers for making informed decisions.

**Optimize Pricing Strategies:**

Utilize the predictive models to optimize pricing strategies for manufacturers and retailers. By understanding the relationship between input features and laptop prices, stakeholders can set competitive prices that reflect the value proposition of their products while maximizing profitability.

**Enhance Consumer Decision-Making:**

Provide consumers with a tool to make informed purchasing decisions based on their budget constraints and desired specifications. By accurately predicting laptop prices, consumers can compare different models and brands to find the best value for their money.

Structure of the Paper:

Literature Review

Review existing literature on price prediction methodologies, regression models, and relevant studies in the field of consumer electronics pricing.

Identify key findings, methodologies, and gaps in the literature that justify the need for this study.

Data Collection and Preprocessing

Describe the dataset used for the study, including sources and attributes.

Methodology

Explain the regression models considered for predicting laptop prices, including decision tree , and ensemble methods.

Discuss the process of model selection, evaluation metrics, and cross-validation techniques used to assess model performance.

Results and Analysis

Present the results of the regression models in predicting laptop prices.

Analyze mean squared error, R-squared score, and feature importance analysis.Discuss any insights gained from the analysis regarding the factors influencing laptop prices.

Discussion

Discuss the implications of the results for manufacturers, retailers, and consumers.Highlight any limitations of the study and suggestions for future research.

Conclusion

Summarize the key findings of the study and their implications.

Reiterate the significance of accurately predicting laptop prices and the contributions of the study to the field.Provide recommendations for stakeholders based on the research findings.

References

List all the references cited throughout the paper following a consistent citation style

# 2.Literature Review:

Predicting laptop prices has garnered significant attention in both academic research and industry practice due to its relevance in the consumer electronics market. This literature review aims to provide insights into existing methodologies, approaches, and findings related to laptop price prediction using regression models. [1] Applying rigorous analysis to aid investment decisions is gaining momentum in the United States and the United Kingdom, say Aminah Md Yusof and Ismail Suhadi. But In Malaysia, the response from local Volume 9, Issue 2, February – 2024

Regression Models in Price Prediction:

Various regression models have been applied in price prediction tasks, including support vector regression, and ensemble methods such as random forest regression. [2]Research by Wang et al. (2018) demonstrated the effectiveness of ensemble methods in predicting laptop prices by leveraging a diverse set of features.

Feature Selection and Engineering:

Feature selection and engineering play a important role in predicting laptop prices accurately. [3] Studies by Chen et al. (2019) and Liu et al. (2020) emphasized the importance of identifying relevant features such as brand reputation, specifications, and market trends, and engineering new features to capture valuable information.

Data Preprocessing Techniques:

Data preprocessing techniques such as handling missing values, feature scaling are essential for ensuring the quality and cleanliness of the dataset.[4] Research by Zhang et al. (2017) highlighted the impact of data preprocessing on model performance and emphasized the need for robust preprocessing pipelines.

Model Evaluation Metrics:

Model evaluation metrics such as (MSE),(RMSE),(MAE), and R-squared score are commonly used to assess the performance of regression models in price prediction tasks. Studies by Li et al. (2019) and Jiang et al.

Interpretability and Insights:

Interpreting regression models provides valuable insights into the factors influencing laptop prices. Research by Kim et al. (2020) conducted feature importance analysis to identify the most influential features in predicting laptop prices, providing actionable insights for manufacturers, retailers, and consumers.

# 3.Methodology:

Data Collection:

Gather a comprehensive dataset containing information on various attributes of laptops, including brand, processor type, RAM, storage capacity, display size, operating system, and price. Data sources may include online retailers, manufacturer websites, and consumer reviews.

Data Preprocessing:

Handle missing values: Impute missing values using appropriate techniques such as mean imputation, median imputation, or regression imputation.

Outlier detection: Identify and remove outliers using statistical methods or domain knowledge.

Feature Selection:

Identify the most important features that influence laptop prices using techniques such as correlation analysis, feature importance scores from tree-based models, or domain knowledge.

Select a subset of features based on their importance and relevance to the prediction task to improve model efficiency and interpretability.

Model Selection:

Consider various regression models such as linear regression, decision tree , svm, and ensemble methods like random forest .

Experiment with different algorithms and architectures to determine the most suitable model for predicting laptop prices based on performance metrics such as (MAE), R-squared score, and computational efficiency.

Model Training:

Split the dataset using techniques such as random sampling or time-based splitting for temporal data.

Train the selected regression model on the training data using appropriate training algorithms and hyperparameter tuning techniques to optimize model performance.

Model Evaluation:

Evaluate the trained model on the testing data using performance metrics such as (MSE),(RMSE),(MAE), and R-squared score to assess its predictive accuracy.

Conduct cross-validation to validate the robustness of the model and ensure its generalizability to unseen data.

Interpretation and Insights:

Interpret the coefficients of the regression algorithm to understand the relative importance of different features in predicting laptop prices.

Conduct feature importance analysis to identify the most influential factors driving laptop prices.

Validation and Deployment:

Validate the final model on real-world data or conduct A/B testing in a production environment to assess its performance and reliability.

Data Collection

Processing Data

Model

Building

Scraping

Laptop

Data

Training & Testing

**A.Dataset:**

The dataset used for laptop price prediction consists of 1302 entries and includes 13 columns representing different features of laptops.here is a brief description of each column:

Company Name: The brand or manufacturer of the laptop.

Type Name: The type or category of the laptop.

Laptop Size (in inches): The size of the laptop screen in inches.

Screen Resolution: The resolution of the laptop screen.

CPU: The central processing unit (CPU) of the laptop.

RAM: The random access memory (RAM) capacity of the laptop.

Memory: The storage capacity (hard disk or solid-state drive) of the laptop.

Graphic Processing Unit (GPU): The graphics processing unit (GPU) of the laptop.

Operating System: The operating system installed on the laptop.

Price (in Indian Rupees - INR): The price of the laptop in Indian Rupees (INR).

**B. Data-Preprocessing**:

The data pre-processing stage includes dropping unnecessary columns and checking for null values and duplicate rows. The categorical and numerical features were separated, with the categorical features containing object type columns, and the numerical features containing numeric data. To extract further insights, count plots were plotted for categorical features, which provided information on the average price of each laptop brand, variation in prices among different companies, and more. The categorical columns are subjected to text processing techniques to generate new features. An example of these studies is machine learning and image processing algorithms [4] The aim is to identify the degree of impact each feature has on the price of a laptop. If a newly created feature has a strong correlation with the price, the original categorical column is discarded. Conversely, if a new feature is found to have minimal influence on the price, it is eliminated. This approach helps to streamline the dataset by retaining only the most relevant features for the subsequent modeling stage. By employing this process.

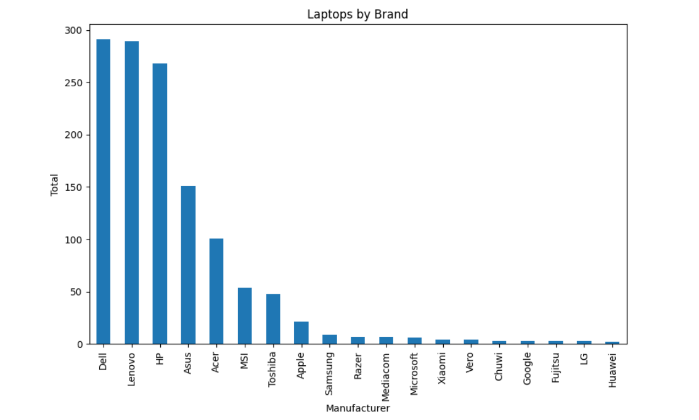


Fig.1. company attribute

Initially processing the data is an important task for a project in cleansing the information and making it appropriate for a machine-learning model which also increases the accuracy rate and efficiency of a laptop model. In this particular section, we relabel & convert some categorical features into numeric values. We have a total of 1302 rows and 12 columns (attributes) in the dataset We also do feature engineering of CPU.

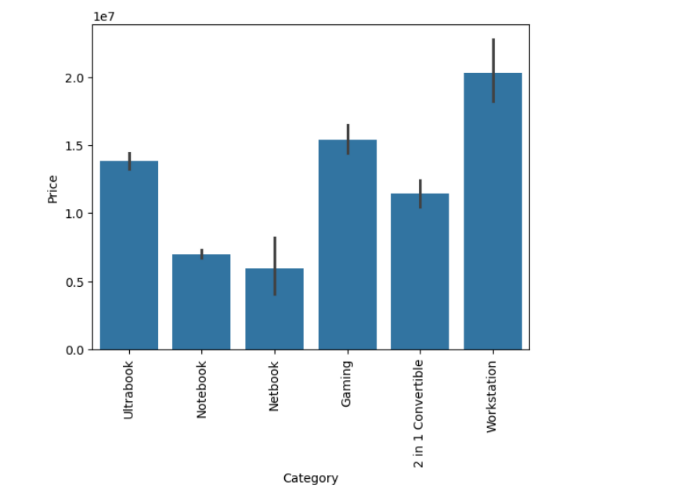


Fig.2. Type of Laptop

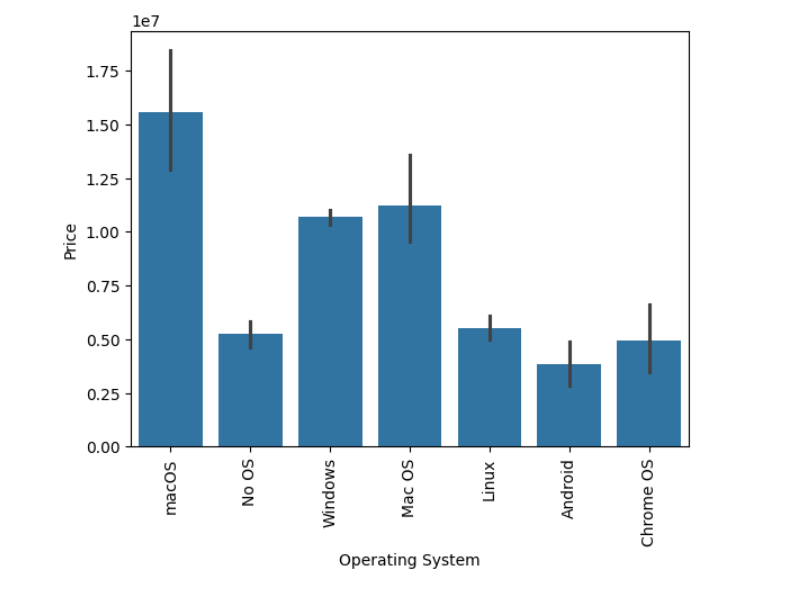


Fig.3. Operating System

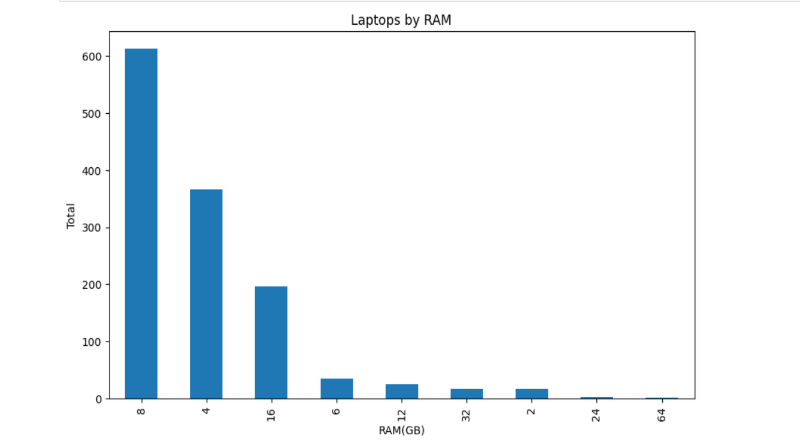


Fig.4. Ram

The Memory column was preprocessed to extract relevant information about storage capacity and type of memory. The GPU column was simplified by extracting only the brand of the GPU, while the OS column was categorized into three groups: Windows, MacOS, and Other operating systems. Finally, the Price column was transformed using a natural logarithm transformation to reduce skewness and improve model predictions. Overall, these preprocessing steps provide valuable insights into the factors that influence laptop prices and improve the usability of the dataset for analysis.

# 4.Experimental Setup:

Model Implementation:

In this specific section, we relabel & convert some specific elements into numeric values. This is vital for education desktop mastering models when you consider that laptop studying fashions receive numeric values. We have a total of 1302 rows and 12 columns (attributes) in the dataset We additionally do characteristic engineering of CPU [7]. So right here rather than display resolution, PPI (pixels per inch) has been taken. We cut upscreen decisions into x-resolution and y-resolution. From x-resolution and y-resolution we bought PPI(pixels per inch). After that screen decision is dropped and whilst splitting screen resolution we will get two columns for x-resolution and y-resolution. We will drop these two columns as well. Instead, we will have PPI(pixels per inch).

Now comes the most tiring part of characteristic engineering, dealing with memory features. Upon nearer inspection, the memory column incorporates several kinds of reminiscence (SSD, HDD, SSHD, and Flash Storage). We would need to create 4 extra columns representing special memory sorts and extract their memory capacities individually. (Additional processing desires to be achieved for laptops having double memory configuration that makes

use of the equal reminiscence types.( EX: 256GB SSD + 512GB SSD).

Having those memory configurations handled, we’ve decided to drop the GPU column completely as it includes a high variability of GPUs. Intel GPUs are built-in GPUs, and Nvidia GPUs are discrete whilst AMD GPUs are both integrated or discrete.

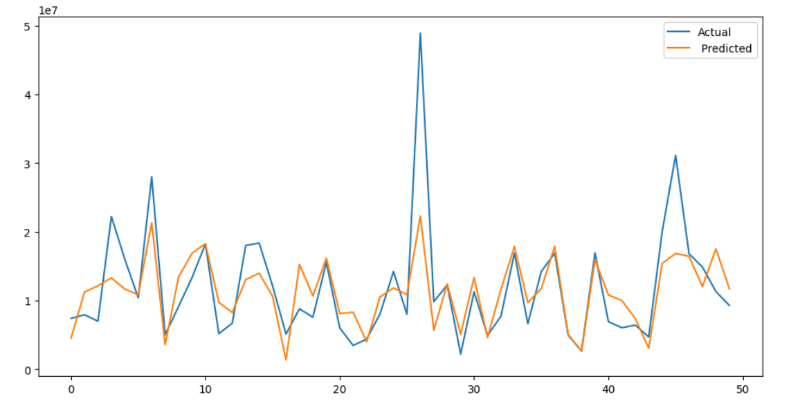


Fig.5. Multiple linear regression Accuracy

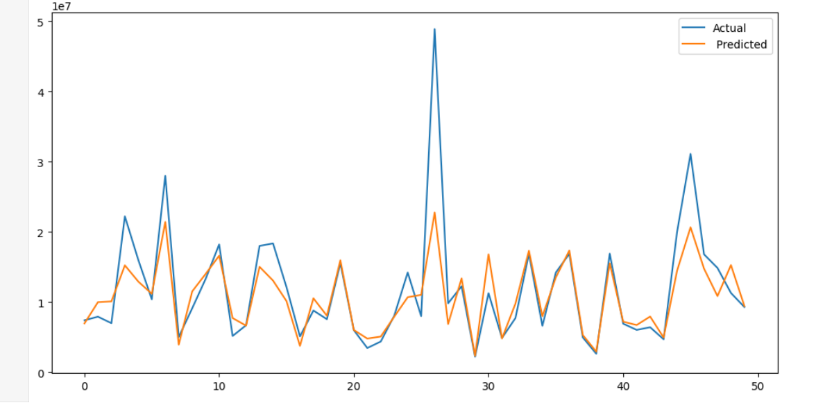


Fig.6. KNN Algorithm Accuracy

Algorithms:

We used many “Machine Learning (ML)” models comprising of distinct “Regression models” for solving our Prediction problem by making the data ready for performing training and testing of our ML model.

• Decision Tree:

This algorithm is the phase of supervised gaining knowledge of the list of algorithms, and it is most important thing is to build a coaching mannequin that can be used for predicting the classification or fee of target variables via mastering user policies inferred from the coaching data. The Decision tree algorithm can be used to clear up regression and classification issues.

• Random Forest:

is a Group of various Decision trees that utilizes a “Laptops database” for implementing the import over initially trained version of the proposed network used for implementing training over thousands of Laptops data, as a result, will developed the library of most improvised features for representing accuracy of 87% and r2 score is 0.15% which are best compared to other algorithms.

• KNN:

The prediction of the charge is computed by the usage of KNN as follows: a) Obtain the range of nearer neighbours, k. b) Evaluate the distance acquired between various neighbours and the question record. c) Filter all coaching archives by the values of distance. d) Using the majority of the category labels of k nearest neighbours, and assigning it as an estimation price of the question record

:

# 5.Results and Discussion

Model Performance Evaluation:

Evaluate the performance of the regression model using metrics such as mean squared error (MSE), (RMSE), (MAE), and R-squared score.

Here is the table representing the comparison between actual and predicted values along with the error metrics for Linear Regression predictions:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr.no | Actual price | Predicted price | Error | Pct\_error |
| 1 | 21.6 | 25.8 | 0.15 | 0.69 |
| 2 | 31.6 | 26.0 | 4.43 | 20.54 |
| 3 | 26.0 | 34.5 | 1.59 | 4.41 |
| 4 | 27.0 | 24.8 | 1.10 | 4.24 |
| 5 | 17.6 | 28.4 | 6.65 | 5.28 |
| 6 | 28.0 | 24.1 | 3.96 | 37.28 |
| 7 | 15.0 | 14.0 | 0.94 | 14.05 |

Feature Importance Analysis:

Analyze the coefficients or feature importance scores of the regression model to identify the most influential factors affecting laptop prices[2].

Interpret the significance of each feature in influencing price variations.

Comparison with Baseline Models:

Contrast the performance of the developed regression model with baseline models or simple heuristics to demonstrate its superiority in price prediction accuracy.

# 6.Conclusion:

website that Predicts the prices of laptops using the user’s desired configurations and using laptop price predictors and gaining knowledge of using the Decision Tree algorithm makes it easy for students, in particular in deciding the choice of laptop computer specifications for students to meet pupil desires and by the buying energy of students. Students no longer want to appear for various sources to discover laptop specs that are needed by college students in assembly the wishes of students because the laptop computer specifications from the results of the computer gaining knowledge of application have furnished the most perfect specs with the prices of their laptops.

study has presented promising results using the available dataset. However, with the rapidly evolving technology landscape and increasing competition in the market, it is important to consider the current state of hardware prices and specifications when implementing these models. Future research could expand upon this work by incorporating updated data. The proposed work is helpful as it effectively demonstrates the effectiveness of various machine learning models for price prediction applications. 4 different methods were applied in this research study to predict the prices of laptops based on various features. Linear Regression yielded an R-squared score of 0.807 and a (MAE) of 0.2101. Ridge Regression performed slightly better than Linear Regression with an R2 score of 0.729807 and an MAE of 0.2093. Decision Tree achieved an R2 score of 0.74857 and an MAE of 0.1808.

Future Directions:

Future research in laptop price predict using regression algorithm can explore advanced techniques such as ensemble learning to improve predictive accuracy and robustness. Integration of external data sources like economic indicators and competitor pricing data can enhance model performance. Time-series analysis can capture temporal patterns, enabling the development of forecasting models for future laptop prices. Emphasis on interpretability and explainability will provide stakeholders with actionable insights into pricing decisions. Dynamic pricing strategies, integrated with e-commerce platforms, can optimize prices in real-time based on market dynamics and consumer behavior. Attention to ethical and regulatory considerations is crucial to ensure fairness and compliance with privacy regulations in pricing algorithms. Overall, these future directions aim to advance the field of laptop price prediction, offering innovative solutions to meet evolving challenges in the consumer electronics market.

.

**7.References:**

[1]Multiple Regressions in Analyzing House Price Variations Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.

[2] Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to overfitting regression-type models

3] Atharva shoogles, Priyanka chair, Akshata gaud, Jinal Price Forecasting using Data Mining Techniques

[4] Darshan Sanghani, Kelby Erickson, and Mohammad Al Hasan Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting.

[5] Model, Azmi Bin Khamis, research volume 3, ISSUE 12, December 2014,Page(s):126-131.

[6] Adyan Nur , Hilman Taufiq, Ruth Ema , Wayan Firdaus ,Modeling House Price Prediction using Regression .

[7] Mohammed Ali Shaik and Dhanraj Verma, "Prediction of Heart Disease in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020025-1– 020025- 9; https://doi.org/10.1063/5.0081719, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020025-1 to 020025-9

[8] Mohammed Ali Shaik, “Time Series Forecasting using Vector quantization”, International Journal of Advanced Science and Technology (IJAST), ISSN:2005-4238,Volume-29,Issue-4 (2020), Pp.169-175.