IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

Advanced Forecasting of Demandable Products Prices using Machine Learning Algorithm

Gayatri Raghuwanshi[1], Gayatri Zamare[2], Rupesh Apar[3], Rupesh Dabhade[4], Vaibhav Wankhade[5]
Students, Department of Computer Science & Engineering[1,2,3,4,5]
Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

Abstract: *Knowing which items would be the most affordable is crucial for the organization. At this stage, categorization and prediction issues, such as price prediction, have been resolved using machine learning technology. This project seeks to produce timely and accurate price forecasts to assist the organisation in switching between neighboring markets to assist the organisation in switching between various neighbouring markets in order to sell their goods and obtain competitive rates. The data can be used by the company to make decisions regarding the timing of marketing. The machine Learning technique allows for predicting the number of products/services to be purchased during a defined period. Demand forecasting is used in which first raw data is collected from the market, then according to the data the product prices are forecasted. This model is a catch-all phrase for the shopping process that establishes product prices in accordance with the level of supplier competition, the hour of the day, and the weather. This model will help to forecast the prices of products according to their historical data. At an organizational level, forecasts of product prices are an essential input to many decision-making activities in various functional areas such as operations, marketing, sales, production, and finance.*

Keywords: Product prices forecasting, Machine Learning, Linear Regression, Lasso Regression, XG Boost Algorithm, Gradient Boosting Algorithm, Random Forest Regressor, Streamlit, SkLearn

I. INTRODUCTION

Effective pricing forecasting assists organizations in anticipating price increases or cuts that may impact customer demand. Previous year's data on different products are being collected and we will predict the prices of products so that we will be able to make good marketing strategies. Using machine learning the system can predict what will be the price of a particular product today or after a certain day. Due to its striking advantages over conventional methods, machine learning techniques have recently become frequently used for price prediction. ML algorithms create models using training and test data, and then use these models to make predictions [1].A prediction algorithm will be used to predict prices. Price and arrival data information strengthens the organization's bargaining position and increases the competitiveness among dealers. The organization can switch between neighboring markets more easily when price information is provided. The information can be used by the organization to make marketing timing decisions.

The majority of machine learning (ML) algorithms that were developed within the context of data science have dominated in recent years. It has previously been used to predict time series in the financial and economic sectors. Numerous empirical studies have demonstrated that machine learning methods are more effective than time series models at forecasting various financial asset values.

II. LITERATURE SURVEY

Julakha Jahan JuiIn et al.,[1] research focuses on, two machine-learning regression-based methods for predicting flat pricing—linear regression and random forest regression—wasgiven. Data has been scraped from a number of real estate websites using the web scrapper (Data Toolbar) software. When developing the model, seven factors that can affect flat pricing were taken into consideration. Here, the data quality has been investigated using the histogram, residual charting, and ANOVA. The linear and random forest model has been created after preprocessing the dataset. MSE, RMSE, MAE, and MRE have all been computed in order to evaluate the performance of both techniques. The measured error rate has led to the conclusion that the random forest regression model performs well.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

Yige Wang et al.,[2] researchstate that clearly more practical in terms of price prediction is the decision tree fitting effect using Random Forest, the order of variable importance as opposed to OLS when dealing with complicated and irregular data. Therefore, we advise choosing a random forest in these two scenarios—one in which there are many observations in the dataset and the other in which there are complex samples with noise.

Xinshu Li et al.,[3] focus on demand for commercial housing falls into three primary categories: speculative, investment, and owner-occupied. The investment need for self-employment is to purchase and lease commercial real estate to generate rental revenue. Hypothetical demand is bought.

UjjawalSonkambale et al.,[4] research machine learning techniques to predict the price of used cars based on historical data from the Kaggle and Car Dekho websites. To determine which predictions offer the best performance and accuracy, the predictions are compared and examined. Delay filters, delay lines, power amplifiers, coaxial resonators, and ceramics are index terms.

Subba Rao Polamuri et al.,[5] this paper focuses on using ML techniques to anticipate the behaviours tracking of the stock market sensex. It compares the accuracy of various models and selects an algorithm with high accuracy. The main aim is to apply innovative work to predict the behaviour tracking of the stock market Sensex.

Mohamed Ali Mohamed et al.,[6] This work presents a promising approach for predicting pricing for retail goods, specifically seasonal Christmas items. Machine learning-based models, such as random forest and ARIMA, are effective in predicting the prices of these items. The results demonstrate that the irregular forest model outperforms other models. The study recommends using the random forest and ARIMA models, building hybrid models, defining the problem as a time-series problem and incorporating date and time input characteristics into the suggested models.

Rohit Joshi et al.,[7] this research paper highlights the importance of understanding customer shopping preferences in the growing online retail industry in India. It gathered information from 124 Indian respondents spread across 18 states and constructed and verified Random Forest prediction models for a number of product categories. The results showed that the model had a high sensitivity for products like books and electronics, while having a low sensitivity for products like movies, sporting goods, and bags.

Ameena sherin et al.,[8] this paper states that Linear Regression, Decision Tree, and Random Forest are three unique algorithms, each of which is based on a different component of the data, that may be used to predict home values. This research study, "House Price Prediction," describes how to utilize these three algorithms to do so. The problem is that predicted prices vary depending on how accurate they are. To get around this problem, we calculate the average of the projections. As a result, it helps with mistake prevention. This average price is a fair representation of the worth of a house. Clients will be pleased with the approach since it may produce reliable results and reduce the possibility of error.

Liu et al.,[9] this research focus on comparing the performance of different machine learning models for prediction of housing prices. The authors compared different algorithm like Linear Regression, Decision Tree, and Random Forest Regression whereas the CNN Random Forest model gives wrong output as compared to the other models in terms of prediction accuracy, suggesting it could be a useful tool for property valuation. Future research should incorporate additional data beyond housing development to improve the model's accuracy.

III. METHODOLOGY

The methodology we employed in this proposed study is one that is frequently used in machine learning project work. Python programming language, Streamlit, Visual Studio Code, Jupyter, and Anaconda are used to implement the entire system. Figure 1 shows how to choose the method to utilise in the model. To forecast the price of the goods, Streamlit is being employed. With Streamlit, a free and open-source platform, users can produce beautiful machine learning and data science models and share them with others. This Python-based library was created in collaboration with ml engineers. While supervised learning develops a model that uses well-known input and output data to predict future results, unsupervised learning uncovers hidden patterns or internal structures in input data. if we need to train the data for making predictions or forecasting such as future values of a continuous variable i.e., temperature or prices we can use supervised learning.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

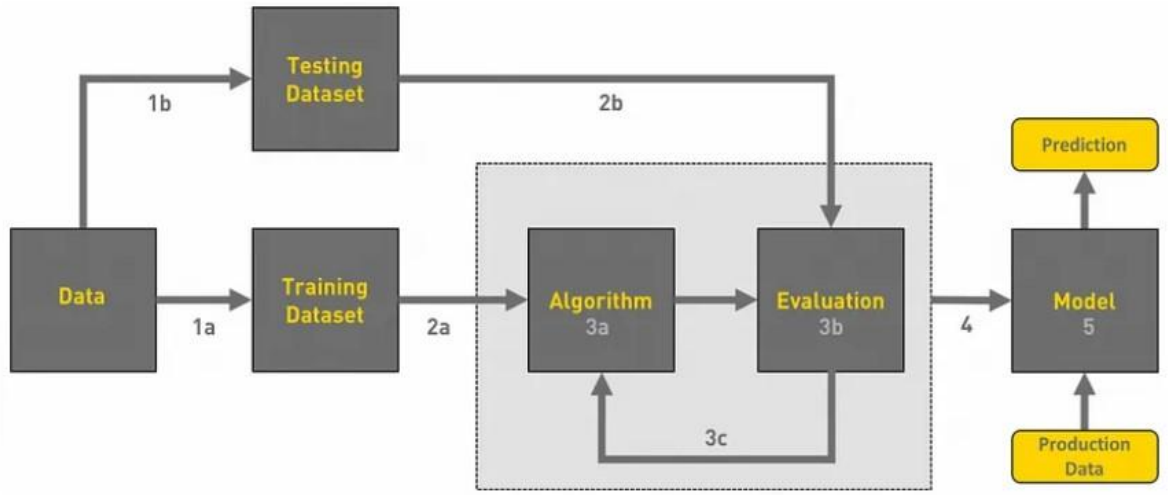

Figure 1 :Workflow of Machine Learning Model

3.1 Data Gathering

Data can be gathered from a variety of sources, including files, databases, sensors, and many other types of data sources. However, the data collected cannot be used directly for the analysis process because there may be a significant amount of missing data, extremely large values, unorganized text data, or noisy data. The process of data collection depends on the type of project we want to make. Kaggle or GitHub are some repositories that are used to collect the dataset for ML model building.

Here we have the dataset of laptop as product. Here historical data of laptop is used for the predictions. In this process we are using total 19 different company's laptops. There are total 1303 columns.

```
df.head()
```

	Unnamed: 0	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price
0	0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	71378.6832
1	1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	47895.5232
2	2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	30636.0000
3	3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	135195.3360
4	4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	96095.8080

Figure 2. Screenshot of first five rows in dataset

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1303 entries, 0 to 1302
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Unnamed: 0        1303 non-null   int64
 1   Company           1303 non-null   object
 2   TypeName          1303 non-null   object
 3   Inches            1303 non-null   float64
 4   ScreenResolution  1303 non-null   object
 5   Cpu               1303 non-null   object
 6   Ram               1303 non-null   object
 7   Memory            1303 non-null   object
 8   Gpu               1303 non-null   object
 9   OpSys             1303 non-null   object
 10  Weight            1303 non-null   object
 11  Price             1303 non-null   float64
dtypes: float64(2), int64(1), object(9)
memory usage: 122.3+ KB
```

Figure 3. Screenshot of datatypes in dataset

3.2 Data Preprocessing

Cleaning raw data, or transforming data that has been gathered from the real world into a clean data set, is known as data preprocessing. The goal of data pre-processing is converting raw data into clean that can be utilized to train a model. The (figure 3.6) shows the duplicate and null values in dataset using python functions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

```
df.duplicated().sum()
0

df.isnull().sum()
Unnamed: 0          0
Company             0
TypeName            0
Inches              0
ScreenResolution    0
Cpu                 0
Ram                 0
Memory              0
Gpu                 0
OpSys               0
Weight              0
Price               0
dtype: int64
```

Figure 4. Screenshot showing the duplicate/null values in dataset

IV. TRAINING AND TESTING DATASET

Following are the three sections that we initially divide the model into for training:-

- Training set: The training set refers to the data used to instruct the computer on handling is used to instruct the computer how to handle data. Machine learning employs algorithms to carry out the training phase. a set of data that is employed for learning, i.e., to match the classifier's parameters.
- Validation set: Cross-validation is frequently used in applied machine learning to evaluate a model's performance on untested data. The classification parameters are adjusted by synthesizingof classification are adjusted through the synthesis of unknown data from the training set.
- Test set: A collection of unobserved data that is only used to evaluate how well a fully expressed classifier performed.

```
sns.barplot(x=df['Company'],y=df['Price'])
plt.xticks(rotation='vertical')
plt.show()
```

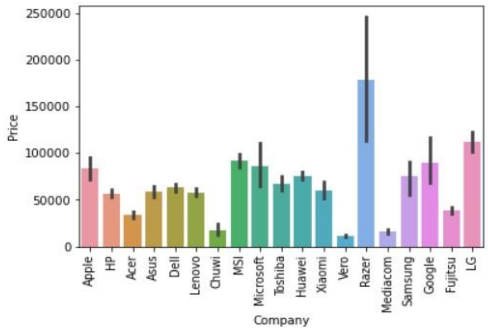

Figure 5. Barplot showing the company names along with price rate of laptop

```
sns.heatmap(df.corr())
<AxesSubplot:>
```

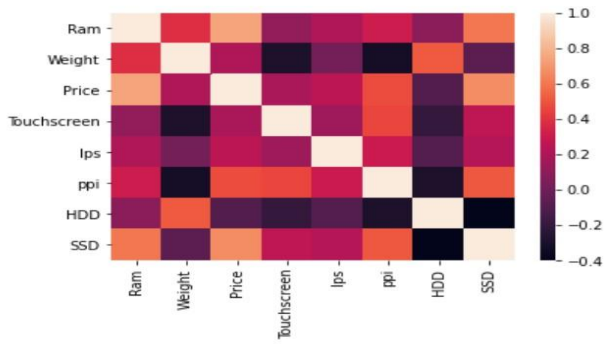

Figure 6. Heatmap for training data

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

V. EVALUATION AND MODEL DEVELOPMENT

5.1 Choosing an Efficient Model

Here we have to choose an efficient algorithm for the model building as it is based on the amount and quality of the dataset we have used. There are mainly two algorithms that can be used for price predictions as Linear Regression and Random Forest regressor.

5.2 Linear Regression

It ranks among the most used machine learning regression algorithms. The output variables (future values) are predicted using a significant variable from the data set. If the labels are continuous, such as the number of planes departing from an airport each day, etc., the linear regression algorithm is utilized.

$$Y = b*x + c$$

is the formula for linear regression.'Y' is the independent variable in the illustration above, whereas 'x' is the dependent variable. The slope of the line that gives us the output variables is referred to as "b" when you plot the linear regression, and its intercept is referred to as "c" [3]. The assumptions made by linear regression algorithms are that the relationship between the input and the output is linear. The first method was chosen because it is straightforward and takes relatively little time to train and test. Because of the feature vectors, the features were used directly without any feature mapping. For improved accuracy, we also apply regularization techniques [4].

5.3 Random Forest Regressor

A random forest is a meta-estimator that employs averaging to improve projected accuracy and decrease overfitting after fitting numerous classification decision trees to various dataset subsamples. The accuracy of the Random Forest Regression is high. It typically produces better results in prediction models. A data estimator for data about data is the rambling forest [7]. Numerous decision-makers were used for various subsamples of the provided data. The limit is exceeded. It increases predictability.

Algorithm:

Step 1: Select N chose records from the dataset.
Step 2: Create a decision tree based on N records.
Step 3a: Repeat steps 1 and 2 after selecting the number of trees from your algorithm.
Step 3b: In the case of a regression issue, each tree in the forest forecasts a value for Y (output) for a new record.

XG Boost Regressor

An excellent and efficient implementation of the gradient boosting approach is provided by the open-source Extreme Gradient Boosting (XGBoost) program. Effective gradient augmentation for regression-based predictive modeling is provided by the XGBoost.

Technology Used

- **Streamlit:** With Streamlit, a free and open-source platform, brilliant machine learning and data science web apps can be produced and distributed quickly. The use of a Python-based library is made.
- **Pandas:** Dataset manipulation is done using the Python package Pandas. It provides tools for data search, cleanup, analysis, and manipulation [8]. Pandas assist us in analysing huge data sets and coming to conclusions based on statistical concepts. Pandas can organize disorganized data sets, making them useful and readable.

VI. CONCLUSION

In this research paper, we predict the prices of products from their historical data. In this model we used laptop as example for prediction of prices. Different regression models such as Linear Regression and Random Forest Regression Lasso Regression, etc., have been used for the product price predictions. During the study, it is observed that the accuracy is increased with decrement in MAE "Mean Absolute Error". Here the MAE count of Random forest

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

Regression is lower than that of other algorithms and R^2 score of Random Forest regression , XGBoost, and Gradient Boost 0.8873, 8811, and 8823 resp. Hence we can conclude that the prediction algorithms i.e., Random Forest Regression, XGBoost Regressor gives better results in the advanced forecasting of product prices.

VII. FUTURE WORK

- As a part of future work, we aim to build the URL so that the user can access it automatically.
- In future we aim to bind the Machine Learning model with different websites on large scale and also to upload a large dataset of different kinds of products.

REFERENCES

[1]. Julakha Jahan Jul, M. M. Imran Molla, Bifta Sama Bari, Mamunar Rashid, "Flat Price Prediction Using Linear and Random Forest Regression Based on Machine Learning Techniques", 2020. (URL:https://www.researchgate.net/publication/342793117)

[2]. Yige Wang, "House-price Prediction Based on OLS linear Regression and Random Forest".

[3]. Xinshu Li, "Prediction and Analysis of Housing Price Based on the Generalized Linear Regression Model", 2022 Computational Intelligence and Neuroscience Volume 2022, Article ID 3590224.

[4]. UjjawalSonkambale, Suraj Kudale, Ojas Pawar, Tushar Salunke, "Use car price prediction", 2022 International Journal of Scientific Research in Computer Science, Engineering and Information Technology ISSN: 2456-3307 (www.ijsrcseit.com) Doi: https://doi.org/10.32628/IJSRCSEIT.

[5]. Subba Rao Polamuri, K. Srinivasi, A. Krishna Mohan, "Stock Market Prices Prediction Using Random Forest and Extra Tree Regression", 2019 International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019.

[6]. Mohamed Ali Mohamed, Ibrahim Mahmoud El-Henawy and Ahmad Salah, "Price Prediction of Seasonal Items Using Machine Learning and Statistical Methods", 2021 Tech Science Press, doi:10.32604/cmc.2022.020782.

[7]. Rohit Joshi, Rohan Gupte, Palanisamy Saravanan, "A Random Forest Approach For Predicting Online Buying Behaviour of Indian Customers", 2018 Scientific Research Publishing, Theoretical Economics Letters, 2018, 8, 448-475(URL:http://www.scirp.org/journal/tel).

[8]. Ameena Sherin. V, Ms Ankitha Philip, "House Price Prediction Using Linear Regression, Random Forest and Decision Tree Algorithms", Proceedings of the National Conference on Emerging Computer Applications (NCECA)-2022 Vol.4, Issue.1.

[9]. Mr B. Saireddy, A. Vamshikrishna, G. Abhilash, D. Vinith Srinivas, "Car Price Prediction Using Machine Learning", 2022 International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:02/February-2022.

[10]. Rola M. Elbakly, Magda M. Madbouly, and Shawkat K. Guirguis, "A Hybrid Approach for Product Price Prediction", 2022 European Journal of Engineering and Technology ResearchISSN: 2736-576XDOI: http://dx.doi.org/10.24018/ejeng.2022.7.5.2883.