APPLIED DATA ANALYTICS: A PRACTICAL APPROACH

PROJECT REPORT ON

LAPTOP PRICE PREDICTION MODEL

UNDER THE GUIDANCE OF

DR. MRIDU SAHU

CONTINUING EDUCATION CELL, NIT RAIPUR

PARTICIPANTS NAME AND CONTRIBUTION:

1. **Aahna Mishra (M-ADA-001):** Ridge and Lasso Regression Model Development
2. **Aditi Rai (M-ADA-006):** EDA and Data Pre-processing
3. **Bharat Kotwani (M-ADA-017):** Linear Regression and Random Forest Model Development
4. **GSS Bhishma Rao (M-ADA-027):** Solution Design, Results and Conclusions
5. **Sudhir Singh (M-ADA-067):** Gradient Boost Regression and SVM Model Development
6. **Tejaswi Kumar Tripathi (M-ADA-072):** Feature Extraction and Data Visualization

INTRODUCTION

The world of technology is constantly evolving, and one of the most ubiquitous devices in today's society is the laptop. Laptops have become an essential tool for work, education, entertainment, and communication. As the demand for laptops continues to soar, it becomes increasingly important for consumers and manufacturers to have insights into future laptop prices. This is where machine learning and predictive analytics come into play.

The project aims at solving the problem of predicting laptop prices as per their features. It involves using historical data, various market trends' analysis, and an analysis of the laptops, their specifications and prices, to predict correct and appropriate price values for laptops. In this project, we explore how machine learning can be used to predict the laptop prices accurately, by delving into and analysing the features and specifications data.

DATA

The project requires and involves the use of data on market trends, specifications and prices related to laptops and such devices. Here we have used the data scraping from various e-commerce websites that sell laptops. We have gathered data regarding the technologies involved, the exact specifications of the laptops, and their prices as per the features, to develop an algorithm to predict the prices.

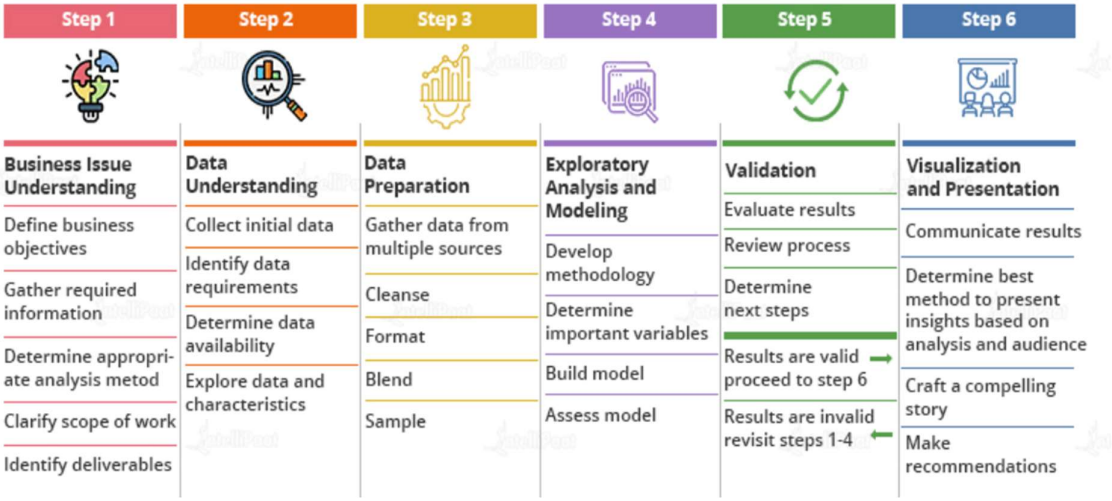
The dataset mainly includes four extracted features: product, MRP (target variable), rating, and features and specifications. The MRP serves as target variable, representing the price of the laptop. This is what our project aims at predicting.

Here we show a small example of the dataset derived and used in the project-

Product	Rating	MRP	Feature
Lenovo IdeaPad 3 Core i3 11th Gen - (8 GB/256 GB SSD/Windows 11 Home) 14ITL05 Thin and Light Laptop	4.2	₹36,990	Intel Core i3 Processor (11th Gen)8 GB DDR4 RAM64 bit Windows 11 Operating Sy
Lenovo IdeaPad 3 Core i3 11th Gen - (8 GB/512 GB SSD/Windows 11 Home) 82H801L7IN | 82H802FJIN | 82H802...	4.2	₹39,990	Intel Core i3 Processor (11th Gen)8 GB DDR4 RAM64 bit Windows 11 Operating Sy
ASUS VivoBook 15 (2022) Core i3 10th Gen - (8 GB/512 GB SSD/Windows 11 Home) X515JA-EJ362WS | X515JA-E...	4.3	₹32,990	Intel Core i3 Processor (10th Gen)8 GB DDR4 RAM64 bit Windows 11 Operating Sy
HP Pavilion Ryzen 5 Hexa Core AMD R5-5600H - (8 GB/512 GB SSD/Windows 10/4 GB Graphics/NVIDIA GeForce ...	4.4	₹49,990	AMD Ryzen 5 Hexa Core Processor8 GB DDR4 RAM64 bit Windows 10 Operating S
ASUS TUF Gaming F15 Core i5 10th Gen - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVIDIA GeForce G...	4.4	₹49,990	Intel Core i5 Processor (10th Gen)8 GB DDR4 RAMWindows 11 Operating System5
HP 14s Intel Core i3 11th Gen - (8 GB/256 GB SSD/Windows 11 Home) 14s - dy2507TU Thin and Light Laptop	4.2	₹36,990	Intel Core i3 Processor (11th Gen)8 GB DDR4 RAM64 bit Windows 11 Operating Sy
DELL Inspiron Core i3 11th Gen - (8 GB/1 TB HDD/256 GB SSD/Windows 11 Home) D560841WIN9S Thin and Ligh...	4.3	₹38,990	Processor: Intel i3-1115G4 (Base- 1.70 GHz & Turbo up to 4.10 GHz) 2 CoresRAM 8
RedmiBook Pro Core i5 11th Gen - (8 GB/512 GB SSD/Windows 11 Home) Thin and Light Laptop	4.1	₹39,990	Intel Core i5 Processor (11th Gen)8 GB DDR4 RAMWindows 11 Operating System5
realme Book (Slim) Core i3 11th Gen - (8 GB/256 GB SSD/Windows 10 Home) RMNB1001 Thin and Light Laptop	4.4	₹46,990	Stylish & Portable Thin and Light Laptop14 inch 2K QHD, IPS LCD Display (400nits
ASUS VivoBook 14 (2022) Ryzen 7 Quad Core AMD R7-3700U - (16 GB/512 GB SSD/Windows 11 Home) M415DA-EB5...	4.1	₹44,990	AMD Ryzen 7 Quad Core Processor16 GB DDR4 RAM64 bit Windows 11 Operating
HP Pavilion Ryzen 5 Hexa Core AMD R5-5600H - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVIDIA GeF...	4.4	₹59,990	AMD Ryzen 5 Hexa Core Processor8 GB DDR4 RAM64 bit Windows 11 Operating S
Lenovo IdeaPad 1 Ryzen 5 Quad Core 3500U - (8 GB/512 GB SSD/Windows 11 Home) 15ADA7 Thin and Light Lap...	4.2	₹38,705	AMD Ryzen 5 Quad Core Processor8 GB DDR4 RAM64 bit Windows 11 Operating S
ASUS Core i3 11th Gen - (8 GB/256 GB SSD/Windows 11 Home) X515EA-EJ312WS Thin and Light Laptop	4.3	₹32,990	Intel Core i3 Processor (11th Gen)8 GB DDR4 RAM64 bit Windows 11 Operating Sy
ASUS TUF Gaming A17 with 90Whr Battery Ryzen 5 Hexa Core AMD R5-4600H - (8 GB/512 GB SSD/Windows 11 Ho...	4.4	₹52,990	AMD Ryzen 5 Hexa Core Processor8 GB DDR4 RAM64 bit Windows 11 Operating S
acer Extensa Core i3 11th Gen - (8 GB/256 GB SSD/Windows 11 Home) EX 215-54 Thin and Light Laptop	4.3	₹29,990	Intel Core i3 Processor (11th Gen)8 GB DDR4 RAM64 bit Windows 11 Operating Sy
ASUS ROG Strix G15 (2022) with 90Whr Battery Ryzen 9 Octa Core 6900HX - (16 GB/1 TB SSD/Windows 11 Hom...		₹1,96,990	AMD Ryzen 9 Octa Core Processor16 GB DDR5 RAMWindows 11 Operating Systen
ASUS VivoBook 15 (2022) Core i5 11th Gen - (8 GB/1 TB HDD/256 GB SSD/Windows 11 Home) X515EA-EJ502WS T...	4.3	₹47,990	Intel Core i5 Processor (11th Gen)8 GB DDR4 RAM64 bit Windows 11 Operating Sy

METHODOLOGY

- To begin with, we first understand the problem and the objectives that we have to achieve. Then we gather the needed information accordingly and try to figure out an appropriate method for the analysis.
- Then we identify the data and datatype required for the analysis. We determine whether the required data is available or not, and look for data sources, and collect it. We then explore the data and characteristics.
- After gathering the data from various sources, it is then filtered, sampled and organised as per our needs.
- Then we develop a methodology for price analysis, determine important variables, and a build a suitable model as per the needs of the project.
- Then we evaluate the results and review the whole process. If the results are correct then we have to move further, but if the results are invalid we need to reconsider the initial steps taken.
- The results are then presented and recommendations are made accordingly.



DATA PROCESSING

Considering the data extracted from various sources, first certain features are derived as per our needs, and need of the project.

	Brand Name	OS	Ram Type	RAM	Processor	GPU	Warranty	ScreenSize	Disk Type	DISK SIZE	Price
0	Lenovo	Windows	DDR4	8	Intel	Others	2	14.0	SSD	256	10.518430
1	Lenovo	Windows	DDR4	8	Intel	Others	2	15.6	SSD	512	10.596410
2	ASUS	Windows	DDR4	8	Intel	Others	1	15.6	SSD	512	10.403990
3	HP	Windows	DDR4	8	AMD	AMD	1	15.6	SSD	512	10.819598
4	ASUS	Windows	DDR4	8	Intel	Others	1	15.6	SSD	512	10.819598

From the above data, following features have been added to the original data –

- Brand name
- Operating System
- RAM Type
- RAM Size
- Processor
- GPU
- Warranty
- Screen Size
- Disk Type
- Disk Size

The original data combined with the above derived features has been used for exploratory data analysis and modelling. Using the above data, analysing it, and making comparisons between prices and specifications, prices for any new models of laptops are derived, using a suitable machine learning algorithm.

MODEL DEVELOPMENT

Linear Regression Model:

It works by finding the line of best fit that minimizes the distance between the predicted values and the actual values in the training data.

Linear Regression is one of the easisest and among the most popular machine learning algorithms. It is a statistical method which is used to make predictions for continuous real or numerical values, such as sales, salary, prices, etc.

This method shows a linear relationship between a dependent (x) and one or more independent variables (y).

Random Forest Model:

This model combines multiple decision trees to reduce overfitting and improve performance. The greater is the number of trees in the model, the more will be the accuracy, and lesser will be the problem of overfitting.

Random Forest technique is a supervised learning technique which can be used for both classification and regression problems.

Ridge Regression Model:

This technique works by adding a penalty term to the least squares equation that minimizes the sum of the squared residuals.

This is a tuning method that is used to analyse any data suffering from multi-collinearity. It performs L2 regularization.

Lasso Regression:

This model works by adding a penalty term to the least squares equation that minimizes the sum of the squared residuals but useful when there is multicollinearity

between the independent variables, and it can also help to prevent overfitting in cases where the number of independent variables is greater than the number of observations.

Lasso stands for 'least absolute shrinkage and selection operator', and it is an algorithm which helps in elimination of irrelevant parameters.

Gradient Boost Regression:

This method combines multiple weak models into a strong model.

In this method every new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent. Multiple such iterations are done to compute the gradient of the loss function corresponding to the predictions of the current ensemble and then a new model is trained to minimize this gradient.

SVM:

This method works by finding a hyperplane that separates the data into two classes with the largest margin possible.

SVM stands for Support Vector Machine algorithm. It is mainly used for classification problems in machine learning.

Out the above six models developed, Gradient Boost Regression has the highest R2 score of 0.87 followed by Random Forest Model.

```
print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.8708143222009221
```

MODEL NAME	R2 SCORE
Gradient Boost Regression	0.870
Random Forest	0.869
Linear Regression	0.861
Lasso Regression	0.859
SVM	0.853
Ridge Regression	0.814

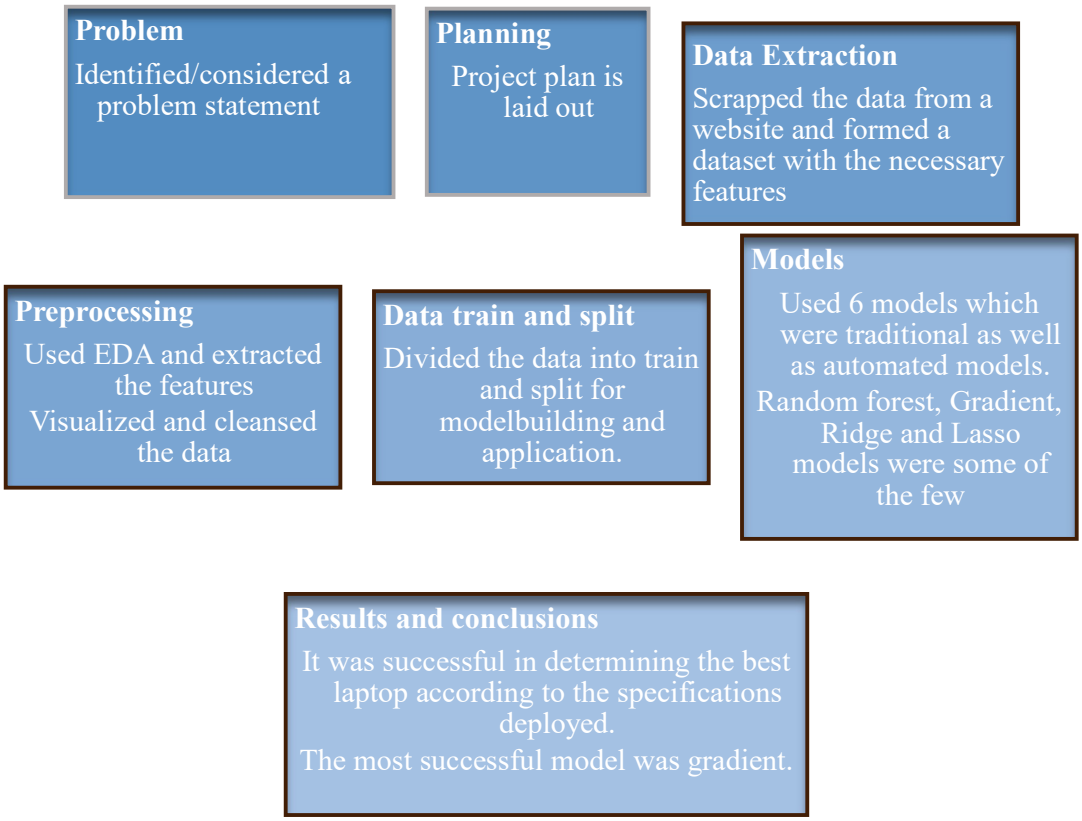
RESULT

Gradient Bosst Regression Method is considered for this project on estimation of new laptop prices, as it gives us the highest R2 value.

```
print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.8708143222009221
```

Following was the methodology used-



GRADIENT BOOST REGRESSION MODEL

```
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,2,3,4,5,6,7,8,9])
],remainder='passthrough')

step2 = GradientBoostingRegressor(n_estimators=500)

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])

pipe.fit(X_train,y_train)

y_pred = pipe.predict(X_test)

print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.8708143222009221
MAE 0.1496866073468099
```

FURTHER IMPROVEMENTS AND SCOPE

- Incorporating deep learning models, such as neural networks, can potentially capture complex patterns and relationships in laptop specifications and prices.
- Including natural language processing (NLP) techniques can enable better understanding and extraction of features from unstructured data sources like customer reviews and descriptions.
- Continued improvement in data collection and availability will contribute to enhanced model performance and generalizability.
- Integration of real-time data feeds and market trends can provide up-to-date insights for more precise laptop price predictions.

REFERENCE

Theme	Research Article
Laptop Price Prediction using Machine Learning Algorithms	Shaik, Mohammed Ali, et al. "Laptop Price Prediction using Machine Learning Algorithms." *2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*. IEEE, 2022.
Laptop Price Prediction using Machine Learning	Surjuse, Vaishali, et al. "Laptop Price Prediction using Machine Learning." (2022).
Laptop Price Prediction Using Real Time Data	Reddy, Chada Lakshma, et al. "Laptop Price Prediction Using Real Time Data." *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*. IEEE, 2023.
Laptop Price Prediction with Machine Learning Using Regression Algorithm	Siburian, Astri Dahlia, et al. "Laptop Price Prediction with Machine Learning Using Regression Algorithm." *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)* 6.1 (2022): 87-91.