LAPTOP
PRICE
PREDICTION

Group members

Neeraj Nishad (2100980140031)

Vikas Prajapati (2100980140058)

Contents

Introduction

Laptop price prediction is a challenging task due to the dynamic nature of the market and the many factors that can affect pricing. The goal of this model is to accurately predict the prices of laptops based on various features.

Laptop price prediction is important for both buyers and sellers. Buyers can use the model to predict the prices of laptops and make informed purchasing decisions, while sellers can use the model to set appropriate prices for their products.

Problem Statement

We will make a project for Laptop price prediction. The problem statement is that if any user wants to buy a laptop then our application should be compatible to provide a tentative price of laptop according to the user configurations. Although it looks like a simple project or just developing a model, the dataset we have is noisy and needs lots of feature engineering, and preprocessing that will drive your interest in developing this project.

Data Exploration

Most of the columns in a dataset are noisy and contain lots of information. But with feature engineering you do, you will get more good results. The only problem is we are having less data but we will obtain a good accuracy over it. The only good thing is it is better to have a large data. we will develop a website that could predict a tentative price of a laptop based on user configuration.

Dataset

|		Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	O
1										
2	0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	m
3	1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	m
4	2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	N
5	3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	m
6	4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	m
7	5	Acer	Notebook	15.6	1366x768	AMD A9-Series 9420 3GHz	4GB	500GB HDD	AMD Radeon R5	W
8	6	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.2GHz	16GB	256GB Flash Storage	Intel Iris Pro Graphics	M
9	7	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	256GB Flash Storage	Intel HD Graphics 6000	m
10	8	Asus	Ultrabook	14.0	Full HD 1920x1080	Intel Core i7 8550U 1.8GHz	16GB	512GB SSD	Nvidia GeForce MX150	W
11	9	Acer	Ultrabook	14.0	IPS Panel Full HD 1920x1080	Intel Core i5 8250U 1.6GHz	8GB	256GB SSD	Intel UHD Graphics 620	W
12	10	HP	Notebook	15.6	1366x768	Intel Core i5 7200U 2.5GHz	4GB	500GB HDD	Intel HD Graphics 620	N
13	11	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i3 6006U 2GHz	4GB	500GB HDD	Intel HD Graphics 520	N
14	12	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.8GHz	16GB	256GB SSD	AMD Radeon Pro 555	m
15	13	Dell	Notebook	15.6	Full HD 1920x1080	Intel Core i3 6006U 2GHz	4GB	256GB SSD	AMD Radeon R5 M430	W
16	14	Apple	Ultrabook	12.0	IPS Panel Retina Display 2304x1440	Intel Core M m3 1.2GHz	8GB	256GB SSD	Intel HD Graphics 615	m
17	15	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	256GB SSD	Intel Iris Plus Graphics 640	m
18	16	Dell	Notebook	15.6	Full HD 1920x1080	Intel Core i7 7500U 2.7GHz	8GB	256GB SSD	AMD Radeon R5 M430	W
19	17	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.9GHz	16GB	512GB SSD	AMD Radeon Pro 560	m

Dataset

Exploratory Data Analysis

Exploratory analysis is a process to explore and understand the data and data relationship in a complete depth so that it makes feature engineering and machine learning modeling steps smooth and streamlined for prediction. EDA involves Univariate, Bivariate, or Multivariate analysis. EDA helps to prove our assumptions true or false. In other words, it helps to perform hypothesis testing. We will start from the first column and explore each column and understand what impact it creates on the target column. At the required step, we will also perform preprocessing and feature engineering tasks. our aim in performing in-depth EDA is to prepare and clean data for better machine learning modeling to achieve high performance and generalized models. so let's get started with analyzing and preparing the dataset for prediction.
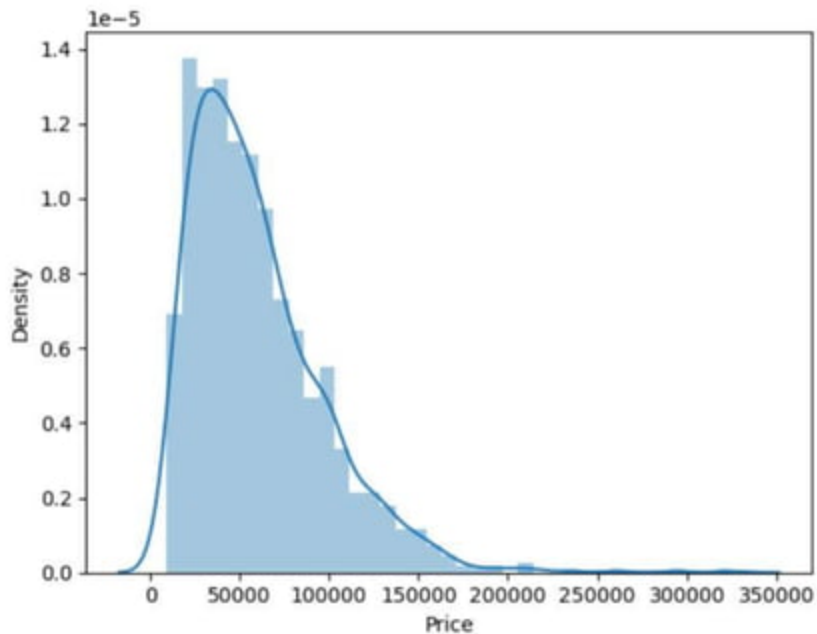
Data Distribution
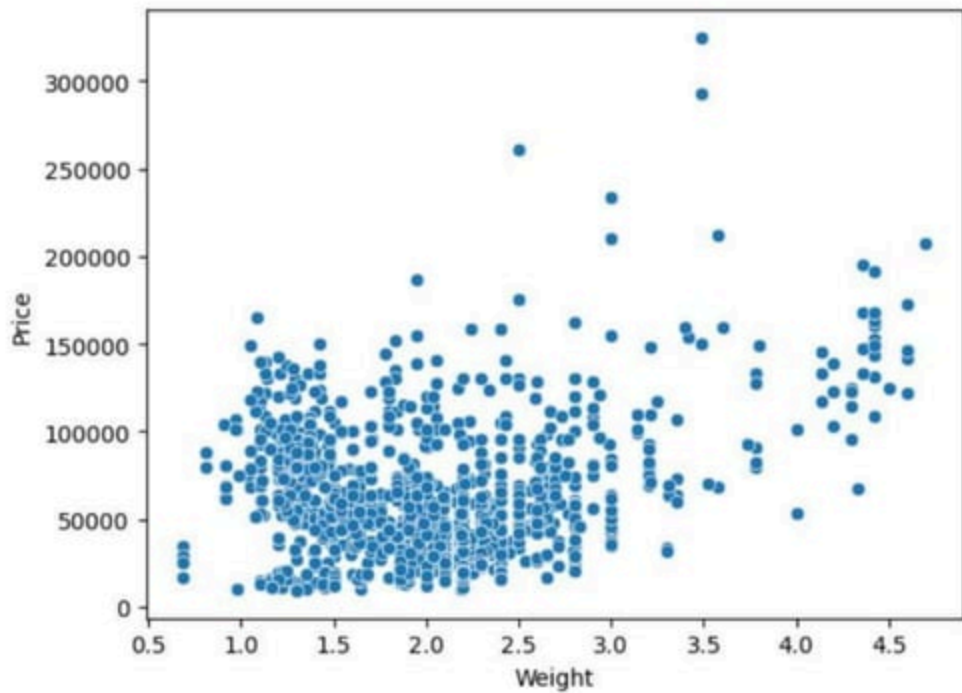
Price corr with weight

Feature Selection

Feature selection is the process of selecting a subset of the most relevant features from a larger set of features for use in building a prediction model. The goal of feature selection is to identify the most informative and relevant features that have the greatest impact on the target variable, while reducing the dimensionality of the data and decreasing the risk of over fitting.
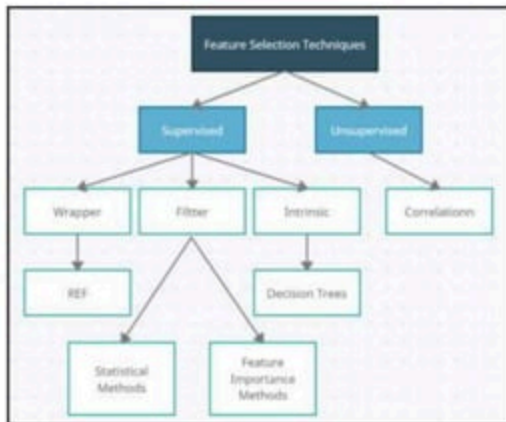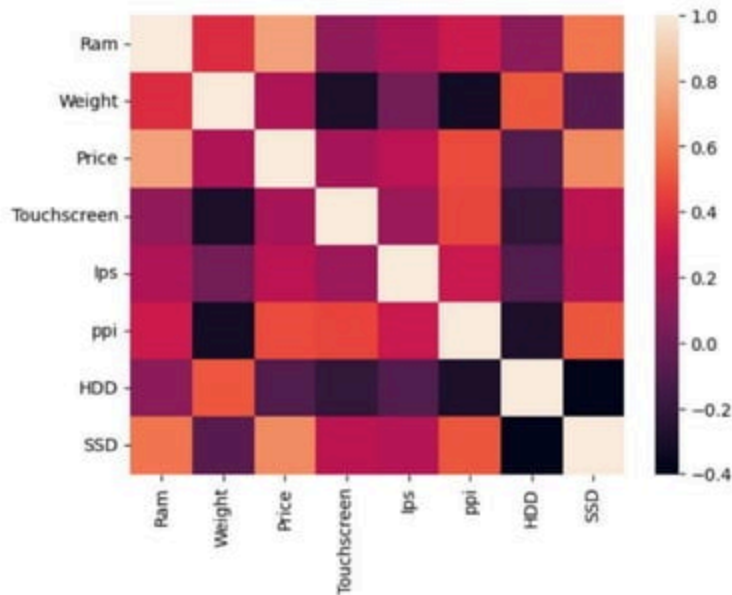
Methods of Feature selection:

- Correlation Analysis
- Mutual Information
- Feature Importance
- Lasso and Ridge regularization
- Dimensionality reduction

Reasons why Feature Selection is imp:

- Reducing the dimensionality of data
- Improving Model performance
- Identifying importance feature
- Reducing over fitting

Correlation analysis

Model Selection

In the first step for categorical encoding, we passed the index of columns to encode, and pass-through means pass the other numeric columns as it is. The best accuracy I got is with all-time favorite Random Forest. But you can use this code again by changing the algorithm and its parameters. I am showing Random forest. you can do Hyper parameter tuning using Grid search CV or Random Search CV. we can also do feature scaling but it does not create any impact on Random Forest.

Different ML models use for Model Selection:

- Linear Regression
- Ridge Regression
- Lasso Regression
- KNN
- Decision Trees
- Random Forest
- SVM
- Ada Boost
- Gradient Boost
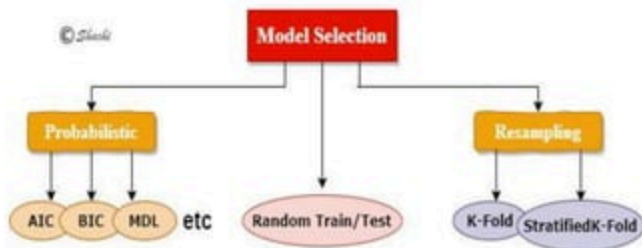
Model Evaluation

Random forest

```
In [108]: step1 = ColumnTransformer(transformers=[
               ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
          ],remainder='passthrough')

          step2 = RandomForestRegressor(n_estimators=100,
                                        random_state=3,
                                        max_samples=0.5,
                                        max_features=0.75,
                                        max_depth=15)

          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 0.8873402378382488
          MAE 0.15860130110457718
```

Selected Model with R2 Score:

Feasibility Studies

- Data availability
- Data quality
- Computational resources
- Model complexity
- Model interpretability
- Model performance
- Legal and ethical considerations
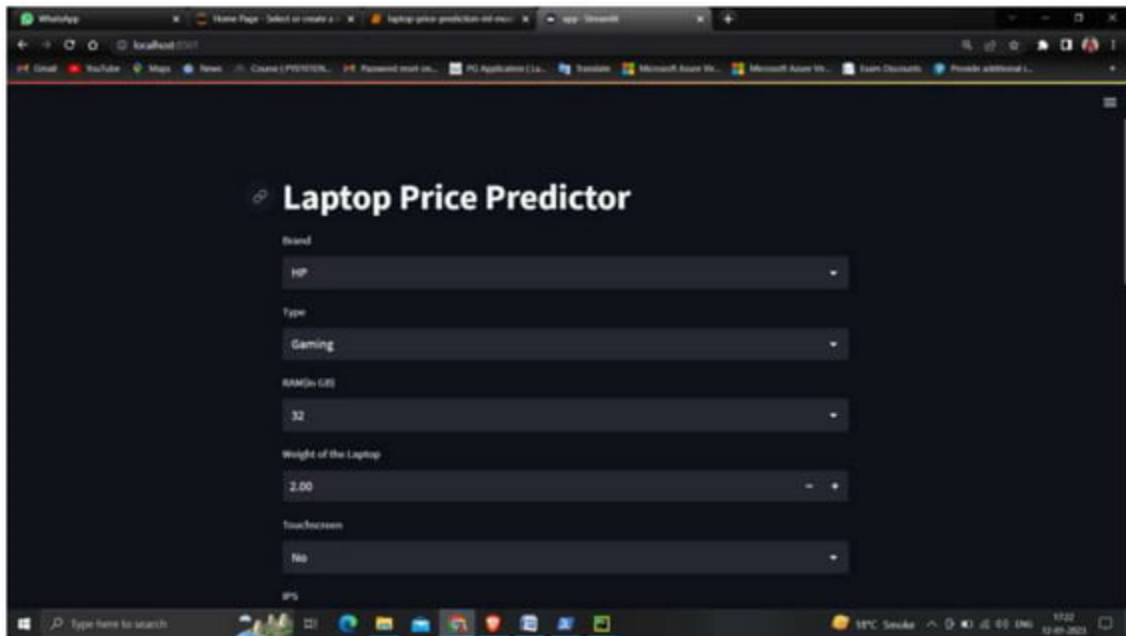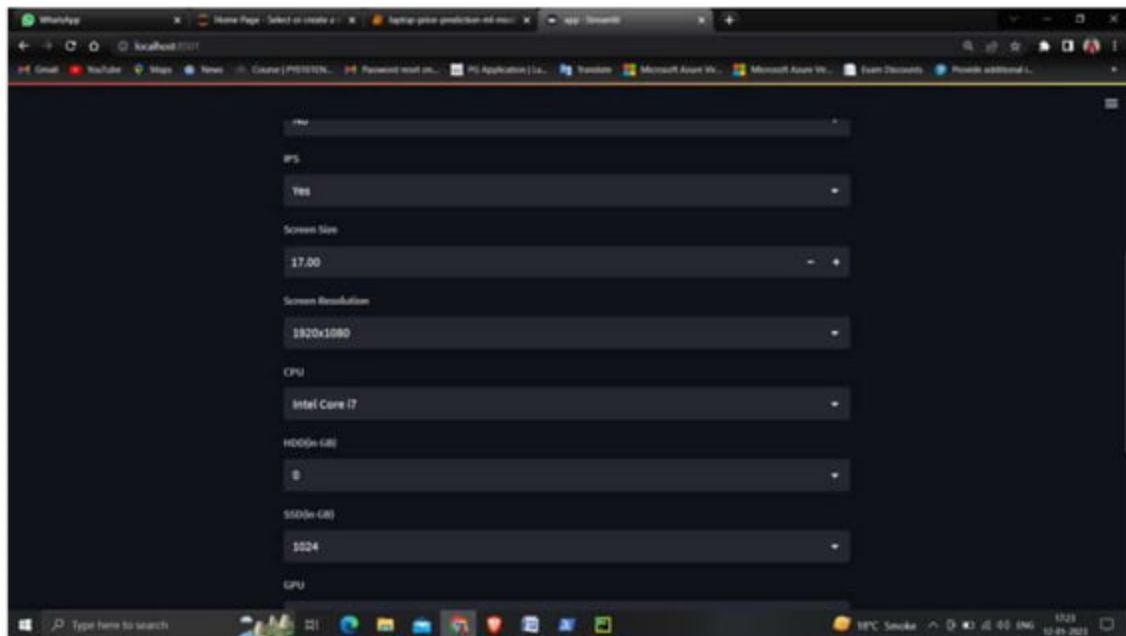- Cost and benefit analysis

Limitations

- Data availability
- Data quality
- Computational resources
- Model complexity
- Model interpretability
- Real-world variability
- Non-linear relationships
- Over fitting

User Interface

Conclusion

- **Summary of key findings:** Summarize the key findings of the model development process, including the performance of the final model and any insights into the factors that affect laptop prices.
- **Model performance:** Provide a summary of the model's performance, such as the accuracy, recall, precision, F1-score, and AUC-ROC, and compare it with the benchmark models or with other models that were considered.
- **Model interpretability:** Discuss how the model provides insights into the factors that affect laptop prices, and how it can be used to make more informed decisions.
- **Potential Impact:** Discuss the potential impact of the model on the domain, such as how it can be used to improve pricing strategies and make better decisions for buyers and sellers.
- **Limitations:** Summarize the limitations of the model, such as the availability of data, the quality of data, or the complexity of the model, and discuss how they may have affected the results.
- **Future work:** Discuss any future work that could be done to improve the model, such as collecting more data, incorporating additional features, or testing different model architectures.
- **Conclusion:** Sum up the main takeaways from the model development and results and its potential impact on the domain.