Electric Device Rating Prediction

TA: Ala

|  |  |  |  |
| --- | --- | --- | --- |
| Names | Seat Number | Department | Sections |
| منار محمد حمدي عبداللطيف | 2021170545 | CS | 7 |
| رنيم محمد عبدالرازق محمود | 2021170197 | CS | 3 |
| شهد السيد احمد محمد | 2021170259 | CS | 4 |
| علياء أسامة إسماعيل محمد | 2021170343 | CS | 4 |
| منة الله مصطفي أحمد رسلان | 2021170549 | CS | 7 |
| ندى أحمد فتح الله محمد إسماعيل | 2021170580 | CS | 7 |

2024

****

**Preprocessing:**

* **Label Encoder:**

Columns encoded using Label Encoder:

* Brand
* Processor\_Brand
* Processor\_Name
* Ram\_Type
* OS
* **Ordinal Encoder:**

Ordinal Encoder is used here to encode the categorical columns with ordinal relationships into numerical format by assigning a unique integer value to each category based on the order or ranking of these categories.

Columns encoded using Ordinal Encoder:

* Ram\_GB
* SSD
* HDD
* Graphic\_Card\_GB
* Warranty
* Processor\_Gnrtn
* Rating
* **One Hot Encoding:**

One Hot Encoder is used here because the following columns are of binary nature with Yes and No values.

Columns encoded using One Hot Encoder:

* Touchscreen
* MSOffice
* Weight
* **Scaling:**

Standard Scaler is used to standardize the numbers so that the weights of the columns Price, Number of Reviews and Number of Ratings will not have a higher effect than the other columns on the model.

* **Train Test Splitting:**

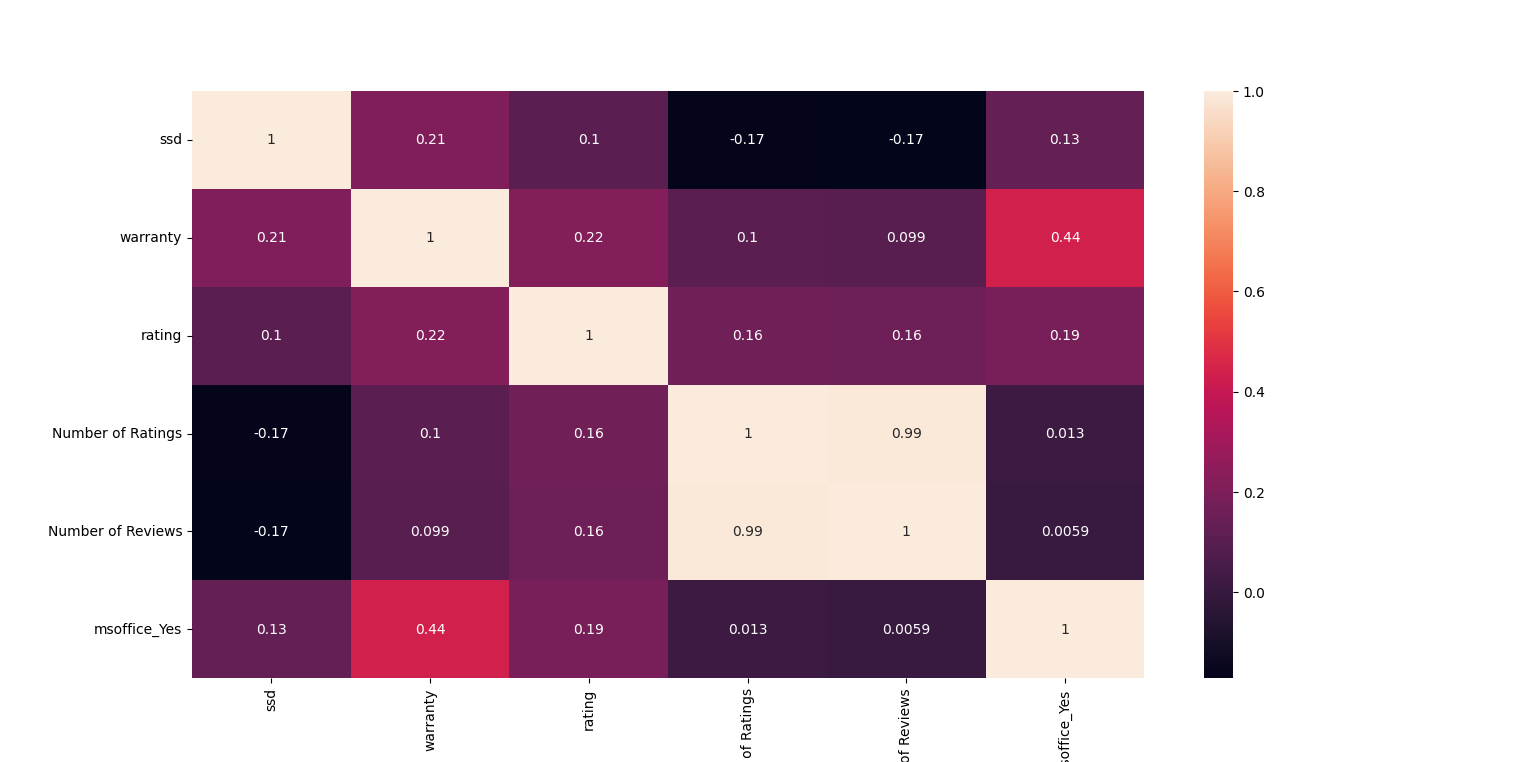
Training Data Size: 80%

Test Data Size: 20%

**Feature Selection and Model Evaluation:**

* **Correlation:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Correlation Value | Columns Selected | Model | Mean Square Error  Train | Mean Square Error Test |
| > 0.2 | Warranty | Linear Regression | 0.9889784935833783 | 0.8039653953542444 |
| > 0.2 | Warranty | Polynomial Regression  (Degree = 2) | 0.9859545163365274 | 0.7840866388843362 |
| > 0.2 | Warranty | Polynomial Regression  (Degree = 3) | 0.958924136098229 | 0.8020200169714738 |
| > 0.1 | SSD,  Warranty, Number of Ratings,  Number of Reviews, MSOffice\_Yes | Linear Regression | 0.949997878082261 | 0.7791479286137523 |
| > 0.1 | SSD,  Warranty, Number of Ratings,  Number of Reviews, MSOffice\_Yes | Polynomial Regression  (Degree = 2) | 0.8700011236847669 | 0.7786521456407908 |
| > 0.1  (Overfitting Case) | SSD,  Warranty, Number of Ratings,  Number of Reviews, MSOffice\_Yes | Polynomial Regression  (Degree = 3) | 0.7778776154851429 | 0.8194912339637208 |

****

Correlation > 0.1

**A black and white squares

Description automatically generated**

Correlation > 0.2

* **Wrapper Methods:**
  + **Forward Selection:**

**Hyper-Parameters:**

* Model used for Selection: Linear Regression
* K\_features: Best
* Scoring Function: neg\_mean\_squared\_error.
* CV: 5

**Selected Columns:**

* processor\_brand
* processor\_name
* ram\_gb
* hdd
* warranty
* Price
* Number of Ratings
* msoffice\_Yes
* weight\_Gaming
* **Model Evaluation On The Selected Columns:**

|  |  |  |
| --- | --- | --- |
| Model | Mean Square Error  Train | Mean Square Error Test |
| Linear Regression | 0.9220701759889802 | 0.810914530105409 |
| Polynomial Regression  (Degree = 2) | 0.747782803177376 | 0.653859399732828 |
| Polynomial Regression  (Degree = 3)  (Overfitting Case) | 0.5331824600760237 | 9.892327069433877 |
| SVR | 0.7974028690928058 | 0.7274617686838897 |
| Decision Tree Regressor  (Overfitting Case) | 1.9216115962766376e-30 | 0.5450601128223309 |

* + **Backward Elimination:**

**Hyper-Parameters:**

* Model used for Selection: Linear Regression
* K\_features: Best
* Scoring Function: neg\_mean\_squared\_error.
* CV: 5

**Selected Columns:**

* processor\_brand
* processor\_name
* ram\_gb
* hdd
* warranty
* Price
* msoffice\_Yes
* weight\_Gaming
* **Model Evaluation On The Selected Columns:**

|  |  |  |
| --- | --- | --- |
| Model | Mean Square Error  Train | Mean Square Error Test |
| Linear Regression | 0.9376583835980458 | 0.8337491645809688 |
| Polynomial Regression  (Degree = 2) | 0.806326705870765 | 0.7168159545532719 |
| Polynomial Regression  (Degree = 3)  (Overfitting Case) | 0.6461094364758808 | 6.575147164706286e+18 |
| SVR | 0.8695870913182452 | 0.834543305476701 |
| Decision Tree Regressor  (Overfitting Case) | 0.04740199827774366 | 1.1524993274454396 |

**Conclusion:**

* Model with Best Performance: Polynomial Regression
* Feature Selection Technique used with it: Forward Selection (Wrapper Method)
* Mean Square Error Train: 0.747782803177376
* Mean Square Error Test: 0.653859399732828