FLIP ROBO

REVIEWS & RATINGS PROJECT

Submitted by:

NATASHA PODDAR

.

INTRODUCTION

- Business Problem Framing

  In today's world where reviews are the back bone of sales for any product, it is extremely essential to have a model that could predict the product ratings for easier and faster decision making for customers.
- Conceptual Background of the Domain Problem

  Basic knowledge on comments and types are needed
- Motivation for the Problem Undertaken

  This issue is very realistic and common in today's world and one should know to deal with such situations in the future

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  Count & Unique values were checked

  Correlation with all independent variables and wrt target were checked

  Skewness was checked and tools were applied to control them and scale the data

  Models were applied to train and test the model

- Data Sources and their formats

Detailed data was scraped from Amazon & Flipkart site

Data Preprocessing Done

What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

1. Duplicate values check
2. Unique & Count of all columns were checked
3. Correlation check
4. Graphical Univariate Analysis
5. Feature Extraction

- Hardware and Software Requirements and Tools Used

1. Pandas – For Data Reading and understanding
2. Duplicate- To check for duplicate Values
3. CORR-To check Correlation
4. Numpy- For mathematical operations
5. KNN CLASSIFIER (SKLEARN) – Training & Testing the model
6. DECISION TREE CLASSIFIER (SKLEARN) – Training & Testing the model
7. MULTINOMIAL NB (SKLEARN) – Training & Testing the model
8. SVC CLASSIFIER- Training & Testing the model
9. RANDOM FOREST CLASSIFIER- Training & Testing the model
10. GRADIENT BOOSTING CLASSIFIER- Training & Testing the model
11. CROSS VAL SCORE – Regularizing the model
12. GRID SEARCH CV- Hyper Tuning the Model for higher accuracy
13. SEABORN- VISUALIZATION LIBRARY –, COUNTPLOTS, BOXPLOTS and other graphs
14. MATPLOTLIB.PY PLOT -Visualization tool

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Correlation with all independent variables and wrt target were checked
  Skewness was checked and tools were applied to control them and scale the data
  Models were applied to train and test the model

- Testing of Identified Approaches (Algorithms)
  1. SVC CLASSIFIER
  2. KNN CLASSIFIER
  3. DECISION TREE CLASSIFIER
  4. MULTINOMIAL NB CLASSIFIER
  5. RANDOM FOREST CLASSIFIER
  6. GRADIENT BOOSTING CLASSIFIER

- Key Metrics for success in solving problem under consideration
  1. ACCURACY SCORE
  2. CONFUSION MATRIX
  3. CLASSIFICATION REPORT
  4. F1 SCORE
  5. PRECISION
  6. RECALL SCORE
  7. AUC-ROC SCORE

- Visualizations

Seaborn Library was used along with matplotlib Library for visualizations

- Interpretation of the Results

All the models predicted an accuracy in the range of 40-55% , Random Forest Classifier was selected and hyper tuned and implemented. The accuracy was second highest and the difference was comparatively lower

CONCLUSION

- Key Findings and Conclusions of the Study

  All the models predicted an accuracy in the range of 40-55% , Random Forest Classifier was selected and hyper tuned and implemented. The accuracy was second highest and the difference was comparatively lower

- Learning Outcomes of the Study in respect of Data Science

  Random Forest Classifier was selected and hyper tuned and the same model was implemented.

  With unique feature we realized the type of data all the columns had, The various visualization tools helped in understanding the different relationships between the variables .Cross Val score helped in regularizing the model