

**Rating Prediction Project**

**Submitted By:**

**Pankaj Pagare**

**ACKNOWLEDGMENT**

I would like to express my gratitude towards FlipRobo Technologies for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons and my mentor **Ms. Swati Mahaseth** for giving me such attention and time as and whenever required.

**INTRODUCTION**

**Business Problem Statement**

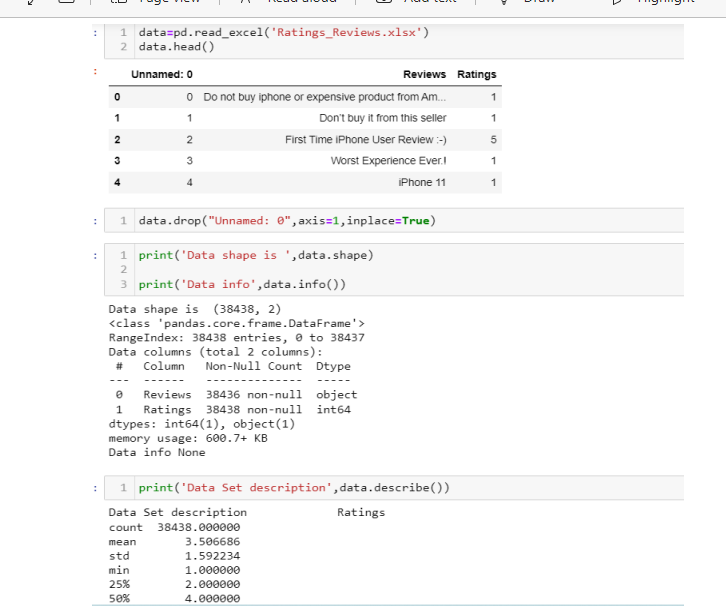
* We have a client who has a website where people write different reviews for technical products.
* Now they are adding a new feature to their website i.e. The reviewer will have to add stars(rating) as well with the review.
* The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars.
* Now they want to predict ratings for the reviews which were written in the past and they don’t have a rating.
* So, we have to build an application which can predict the rating by seeing the review.

**Conceptual Background of the Domain Problem**

* There are many users who purchase products through E-commerce websites.
* Through online shopping many E-commerce enterprises were unable to know whether the customers are satisfied by the services provided by the firm.
* This boosts us to develop a system where various customers give reviews about the product and online shopping services, which in turn help the E-commerce enterprises and manufacturers to get customer opinion to improve service and merchandise through mining customer reviews.
* An algorithm could be used to track and manage customer reviews, through mining topics and sentiment orientation from online customer reviews. In this system user will view various products and can purchase products online.
* Customer gives review about the merchandise and online shopping services.
* Certain keywords mentioned in the customer review will be mined and will be matched with the keywords which are already exist in the database based on the comparison, system will rate the product and services provided by the enterprise.

**Analytical Problem Framing**

**Dataset Representation**

****

* **It Is Classification Problem**

**Data Sources and their formats & inferences**

* Reviews of the product
* Rating of the products

**Observation**

* One Independent variable with one target variable.
* From the dataset we can infer that it is clearly a classification problem.

**Assumption For The Problem**

* From the dataset we can infer that it is clearly a Classification problem.
* This system will use text mining algorithm in order to mine keywords. The System takes review of various users, based on the review, system will specify whether the products and services provided by the E-commerce enterprise is good, bad, or worst.
* We use a database of sentiment-based keywords along with positivity or negativity weight in database and then based on these sentiment keywords mined in user review is ranked.
* This system is a web application where user will view various products and purchase products online and can give review about the merchandise and online shopping services.
* This system will help many E-commerce enterprises to improve or maintain their services based on the customer review as well as to improve the merchandise based on the customer review.

**Hardware and Software Requirements**

**Software Used :**

* Jupiter Notebook
* MS-Office

**Hardware Used :**

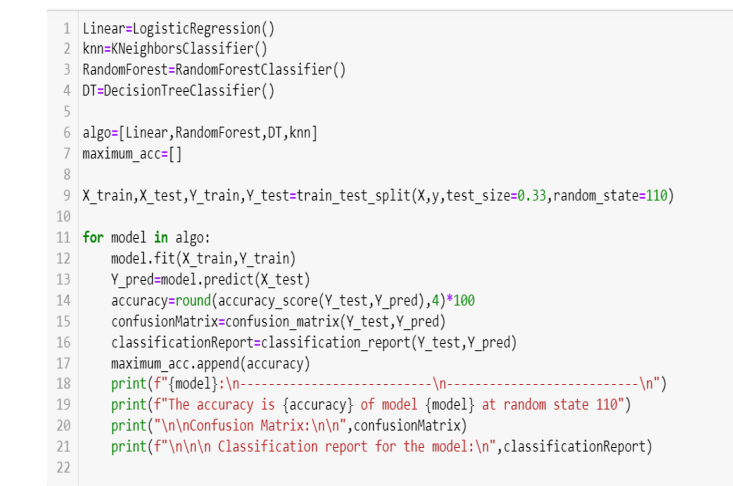
* PC **(Personal Computer)**

**Modes Development And Evaluation**

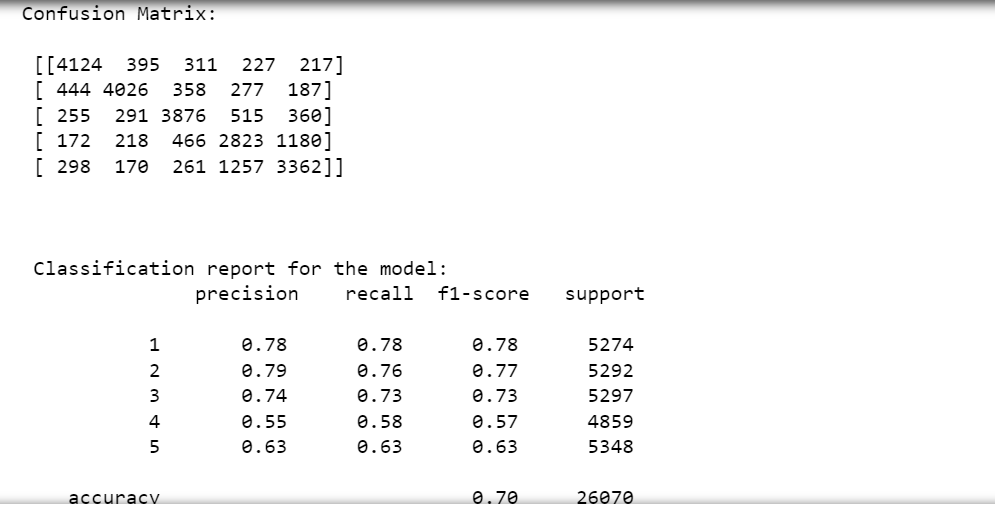
**Testing of Identified Approaches**

* LogisticRegression()
* KNeighborsClassifier()
* RandomForestClassifier()
* DecisionTreeClassifier()

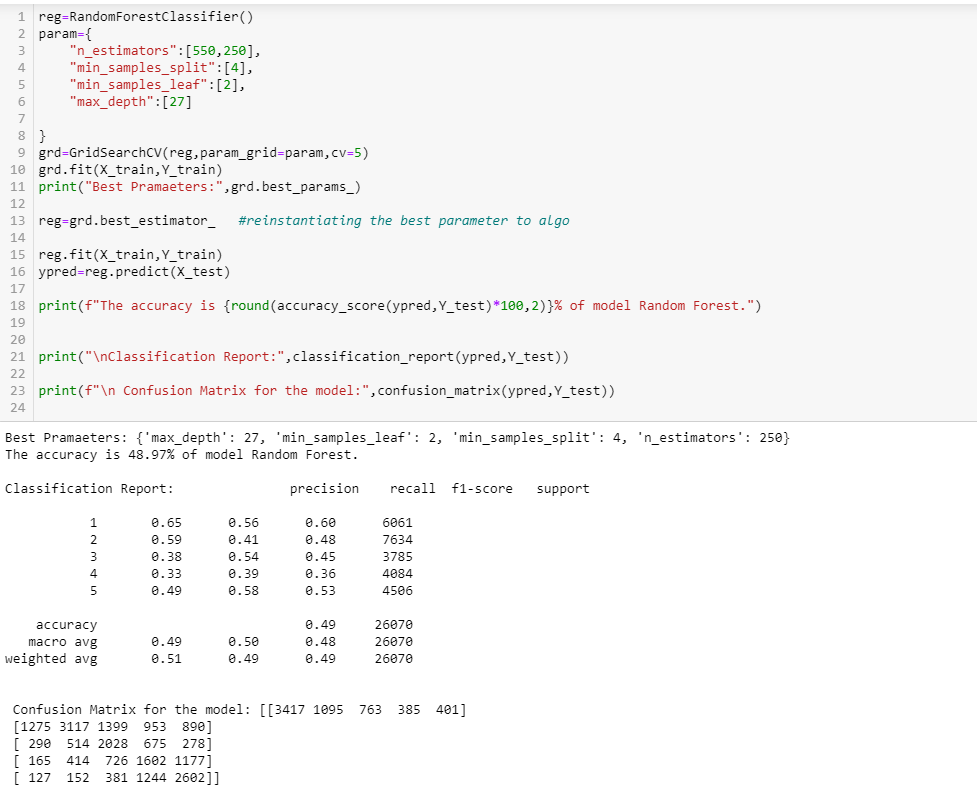
**Running the selected Models:**

****

**Accuracy of the Best Model (RFC Model)**

****

**Hyper Tuning The Random Forest Model**

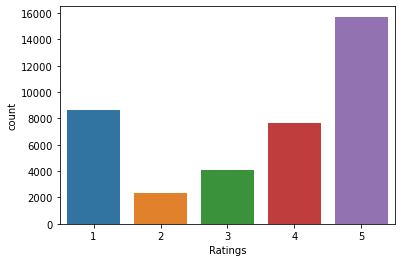
****

**OBSERVATION**

* Finally, we have saved the Random Forest Classifier Model

**Visualizations**

* Distribution of Original Ratings:

****

**Observation:**

* The ratings are imbalanced.
* 5-star ratings are highest in number.
* It seems customers are quite satisfied with the products and giving good ratings.

**Distribution of Ratings after oversampling**

****

**Interpretation of the Results:**

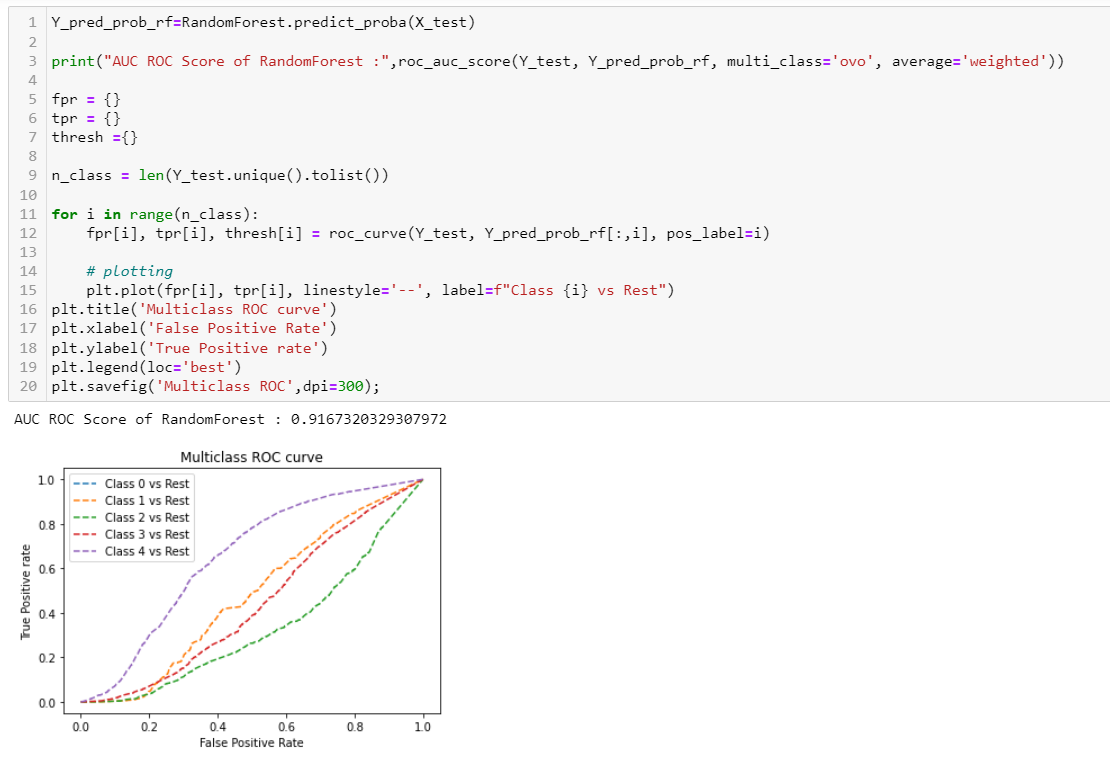
* Hence, we can go with normal Random Forest Classifier model with is also giving good accuracy.
* As we have decrease in accuracy after hyper tuning.
* So, it's better to use Random Forest with default parameters.

**CONCLUSION:**

* Finally, we have saved the Random Forest Classifier Model.

**Learning Outcomes of the Study in respect of Data Science**

* Hence we can go with normal Random Forest Classifier model which is also giving good accuracy.
* Finally, we have saved the Random Forest Regression Model.

**ROC \_AUC CURVE**

**Limitations of this work and Scope for Future Work**

* In future this machine learning model using NLP may bind with various website which can provide real time data for Rating Review prediction.
* This system is a web application where user will view various products and purchase products online and can give review about the merchandise and online shopping services.
* This system will help many E-commerce enterprises to improve or maintain their services based on the customer review as well as to improve the merchandise based on the customer review.

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***