iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57298

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Credit Score Prediction

Krishna Balaji[1], Aashima Gupta[2], Shristy Goswami[3]
Computer Science Engineering, HMRITM

Abstract: ***A credit score is the numerical representation of a person's credit worthiness, which is the likelihood that they will replay the borrowed money. Credit scores are used by lenders, such as banks and credit companies, to evaluate the risk of lending money or extending credit to an individual. Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without being explicitly programmed. Learning algorithms in many applications that we make use of daily. These algorithms are used for various purposes like data mining, image processing, predictive analytics, etc. to name a few. The main advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its work automatically. Various algorithms used in machine learning, along with their drawbacks and usages, have been discussed briefly. A model is also made using one of these algorithms which provide minimum error of provide correct predictions based on the previous data available. The model is made using random forest classifier algorithm which gave the highest accuracy ,out of all the machine learning algorithm, up to 79.97%. The result was printed at last containing the predictions made on the test data , id and customer id.***

I. INTRODUCTION

Credit score prediction involves utilizing data analytics and machine learning algorithms to forecast an individual's future creditworthiness. By analyzing historical financial behaviors, payment patterns, and other relevant factors, these models generate predictions to estimate a person's potential credit score. This innovative approach allows financial institutions to make more accurate lending decisions, tailor loan terms, and assess risk effectively. Predictive models leverage vast datasets to identify patterns that may impact creditworthiness, providing valuable insights for both lenders and borrowers. As technology · continues to advance, credit score prediction contributes to a more nuanced and precise evaluation of financial reliability. Credit scores typically range from 300 to 850, with higher scores indicating better creditworthiness. Credit score in the range of 300 to 560 is considered as poor, between 561 to 759 is considered as standard and between 760 to 850 is considered as good.

II. METHODS AND PROCEDURE

The dataset was taken from Kaggle for making the model. The model was made on google cola which is a free cloud-based platform provided by google that allows the users to write and execute python code in a collaborative environment. The first step is to import necessary libraries such as NumPy for supporting large, multi-dimensional arrays and matrices, pandas for read csv in a dataframe, seaborn for plotting graphs and matplotlib for visualization. The train dataset was read using read_csv() method of pandas. Data exploration was done using info() and head() methods of pandas. The following data-preprocessing techniques were applied[2]:

1) Filling missing values by taking mode of the column
2) Manually mapping string values to integer values using map() method for some columns
3) LabelEncoder() was used to convert the datatype of some columns into integer-type

After performing the data-preprocessing techniques, the train data was divided into x & y. The x consisted of all the columns which would use to make predictions. The y consisted of only that column on which the prediction was to be made. The x is known as feature and y is known as target. A train-test split was performed on the x & y to divide the data into training and testing the data. The training data consisted of 80% of the data in x & y . The testing data consisted of 20% of the data in the x & y. So, the original data was partitioned into training and testing data in the ratio of 80:20 [2]. The training data was used to training the model and testing data was used for evaluating the made model . This step was done so that the best machine learning algorithm would be chosen with the highest accuracy[1] . After selecting the best machine learning model, the test dataset was read using read_csv() method of pandas. All the data preprocessing techniques which were applied on train dataset was applied on test dataset as well. The selected model was then used to get predictions for the test dataset. The result which contained the predictions made along with id and customer id of the individuals were printed out.

©IJRASET: All Rights are Reserved

III. RESULT

Various machine learning algorithms were used which included base algorithms and ensemble tree algorithms as well. The accuracy obtained from the various algo is given below:



As we can see from the graph, random forest classifier had the highest accuracy. So, the random forest classifier was chosen and the model made using this algo was used on test data to get predictions.

IV. CONCLUSION

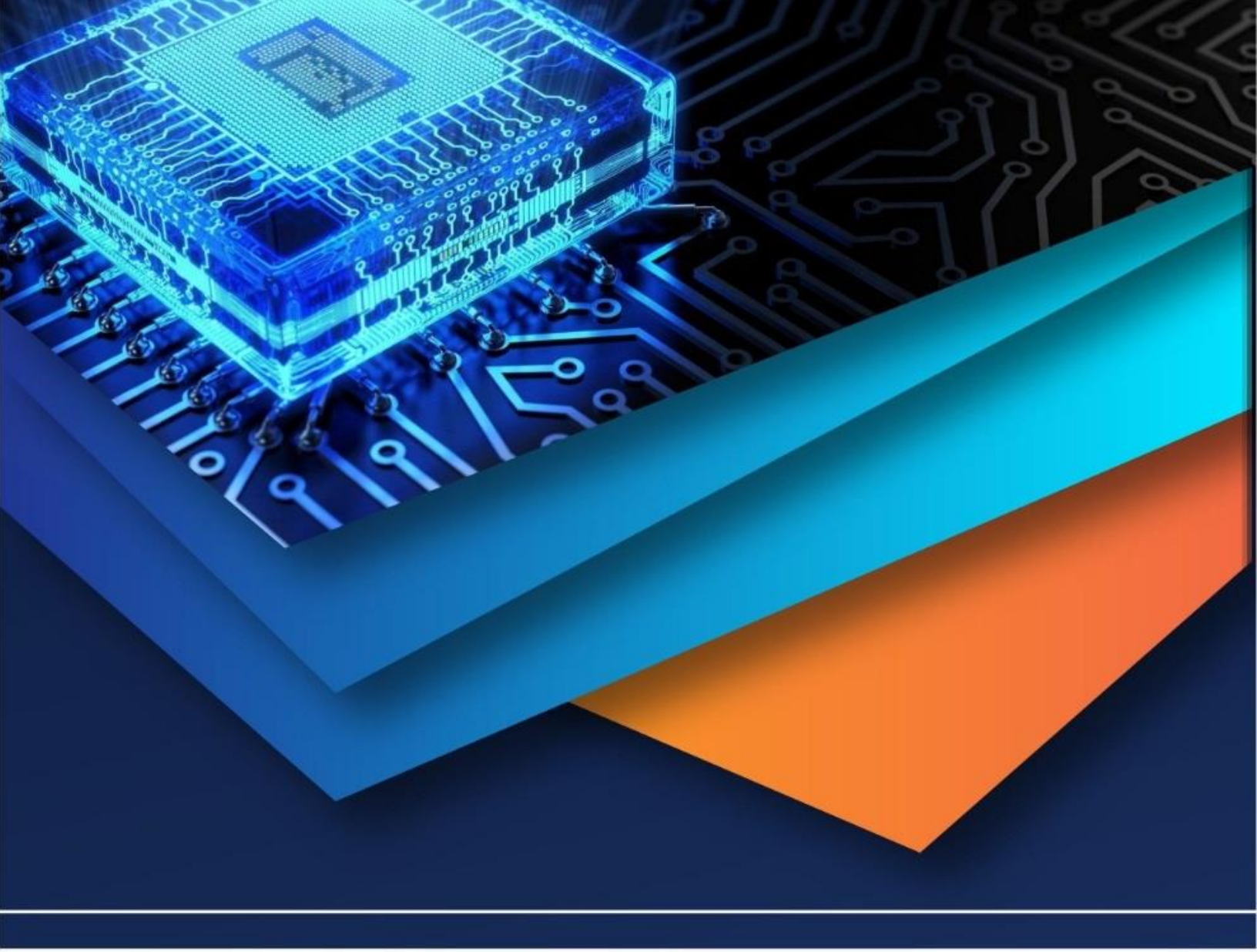
The ensemble tree algorithms were predicting more accurately than the base algorithms. Random Forest Classifier was chosen because it had the highest accuracy of 79.9%. The result consisting of the predictions made by the model on test dataset along with ids and customer ids of persons were concatenated into a dataframe and was printed out row-wise.

V. FUTURE SCOPE

1) Companies use credit scores to make decisions on whether to offer mortgage, credit card, auto loan and other credit products.
2) It can also be used for tenant screening and insurance.

REFERENCES

[1] https://libgen.li/edition.php?id=90014421
[2] An Ensemble Classifier Model to Predict Credit Scoring - Comparative Analysis | IEEE Conference Publication | IEEE Xplore



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)