Financial Oracle: Unlocking Credit Scores

Article · March 2024

DOI: 10.48001/joitml.2024.119-11

2 authors:

T. Aditya Sai Srinivas

188 PUBLICATIONS **390** CITATIONS

SEE PROFILE

Bharathi Munnuru

Jayaprakash Narayan College of Engineering

90 PUBLICATIONS **18** CITATIONS

SEE PROFILE

Research Article

Volume-1 | Issue-1| Jan-Jun-2024

JOURNAL OF
IoT and Machine Learning
Double Blind Peer Reviewed Journal
DOI: https://doi.org/10.48001/JoITML

Financial Oracle: Unlocking Credit Scores

M. Bharathi[1], T. Aditya Sai Srinivas[1*]

[1]Department of Artificial Intelligence and Machine Learning, Jayaprakash Narayan College of Engineering, Dharmapur, Telangana, India

[*]Corresponding Author's Email: taditya1033@gmail.com

ARTICLE HISTORY:

Received: 24th Jan, 2024

Revised: 12nd Feb, 2024

Accepted: 28th Feb, 2024

Published: 11th Mar, 2024

KEYWORDS:

Credit score predictor, Creditworthiness, Financial attributes, Machine Learning (ML), Random forest classifier

ABSTRACT: This paper provides individuals with a reliable tool for predicting their credit scores based on various financial attributes. Leveraging ML models, specifically a Random Forest classifier, our system analyzes features such as yearly income, salary, bank accounts, credit card usage, loan history, and payment behavior. Users can input their financial information, and the system generates a predicted credit score. This predictive tool empowers individuals to gain insights into their creditworthiness, enabling better financial planning and decision-making. By offering a user-friendly interface and accurate predictions, our project seeks to assist users in understanding the factors influencing their credit scores and ultimately improving their financial well-being. With the increasing importance of credit scores in financial transactions and lending decisions, our tool serves as a valuable resource for individuals seeking to manage their finances more effectively.

1. INTRODUCTION

In today's financial world, credit scores are super important. They decide if you can get loans, credit cards, or mortgages. Your credit score is like a grade that shows how good you are at paying back money. Knowing your score helps you make smart financial choices and plan ahead. To address the need for reliable credit score prediction, we have developed a project titled "Credit Score Predictor." Our system uses advanced machine learning, like Random Forest, to predict credit scores (Tripathi et al., 2021). It analyzes financial info like income, bank accounts, and loan history to make accurate predictions. The "Credit Score Predictor" project aims to empower individuals to gain insights into their creditworthiness and make informed financial decisions. Our tool offers an easy-to-use interface and accurate predictions, helping people effectively manage their finances and enhance their financial health. We'll look at why credit scores matter, what affects them, the value of predicting credit scores, and how our project tackles these issues. Additionally, we will discuss the methodology behind our predictive model, its features, and the potential impact on individuals' financial lives (Arram et al., 2023).

1.1. Dataset Overview

The dataset contains various features related to individuals' financial attributes and credit behavior. The goal is to predict a person's credit score using provided features. This dataset helps understand what factors influence credit scores and develop models to assess creditworthiness. The "Credit_Score" column is the target variable we're trying to predict (Hand & Henley, 1997).

1.2. Credit Score Prediction

Import Libraries: Begin by importing the necessary libraries. In this case, the code imports pandas, numpy, and plotly libraries for data manipulation, array operations, and visualization.

Set Plotly Template: Set the default template for Plotly visualizations to "plotly_white" for better visualization aesthetics.

Read Dataset: Read the dataset "credit.csv" into a pandas DataFrame named "data".

Preview Dataset: Print the first few rows of the dataset to understand its structure and contents.

```
     ID  Customer_ID  Month           Name   Age          SSN Occupation  \
0  5634         3392      1  Aaron Maashoh  23.0  821000265.0  Scientist
1  5635         3392      2  Aaron Maashoh  23.0  821000265.0  Scientist
2  5636         3392      3  Aaron Maashoh  23.0  821000265.0  Scientist
3  5637         3392      4  Aaron Maashoh  23.0  821000265.0  Scientist
4  5638         3392      5  Aaron Maashoh  23.0  821000265.0  Scientist

   Annual_Income  Monthly_Inhand_Salary  Num_Bank_Accounts  ...  Credit_Mix  \
0       19114.12            1824.843333                3.0  ...        Good
1       19114.12            1824.843333                3.0  ...        Good
2       19114.12            1824.843333                3.0  ...        Good
3       19114.12            1824.843333                3.0  ...        Good
4       19114.12            1824.843333                3.0  ...        Good

   Outstanding_Debt  Credit_Utilization_Ratio Credit_History_Age  \
0            809.98                 26.822620              265.0
1            809.98                 31.944960              266.0
2            809.98                 28.609352              267.0
3            809.98                 31.377862              268.0
4            809.98                 24.797347              269.0

   Payment_of_Min_Amount  Total_EMI_per_month  Amount_invested_monthly  \
0                     No            49.574949                 21.46538
1                     No            49.574949                 21.46538
2                     No            49.574949                 21.46538
3                     No            49.574949                 21.46538
4                     No            49.574949                 21.46538

                 Payment_Behaviour Monthly_Balance  Credit_Score
0   High_spent_Small_value_payments      312.494089          Good
1    Low_spent_Large_value_payments      284.629162          Good
2   Low_spent_Medium_value_payments      331.209863          Good
3    Low_spent_Small_value_payments      223.451310          Good
4  High_spent_Medium_value_payments      341.489231          Good
```

1.3. Check Dataset Information

```
Data columns (total 28 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   ID                        100000 non-null  int64
 1   Customer_ID               100000 non-null  int64
 2   Month                     100000 non-null  int64
 3   Name                      100000 non-null  object
 4   Age                       100000 non-null  float64
 5   SSN                       100000 non-null  float64
 6   Occupation                100000 non-null  object
 7   Annual_Income             100000 non-null  float64
 8   Monthly_Inhand_Salary     100000 non-null  float64
 9   Num_Bank_Accounts         100000 non-null  float64
 10  Num_Credit_Card           100000 non-null  float64
 11  Interest_Rate             100000 non-null  float64
 12  Num_of_Loan               100000 non-null  float64
 13  Type_of_Loan              100000 non-null  object
 14  Delay_from_due_date       100000 non-null  float64
 15  Num_of_Delayed_Payment    100000 non-null  float64
 16  Changed_Credit_Limit      100000 non-null  float64
 17  Num_Credit_Inquiries      100000 non-null  float64
 18  Credit_Mix                100000 non-null  object
 19  Outstanding_Debt          100000 non-null  float64
 20  Credit_Utilization_Ratio  100000 non-null  float64
 21  Credit_History_Age        100000 non-null  float64
 22  Payment_of_Min_Amount     100000 non-null  object
 23  Total_EMI_per_month       100000 non-null  float64
 24  Amount_invested_monthly   100000 non-null  float64
 25  Payment_Behaviour         100000 non-null  object
 26  Monthly_Balance           100000 non-null  float64
 27  Credit_Score              100000 non-null  object
dtypes: float64(18), int64(3), object(7)
```

1.4. Check for Missing Values

```
ID                          0
Customer_ID                 0
Month                       0
Name                        0
Age                         0
SSN                         0
Occupation                  0
Annual_Income               0
Monthly_Inhand_Salary       0
Num_Bank_Accounts           0
Num_Credit_Card             0
Interest_Rate               0
Num_of_Loan                 0
Type_of_Loan                0
Delay_from_due_date         0
Num_of_Delayed_Payment      0
Changed_Credit_Limit        0
Num_Credit_Inquiries        0
Credit_Mix                  0
Outstanding_Debt            0
Credit_Utilization_Ratio    0
Credit_History_Age          0
Payment_of_Min_Amount       0
Total_EMI_per_month         0
Amount_invested_monthly     0
Payment_Behaviour           0
Monthly_Balance             0
Credit_Score                0
dtype: int64
```

Explore Target Variable: Investigate the distribution of the target variable "Credit_Score" by printing the value counts (Kuppili et al., 2020).

```
Standard    53174
Poor        28998
Good        17828
```

Data Preprocessing: Map categorical values in the "Credit_Mix" column to numerical values for model training.

Prepare Features and Target: Separate the features (X) and target variable (y) from the dataset.

Split Data into Train and Test Sets: Split the dataset into training and testing sets to evaluate the model.

Initialize and Train Model: Initialize a Random Forest classifier model and train it on the training data.

```
▾ RandomForestClassifier
RandomForestClassifier()
```

Predict Credit Score: After training the model, you can use it to predict the credit score for new data. Below is an example of how you can input values for various features and predict the credit score using the trained model:

DOI: https://doi.org/10.48001/JoITML.2024.119-11

```
Credit Score Prediction :
Annual Income: 200000
Monthly Inhand Salary: 24000
Number of Bank Accounts: 1
Number of Credit cards: 2
Interest rate: 8
Number of Loans: 2
Average number of days delayed by the person: 12
Number of delayed payments: 3
Credit Mix (Bad: 0, Standard: 1, Good: 3) : 1
Outstanding Debt: 300
Credit History Age: 150
Monthly Balance: 299
Predicted Credit Score =  ['Standard']
```

Evaluate Model Performance: Optionally, Evaluate the model using the testing data and metrics like accuracy, precision, recall, and F1-score (Putra et al., 2020).

	precision	recall	f1-score	support
Good	0.78	0.77	0.77	5866
Poor	0.79	0.83	0.81	9633
Standard	0.83	0.81	0.82	17501
accuracy			0.81	33000
macro avg	0.80	0.80	0.80	33000
weighted avg	0.81	0.81	0.81	33000

2. CONCLUSION

This paper offers a valuable tool for individuals to gain insights into their creditworthiness and make informed financial decisions. By leveraging machine learning techniques, specifically a Random Forest classifier, we have developed a predictive model capable of accurately forecasting credit scores based on various financial attributes. Through data analysis and model training, we have identified key factors influencing credit scores, enabling users to understand their financial standing better. With a user-friendly interface and precise predictions, our tool empowers individuals to manage their finances effectively and improve their overall financial well-being. Moving forward, our project contributes to enhancing financial literacy and assisting individuals in navigating the complexities of credit management, ultimately fostering greater financial stability and success.

REFERENCES

Arram, A., Ayob, M., Albadr, M. A. A., Sulaiman, A., & Albashish, D. (2023). Credit card score prediction using machine learning models: A new dataset. *arXiv preprint arXiv:2310.02956*. https://doi.org/10.48550/arXiv.2310.02956.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *160*(3), 523-541. https://doi.org/10.1111/j.1467-985X.1997.00078.x.

Kuppili, V., Tripathi, D., & Reddy Edla, D. (2020). Credit score classification using spiking extreme learning machine. *Computational Intelligence*, *36*(2), 402-426. https://doi.org/10.1111/coin.12242.

Putra, S. G. P., Joshi, B., Redi, J., & Bozzon, A. (2020, June). A credit scoring model for smes based on social media data. In *International Conference on Web Engineering* (pp. 113-129). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-50578-3_9.

Tripathi, D., Edla, D. R., Bablani, A., Shukla, A. K., & Reddy, B. R. (2021). Experimental analysis of machine learning methods for credit score classification. *Progress in Artificial Intelligence*, *10*(3), 217-243. https://doi.org/10.1007/s13748-021-00238-2.

.

.

DOI: https://doi.org/10.48001/JoITML.2024.119-11