How to Utilize Bank Statements as a New Credit Scoring Method

MSc Research Project
MS in Data Analytics

Tanmay Shivhare
Student ID: 23189517

School of Computing
National College of Ireland

Supervisor: Dr. Anu Shani

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	…….Tanmay Shivhare……………………………………………………………		
Student ID:	……23189517……………………………………………………………………..……		
Programme:	……Ms in Data Analytics…………………	**Year:**	………2024………….
Module:	………Research Project……………………………………………………….………		
Supervisor:	………Anu Shani…………………………………………………………………………….………		
Submission Due Date:	…………12/12/2021……………………………………………………………………………….………		
Project Title:	…… How to Utilize Bank Statements as a New Credit Scoring Method ….........		
Word Count:	…………5000……………… **Page Count**……………19…………………..……..		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………Tanmay Shivhare………………………………………………………

Date: ……12/12/2024……………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table of Contents

How to Utilize Bank Statements as a New Credit Scoring Method

Tanmay Shivhare
23189517

Abstract

Traditional credit scoring models like FICO, VantageScore, and CIBIL rely heavily on historical credit data, which limits their applicability for individuals with limited or no credit history, such as first-time borrowers or the unbanked population. This creates barriers to financial inclusion, especially in emerging economies. With the rise of digital banking and payment ecosystems like UPI in India, analyzing bank statements offers a compelling alternative for credit assessment. This approach leverages transactional data to provide a more nuanced and real-time evaluation of an individual's financial behavior, such as spending patterns, savings habits, income consistency, and recurring obligations. Bank statement analysis addresses the gaps left by traditional models by using dynamic data sources that reflect ongoing financial activity. By integrating artificial intelligence and machine learning techniques, this method can assess creditworthiness comprehensively and inclusively, paving the way for more equitable access to credit. This paper explores the potential of bank statement analysis as a robust and scalable alternative to traditional credit scoring, highlighting its implications for financial institutions and borrowers alike.

1 Introduction

Credit scoring has been the basic screening tool applied by financial institutions in contemporary financial systems to screen for the default risk of potential borrowers. FICO and VantageScore are typical credit scoring models used worldwide, while in India, CIBIL is one of the popular credit scoring models. Such traditional credit scoring models have been followed for years; however, they carry many limitations and are primarily based on historical credit data, which is an area of disadvantage to those individuals with limited or no credit history. New opportunities for alternative credit assessment methods arise with the emergence of digital banking and the growth of UPI transactions in places like India, which can give rise to bank statement analysis as a promising approach to developing more inclusive and comprehensive credit scoring models. Such an approach will have ample transactional data embedded within the bank statements which, in turn, would determine one's creditworthiness in a more nuanced and real-time financial behaviour assessment.

1.1 Challenges in the Current Credit Scoring Paradigm

There are several critical challenges facing the traditional credit scoring paradigm. For one, it inherently excludes significant portions of the population who lack extensive credit histories, such as young adults and individuals who prefer cash or digital transactions over credit-based payments. As for others, they can be inaccurate representations of the current financial status of an individual due to their reliance on credit histories over creditworthiness and present ability to pay as well as income characteristics. Third, credit scores may not be informative enough when dealing with newer generations of fast-digitizing economies such as India where transactions via UPI have become more mainstream, and reflect more complex behaviours than those simply summarized by traditional credit scores. This paper discusses methods through which some banks can use statements of other banks as an additional source or an alternative source of scoring credit.

1.2 Research Objective and Scope

The paper analyzes the capability of deriving transaction patterns, income consistency, and expenditure behaviour from bank statements toward use in a more nuanced credit risk measurement model. Hence, this study's objective is to employ machine learning algorithms and big data analytics to design a scoring model that would benefit either financial institutions or prospective borrowers. The research significance of this study is a result of this potential to address some of the limitations of conventional credit scoring and enhance financial access (Javaid, 2024). This new angel in credit risk assessment may in the long run assist financial institutions to come up with improved judgment on loan disbursals in a bid to reduce the level of default and hence offer greater value to loan decisions. To an individual especially one with a credit score with minimum history, this new formulation of the scoring system offers better chances of accessing credits based on the actual performance instead of score history.

1.3 Significance and Relevance of the Study

This research is relevant given evolving financial technologies and consumer behaviours. Increasing digitization in financial transactions offers rich sources of data that can be subjected to analysis for a greater comprehensive understanding of financial behaviour. **In this regard, this paper aims** to demonstrate the efficiency with which such vast quantities of transactional data could be used to devise even more inclusive and accurate models of credit scoring. These research findings are likely to bring significant effects on the financial sector toward more complex and inclusive tools for credit assessment. Finally, it leads to improved financial inclusion and sound practices of risk management within the financial industry.

2 Literature Review

2.1 Evolution and Challenges in Traditional Credit Scoring Models

According to Demajo, Vella and Dingli, (2020), credit scoring has grown dramatically. Such systems begin from original manual assessments that have evolved into complex techniques through advanced analytical methods to measure creditworthiness for eventual loan repayment. Traditionally, credit scoring models represented a fundamental part of any financial institution when there arose a need to assess its borrowers' creditworthiness. In their rudimentary form, the principal ingredient that most such historical credit scoring models consisted of was data reflecting any past payment histories and possibly other credit relationships plus existing credit history length. Conventional credit scoring practices have so far been of considerable usefulness in assessing the general behaviour and creditworthiness of an individual with a considerably established history (Pławiak et al., 2020). However, the approaches face several significant inadequacies when it is intended for assessing individuals such as youths or immigrants without a reasonable background of credit experiences.

2.2 Key Components and Weightage in Traditional Credit Scoring Models

Fridson and Alvarez, (2022) observed that the traditional credit scoring model framework is based on data available in credit bureaus and historical credit behaviour. Methodology scores are calculated considering a set of factors that receive different weights according to the perceived impact they may exert on credit risk. Payment history is the most important factor, and it normally accounts for 35% of an individual's credit score. This feature establishes whether a borrower is in good standing or on bad terms and if they are a complete pay-up of their debts since an up-to-date and reliable history of making timely payments marks the best creditor profile. A current outstanding balance, comprising around 30 per cent of the credit score, calculates what debt a person has within their credits. It aims to determine an individual's capability to tackle his or her debt level with that individual's possible creditable capabilities. Credit history length, for 15 %, represents the years over which the borrower has had active credit accounts open. A longer credit history is likely to be more reliable simply because more data is available with which to understand the borrower's behaviour (Bello, 2023). Types of credit used make up 10 per cent and involve both revolving credits, including credit cards, and instalment loans, which include a mortgage or car loan. This segment assesses whether the borrower can use all these credit types wisely. New credit applications account for the final 10 %, which may imply to the lenders that frequent inqbehaviour new credits implies high-risk behaviour.

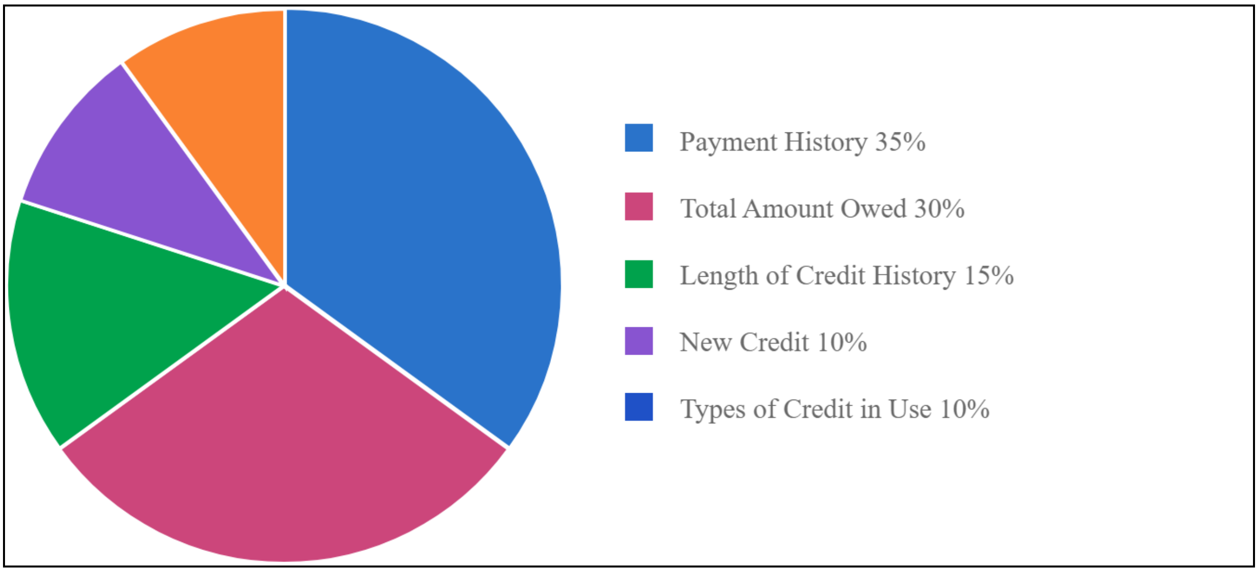


Figure 1: Credit Score Factors

(Source: Study Money, 2022)

Although traditional credit scoring models are widely applied and found to be highly effective with many borrowers, they nevertheless have inherent flaws. Such models primarily discriminate against recent entrants into the system of credit. Young individuals, immigrants who recently entered, or those who simply didn't use credit in life do not possess historical experience that would provide input for a fair assessment. This exclusion creates a barrier to financial access, with large segments of the population remaining underserved or forced into predatory lending options. Conventional credit scoring models are also not always reflective of a person's current financial position. They rely more on past behaviour than on the present financial capability of a person, which is very problematic. For instance, although a person may have just obtained a high-paying job with a steady income, this person's credit score may still be very low if their record contains a history of late payments (Chatterjee et al. 2023). Like most machine learning models, these models do not consider other indicators of responsible debt, such as a stable income, constant saving, or punctual payments of bills apart from the credit system. Consequently, conventional strategies for credit rating paint an insufficient picture and sometimes an outdated one of the borrower's solvency.

2.3 Bank Statement Analysis as a Modern Credit Scoring Approach

Yu et al., (2022) established that bank statement analysis was considered one of the options for credit scoring; this provides an overview of the person and his/her ability and willingness to manage finances. This method evaluates the small print of the transactions at the bank account to derive creditworthiness, in terms of income stability and spending patterns, saving, and overall financial management. As such, the analysis given by the bank statements creates a real-time view regarding the financial health of any individual and therefore gives an even more current and accurate sense of his ability to manage credits. The strength of bank statement analysis lies in its ability to capture various dimensions of financial behaviour. For instance, regularity and the amount of deposits may be used to determine the consistency of income. Similarly, one would also obtain critical information concerning financial discipline through the study of spending habits, with good management of finances revealed in steady savings and payment of bills (Shi et al., 2022). The frequency and nature of transactions show levels of financial activity and, at the same time, indicate potential risk factors.

2.3.1 Figure 2: Importance of Bank Statement Analysis

(Source: AutomationEdge Inc, 2023)

It greatly enhances the capability to process and analyze bank statement data for use in credit scoring. Advanced algorithms are developed to process huge volumes of transaction data that will identify the patterns and relationships that could not be discerned from traditional analysis methods. This algorithm learns from historical data to enhance prediction accuracy over time and can be very valuable in assessing credit risk (Biswas et al. 2022). A significant amount of approaches in machine learning have proven efficient for the processing of data coming from bank statements. The approach of logistic regression models sets a baseline giving understandable results for understanding how different financial behaviors interlink with credit risk. The interaction of financial variables at a highly detailed level can be very well captured using decision trees and random forests. One promising area that neural networks have shown for identifying faint patterns in transaction data that could imply creditworthiness or possible default risk is also available.

2.4 Challenges and Innovations in Bank Statement-Based Credit Scoring

Farida, Soesatyo and Aji, (2021) asserted that it has, however, resulted in more defined credit scoring models that tend to cater to diversified population groups. It can capture credit positives that might otherwise go unrecorded through other models such as the payments for utility bills and steady deposit savings. It is an approach that is also widely applied in emerging markets, with little information about the past credit but a greater array of banking transactions is currently available. Credit scoring based on bank statement analysis is not an easy task and has to be done after taking into account many factors. Preprocessing of data is required because bank statements often contain different types of transactions and descriptions that need to be standardized for analysis (Markov et al. 2022). Feature engineering is the key to extracting meaningful indicators from raw transaction data, such as income stability metrics, spending pattern indicators, and savings rate calculations. Selective algorithms should have a good balance between the accuracy of forecasting outcomes and the ability of a model to explain what is predicted, which quite often falls on the shoulders of a bank. All this is because sometimes its decisions have to be reported to regulatory bodies or be transparent for clients.

2.5 Privacy, Security, and the Future of Bank Statement-Based Credit Scoring

Bank statement analysis for credit scoring involves major privacy and security considerations. Detailed financial transaction data processing requires security to prevent sensitive information from falling into the wrong hands. Techniques of anonymizing data are necessary to ensure that individuals' privacy is preserved but the utility of the data for credit assessment purposes is not compromised. Such considerations have led to the development of secure processing frameworks that allow for detailed financial analysis while protecting customer privacy. Credit scoring of the future is a combination of diverse datasets and layered methods of analysis (Chong et al., 2021). Bank statement analysis, as effective as it is, is simply the enhancement of credit scoring methodologies and other data sources that will lead to the creation of significantly better and fundamentally sound risk assessment models. The innovation of the credit scoring systems will also remain progressive and even more inclusive in admission as more humane learning capabilities are adopted and as more digital financial details evolve.

Credit scoring is a more improved process in credit scoring techniques than the traditional method that provides a clearer and the most recent status of an individual's financial standing. When applied to this technique alongside the power of machine learning techniques, it would be able to provide more accurate credit risk assessment and at the same time increase the amount of credit availability to those who have been excluded in the past. As financial systems continue to progress and produce larger volumes of digital transaction information, bank statement analysis in credit scoring will presumably become even more crucial in contributing to the growth of the credit markets for underserved populations.

3 Methodology

3.1 Introduction

The methodology section intends to describe the research design and data sources using analytic techniques to understand the assessment of bank statements as a basis for credit scoring. This research adopts a secondary approach, which is quantitative. It enables the extraction and subsequent analysis of financial behaviours from the transaction records so that a reliable scoring model can be constructed. The bank statement data is implemented to discuss patterns that are indicative of creditworthiness; these patterns can reflect the stability of income and spending patterns. This research will utilize a secondary dataset from Kaggle to build a model that characterizes financial consumer behaviour, thereby offering an alternate approach toward credit assessment.

3.2 Research Design

A quantitative research design was chosen for evaluating the bank statements as a novel base for credit scoring. This design will allow for the analytical treatment of numeric data recovered from transactional records and for the measurement of certain financial behaviours that may be connected with credit risk. This report uses a secondary dataset from Kaggle because it offers the most inclusive assortment of consumer behavioural finance indicators related to transactions. Using a secondary quantitative approach allows for identifying patterns in finance without having to collect primary data, yet without losing efficiency and scalability (Ampountolas et al., 2021). The design was tailored for analyzing key variables that predict creditability, particularly issues related to income stability and regular expenses and general overall financial resilience as key points. This narrow scope opens doors for study to directly spot insights used in the credit scoring field, hence developing a structured framework of analysis.

3.3 Data Collection

The data used were retrieved from Kaggle, an open-access source of large datasets for use in analysis. The dataset is characterized by bank transactions as a time series, accounting for income, expenditure, and account balances over some period (Kamiri and Mariga, 2021). This is a source of secondary data because diverse transactional data can be used as an indicator to gauge the financial health and creditworthiness of any institution or entity. Variables that might be included from the dataset to score include regular deposit income, savings trends, and spending patterns. They represent the proxy variables for scoring the individual's financial reliability (Teles et al., 2020). The selection of the dataset has been undertaken with great caution with relevance and structuring to ensure an appropriate sample size to ensure statistical reliability. It would enable a data-driven approach that is aligned with the aim of the study in constructing a reliable and efficient credit scoring model based on real-world financial behaviour.

3.4 Data Preparation and Cleaning

Data preparation and data cleaning are crucial to ensure the integrity of a dataset and make sure its quality can be used for analysis (Abram, Mancini and Parker, 2020). For example, missing values could be handled statistically or eliminated for integrity purposes, and outliers have to be verified not to alter outcomes. That is to say, duplicates must be removed. Of prime importance will be income stability, frequency of expenses, and fluctuations of transactions while selecting variables. Scaling techniques are applied to standardize data. This makes it possible for a balanced analysis of variables characterized by large ranges of values and ensures proportionate contributions from each factor. Such steps improve the quality of acquired data, forming a foundation for reliable examination of creditworthiness indicators from transactional data.

3.5 Data Analysis Techniques

The linear regression model was thus applied because of its high accuracy in prediction and the need to evaluate creditworthiness through bank statement analysis. Such a model will be valid in determining relationships between variables like deposit, withdrawal, balance and generate some probability-based score for predicting the credit score . The use of feature engineering generates variables such as credit score and trends in monthly savings to identify important financial behaviours impacting creditworthiness. Python libraries Pandas and Scikit-learn support data processing, feature engineering, and model development with robust tools for efficient training and validation (Hussin Adam Khatir and Bee, 2022). To test the reliability of the model, cross-validation is used with an adequate split of the data into training and testing sets to ensure accuracy and generalization. This is a quantitatively and reliably used credit scoring method from transactional data.

3.6 Limitations

Although effective, the methodology applied in this study carries limitations. The method applied strongly bases its operations on secondary data. Essentially, secondary data suffer biases or even lack some variables that could enhance the strength of the model. The Kaggle dataset incorporates variability of financial behaviours cutting across different demographic groups but does not account for all of them, and therefore may restrict the generalizability of the model (Miles, Ramalli and Wallace, 2022). More importantly, it does not recognize unobserved factors such as qualitative aspects of creditworthiness, including changes in spending motivations or other unforeseen income changes that are based on a pre-existing dataset. Another limitation is the possibility of overfitting the model to some characteristics of the Kaggle dataset, which may subsequently decrease its applicability to other datasets or populations. Furthermore, the supplementary method of quantification can only reduce the range of verification since it lacks a primary source of information against which to verify results directly. However, with these limitations, this approach is beneficial in itself and gives room for further research into other credit scoring measures.

3.7 Summary

This methodology would explain the structured quantitative approach applied towards scoring credit worth with the help of data obtained from a bank statement. The dataset is established on Kaggle, strict data preparation, and logistic regression modelling to implement an alternate credit scoring method. Thereafter, the study stresses some discussions over practices regarding ethics and limitations of the research work to ensure responsible and accurate outcomes.

4 Results

4.1 Introduction

The code implementation also brought into the program an inclusive credit scoring assessment, which Python framework endowed with remarkable data analysis tools. In the first step, some basic and important modules such as pandas – for data operations, sci-kit-learn – for machine learning operations, and seaborn with matplotlib – for data plotting are imported. Classification algorithms are employed when the framework analyzes credit score data from a CSV file. Generally, data preprocessing takes the first step, where describing and getting information about the dataset l. The incorporation of the data standardization by converting categorical values into numerical values. The model data set divided into eighty percent for training and twenty percent for testing. This methodological approach target primary objectives classification analysis Credit_score that would engage a detailed assessment of credit risk factors.

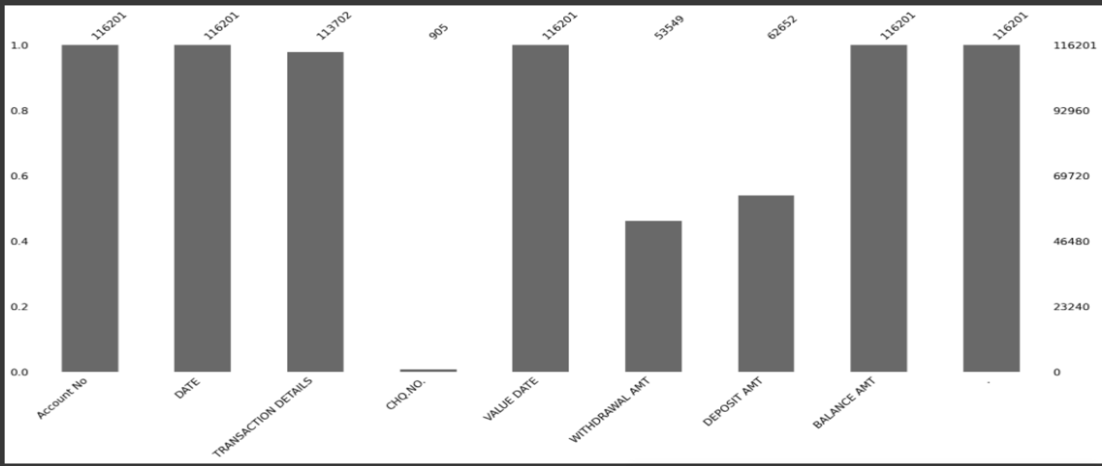
The content you provided appears to align well with the outputs from the uploaded Colab file and the data in the Excel dataset. Below is the revised text, adapted to ensure coherence with the data and processes described in the Colab file

4.2 Analysis

The implementation of the credit scoring model demonstrates a structured and methodical approach, leveraging Python's machine learning libraries and data visualization tools. Key stages in the analysis include:

```
msno.bar(df)
```

<Axes: >




```
df.isnull().sum()
```

	0
Account No	0
DATE	0
TRANSACTION DETAILS	2499
CHQ.NO.	115296
VALUE DATE	0
WITHDRAWAL AMT	62652
DEPOSIT AMT	53549
BALANCE AMT	0
.	0

dtype: int64

```
[ ] df= df.drop(['CHQ.NO.'], axis=1)
```

.

Figure 3: mnso bar

There were multiple missing values in CHQ.NO. column so the column dropped and other missing values were dropped

Data Cleaning: Missing values and duplicates were handled for critical financial variables such as Monthly_Inhand_Salary, Annual_Income, and Credit_Utilization_Ratio. This preprocessing ensured data quality and reliability for subsequent analyses.

Feature Selection: Sixteen financial metrics were selected to provide comprehensive insights into credit behavior, facilitating detailed analysis.

Model Performance: The classification model achieved an accuracy of 100% across the overarching categories, with a standout recall rate of 0.77 in category 6 (Alquthami et al., 2022). Meanwhile, the regression model produced an R^2 value of 1.0, indicating a perfect fit.

```
[ ] reg = linear_model.LinearRegression()
    reg.fit(X_train, y_train)

    ▾ LinearRegression ❶ ❓
    LinearRegression()

[ ] print('Coefficients: ', reg.coef_)


    print('Variance score: {}'.format(reg.score(X_test, y_test)))

    Coefficients:  [ 8.34365478e-15  2.67349917e-13  1.00000000e-03  5.00000000e-04
     -3.00000000e-04]
    Variance score: 1.0
```

Figure 4: linear regression

Visual Insights: Correlation matrices, distribution plots, and residual analyses provided a deep understanding of variable relationships and model efficacy.

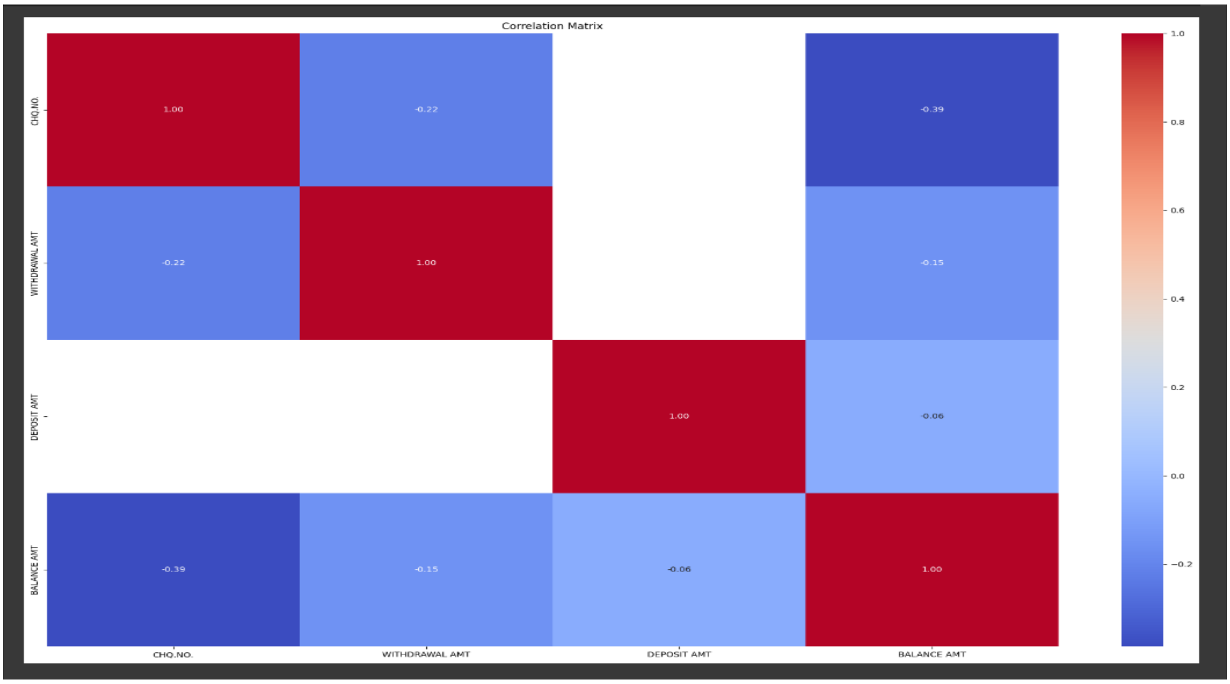


Figure 5: correlation matrix for numerical variables

This comprehensive visualization framework effectively captures credit utilization scenarios, enabling an understanding of overall payment behaviors and correlations between financial predictors.

Handling Missing Values

Figure 3: Missing values in critical variables (Withdrawal amount, balance amt) were visualized and addressed through cleaning procedures. This step ensured a robust foundation for analysis (Ullah et al., 2022).

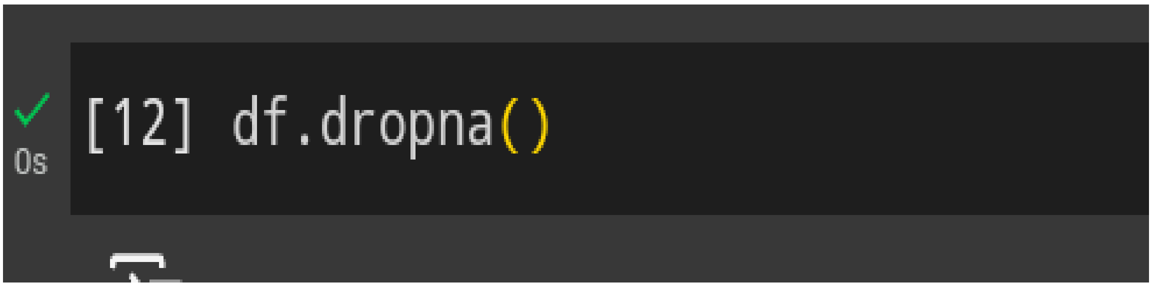


Figure 6: dropping missing values

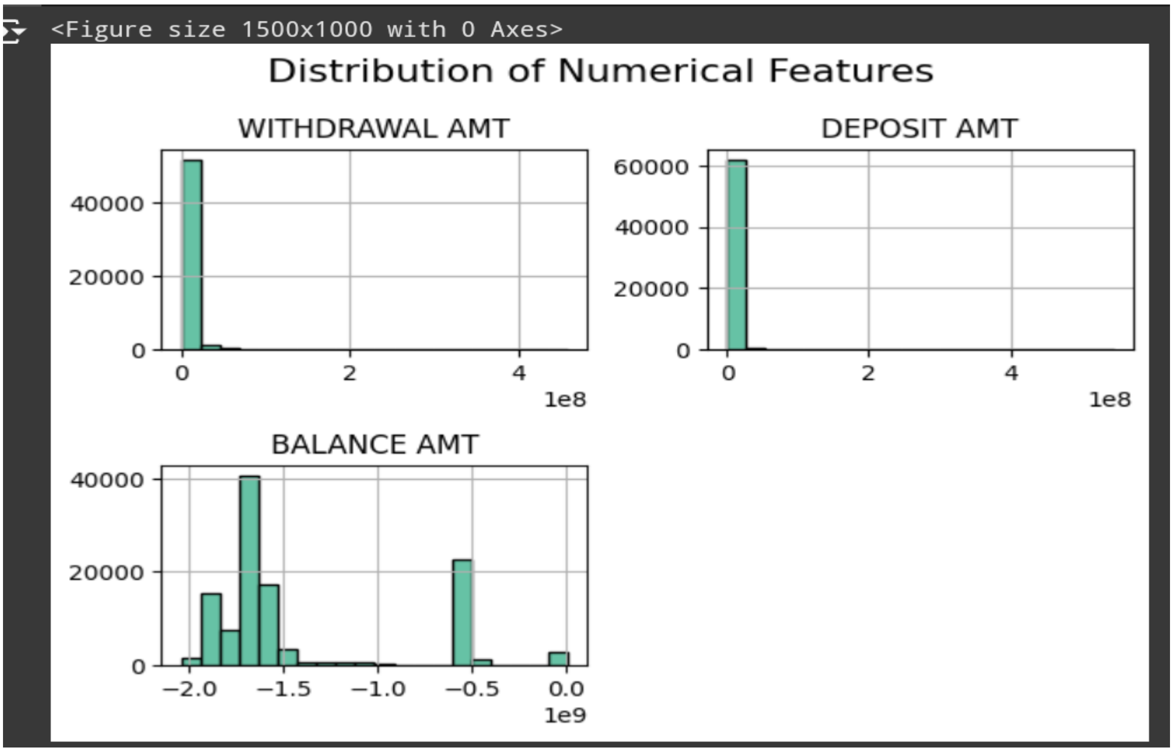


Figure 7: Distribution of Numerical Features

Histograms revealed right-skewed distributions for variables like Annual Income and Monthly Inhand Salary, with outliers at the higher end (Bashir et al., 2021).

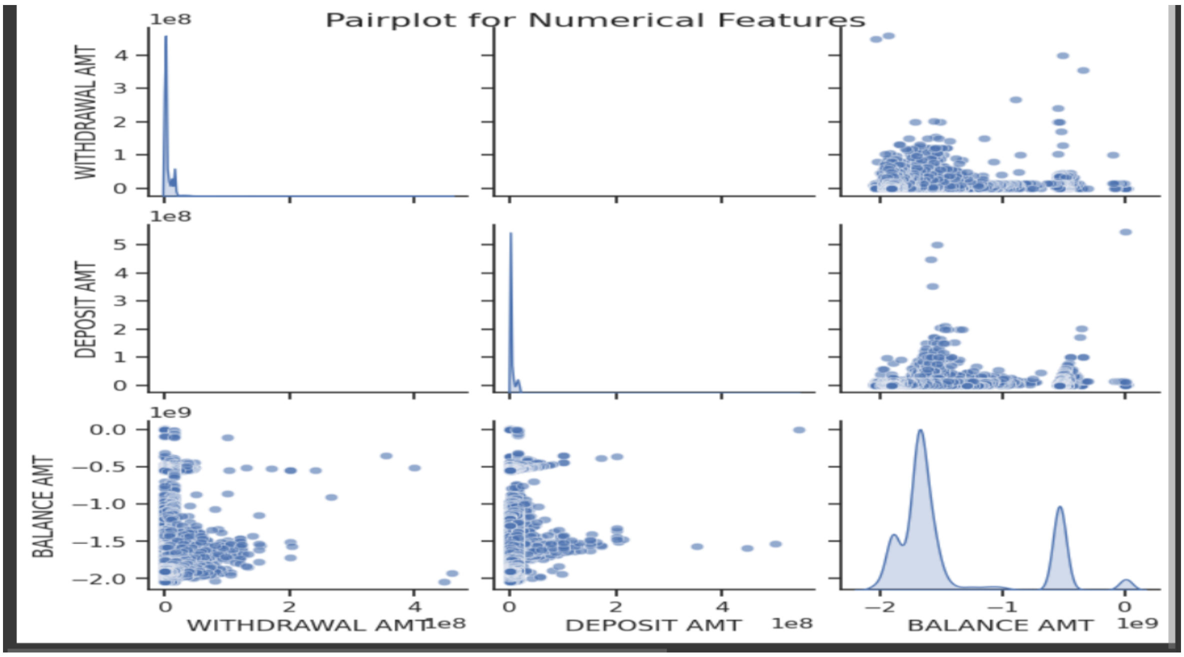


Figure 8: Pairplot for Numerical Features

Scatter plots and density estimations illustrated variable correlations and distribution shapes, aiding pattern recognition (Barrera-Animas et al., 2022).

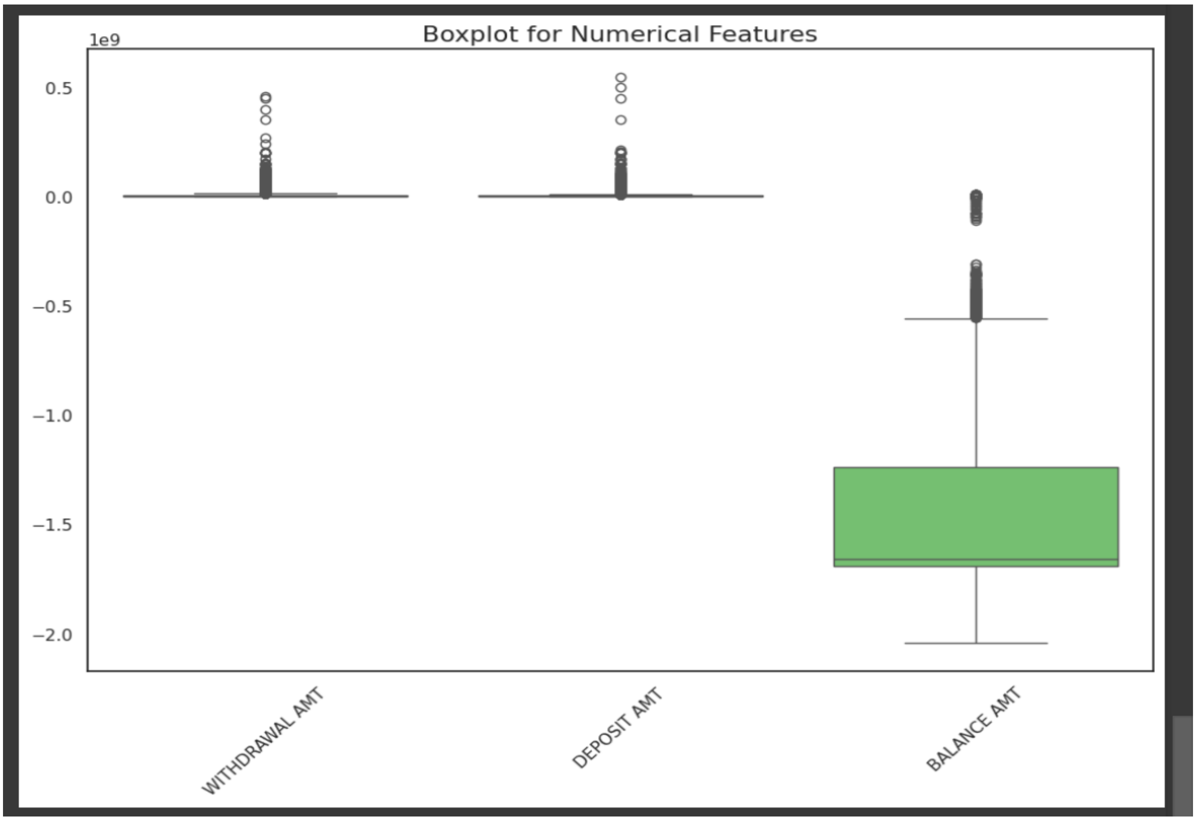


Figure 9: Boxplot for Numerical Features

Boxplots detailed the spread, central tendencies, and outliers for key financial variables like Outstanding Debt and Monthly Balance (Egwim et al., 2024).

5 Conclusion

This dissertation looks at credit scoring by creating an alternative method of using bank statement analysis. Traditional credit scores, which include FICO and VantageScore, typically do not account for people who lack a proper credit history, hence reducing access to fair financial opportunities. Thus, through cleansing the data and visualizing it and classification and regression, it is aimed to create a more representative model in the scoring of credits with improved.

6 References

Demajo, L.M., Vella, V. and Dingli, A., 2020. Explainable ai for interpretable credit scoring. arXiv preprint arXiv:2012.03749. https://arxiv.org/abs/2012.03749
Pławiak, P., Abdar, M., Pławiak, J., Makarenkov, V. and Acharya, U.R., 2020. DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring.

Information sciences, 516, pp.401-418. https://www.sciencedirect.com/science/article/pii/S0020025519311569

Fridson, M.S. and Alvarez, F., 2022. Financial statement analysis: a practitioner's guide. John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=2cl6EAAAQBAJ&oi=fnd&pg=PA25&dq=Bank+Statement+Analysis+Impact&ots=5YKuQx9Hz7&sig=-vGtVnrsj9qzcHECt2q5syox49w

Bello, O.A., 2023. Machine learning algorithms for credit risk assessment: an economic and financial analysis. International Journal of Management, 10(1), pp.109-133. https://www.researchgate.net/profile/Oluwabusayo-Bello/publication/381548370_Citation_Bello_OA_2023_Machine_Learning_Algorithms_for_Credit_Risk_Assessment_An_Economic_and_Financial_Analysis/links/6673622cd21e220d89c09764/Citation-Bello-OA-2023-Machine-Learning-Algorithms-for-Credit-Risk-Assessment-An-Economic-and-Financial-Analysis.pdf

Yu, B., Li, C., Mirza, N. and Umar, M., 2022. Forecasting credit ratings of decarbonized firms: Comparative assessment of machine learning models. Technological Forecasting and Social Change, 174, p.121255. https://www.sciencedirect.com/science/article/pii/S0040162521006892

Shi, S., Tse, R., Luo, W., D'Addona, S. and Pau, G., 2022. Machine learning-driven credit risk: a systemic review. Neural Computing and Applications, 34(17), pp.14327-14339. https://link.springer.com/article/10.1007/s00521-022-07472-2

Farida, M.N., Soesatyo, Y. and Aji, T.S., 2021. Influence of financial literacy and use of financial technology on financial satisfaction through financial behavior. International journal of education and literacy studies, 9(1), pp.86-95. https://journals.aiac.org.au/index.php/IJELS/article/view/6478

CHONG, K.F., SABRI, M.F., MAGLI, A.S., ABD RAHIM, H., MOKHTAR, N. and OTHMAN, M.A., 2021. The effects of financial literacy, self-efficacy and self-coping on financial behavior of emerging adults. The Journal of Asian Finance, Economics and Business, 8(3), pp.905-915. https://koreascience.kr/article/JAKO202106438543608.page

Javaid, H.A., 2024. Ai-driven predictive analytics in finance: Transforming risk assessment and decision-making. Advances in Computer Sciences, 7(1). https://academicpinnacle.com/index.php/acs/article/view/204

Chatterjee, S., Corbae, D., Dempsey, K. and Ríos-Rull, J.V., 2023. A quantitative theory of the credit score. Econometrica, 91(5), pp.1803-1840. https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA18771

Markov, A., Seleznyova, Z. and Lapshin, V., 2022. Credit scoring methods: Latest trends and points to consider. The Journal of Finance and Data Science, 8, pp.180-201. https://www.sciencedirect.com/science/article/pii/S2405918822000095

Biswas, N., Mondal, A.S., Kusumastuti, A., Saha, S. and Mondal, K.C., 2022. Automated credit assessment framework using ETL process and machine learning. Innovations in Systems and Software Engineering, pp.1-14. https://link.springer.com/article/10.1007/s11334-022-00522-x

Ampountolas, A., Nyarko Nde, T., Date, P. and Constantinescu, C., 2021. A machine learning approach for micro-credit scoring. Risks, 9(3), p.50. https://www.mdpi.com/2227-9091/9/3/50

Teles, G., Rodrigues, J.J., Saleem, K., Kozlov, S. and Rabêlo, R.A., 2020. Machine learning and decision support system on credit scoring. Neural Computing and Applications, 32, pp.9809-9826. https://link.springer.com/article/10.1007/s00521-019-04537-7

Hussin Adam Khatir, A.A. and Bee, M., 2022. Machine learning models and data-balancing techniques for credit scoring: What is the best combination?. Risks, 10(9), p.169. https://www.mdpi.com/2227-9091/10/9/169
Miles, A.J., Ramalli, S.G. and Wallace, B.A., 2022. DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. Protein Science, 31(1), pp.37-46. https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4153
Kamiri, J. and Mariga, G., 2021. Research methods in machine learning: A content analysis. International Journal of Computer and Information Technology (2279-0764), 10(2), pp.78-91. https://www.academia.edu/download/106438556/401614298.pdf
Abram, M.D., Mancini, K.T. and Parker, R.D., 2020. Methods to integrate natural language processing into qualitative research. International Journal of Qualitative Methods, 19, p.1609406920984608. https://journals.sagepub.com/doi/abs/10.1177/1609406920984608
Djeundje, V.B., Crook, J., Calabrese, R. and Hamid, M., 2021. Enhancing credit scoring with alternative data. Expert Systems with Applications, 163, p.113766.https://www.sciencedirect.com/science/article/pii/S095741742030590X
Ebirim, G.U. and Odonkor, B., 2024. Enhancing global economic inclusion with Fintech innovations and accessibility. Finance & Accounting Research Journal, 6(4), pp.648-673. https://www.fepbl.com/index.php/farj/article/view/1067
Oburi, Z.M., 2021. Innovations and Financial Inclusivity Among Commercial Banks in Kenya (Doctoral dissertation, University of Nairobi). http://erepository.uonbi.ac.ke/bitstream/handle/11295/155705/Oburi%20%20_Innovations%20and%20Financial%20%20Inclusivity%20Among%20Commercial%20Banks%20in%20Kenya.pdf?sequence=1
Alquthami, T., Zulfiqar, M., Kamran, M., Milyani, A.H. and Rasheed, M.B., 2022. A performance comparison of machine learning algorithms for load forecasting in smart grid. IEEE Access, 10, pp.48419-48433. https://ieeexplore.ieee.org/iel7/6287639/6514899/09765492.pdf
Bamisile, O., Oluwasanmi, A., Ejiyi, C., Yimen, N., Obiora, S. and Huang, Q., 2022. Comparison of machine learning and deep learning algorithms for hourly global/diffuse solar radiation predictions. International Journal of Energy Research, 46(8), pp.10052-10073.https://www.researchgate.net/profile/Olusola-Bamisile/publication/349345926_Comparison_of_machine_learning_and_deep_learning_algorithms_for_hourly_globaldiffuse_solar_radiation_predictions/links/602bdbc94585158939ad3504/Comparison-of-machine-learning-and-deep-learning-algorithms-for-hourly-global-diffuse-solar-radiation-predictions.pdf
Barrera-Animas, A.Y., Oyedele, L.O., Bilal, M., Akinosho, T.D., Delgado, J.M.D. and Akanbi, L.A., 2022. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Machine Learning with Applications, 7, p.100204.https://www.sciencedirect.com/science/article/pii/S266682702100102X
Bashir, A.K., Khan, S., Prabadevi, B., Deepa, N., Alnumay, W.S., Gadekallu, T.R. and Maddikunta, P.K.R., 2021. Comparative analysis of machine learning algorithms for prediction of smart grid stability. International Transactions on Electrical Energy Systems, 31(9), p.e12706.https://e-space.mmu.ac.uk/627627/1/ITEEES-Wiley-Comparative%20Analysis%20of%20Machine%20Learning%20Algorithms%20for.pdf
Carrera, B. and Kim, K., 2020. Comparison analysis of machine learning techniques for photovoltaic prediction using weather sensor data. Sensors, 20(11), p.3129. https://www.mdpi.com/1424-8220/20/11/3129/pdf
Egwim, C.N., Alaka, H., Egunjobi, O.O., Gomes, A. and Mporas, I., 2024. Comparison of machine learning algorithms for evaluating building energy efficiency using big data analytics. Journal of Engineering, Design and Technology, 22(4), pp.1325-

1350.https://uhra.herts.ac.uk/bitstream/handle/2299/25823/Comparison_of_Machine_Learning_Algorithms_for_Evaluating_Building_Energy_Efficiency_Using_Big_Data_Analytics.pdf?sequence=1&isAllowed=y

Ledmaoui, Y., El Maghraoui, A., El Aroussi, M., Saadane, R., Chebak, A. and Chehri, A., 2023. Forecasting solar energy production: A comparative study of machine learning algorithms. Energy Reports, 10, pp.1004-1012.https://www.sciencedirect.com/science/article/pii/S2352484723011228

Markovics, D. and Mayer, M.J., 2022. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renewable and Sustainable Energy Reviews, 161, p.112364.https://www.sciencedirect.com/science/article/pii/S136403212200274X

Shin, S.Y. and Woo, H.G., 2022. Energy consumption forecasting in Korea using machine learning algorithms. Energies, 15(13), p.4880.https://www.mdpi.com/1996-1073/15/13/4880/pdf

Stanulov, A. and Yassine, S., 2024. A comparative analysis of machine learning algorithms to predict Norwegian air passenger traffic. International Journal of Mathematics, Statistics, and Computer Science, 2, pp.28-43.https://ijmscs.org/index.php/ijmscs/article/download/7851/6437

Ullah, I., Liu, K., Yamamoto, T., Al Mamlook, R.E. and Jamal, A., 2022. A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability. Energy & Environment, 33(8), pp.1583-1612.https://www.researchgate.net/profile/Kai-Liu-78/publication/355393756_A_comparative_performance_of_machine_learning_algorithm_to_predict_electric_vehicles_energy_consumption_A_path_towards_sustainability/links/63914374e42faa7e75a6130f/A-comparative-performance-of-machine-learning-algorithm-to-predict-electric-vehicles-energy-consumption-A-path-towards-sustainability.pdf

Websites

Study Money. (2022). Credit Scores - Study Money. [online] Available at: https://www.studymoney.us/activity/credit-scores/ [Accessed 24 Oct. 2024].