University for the Creative Arts

BERLIN SCHOOL OF BUSINESS & INNOVATION

Essay / Assignment Title: Decoding the financial futures:

Crafting a simple credit scoring model to illuminate the creditworthiness of individuals or businesses.

Program title: MSc. Data Analytics

Name: RIZWIN PALAKKAL ABDUL SALAM

Year: 2022 - 2024

Abstract

Purpose

The study aims to examine the concept and role of the credit scoring model to develop a simple credit scoring process for predicting the creditworthiness of individuals or a business.

Scope

Articulation of the benefits and roles of the credit scoring model in evolving the structure of individuals and businesses has improved the further scope of determining how easy the techniques are to manage sustainability in financial risk management. Apart from that, the application of Pecking order theory has also emphasized the area of how debt financing, and equity financing structures improvement by advanced technological tools would be beneficial for financial institutions to minimize their risk from business processes.

Method

The secondary data analysis method is used to collect the data from secondary resource. The collected data contains the credit score details of multiple customers. The mix method is used to evaluate the collected data. This defines the implementation of thematic analysis and the prediction model construction approach. The prediction model evaluation method defines the implementation of Python coding which is used to implement different machine learning models. Those models are used to predict the credit score for multiple customers.

Recommendations

The main recommendation for the credit score prediction research highlights the involvement of more accurate model implementation. This will help to increase the accuracy level of multiple models and increase the effectiveness of the analysis process. AI, automation, deep learning, and many advance analysis methods can be implemented to increase the level of accuracy as well as effectiveness of the system or model. This process will provide more effectives and accurate research outcome.

CONTENTS

Statement of compliance with academic ethics and the avoidance of plagiarism

I honestly declare that this study is entirely my own work and none of its parts has been copied from printed or electronic sources, translated from foreign sources, and reproduced from essays of other researchers or students. Wherever I have been based on ideas or other people's texts I clearly declare it through the good use of references following academic ethics.

(In the case that is proved that part of the essay does not constitute an original work, but a copy of an already published essay or from another source, the student will be expelled permanently from the postgraduate program).

Name and Surname (Capital letters):

RIZWIN PALAKKAL ABDUL SALAM

....

Date: ...23... /...01.../…2024....

INTRODUCTION

Background of Study

(Dastile *et al.,* 2020: 21) stated that Credit scoring model is the mathematical model that helps to evaluate the probability and support the customers in terms of any kind of credit event. The study focuses on the importance of this model to identify the common factors of loan characteristics and demonstrate the importance of effective consumer behaviour for financial or economic cycles. According to the research of (Shen*et al.,* 2020: 15), the simple credit scoring model can help to assess the criteria of loans from different financial institutions to reduce the off rate of fraud. In terms of evaluating the importance of this credit scoring model an individual can focus on the financial products like auto loan credit cards and the mortgage related issues. In terms of business development, the model can help to determine the investment trustworthy process identify the length of credit history and payment process for the business (Zhang *et al.,* 2020: 25).

Research aims and Objectives.

The aim of the study is to the examine the concept and role of credit scoring model in order to develop simple credit scoring process for predicting the creditworthiness of individuals or in a business.

By considering the above-mentioned aim here are the objectives-

- To identify the concept of credit scoring model
- To evaluate the benefits of credit scoring model for predicting credit worthiness of individuals or businesses
- To evaluate the possible challenges of implementing credit scoring model
- To identify the role of credit scoring model in improving the financial landscape of an individual or in a business

Research Question

The research questions of the study related to credit scoring model are-

1. What is the concept of credit scoring model for developing creditworthiness?
2. What are the possible benefits of credit scoring model to predict the creditworthiness of individuals and a business?
3. What kinds of challenges have been seen for implementing credit scoring model?
4. What are the possible roles of credit scoring model to improve the financial landscape of an individual or in a business?

Research Rationale

The major issue of the study is to identify the importance of credit scoring model and the role of the prediction process of this credit scoring model in terms of supporting a business and individual. *(Anarfo and Abor et al., 2020: 19)* articulated that to improve the financial stability and develop the creditworthiness in business the credit scoring model can include big data related information and tender the process of financial issues by using green credit process. "On the other hand, *(Niu et al., 2019: 55)* demonstrated that application of an effective credit scoring model helps in the process of lending loans and payment through the online applications that could help to evaluate the profit-sharing process of the financial institutes."

FICO Score	Rating	What the Score Means
< 580	Poor	• Well below average • Demonstrates to lenders that you're a risky borrower
580 – 669	Fair	• Below average • Many lenders will approve loans
670 – 739	Good	• Near or slightly above average • Most lenders consider this a good score
740 – 799	Very Good	• Above average • Demonstrates to lenders you're a very dependable borrower
800+	Exceptional	• Well above average • Demonstrates to lenders you're an exceptional borrower

Table 1: Credit Score Rates from Low to High

(Source: Berg *et al.,* 2020: 62)

(Berg et al., 2020: 66) asserted that good credit scores can help individuals to get loans related to their home and they can easily borrow money from financial institutes by the development of high credits scores. “The study sheds light on the importance of credit scoring model and how the model has influenced the money transactional process in a business.” According to the rate chart of the credit scoring model it is identified that after taking a loan in case the credit score is below 579 then it shows very poor rating for the user. In terms of justifying the high score or the excellent grade the score needs to be 760 to 849 (*Chatterjee et al., 2023: 33*).

Outline of the Methodology

(*Snyder et al.,* 2019: 28) articulated research methodology is the purpose of identifying effective data sources for a research project. There are two types of data sources primary and secondary the researcher has adopted mixed method by considering primary quantitative and secondary qualitative data collection method. In terms of philosophy the researcher has taken Pragmatism research philosophy to develop testable and untestable analysis of credit scoring model and its implication. The design of the research project which has been selected is experimental research design that could help to evaluate the existing data set also identify the real time sources from involved participants. In the context of the research approach abductive has been selected as primary and secondary data collection methods. For primary quantitative data collection methods survey analysis has been selected for which purposive sampling method has been selected. 60 respondents have been articulated in this study to evolve the importance of credit scoring model to prevent any kind of financial risk. On the other side, for secondary qualitative data collection method thematic analysis has been selected that has helped to identify different published journals articles and books related to credit scoring model and its importance as well as creditworthiness for an individual or in a business.

Structure of the Research

The structure of the research demonstrates the stages of the research chapters in research. In this research project there are four major chapters other than that introduction and conclusion is also included in this study. Considering the handbook structure, it is identified that chapter one refers to Literature Review I which helps to demonstrate the theoretical knowledge as well as contextualise the previous works related to credits coding model by evaluating the credit worthiness. Chapter two refers to Literature Review II that helps to determine the contradictions and themes of the study by evaluating reliability purpose and demonstrating an in-depth investigation according to the research aim and objectives.

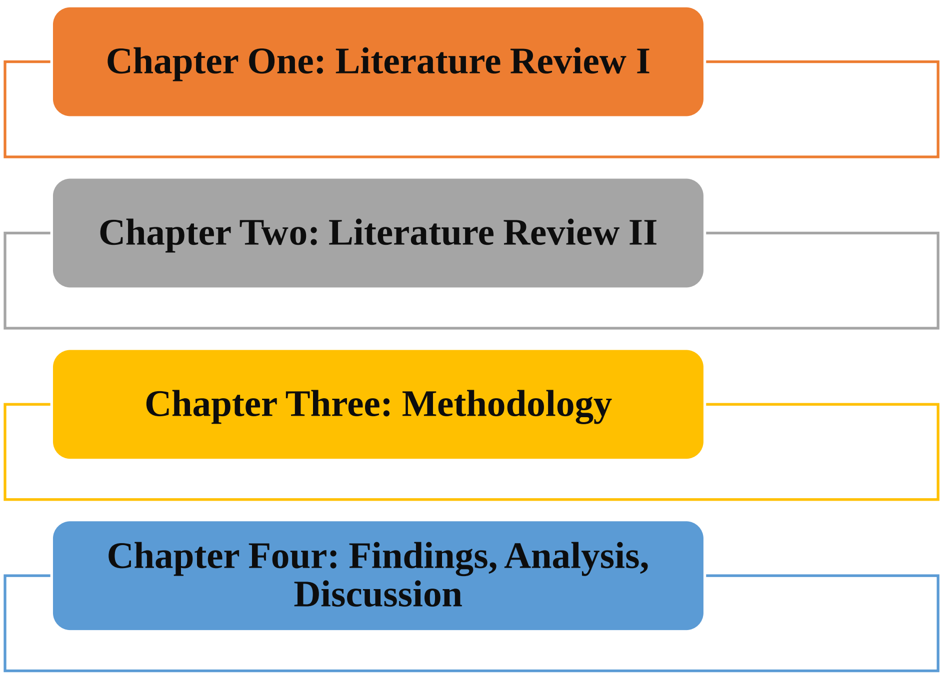

Table 2: Structure of the Research

(Source: Self-Developed)

Chapter III is Research Methodology which examines the importance of appropriate selection regarding data collection approach and design of the study. Chapter 4 refers to findings and discussion by including appropriate data collection method from primary and secondary sources that could be far interview or thematic analysis. “Finally, conclusion and recommendation helps to provide effective analysis of the identified objectives by considering the literature review and also provide smart recommendations for future development of the study. The introduction chapter can identify the main reasons that can be used to deliver the objectives to evaluate the importance of the credit scoring model for an individual and in a business. After the analysis of this introduction chapter, a literature review chapter has been developed in this part and this part can be used to manage the find out the findings related to the topic researched by the previous researchers. Going through the previous researchers' papers can help to develop the base of the research. Studies related to the topic and previous papers can help to find out the theories that can be useful in conducting the research with better understanding.

CHAPTER ONE: Literature review I

The chapter is going to demonstrate the main concept of the credit scoring model while analysing its effectiveness in predicting the creditworthiness of businesses or individuals. Elaboration of appropriate theoretical frameworks and highlighting articulation of scholarly authors according to the reliability of the credit scoring model to decide credit extent or denial. Moreover, prediction-related challenges of implementation issues faced by businesses or individuals while using the credit scoring model are also going to be illustrated in the chapter depending on the contradictory analyses of authors. Financial sustainability and institutional factors of companies to emphasize their business practices effectively, detailed elaboration of different literatures as the secondary sources of articulation are also going to be critically evaluated in the section.

1.1 The concept of credit scoring model

The credit scoring model is contemplated as the statistical analysis model that helps to estimate the probability of further defaults in lending. Alongside, different credit scoring models are used by lenders or financial institutes to analyse the creditworthiness of businesses or individuals at the segment of their transaction at a particular point in time. (Kozo Doi *et al.*, 2022: 1084) have opined that the credit scoring model helps to increase fairness in financial transactions where proper implementation of devices, and profit implication of companies are also started increasing. Identifying suitable options for credit scoring, 'profit-fairness trade-off in lending decisions' are also developed with the fair structure of the model by financial institutions. In the article, considering the aspect of business, the application of 'ML-based scoring models' and $4,161 billion of retail credit increase in 2020 has been highlighted. Based on that economic sustainability of the industry and data protection process during credit has also been found as secure and fair as per the necessity. On the contrary, (Wang *et al.*, 2020: 925) have stated that the main concept of the credit scoring model is highlighting peer-to-peer lending to recommend the prediction of businesses about their risks to taking loans or financial decisions. The structure of misclassification cost matrix for credit rating and management is used by most banking companies depending on the market risk of the factors before providing loans can be measured. Additionally, protecting the transaction of personal loans or online lending is also measured depending on the sustainable structure of the credit scoring model.

Prediction of making decisions on transactions by individuals and lending by owners or financial institutes about its risks, application of credit scoring model is viable. According to (Niu *et al.*, 2019: 398), to develop a peer-to-peer lending process, an effective credit scoring structure evolving is easier for companies. In evaluating potential risks when considering loans, the structure of the credit scoring model is used by which predicting the future losses in financial institutions can be determined. Apart from that, companies are currently using LightGBM, AdaBoost, and Random Forest to predict the performance of social networking information management to minimize the direct risks of credit scoring in business structures. On the other hand, (Giudici *et al.*, 2019: 3) have argued that a network-based scoring model helps to analyse the current risk factors of companies to manage their financial transaction procedures as well as emphasizing their capacity to estimate social data about transaction decisions. Henceforth, the prediction of further risks of credit can be minimized, and 'peer-to-peer lending platforms' of businesses, as well as individuals, are also started to improve with sustainability.

1.2 The benefits of the credit scoring model for predicting the worthiness of individuals or businesses.

As per the business context, the application of the credit scoring model in operations helps to control the structure of cost functions and risk allocation of financial institutes that improves their customers' trust management goals. Apart from that, in the individual segment, the credit scoring model is beneficial in allowing them to access personal loans as per their necessity from the market. (Ampountolas *et al.*, 2021:52) have articulated that machine learning tools are used to estimate the automated process of transactions by which micro-credit scoring functions of business are improving. A sustainable structure to fruitfully manage micro-credit scoring structures of companies by predicting the areas of their risks and minimizing their operating threats is also effective in ensuring the creditworthiness of transaction decisions. Moreover, reflecting level of credit risk and financial obligations for individuals' transaction decisions are also improved depending on the machine learning structure of credit scoring currently. On the other hand, (*Goel and Rastogi et al.*, 2023:206) have contradicted that identifying the psychological and behavioural traits of borrowers is effective in minimizing the difficulties during credit scoring for individuals. The character and reliability of the borrower determining and predicting the further risk of fraud can be easily estimated depending on the psychological and behavioural traits of each of the borrowers in the current market.

In the words of (*Mhlanga et al., 2021:2*), the credit scoring model helps to analyse the financial inclusion of companies in emerging economic structures depending on which sustainability on lending risk minimization can also be evolved. Developing the process of automated marketing functions, Artificial Intelligence, and machine learning tools are also effectively used to ensure the structure of credit risk management. Accordingly, it benefited businesses to predict their potential risks in their operating market and identify further credit risks during the transaction, the tools of machine learning and AI are also used. Approximately 95% of companies are currently using machine learning tools for predicting financial-related risks and managing their data sources as per the changing transformational practices effectively. On the other hand, (*Shen et al., 2020: 407*) have mentioned that optimizing multi-approaching structures of companies to analyse their different financial risk areas and determining their business structures developing objectives are beneficial to predict business segments. In business process management and transaction-related decision-making, the 'cost-sensitive learning algorithm' introduces the structure of cost-sensitiveness as well as betterer learning procedures to minimize further risk issues of companies in their current competitive environment. Henceforth, the benefits of measuring the risks in taking loans or conducting a payment after estimating the losses are reliable to evolving the credit scoring structure of businesses and individuals by emphasizing risk management functions.

1.3 The role of the credit scoring model in improving the financial landscape of an individual or in a business.

The credit scoring model helps to improve the financial sustainability goals of companies to meet their existing market needs as well as emphasize the payment function management capability of individuals as per their sustainability betterment. In the transformed environmental functions, creditworthiness is increasing based on advanced technological functions as most customers or employers prefer to conduct payment with an online process. In this context, (*Hasan et al., 2020: 3)* have stated that the current landscape of business is changing with the effective application of big-data analytic tools in the current market by which increasing the reliability of customers on business processes has also evolved. Big data in finance helps to predict financial losses and generates the process of automated estimation about the profit and losses of companies after conducting a tendering process. Hence, the creditworthiness of companies to manage their customers and improving the current landscape of financial issues management, internet marketing procedures have also started improving as per the sustainability prospects of the companies in their current operating environment. On the contrary, (*Nabeeh et al., 2021: 2*) have argued that the structure of green credit rating emphasizes the structure of sustainable transaction management and individual payment practices depending on their financial sustainability goals. Considering economic activities and social processes or the functional progress index structure of the business with proper implementation of the green credit scoring model in the business transactional process is viable. Apart from, improving the environmental structure of businesses to ensure their operating sustainability in their current competitive environment, the green credit scoring process helps to evolve their trading practices. Alongside, data transferring system security management goals by which further failure of any payment during the transaction, it would be able to predict initially.

(Óskarsdóttir et al., 2019:3) have amplified that managing effective financial inclusion structure of companies to secure their social networking data and mobile data before transferring, credit scoring model playing an effective role. Beyond the structure of lending and taking loans with security, data transferring process management, and automated financial transaction management to secure business and individuals of companies are also considered as the most effective processes. Social networking data management, banking data combination, and maintaining over two million of banking customers' information with security are contemplated as effective processes developed with the structure of the traditional credit scoring model of companies. On the other hand, (*Bazar bash et al.*, 2019:2) has contradicted that machine learning tools in financial institutes and business segments play a major role in minimizing their credit risk issues. Accordingly, in online or automation transactions, machine learning tools help to determine the accurate source of data estimation as well as fintech financial inclusion structure management by predicting further data losses of that business. Mostly with the structure of businesses, fintech economic structure management, and securing the process of data transaction, the application of credit scoring model is reliable.

1.4 The possible challenges of implementing the credit scoring model.

In the business or individual financial security improving, some of the challenges are affecting proper operating aspects of credit scoring models in financial institutions. The credit scoring model is used as an innovative method for predicting financial losses and emphasizing data history management. However, some of the barriers to managing credit scoring practices as per the changing prospects of the companies or individuals in the recent environment, issues of fraud and instability have started increasing. (*Suryono et al., 2019:205*) have opined that a lack of accurate analysis of the financial behaviour of individuals and understanding of their wealth make the structure of overall transactions affected by peer-to-peer lending structure. Problems in using any alternative applications to manage the conducted payment structure as well as implementation of technological procedures are also affected, as companies are lacking to determine their customers' financial capability appropriately. On the contrary, (*Alazab et al., 2021:1*) have contradicted that cybersecurity issues are one of the major challenges in developing the structure of the credit scoring model while conducting any traditional practices by businesses or individuals. In industry functions, customer engagement, and financial data transferring, frequent cyber-attacks have increased based on which real-time estimation of information factors is quite difficult for companies. Furthermore, preserving privacy by predicting the accurate area of cyber-attacks as well as random development of central learning concepts of companies has also started affecting the structure of the credit scoring process of models effectively. Whether minimizing the structure of data decentralization, Federated learning tools have been used by companies as part of AI to evolve their structure of data security.

(Paleyes et al., 2022:2) have illustrated that the deployment of machine learning tools in credit scoring and the excessively evolved business structure of companies has increased challenges of potential data security management. Poor interpretability of the financial data transferring modes and lack of development of the structure of perpetual history management functions have also started affecting the sustainability of credit scoring functions of companies. Apart from that, developing individuals' payment transaction areas, lack of managing histories, or failure during payment are also increasing the risk of online transactions at any point in time regularly.

1.5 Alternative strategic approaches to address barriers and improve the structure of implementing the credit scoring model effectively.

In terms of minimizing potential challenges that affect the sustainability of the credit scoring model in the current business structures, the implementation of appropriate strategic functions and effective business goals are essential. Improving green credit scoring structure and evolving the sustainability of traditional credit scoring procedures of companies or individuals, the application of appropriate strategic function is reliable. As per the view of (*Siskin et al., 2021:1*), in financial institutions, FICO Scores are used to improve the prediction of default risks that could affect consumer transaction-related areas in the future. Implementation of FICO Scores in financial practices helps to evolve statistical analyses on the sustainability relevance of automated payment transaction tools as well as improving fair lending practices effectively. The countries like UK, the US, and Canada are mostly using FICO Scores to develop the quality of credit scoring in their business and emphasise the security of loan borrowing as per the structural process betterment. On the other hand, (*Munkhdalai et al., 2019:251*) have demonstrated that the application of a hybrid credit scoring model would help to regression of logistics data management structure of companies and design a neutral networking structure of data collection. Better performance managing and securing the financial transaction practices of individuals or businesses in the online segment of data transferring, the application of a hybrid credit scoring model is viable. Determining the bad borrowing zones to secure the further failure issues of individuals while conducting any copayment procedures online, the structure of the hybrid credit scoring model is reliable.

(Yang et al., 2023:253) has elicited that implementation of Vantage Score as the advanced credit scoring model is rational to evolve credit score-based digital transitioned practices of companies effectively. Identifying the issues of companies borrowing loans or mortgage housing loans of individuals has also started developing as secured based on the structure of Vantage Score in the sustainable credit scoring model. Developing the structural process application in the current operating environment to improve the sustainability of digital footprint by maintaining the security system of individuals in their mobile applications. As 3.95 billion users of smartphones mostly rely on online lending or money transferring to sustain their further security goals as per their operating practices effectively. Thus, the application of Vantage Score as the advanced credit scoring model is reliable for emphasizing individual engagement as well as business companies' sustainable payment process growth effectively.

1.6 Theoretical framework

Pecking order theory

The main ideology of the theory is highlighted as the effective financial management theory that helps to recognize the funding capability and sustainability procedures of a company or individuals. Measuring financial debts and affordable possibilities of individuals, the main ideologies of the theory are also evolving properly. (*Rahman et al., 2019:64*) has mentioned that the components of the theory demonstrate how different sources of companies are used to improve the structure of internal business funding and also evolve new equity goals of companies effectively. As per companies' structure, prioritizing their internal financing sources to manage their current business equity procedures as well as considering the further sustainability of equity goals to improve customer reliability are defining the effectiveness of theory-based components in a practical environment. Apart from that, the ideologies of the theory help to evolve a positive relationship between the lenders or owners with the individuals while they are conducting transformation in their current marketing segment. Based on that financial security, and transparency in managing more sustainable operating goals during any necessary transactional segment, the components of the theory are also effectively used. Debt financing, equity financing, and internal financing are the main three components of the theory which denote risk assessment issues to manage transactions in a practical environment (Naranjo *et al.*, 2022:729).

Figure 1.1: Pecking order theory.

Source: (Naranjo *et al.*, 2022:729)

In terms of improving the credit scoring model implication by businesses and individuals to reduce the potential risks of loan taking or lending, the implication of three different components of financial management could be beneficial.

Internal financing- The structure of internal financing in the current operating environment is contemplated as the lowest cost and risk management function depending on which internal operating capability of companies to improve their further business goals are reliable. The application of machine learning tools in the current credit scoring structure of a business can be effective in determining how the internal funding structure of companies would be improved. Alongside, for individuals' financial risk management, machine learning tools would be effective in emphasizing the process of analysing financial behaviour, understanding that individual, and improving their security during their traction.

Debt financing- In financial institutions or creditworthiness structures of individuals, the debt financing process is contemplated as the way to develop medium cost-based and risk functions.

Estimating the debt score of business or individual payments and improving sustainability by minimizing risks of debt financing, utilization of big-data analytical tools can be reliable. In online lending and financial transactions, big-data analytical tools would be beneficial to manage the high data security of the lender while payment and ensure the sustainability of history recording effectively.

Equity financing- In financial structure inquiry financing refers to the financial institution's practices depending on which high risks of their business progress and sustainability goals are relying. In terms of improving equity financing structure and operating sustainability goals of institutes with proper management of risk factors, the application of AI-based feasible learning tools can be effective. Based on that, determining the area of automated data transaction decision-making, and evolving the process of risk management functions can be ensured as well and the security of individuals would also be improved.

The section has highlighted how the creditworthiness of businesses and individuals is managed with the proper application of the credit scoring model in the current transformed market, along with different perspectives of authors on managing credit scoring functions. Moreover, detailed articulation of the Pecking order theory has also helped to determine how sustainable strategies are effective in minimizing the risks.

CHAPTER TWO: Literature review II

Credit ratings are essentially a technique of predicting how much money will be lost due to default, delinquency, or arrears. Credit decisions, or binary variables, were previously utilised to make decisions on credit applications and underwriting (Chao et al., 2021). A multinomial variable has just been introduced, which categorises clients into low-risk, medium-risk, and high-risk groups based on credit evaluation criteria. Risk-based pricing considers a consumer's past financial transactions and, in some situations, ongoing conduct when making credit decisions (Dastile et al., 2020). Their credit score is calculated using this info.

A complete framework is used to introduce credit scoring with a focus on regulation (Lima et al., 2020). After that, “behavioural finance” will be defined and what it means to include acceptable traits in borrowers' credit ratings will be studied. Following an examination of the implications of behavioural finance on borrower decisions, this section will move on to social networks, as well as the numerous recent breakthroughs and applications of this idea. In this post, it will examine the utility and reasoning power of parametric and non-parametric credit risk models. Finally, it will address the literature/research gap *(Doumpos and Figueira et al., 2019)*.

2.1 Rules Governing Creditworthiness Evaluation

In 1974, the British government passed the Consumer Credit Act (CCA), which governs how businesses assess the creditworthiness of their clients. The FCA is in charge of putting the regulation into effect *(Cucinelli et al., 2018)*. When someone requests credit or has their credit limit increased, the creditworthiness evaluation procedure is constantly watched. Borrowing money and repaying it over time benefits both businesses and the economy (*Dhliwayo et al., 2015)*. However, the Financial Conduct Authority (FCA) has recognised that certain borrowers are subprime. Lenders can now access industry-specific macroeconomic studies on their loans. To limit reckless lending, the FCA (2018) established prudential procedures such as stress testing to monitor lenders' liquidity and reserves.

Because of the rapid advancement of financial services technology, authorities are revising policies to incorporate new fintech models. A variety of options for financial inclusion were provided to assist the unbanked in entering the financial system *(Gastelum et al., 2017)*. The Financial Action

Task Force (FATF), headquartered in Paris at the OECD, monitors things. According to (*Ijadi et al., 2020*) the group has produced risk-based technology standards that encourage honesty and include the unbanked.

The UK FCA advocates open banking, which allows bank clients to access machine learning models that learn from more transactions *(Ignatius et al., 2018).* This entails banks sharing client information with allowed third parties. Lenders can trust their credit ratings since affordability studies ensure that applicants can repay loans *(Kahraman et al., 2015)*. The UK economy will become less financially fragile. In this era of digital disruption, industrialised countries' open banking capabilities necessitate accurate credit reporting (CCR). If their credit options are questioned, lenders must have a backup *(Dias et al., 2017)*.

2.1.1 Regulation on the General Data Protection

The General Data Protection Regulation (GDPR) is primarily concerned with protecting the private information and data of EU citizens and EEA residents. The main issue is that lenders are increasingly seeing credit scoring as a useful tool *(Kahraman et al., 2015).* To summarise, collecting the score and acting on it is more crucial than attempting to decipher it and provide an explanation *(Kou et al., 2021)*. Still, under GDPR, businesses cannot collect more data than is strictly necessary, and their privacy rules must specify the data's purpose, retention, and use. "The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects based solely on automated processing, such as the automatic refusal of an online credit application," the directive states in part. To avoid potential GDPR violations, several banks now allow customers to seek a personnel review if they disagree with an automated loan decision *(Shi et al., 2019).*

2.1.2 Basel Accords

The Basel Accords establish the risk and capital standards that financial institutions must meet. To reduce credit portfolio risk, the Basel II agreements required financial institutions to use the most recent and accurate models *(Shi et al., 2019)*. It is becoming increasingly difficult for banks to provide both a personalised credit risk assessment and a composite score for all of their loan products *(Zhang et al., 2019).* Borrowers were able to obtain loans with low interest rates, little or no collateral and minimal documentation requirements as a result of the financial credit crisis *(Kahraman et al., 2015).*

The Basel Committee advised that financial institutions strengthen their internal controls to assist them avoid hazards such as credit risk. Companies accredited by the Advanced Internal Rating Based (A-IRB) system were able to reduce their credit risk coefficients as a result of Basel II's (2010) standards. Because credit risk models are so focused on corporate lending and completely disregard the behavioural component of risk assessment, they have cast doubt on the A-IRB and Basel capital agreements *(Jackowicz and Kozłowski et al., 2019).*

2.2 Credit Scoring

Understanding and assessing the financial, macroeconomic, and situational factors that influence repayment is a lending decision problem *(Doumpos and Figueira et al., 2019)*. Authorities usually enforce creditworthiness standards *(Dias et al., 2017)*. According to the FCA, basic credit scoring in the UK involves the loan type and amount, the cost or interest rate, the amount and frequency of instalments repayments, and the potential repercussions of non-payment, such as default charges *(Dhliwayo et al., 2015)*. Some research suggests that a credit risk score include repayment capacity, basic information, guarantees, life stability, credit history, and financial stability. Credit risk assessors look for opportunities to improve ratings and classifications *(Gastelum et al., 2017)*.

According to (*Hirsch et al., 2018)*, this allows one to study historical data to see whether adding a feature enhances categorization or scoring. (*Chao et al., 2021)* discovered that lenders are more inclined to lend to borrowers who provide better financial information. The creditworthiness of a borrower is determined by the risk of negotiations *(Cucinelli et al., 2018)*. Original methods for evaluating this risk relied on loan proceeds such as principal payments or instalments.

Another technique is to apply the time value of money rule, which reduces future cash flows such as loan payments to the present value of the loan amount plus the margin required to calculate interest. In investment analysis, the discount rate is referred to as the “internal rate of return” (IRR). The ROI of a loan indicates its profitability. Most lenders conduct regular checks with borrowers and update credit ratings as needed *(Cummings and Durrani et al., 2016).*

Renewing credit completely corrects credit scores. Before making a decision on a loan application, underwriters estimate the probability of default *(Dastile et al., 2020)*. Estimating an overdue or several arrears would necessitate a higher interest rate to compensate for payment delays. Predicting the default date is even more significant when deciding on a financing duration *(Lima et al., 2020).*

2.2.1 Scope of Assessment

The people seeking consumer loans are the ones who really matter. Along with credit ratings and default likelihood, affordability is a significant factor in credit scoring *(Doumpos and Figueira et al., 2019)*. A significant affordability risk and its consequences, including the applicants' unwillingness or inability to repay the loans, would harm their financial condition. On the other hand, financial exclusion would occur from denying multiple applications due to predicted high affordability concerns *(Dias et al., 2017).*

This is why one can are investigating the applicant's house to determine if there are any further methods, they can earn money. In the event of newlyweds or married couples living at home with their parents, the joint income is taken into account *(Gastelum et al., 2017)*. One example is how the application procedure evaluates joint account performance. Given this reality, several credit-scoring specialists encourage applicants to connect to an existing account with satisfactory credit. (*Dhliwayo et al., 2015)* believes that a joint account would be quite valuable in this situation. Lenders have the power to decide whether borrowers can access this money when they are experiencing financial difficulties *(Lima et al., 2020).*

According to the FCA, guarantor loans should only be approved after thorough evaluation of the guarantor's potential duties in the case of the borrower defaulting or falling behind on payments. However, this examination does not have to be as detailed as the application.

2.2.2 Credit Pricing

The anticipated return on a loan portfolio was the first criterion used to determine interest rates. The expected monetary value (EMV) theory, commonly known as interest pricing, is used to generate risk assessments for various loan portfolios *(Gastelum et al., 2017)*. It considers the distribution of probability and pay-off. According to the anticipated market value, borrowers should be offered a £100,000 loan inside a portfolio with a 5% default rate and a 10% interest rate. Because, as (*Lyn C. Ijadi et al., 2020)* points out, the expected result is 4.5, which equals -(0.05*100) + (0.95*10) = -(5) + 9.5.

The previously defined approach was unclear because no specific individuals were specified. The focus was on loans with identical terms and conditions that were pooled. Customers with great credit will not apply to this pool for fear of being overcharged or, worse, competing with low-quality applicants *(Ignatius et al., 2018).*

Thus, risk-based pricing was proposed as a method of assessing individuals based on their unique characteristics and circumstances. Experian (2013) said that this technique resulted in customised loan terms, grace periods, guarantors, and total credit charges. Other outcomes included different interest rates, loan amounts, payback kinds, amounts, frequency, total payments, and credit charges *(Jackowicz and Kozłowski* et al., 2019)*.*

As a result, loan pricing is now determined by the specific information collected from customers during the application process, allowing for a more customised approach. Hard information is data that the lender or CRA can extract from the application, whereas soft information is data that the applicant's environment can supply *(Ignatius et al., 2018)*. The first two are more common, but lenders are now considering the third, which is newer, as an option. The modelling section below will discuss how these two types affect a consumer's credit score *(Jackowicz and Kozłowski et al., 2019).*

2.2.3 Traditional Criteria

Lending presume that individuals can translate their financial situation into trustworthiness. According to (*Brockett and Golden et al., 2007),* socioeconomic considerations influence traditional grading practices. A person's financial position can be assessed in a variety of ways. Examples include age, marital status, nationality, gender, number of children, occupation, sector, residency status, job type, length of employment, and loan data. The application gathers payment history, credit score, and current obligations. When applying for a mortgage or other asset, are asked about the location and vehicle ownership.

According to (*Kahraman et al., 2015)*, credit ratings are based on application and transactional data. Financial status at the time of application submission is both demographic and personal *(Kou et al., 2021).* Examples include marital status, number of dependents, age, income, loan purpose, current residence, number of bank accounts and credit cards, bank years, employment category, residency type/home ownership, debt-to-income ratio, and credit history. Mortgage lenders consider property location *(Chao et al., 2021)*.

Aside from work and salaries, family income includes any money received from savings or other sources. Costs can be "non-discretionary" and eligible for reimbursement, or "ineligible" and paid freely *(Chi and Zhang et al., 2017).* The first two are required to establish credit and can be used to replace any legal or contractual obligation. FCA (2018) defines disposable income as the amount of money left over after meeting essential expenses.

Demography, marketing, and macroeconomics are all factors to consider when gathering application data. According to marketing theory, lenders will respond to their borrowers. (*Cucinelli et al., 2018)* define market research and economic conditions as macroeconomic factors. These factors forecast economic indexes such as the consumer price index, average interbank lending rate, FTSE 100 return on log, and GDP growth *(Cummings and Durrani et al., 2016)*. Sometimes regulators are in charge of macroeconomic factors. According to (*Dastile et al., 2020)*, the UK government permitted lenders under the FCA to grant holiday pay for loans to people whose affordability was affected by the coronavirus epidemic without harming their credit scores.

2.3 Theoretical framework of Credit Scoring Criteria

In relation to the earlier analysis of traditional and behavioural criteria, Figure.2.1 depicts lenders' scoring systems. Importantly, the components required are influenced by the lender's capacity, the type of loan, and its size (*Lima et al., 2020)*. For example, traditional brick-and-mortar banks may prioritise thorough transaction records and industry-specific economic studies over behavioural characteristics and CRA data. Alternatively, internal application data and external CRA could serve as the foundation for a new organisation. In addition, it may collaborate with another psychometrics company to develop a behavioural model.

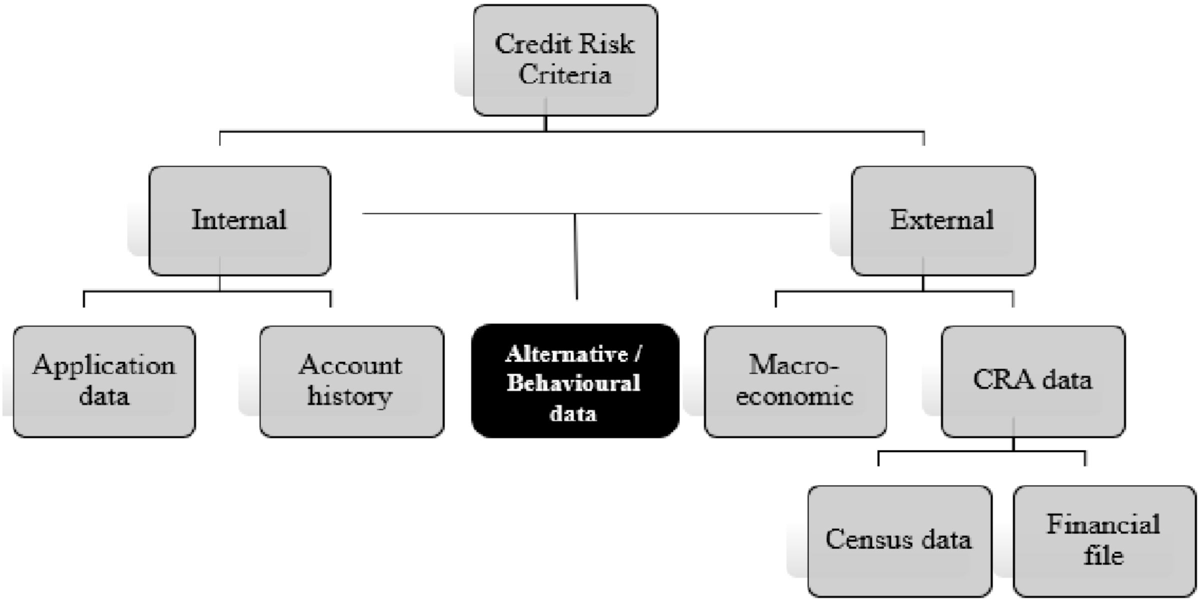

Figure 2.1: Theoretical framework of credit scoring criteria

Source: *(Lima et al., 2020)*

Instead of traditional financial data, lenders can now use behavioural data, often known as “alternative data”, in the literature. Internal data includes the amount of persons who opened a campaign email or went with a borrower to the lender's premises *(Doumpos and Figueira et al., 2019).* Lenders may get access to behavioural datasets such as telephony records as part of service provider agreements. Once developed, the model will be internally maintained to provide behavioural credit scores *(Dias et al., 2017).*

Credit reporting bureaus are one possible source for this type of information. A psychometrics firm could use a borrower's geospatial data, or a personality trait acquired from tests or social media activity to generate an external behavioural score. Any financial institution that works with a borrower can share financial information with the consumer reporting agency. As previously discussed in this chapter, the borrower may also have access to their court records, voting records, and census statistics *(Dias et al., 2017)*.

2.4 Conceptual framework of credit risk models

The challenge of applying the best models to credit risk assessment has been essential for financial institutions. The onus is on the borrowers to keep loan amounts acceptable and affordable *(Dhliwayo et al., 2015)*. To be clear, even if they are experiencing financial difficulties, persons may still be held liable if they fail to return a loan. An individual's capacity to repay loans while considering credit risk is called creditworthiness. Scholars have discussed credit risk prediction methods *(Gastelum et al., 2017).* Three models examined customer default probability (PD), whereas two examined loan exposure at default. Basel agreements establish a formula for loan expected loss by multiplying the three components above.

Credit cards, overdrafts, and other revolving credit are vulnerable to EAD. It can be used to estimate loan debt at any time. One might simulate eads using survival analysis to track state changes between payment due dates. (*Hirsch et al.* 2018) suggest using a cox proportional hazard model for survival analysis to detect borrower situations. A credit risk model is being created to help detect candidates with low creditworthiness; however, the model must be faster, easier to use, and more accurate *(Gastelum et al., 2017).*

Models should preserve transparency and honesty while aiming for maximum accuracy. When automated findings are contested, they must be supported with reasoning. Lenders in the United Kingdom and the European Union insist on the aforementioned conditions, even if they understand their importance. According to (*Ijadi et al., 2020)*, the anticipated loss is calculated by multiplying the EAD ratio, likelihood of default, and loss given default (LGD). The model is considered accurate when the score that minimises this EL is minimised.

The widespread consensus is that a broad area under the receiver operating characteristic (AUROC) curve suggests a good model. The credit risk model's performance will be assessed using the ROC curve *(Ignatius et al., 2018)*. Also, take in mind that credit scores allowed lenders more leeway. A certain score may not be sufficient for numerous lenders due to differing standards. This will be determined by the lender's risk tolerance and the objectives of the credit officers. (*Jackowicz and Kozłowski, 2019)* outlines three goals for credit risk modelling. In an attempt to increase earnings, he highlighted the lender's current concentration on cross-selling. He stated there are two approaches to improve market penetration: reduce customer churn and attrition and attract new borrowers. Finally, calculating the acceptance response rate allows us to determine a convincing return on equity for the borrower *(Kahraman et al., 2015).*

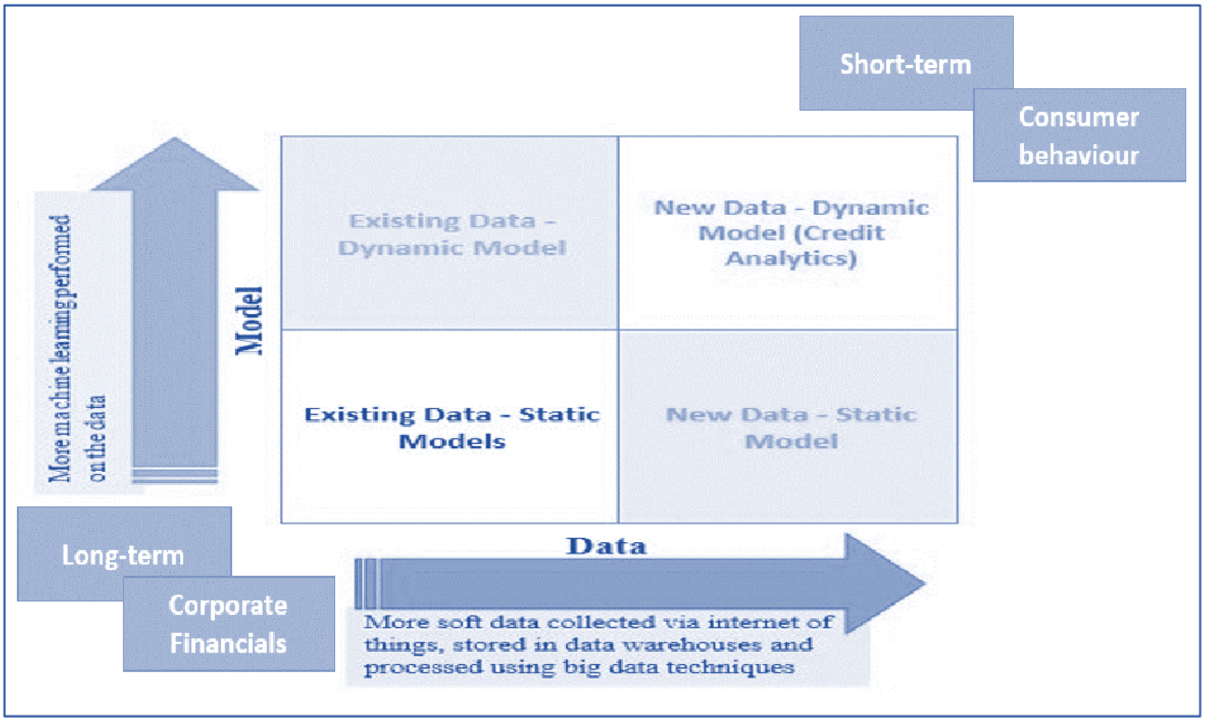

Figure 2.2: Trends in credit scoring models and data used.

Source: (*Chao et al., 2021)*

Figure 2.2 summarises the research on credit scoring algorithms and data kinds. It gives insight on the current state of digital lending and its potential benefits. Graphs are an effective tool for evaluating networks, which are collections of interrelated data. A graph is one type of graphical depiction of a network that uses vertices and edges. Undirected networks can have interactions in either way, whereas directed graphs, such as digraphs, have each edge with a definite orientation *(Kahraman et al., 2015).*

Numerous research articles, including (*Kou et al., 2021)* have presented community discovery methods that leverage current distance functions to calculate node similarity. According to (*Shi et al., 2019)*, one strategy for developing this type of model is to apply the general stochastic model in electronic communications. Another example is the community affiliation model, which introduced the idea of overlapping and nested communities. (*Zhang et al., 2019)* attempted to better understand network communities by distinguishing several types of communities and proposing a method for doing so. Others, such as RSN-BD and MRGC, as well as the socdim model, which considers modularity and employs a discriminative classifier like SVM, were proposed based on the community's definition.

2.5 Research gap

Financial inclusion is critical because half of the global population does not have a bank account. Even with inclusive policies, (*Kahraman et al., 2015)* found that knowledge asymmetry hurts financial institutions. Adverse selection and moral hazard ensued. Once can investigated credit scoring methods to improve categorization accuracy. (*Ignatius et al., 2018)* found that even minor accuracy increases saved lenders money. According to the FCA (2018), lenders overlooked opaque borrowers and meritorious instances, resulting in increased limitations and durations.

(Dhliwayo et al., 2015) looked at new socioeconomic and demographic credit rating criteria. Adjusting for credit score volatility increased the model's fitness. Very little research has been conducted on ideal values or categories within variables other than the two primary streams. The second stream, which analysed diverse data sets, increased. (*Dias et al., 2017*) investigated local consumer credit scores *(Doumpos and Figueira et al., 2019)*. In this significant work, scientists tested the logistic regression model on six data sets, analysed numerous machine learning methods, and examined the findings. To assess evidence levels, social data sets are examined using the Naïve Bayesian technique *(Dias et al., 2017)*.

This study looks into the economical, psychological, behavioural, sociological, and technical implications of iot and big data analytics *(Chao et al., 2021)*. The influence of big data on finance and decision-making has challenged existing frameworks. The properties of social networks were explored in real life. (*Chi and Zhang et al., 2017)* are among the few research that has looked into how social networks affect credit ratings. Probability was used in linear regression and Cox proportional hazard models. All three studies examined real-world peer-to-peer impacts. (*Cucinelli et al., 2018)* reported inaccurate social network data results. (*Cummings and Durrani et al., 2016)* looked on predicting credit scores with phone network and traditional data.

(Dastile et al., 2020) used mathematics and statistics to investigate the effects of social networks on credit, however they offered no data. Under some conditions, the equations that simulate homophily have an impact on credit ratings and default risks. Few writers, like *(Lima et al., 2020)*, have done experimental research of social networks from the perspective of lenders. It distinguishes itself with its lender-centric approach to examining how different types and sizes of social networks affect credit scores *(Doumpos and Figueira et al., 2019)*.

(*Dias et al., 2017*) investigated social media credit rating algorithms. (*Dhliwayo et al., 2015)* investigated social media dynamics. These personality qualities were primarily behavioural. The number of pages or subscriptions was indirectly related to social networks. Subscribe to the same page to organise the social networks. (*Gastelum et al., 2017)* confused social media and networks. This study compared the two categories and investigated how evaluation criteria influence borrowers' credit scores.

This study eliminates information asymmetry and enhances financial inclusion to achieve equitable credit scoring for governments, banks, and borrowers, making loans and company formation easier *(Hirsch et al., 2018)*. Despite on-going credit risk management research involving big data and credit analytics, there are no empirical criteria for determining which sources to check and when *(Ijadi et al., 2020)*.

2.6 Summary

Credit rating predicts default, delinquency, and arrears losses. A multinomial variable divides client into low-risk, medium-risk, and high-risk groups depending on credit evaluation. Credit scores are based on a consumer's past financial transactions and behaviour. The study examines “behavioural finance” and borrower decisions. It tackles the literature gap and evaluates parametric and non-parametric credit risk models' rationale and utility.

Industry-specific macroeconomic research on loans, prudential procedures to check lenders' liquidity, and unbanked financial inclusion solutions are available from the UK FCA. The FATF created risk-based technology guidelines to promote honesty and inclusivity. The UK FCA supports open banking using machine learning. For accurate credit reporting in the digital age, lenders need backups. EU individuals and EEA residents' private data is protected under the GDPR.

The applicant's home is checked for additional income, especially for newlyweds or married couples living with parents. The application considers joint income, and credit-scoring experts advocate connecting to a good account. After assessing default and lateness risks, lenders can issue guarantor loans. The expected monetary value (EMV) hypothesis considers probability and pay-off to set interest rates. Pooling loans with comparable terms and conditions avoids overcharges and low-quality applicants.

Risk-based pricing customises loan terms, grace periods, guarantors, and credit charges based on individual traits and circumstances. Loan price now depends on hard and soft application data. Traditional lending criteria include age, marital status, nationality, gender, occupation, and loan records. Application and transactional data, including demographic and personal characteristics, determine credit ratings. Mortgage lenders evaluate property location when calculating credit scores.

Establishing credit and substituting legal or contractual obligations requires family income, including non-discretionary and ineligible charges. Demographics, marketing, and macroeconomics influence application data. Marketing theory and macroeconomic elements like market research and economic conditions influence lenders' lending decisions. Lenders might give holiday pay for loans to coronavirus victims without hurting their credit scores under UK law. Lenders can now utilise “alternative data,” or behavioural data, instead of financial data.

Financial firms need credit risk models to evaluate borrowers' loan repayment ability. Customer default probability (PD), loan exposure at default (EAD), and estimated loss are credit risk factors. Psychometric organisations can produce external behavioural scores from telephonic records for lenders. Court records, voting data, and census figures are available to borrowers. Revolving credit can be estimated using the Basel agreements' loan expected loss (EL) methodology.

Survival analysis utilising a cox proportional hazard model can discover borrower situations and simulate eads. A credit risk model is being created to identify low-credit candidates. The model's accuracy depends on minimising anticipated loss, default likelihood, and default loss. The ROC curve evaluates model performance. A broad AUROC curve area is required by UK and EU lenders. Credit risk modelling aims to increase earnings through cross-selling, reduce customer turnover and attrition, and calculate the borrower's return on equity.

Researching credit scoring algorithms and data kinds reveals digital lending's current condition and future benefits. For network evaluation and community discovery, graphs are popular. Research has examined community affiliation, RSN-BD, MRGC, and socdim models. Since even small categorization accuracy can save lenders money, there is a research vacuum. The FCA (2018) found that lenders ignored opaque borrowers and meritorious cases, increasing limits and durations.

This study evaluates social media and networks on borrowers' credit scores to reduce information asymmetry and improve financial inclusion. It reduces information asymmetry and improves government, bank, and borrower credit scores. Despite big data and credit analytics studies on credit risk management, there are no empirical criteria for checking which sources and when. Multicriteria outranking for business credit ratings may increase financial inclusion, according to the study.

Studies on collateral-based SME lending, including commercial and personal collateral in developing nations, are discussed. It also covers interorganizational trust and agency costs in savings bank-SME credit arrangements. Implementing interval-valued multi-granular 2-tuple linguistic BWM-CODAS with target-based qualities for building site selection is also covered. The text also examines SME manager-bank staff social relations and fuzzy multicriteria decision-making in SME bankruptcy prediction.

CHAPTER THREE: Methodology

This section provides critical details regarding various techniques and procedures that have been used for the purpose of identifying, selecting, processing, and analysing the collected information regarding the concerned topic of the research. It is a crucial section which allows researchers to evaluate the reliability and validity of a study. The current study focuses on examining the role and concept associated with credit scoring models in order to develop the process of simple credit storing to predict the creditworthiness of a business or individuals. This model is helpful for determining the trust for the process of investment along with identification of the length of the payment process and credit history for a business. In this context, this section is going to provide a detailed discussion regarding the philosophy of the research that has been chosen for the process of Data Collection. In addition, the specific research philosophy approach and design are going to be explained in an appropriate manner along with proper justification behind the selection of the specific ones and rejection of others. The method of Data Collection along with the strategies that would be considered in the study are also going to be discussed. The data analysis process is also going to be discussed which would be incorporated in the concerned research study for analysing the collected information and data for the research.

3.1 Research Philosophy

Research philosophy involves the assumptions, nature, and knowledge of the research study. It focuses on specific methods of knowledge development. Positivism research philosophy has been found to be helpful for quantitative research and quantitative data is involved in this kind of philosophy. Therefore, this has been discarded for the current study. On the other hand, realism focuses on developing theories which is not required for the current study; hence, it has been rejected *(Pillai and Kaushal et al., 2020)*. Conversely, the Pragmatism research philosophy is effective for the purpose of developing testable and untestable analysis in the study, hence, it has been accepted.

3.2 Research Approach

On the contrary deductive research approach has been discarded as it involves a quantitative approach for the purpose of Data Collection which is not suited to the current research. Researchers are allowed to use their perspective along with other perspectives for the purpose of Representation of the data *(Tomaszewski et al., 2020)*. It is beneficial for researchers to make observations for the purpose of collecting data. A broader view of the data is involved in this kind of research which is highly helpful for the current research.

With the help of an abductive research approach, it has been identified that inferring new theories is possible along with developing existing theories which fit with the idea associated with qualitative field study practices *(Brandt and Timmermans et al., 2021).* It also includes ongoing reflection based on the data along with positioning the same against several theories in such a manner that data contributes significantly and develops further. The research approach focuses on cultivating surprising and Anomalous empirical findings based on the background of various existing theories. The research process in the abductive approach has been found to initiate with puzzles or surprising facts and the process of research has been found to be devoted to explanations. There is an incomplete set of data which proceeds to the most likely possible description of the set.

3.3 Research Design

Research Design refers to the plan involved in the research for the purpose of answering the questions developed in the study. It involves strategies that are going to be used for the purpose of implementing the plan. Experimental research design has been found to allow researchers to explore and present new ideas that are crucial to finding solutions to research problems. This is helpful for researchers to solve complicated problems of the research. This research design increases the flexibility of research allowing various hypothesis testing. Experimental research leads to better results in the research, resulting in increasing the validity and reliability of the study *(Pandey and Pandey et al., 2021)*. It makes researchers draw valid conclusions from the information and data. Researchers are able to make informed and good decisions based on collected data. Experimental Research Design has been helpful for the study to evaluate existing sets of data along with identifying the real-time sources by involving participants *(Dawadi et al., 2021).* This research project design has helped researchers to utilise real-time sources and information in the current study.

3.4 Data Collection Method and Strategy

A mixed method has been selected for the purpose of data collection in the study. Pragmatism research philosophy focuses on understanding the different aspects associated with the information and data in the research. It allows researchers to flexibly adapt to the evolving research nature *(Kelly and Cordeiro et al., 2020).* Pragmatism philosophy includes the designs in the research which implement operational decisions on best-suited recommendations to find answers to the Research question. It allows researchers to conduct the study in dynamic and innovative ways for the purpose of finding solutions to the problems of the research.

3.5 Data Analysis

It minimises the time associated with data analysis; hence the data and others are not required to spend a lot of time for families arising themselves with the language of programming *(Hunziker and Blankenagel et al., 2021)*. Both of these methods have been effective and beneficial for the Analysis of data and helped in understanding that the credit scoring model is helpful for estimating the probability of upcoming defaults in the lending process.

3.5.1 Python Data Analysis

Python data analysis is used for the purpose of establishing and evaluating the models of data. Python data analysis is helpful for the prediction of future Trends associated with the collected patterns of data *(Raschka et al., 2020).* Python is increasing its popularity in the data analysis process as it is Highly Effective and beneficial for company managers to make effective and informed decisions by identifying market Trends and making better business-related decisions. Different Python libraries have been found to allow researchers to convert digits into histograms graphics Pie Charts and many others which helps them to make their drive and information and the data comprehensible and visually appealing Python is easily readable and has simple syntax which is an important significance *(Abkenar et al., 2021)*. This is because the aspect reduces the time consumed by data analysts to familiarise themselves with specific programming languages. It has been identified that the gentle curve for learning is helpful to stand out among previous languages of programming having Complex syntax.

3.6 Ethical Considerations

Ethical consideration is an essential aspect associated with research study as it ensures that the research has been conducted responsibly along with respecting the rights of involved studies and participants in the research process *(KANG and Hwang et al., 2021)*. In this context, the study has ensured to follow ethical considerations throughout the research. Data protection regulations have been Incorporated in the study to protect the information of involved individuals *(Abkenar et al., 2021)*. Moreover, the guidelines of the university for the purpose of conducting research have been followed appropriately to ensure the validity and reliability of the research. The study has included a qualitative Data Collection method for which the principles and regulations associated with the copyright act have been implemented and followed appropriately *(Chen et al., 2021).* By addressing the Up 4 mentioned ethical considerations it has been ensured that the study is completed responsibly along with maintaining the trust of the public.

3.7 Summary

The above discussion concludes that the section has included a detailed overview of the Planning regarding the Different techniques and tools used in the process of collecting data and information for the current study. In this context, it has been identified that the section has included an in- depth discussion regarding the philosophy approach and design of the current research with proper justification. Moreover, the method strategy and tools associated with collecting data have also been described in detail. This section has provided the data analysis techniques as well. In this context, it has been identified that Python data analysis has been involved in the study along with thematic analysis.

CHAPTER FOUR: Findings-Analysis-Discussion

The credit scoring model has been found to be involved in the several factors such as age income employment status and many others for the purpose of understanding the credits of the business and individual. At the time of creating the model deep understanding of the statistical techniques are required along with accessing the extensive data that can be helpful for the purpose of building the simple model. The simplicity of the model can be understood with the help of selecting the target variable. However, age, employment, status, debt to income ratio credit history and revolving credit utilisation are also the important data that can be utilised for the purpose of understanding. Cleaning of data is also needed for the purpose of selecting feature and using techniques such as correlation analysis for the purpose of identifying the event variables.

4.1 Data Analysis

4.1.1 Thematic Analysis

Getting the data representation

The selected data has been found to be based upon providing different information such as customer ID month name age occupation annual income monthly salary net bank accounts and different others. These data are found to be selected for the purpose of evaluating it in such a manner that credit scoring model can be developed. This model is designed based upon using the data for cleaning purposes in such a way that the null values are replaced with the mean value of the data *(Ampountolas et al., 2021)*. This replacement has been done for the purpose of enhancing the overall representation of the data. It is also going to be improved is such a manner that the data can be evaluated appropriately followed by enhancing the overall quality of the data. With the help of using this data the credit scoring model is going to be developed in this assessment with proper incorporation of all the factors which are associated with it.

Identifying the factors affecting the credit scoring

The credit score in has been found to be involved in the credit were the Ness of an individual based upon several factors that also provide and information about the financial behaviour of an individual for the purpose of determining their ability for the purpose of managing the responsibility of credit in an effective and advanced manner. However, it is also going to provide an accurate information about the payment history where the timely payment of credit obligations includes the credit cards mortgages loans and many others. However, the credit utilisation is also one of the major factors that play an important role in terms of understanding the credit goodness of an individual. This is illustrated as the ratio of card balance and the current credit to the credit limit provides and propriety information about the credits of an individual. The length of the credit history is also helpful for the purpose of providing and a pro create information regarding the average age of the credit accounts whereas the credit in use also help in having a positive impact on the credit score *(Ndayisenga et al., 2021)*. This includes the credit card, installments, and the retail accounts. The new credit is also the recent application that include the credit inquiries and the number of recently opened account of an individual which help in integrating the sign of financial stress whereas the public records such as negative public record. This has been found to be the bankrupt sees and judgement that can significantly lower the credit score. The recent credit behaviour is also one of the major factors that are included in the credit were the analysis of an individual such as late payment or the account which are going to collect the positive behaviour that can contribute to the score improvement. The credit counselling and credit dead management programs are also helpful in terms of reporting its influence on the credit scoring that help in different credit bureaus and the financial institution that can use different models to analyse these factors.

The benefits of credit scoring model

Credit scoring model has been found to be automating the assessment process of an individual that enable the leaders to understand the effective decision of the loan approval. It also helps in streamlining the landing process and reducing the time which are required for manual reviews. With the help of evaluate in the credit for dinners of an individual the organisational managers can better access the risk which are associated with the extending credit. This can be helpful for the purpose of providing the inform the decision to minimise the lightly good of the financial losses. The credit scoring model has been found to be providing and standardised and consistent approach to evaluate the application of credit which help in ensure in the applicant which are using the same criteria. The scalability is also found to be one of the major factors for the credit application that help in enhancing the value ability of the institution in terms of dealing with the high number of loan application *(Hussin Adam Khatir, and Bee et al., 2022).* The cost reduction is also done with the help of credit scoring as it helps in extensive manual reviews that lead towards leading the cost saving for the landers. Effective games the credit scoring model in order to suit the specific business needs and segments for the customer base. This has been found to be allowed for the more target landing strategies based on the unique characteristics of different demographic groups. The regulatory complaints are also found to be one of the major effective and appropriate factors for the purpose of providing the transparent and standardized approach to credit assessment which is very crucial for fair lending practices and regulatory standards.

Role of credit scoring model and improving financial landscape

It has been identified that the credit scoring model is found to be very helpful for the purpose of financial landscape is multifaceted. It also contributes to the fairness and the efficiency of the financial scale ability. Accessibility is also improved with in the financial landscape with the help of using the credit modelling. The credit score modelling also help the managers on the organisations to allocate the resources in an effective and advanced manner with the help of automating the evaluation of streamline the landing process and allows the financial institution to process the high volume of loan application. It also helps in increasing the access to credits. The credited school model models are also found to be helpful for accessing the credits especially for the individuals with having no traditional credit history with the help of consisting the variety of factors which are beyond the credit history that help to enable more people in order to establish the credit and access the financial products *(Dastile, and Celik et al., 2021)*. The fair lending practices are also provided with the help of credit scoring models as it helps in applying the standardized criteria for the purpose of accessing the credit. It also helps in eliminating the discriminatory and ensured in the individuals are evaluated based upon objective factors that and subjective judgements. It also helps and standardising the decision making with the help of providing a standardised frame in order to make the credit decision consists of and showing that all the applicants are getting chances based on same criteria which help in transparency and fairness in the decision-making process.

It also helps in enhancing the encouragement of the responsible with help in rewarding the individuals with positive financial habits were borrows would a monster that the responsible credit

Management can be qualify and invaded terms with low interest rate it can be provided. The credit education opportunities are also provided it with the help of providing in the visuals with insights into the factors which are influencing the credit scores. It also promotes the financial literacy and empower the terms of making good decision for the financial health. Enhance operations of the financial institution with the help of alternating the credit assessment process. It also reduces the manual workload and minimise the errors *(Bücker et al., 2020)*. The institutions are also allowed to focus on strategic initiative based upon the data driven of credit involvement and also done with the help of credit score in models with the help of that inter visuals are provided with opportunity to activate the improvement of the credit profiles over time with the help of understanding different factors that influence the scores and take steps to address the weaknesses followed by building opposite of credit history. The credit scoring models also contribute to the global financial and fusion by providing the framework for the purpose of analysing the credit for the Ness which can be adapted to different region and financial ecosystems.

Possible challenge is in terms of implementing the credit scoring model.

There can be a data quality and consistency issue and challenge in terms of having inconsistent data and in accurate data that can the earth adversely impact the performance of credit scoring model. The data quality issues can also be arisen from the errors in reporting that are outdated information. The BIOS and fairness can also be there as a challenge at the time of incorporating the concerned model based upon the race general and social economic status. Can be a major challenge at the time of implementing the credit scoring model.

4.1.2 Result and Analysis

The libraries are the main portion of the Python evaluation which provides the necessary base setting of the evaluation. The libraries are used to evaluate the data in an analytical process. The 'numpy' library is used to evaluate the calculative approach in the main frame. The library 'pandas' is used to evaluate the collected data that supports the research process. Some of the libraries like 'matplotlib', 'seaborn', and so on are used to demonstrate the visualisation of the data. The labelling and standardisation libraries are used to implement necessary data-setting factors. The warning module is used to eliminate the code's warnings generated while evaluating the code.

```python
import numpy as np
import pandas as pd
import tensorflow as tf
import matplotlib.pyplot as plt
import plotly.express as px
import seaborn as sns
from sklearn.preprocessing import LabelEncoder, StandardScaler
import warnings
warnings.filterwarnings('ignore')
```

Figure 4.1: Libraries import.

Source: (*Jupyter Notebook*)

The evaluation of the data provides information regarding the evaluation of the collected CSV data. This CSV data is a secondary resource data which contains the details of some credit data of multiple customers. The data contains various necessary attributes that support the evaluation of the data. The customer information is also evaluated in this section which provides the information regarding the credit data of multiple customers *(Teles et al., 2020)*. The customer data represents the customer ID, name, age, SSN number, and occupation. The structure of the evaluated data is elaborated in this component.

```python
credit_score = pd.read_csv("test.csv")
```

```python
credit_score.head()
```

	ID	Customer_ID	Month	Name	Age	SSN	Occupation	Annual_Income	Monthly_Inhand_Salary	Num_Bank_Accounts	...	Num_Credit_Inquiries	Credit_M
0	0x160a	CUS_0xd40	September	Aaron Maashoh	23	821-00-0265	Scientist	19114.12	1824.843333	3	...	2022.0	Gc
1	0x160b	CUS_0xd40	October	Aaron Maashoh	24	821-00-0265	Scientist	19114.12	1824.843333	3	...	4.0	Gc
2	0x160c	CUS_0xd40	November	Aaron Maashoh	24	821-00-0265	Scientist	19114.12	1824.843333	3	...	4.0	Gc
3	0x160d	CUS_0xd40	December	Aaron Maashoh	24_	821-00-0265	Scientist	19114.12	NaN	3	...	4.0	Gc
4	0x1616	CUS_0x21b1	September	Rick Rothackerj	28	004-07-5839	_______	34847.84	3037.986667	2	...	5.0	Gc

5 rows × 27 columns

Figure 4.2: Evaluation of data

Source: (Jupyter Notebook)

Detailed information regarding the types, count values, and numbers of null values present in the columns of the data is highlighted here. This provides the information on the columns of the data that provide the information regarding the handling of the data. The non-null portion provides the information that assists in the data validation and configuration process. The type of the data details represents the involvement of object, float, and integer data.

```
credit_score.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 27 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   ID                        50000 non-null  object 
 1   Customer_ID               50000 non-null  object 
 2   Month                     50000 non-null  object 
 3   Name                      44985 non-null  object 
 4   Age                       50000 non-null  object 
 5   SSN                       50000 non-null  object 
 6   Occupation                50000 non-null  object 
 7   Annual_Income             50000 non-null  object 
 8   Monthly_Inhand_Salary     42502 non-null  float64
 9   Num_Bank_Accounts         50000 non-null  int64  
 10  Num_Credit_Card           50000 non-null  int64  
 11  Interest_Rate             50000 non-null  int64  
 12  Num_of_Loan               50000 non-null  object 
 13  Type_of_Loan              44296 non-null  object 
 14  Delay_from_due_date       50000 non-null  int64  
 15  Num_of_Delayed_Payment    46502 non-null  object 
 16  Changed_Credit_Limit      50000 non-null  object 
 17  Num_Credit_Inquiries      48965 non-null  float64
 18  Credit_Mix                50000 non-null  object 
 19  Outstanding_Debt          50000 non-null  object 
 20  Credit_Utilization_Ratio  50000 non-null  float64
 21  Credit_History_Age        45530 non-null  object 
 22  Payment_of_Min_Amount     50000 non-null  object 
 23  Total_EMI_per_month       50000 non-null  float64
 24  Amount_invested_monthly   47729 non-null  object 
 25  Payment_Behaviour         50000 non-null  object 
 26  Monthly_Balance           49438 non-null  object 
dtypes: float64(4), int64(4), object(19)
memory usage: 10.3+ MB
```

Figure 4.3: Information about data

Source: (*Jupyter Notebook*)

The cleaning process of the data provides information regarding the null presence of the data. The provided information defines the attributes which have null values in the main section of the data. The null values are found in the column name, monthly in-hand salary, number of loans, and so on *(Li, and Chen et al., 2020)*. This portion of evaluation provides the information about the presence of null or 'NAN' values in the columns.

```
credit_score.isnull().sum()
```

```
ID                             0
Customer_ID                    0
Month                          0
Name                        5015
Age                            0
SSN                            0
Occupation                     0
Annual_Income                  0
Monthly_Inhand_Salary       7498
Num_Bank_Accounts              0
Num_Credit_Card                0
Interest_Rate                  0
Num_of_Loan                    0
Type_of_Loan                5704
Delay_from_due_date            0
Num_of_Delayed_Payment      3498
Changed_Credit_Limit           0
Num_Credit_Inquiries        1035
Credit_Mix                     0
Outstanding_Debt               0
Credit_Utilization_Ratio       0
Credit_History_Age          4470
Payment_of_Min_Amount          0
Total_EMI_per_month            0
Amount_invested_monthly     2271
Payment_Behaviour              0
Monthly_Balance              562
```

Figure 4.4: Cleaning data

Source: (*Jupyter Notebook*)

The replacement process of the null section provides the information about the filling of the null places of the column with 0. It provides a detailed structure of the data that supports the evaluation of the information. The intricate configuration furnishes the elimination of the null with the support of 0.

```
credit_score = credit_score.fillna(0)
```

```
credit_score.isnull().sum()
```

```
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
dtype: int64
```

Figure 4.5: Null value replace.

Source: (Jupyter Notebook)

The substitute bloc furnishes the straightforward vision of the substitute of unique essences such as ‘_’ with an expressionless matter. The transformation approach also furnishes facts concerning the involvement of the integer data province. This notion furnishes the involvement of the modification procedure that is operated to reverse the ‘object’ column data into ‘integer’ queue data. The data mutation strategy is manipulated to assemble the queue into an exhaustive data type which is influential for the evaluation.

```
credit_score.Num_of_Delayed_Payment = credit_score.Num_of_Delayed_Payment.str.replace('_', '')
```

```
credit_score['Num_of_Delayed_Payment'] = pd.to_numeric(credit_score['Num_of_Delayed_Payment'], errors='coerce', downcast='integer')
```

Figure 4.6: Replacement of special characters

Source: (Jupyter Notebook)

The staging procedure of the data supplies the recap of the changing of the extraordinary surface with vital pieces. This feeds the dispatch concerning the transformation of the data class. This modification orients some expressionless spaces which can be annihilated by substituting those blank stretches with 0. The final structure of the main frame of the data is evaluated after this conversion of the data *(Varun Kumar et al., 2020)*. The setting of the data parameters provides information about the involvement of the data setting section.

```
credit_score.Outstanding_Debt = credit_score.Outstanding_Debt.str.replace('_', '')
```

```
credit_score['Outstanding_Debt'] = pd.to_numeric(credit_score['Outstanding_Debt'], errors='coerce', downcast='integer')
```

```
credit_score['Amount_invested_monthly'] = pd.to_numeric(credit_score['Amount_invested_monthly'], errors='coerce', downcast='integer')
```

```
credit_score = credit_score.fillna(0)
```

```
credit_score['Monthly_Balance'] = pd.to_numeric(credit_score['Monthly_Balance'], errors='coerce', downcast='integer')
```

Figure 4.7: Data setting

Source: (Jupyter Notebook)

The description of the data provides the detailed structure of the descriptive investigation. This type of investigation provides information on the count value, and other parameter values such as min, percentage, and many more, of the numeric data. The integer or float-type data columns are evaluated in this type of evaluation. The sectional portion provides a detailed evaluation of those parameters of the data columns.

```
credit_score.describe()
```

	Annual_Income	Monthly_Inhand_Salary	Num_Bank_Accounts	Num_Credit_Card	Interest_Rate	Num_of_Loan	Delay_from_due_date	Num_of_Delayed_Payment	Ch
count	5.000000e+04	50000.000000	50000.000000	50000.000000	50000.000000	50000.000000	50000.000000	50000.000000	
mean	1.534927e+05	3554.870928	16.838260	22.921480	68.772640	3.108580	21.052640	28.727960	
std	1.294556e+06	3285.349285	116.396848	129.314804	451.602363	65.487429	14.860397	213.766645	
min	0.000000e+00	0.000000	-1.000000	0.000000	1.000000	-100.000000	-5.000000	-3.000000	
25%	1.744485e+04	1183.200833	3.000000	4.000000	8.000000	1.000000	10.000000	8.000000	
50%	3.489922e+04	2615.258333	6.000000	5.000000	13.000000	3.000000	18.000000	13.000000	
75%	6.995556e+04	5338.967500	7.000000	7.000000	20.000000	5.000000	28.000000	18.000000	
max	2.413726e+07	15204.633333	1798.000000	1499.000000	5799.000000	1496.000000	67.000000	4399.000000	

Figure 4.8: Description of data

Source: (Jupyter Notebook)

The occupation-wise evaluation of the data is elaborated in this portion. This provides the details of the occupations which are present in the dataset. The outstanding debts of those occupations are evaluated here. This provides the initialisation of the group by technique which is used to group the data. The grouping method is used to evaluate the sum of the outstanding debt of multiple occupations.

```
occupation_wise_evaluation = credit_score.groupby('Occupation')['Outstanding_Debt'].sum().reset_index()
occupation_wise_evaluation.head()
```

	Occupation	Outstanding_Debt
0	Accountant	4530059.89
1	Architect	4265514.99
2	Developer	4531705.11
3	Doctor	4396145.30
4	Engineer	4638882.81

Figure 4.9: Occupation-wise data evaluation

Source: (Jupyter Notebook)

The plot provides the construction of the pie that highlights the occupation-wise outstanding debt evaluation. This provides the plotting of different sum values of the outstanding debt for various occupations. The constructive portion provides information on the making of the legends of the plot that defines the types of occupations of the collected dataset *(Xu et al., 2021)*. The plot supports the outcome of the visualization process using Python coding.

```
figure1 = px.pie(occupation_wise_evaluation, values='Outstanding_Debt', names='Occupation', title='Occupation-wise Outstanding Debit')
figure1.show()
```

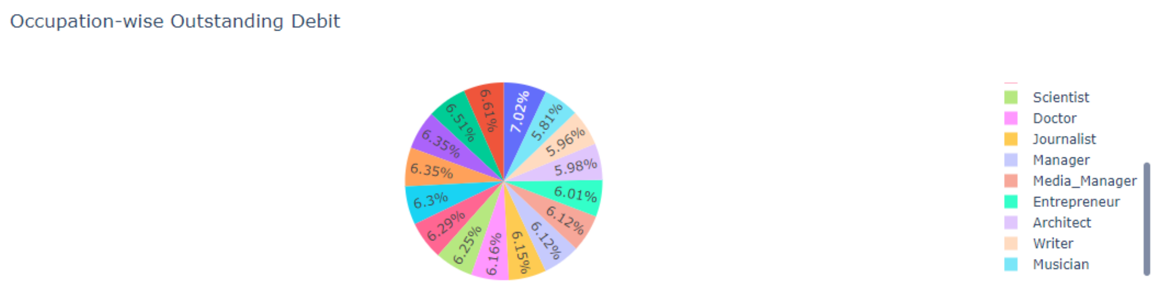

Figure 4.10: Pie plot of Occupation-wise Outstanding Debt

Source: (Jupyter Notebook)

The month-wise data evaluation provides the details of the sum of the total number of bank accounts of the customer in various months.

```
month_wise_evaluation = credit_score.groupby('Month')['Num_Bank_Accounts'].sum().reset_index()
month_wise_evaluation.head()
```

	Month	Num_Bank_Accounts
0	December	210190
1	November	205070
2	October	211499
3	September	215154

Figure 4.11: Month-wise data evaluation

Source: (Jupyter Notebook)

The plot detailed the number of bank sections in various months.

```
figure2 = px.bar(month_wise_evaluation, x='Month', y='Num_Bank_Accounts', title='Number of Bank Accounts with respect to Month')
figure2.show()
```

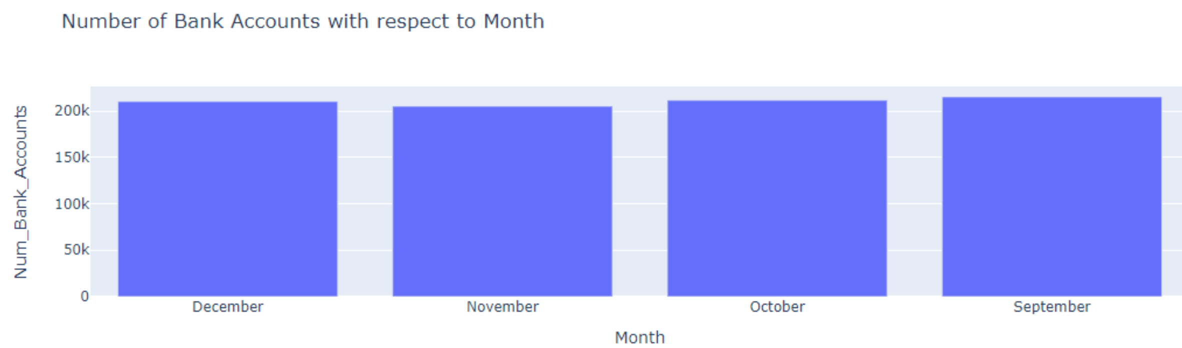

Figure 4.12: Bar plot of Month-wise Number of Bank Accounts

Source: (Jupyter Notebook)

The evaluation of the data with respect to the credit mix is evaluated in this section. This provides the execution of the monthly balance with respect to different credit mixes. As per the evaluated data, there are three major types of credit mixes such as bad, standard, and good. The count value of the monthly balances with respect to the credit mixes is evaluated in this analytical evaluation process. The structure of the data model is defined in this portion of the evaluation which evaluates the detailed value of the data. The necessary parameter such as group by is used to evaluate the monthly balance data with the help of credit mix.

```
credit_mix_wise_evaluation = credit_score.groupby('Credit_Mix')['Monthly_Balance'].count().reset_index()
credit_mix_wise_evaluation.head(3)
```

	Credit_Mix	Monthly_Balance
0	Bad	9556
1	Good	12260
2	Standard	18379

Figure 4.13: Credit Mix-wise Monthly Balance

Source: (Jupyter Notebook)

Doughnut plot provides the information and plotting of the credit mix-wise monthly balance. This plot provides information about the evaluation of the various types of credit mixes. According to the evaluated data, the maximum percentage of the monthly balance is found in the portion of the standard credit mix section. The provided information also represents the minimum percentage of the monthly balance data which is ‘Bad’ credit mix type data. This evaluation also provides the construction of the visualization that is used to implement the ‘plotly’ library that is used in data evaluation.

```
figure3 = px.pie(credit_mix_wise_evaluation, values='Monthly_Balance', names='Credit_Mix', hole=0.6,
             title='Credit Mix-wise Monthly Balance')
figure3.show()
```

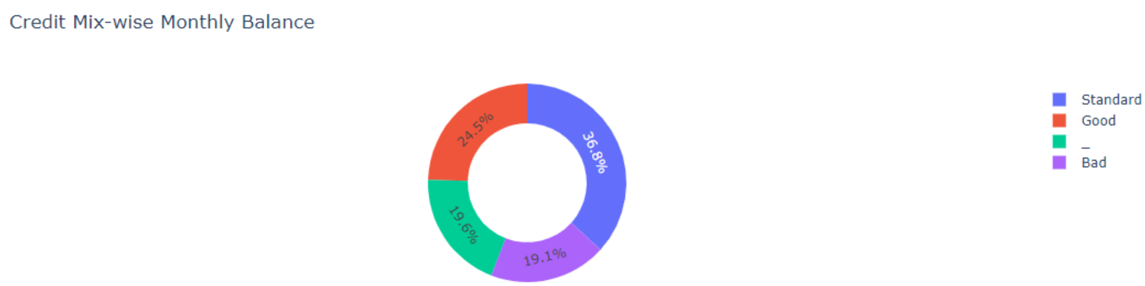

Figure 4.14: Doughnut plot of Credit Mix-wise Monthly Balance

Source: (Jupyter Notebook)

The drop section of the column provides the functionality of the elimination of the unused columns which are not used in the credit evaluation process. There are some unused columns such as ID, customer ID, month, name, age, SSN, occupation, loan type, credit mix, age-wise credit history, and payment behaviour. The 'drop' method is utilised to drop the unused columns of the dataset. The sectional portion provides information that supports the elimination of unused columns that are not implemented in the main analytical frame.

```
credit_score = credit_score.drop(columns=['ID'])
credit_score = credit_score.drop(columns=['Customer_ID'])
credit_score = credit_score.drop(columns=['Month'])
credit_score = credit_score.drop(columns=['Name'])
credit_score = credit_score.drop(columns=['Age'])
credit_score = credit_score.drop(columns=['SSN'])
credit_score = credit_score.drop(columns=['Occupation'])
credit_score = credit_score.drop(columns=['Type_of_Loan'])
credit_score = credit_score.drop(columns=['Credit_Mix'])
credit_score = credit_score.drop(columns=['Credit_History_Age'])
credit_score = credit_score.drop(columns=['Payment_Behaviour'])
```

Figure 4.15: Drop Column Functionality

Source: (Jupyter Notebook)

The below given image showcases the final structure of the credit score framework. The credit score framework depicts 8 fields and their respective values. The 8 fields under this framework are annual income, monthly_inhand_salary, num_bank_accounts, num_credit_card, interest_rate, num_of_loan, delay_from_due_date and number_of_delayed_payment. Annual income of id 4 is 34847.84, similarly monthly inhand salary of id 4 is 3037.086667. Num credit card for the same id is 4 and interest rate of the same id 4 is 6. For field 1, the number of delayed payments is 9.0 and this is higher than any other values of this field. Overall, the final structure represents valuation of different fields visible in the credit score section.

```
credit_score.head()
```

	Annual_Income	Monthly_Inhand_Salary	Num_Bank_Accounts	Num_Credit_Card	Interest_Rate	Num_of_Loan	Delay_from_due_date	Num_of_Delayed_Payment	Change
0	19114.12	1824.843333	3	4	3	4.0	3	7.0	
1	19114.12	1824.843333	3	4	3	4.0	3	9.0	
2	19114.12	1824.843333	3	4	3	4.0	-1	4.0	
3	19114.12	0.000000	3	4	3	4.0	4	5.0	
4	34847.84	3037.986667	2	4	6	1.0	3	1.0	

Figure 4.16: Final structure

Source: (Jupyter Notebook)

The below image depicts final information of the framework credit score. Pandas' library is used to create framework credit scores. This framework can take almost 50k entries and the index of this framework is 0 to 49.9k. The final information represents 16 columns in tabular format. Range and datatype of each column level is visible in the final information section with the condition non-null for all. Three different data types are shown in the 16-column field. The number three different data types are float64(11), int64(4) and object (1). The memory usage of the entire final information credit score framework is 6.1+ MB. Payment of min amount field only holds object type data in this section.

```
credit_score.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 16 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Annual_Income             50000 non-null  float64
 1   Monthly_Inhand_Salary     50000 non-null  float64
 2   Num_Bank_Accounts         50000 non-null  int64  
 3   Num_Credit_Card           50000 non-null  int64  
 4   Interest_Rate             50000 non-null  int64  
 5   Num_of_Loan               50000 non-null  float64
 6   Delay_from_due_date       50000 non-null  int64  
 7   Num_of_Delayed_Payment    50000 non-null  float64
 8   Changed_Credit_Limit      50000 non-null  float64
 9   Num_Credit_Inquiries      50000 non-null  float64
 10  Outstanding_Debt          50000 non-null  float64
 11  Credit_Utilization_Ratio  50000 non-null  float64
 12  Payment_of_Min_Amount     50000 non-null  object 
 13  Total_EMI_per_month       50000 non-null  float64
 14  Amount_invested_monthly   50000 non-null  float64
 15  Monthly_Balance           50000 non-null  float64
dtypes: float64(11), int64(4), object(1)
memory usage: 6.1+ MB
```

Figure 4.17: Final information

Source: (Jupyter Notebook)

The below picture represents the final checking of null concerning the framework credit score. Isnull and sum are two methods that are used for checking null values. These two methods display the existence of a null value within almost every field. The 16-column level has a null value and the Jupyter notebook is used as the environment. None keyword is the most standard form to inspect for a null value. It corresponds if the variable is equivalent to None. If it is, the keyword will return a true value otherwise, it will return a false value.

```
credit_score.isnull().sum()
```

```
Annual_Income               0
Monthly_Inhand_Salary       0
Num_Bank_Accounts           0
Num_Credit_Card             0
Interest_Rate               0
Num_of_Loan                 0
Delay_from_due_date         0
Num_of_Delayed_Payment      0
Changed_Credit_Limit        0
Num_Credit_Inquiries        0
Outstanding_Debt            0
Credit_Utilization_Ratio    0
Payment_of_Min_Amount       0
Total_EMI_per_month         0
Amount_invested_monthly     0
Monthly_Balance             0
dtype: int64
```

Figure 4.18: Final checking of null

Source: (Jupyter Notebook)

The below image depicts data processing and checking of the credit score framework. Payment of min amount is passed under the framework Credit_score. Replace method is used with this aspect while passing parameters. Info method is used with the framework credit_score and panda library is used to build the data frame. Range index of this section is 50k and 16 column level is visible in the data processing and checking section. Data processing and checking shows two different data types such as float64 and int64. The total number of level index float64 is 11 and total number of int64 is 5. The memory usage of this data processing and checking is 6.1 MB.

```
credit_score['Payment_of_Min_Amount'] = credit_score['Payment_of_Min_Amount'].replace({'NM': 0, 'No': 0, 'Yes': 1})
```

```
credit_score.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 16 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Annual_Income             50000 non-null  float64
 1   Monthly_Inhand_Salary     50000 non-null  float64
 2   Num_Bank_Accounts         50000 non-null  int64  
 3   Num_Credit_Card           50000 non-null  int64  
 4   Interest_Rate             50000 non-null  int64  
 5   Num_of_Loan               50000 non-null  float64
 6   Delay_from_due_date       50000 non-null  int64  
 7   Num_of_Delayed_Payment    50000 non-null  float64
 8   Changed_Credit_Limit      50000 non-null  float64
 9   Num_Credit_Inquiries      50000 non-null  float64
 10  Outstanding_Debt          50000 non-null  float64
 11  Credit_Utilization_Ratio  50000 non-null  float64
 12  Payment_of_Min_Amount     50000 non-null  int64  
 13  Total_EMI_per_month       50000 non-null  float64
 14  Amount_invested_monthly   50000 non-null  float64
 15  Monthly_Balance           50000 non-null  float64
dtypes: float64(11), int64(5)
memory usage: 6.1 MB
```

Figure 4.19: Data processing and checking

Source: (Jupyter Notebook)

The below picture illustrates the data setting and splitting process from the credit score framework. Payment of min amount is passed through the credit score framework from the x and y axis. Drop method is used with the framework while setting the axis value 1 for the X axis. Scikit-learn (Sklearn) is the most powerful machine-learning library in Python *(Yang et al., 2021)*. It utilizes a Python consistency interface to deliver a set of efficient instruments for machine learning and statistical modelling like regression, classification, clustering, and dimensionality speculation. Various modules are imported to set and split data. Test and train data from both axis display the passing parameters of the test train split function are test size and random state.

```
x = credit_score.drop(['Payment_of_Min_Amount'], axis=1)
y = credit_score['Payment_of_Min_Amount']
```

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
```

```
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

Figure 4.20: Data setting and splitting.

Source: (Jupyter Notebook)

The below figure signifies use of KNN with the machine learning library sklearn. KNN is imported from sklearn with neighbours method. The valuation of K is 4 and the n-neighbors value of the KNN model is k. Fit method is used with the KNN model to pass the parameter of x train and y train data. As the valuation of k is 4, n_neighbors value of KNN is 4 as well. A KNN classifier is a machine learning algorithm utilized for regression and classification problems *(Boateng et al., 2020).* It operates by locating the K nearest points in the training dataset and utilizes their class to forecast the class or value of a new data point.

```
from sklearn.neighbors import KNeighborsClassifier
```

```
k = 4
knn_model = KNeighborsClassifier(n_neighbors=k)
```

```
knn_model.fit(X_train, y_train)
```

```
▾       KNeighborsClassifier
KNeighborsClassifier(n_neighbors=4)
```

Figure 4.21: KNN use.

Source: (Jupyter Notebook)

The below image depicts the prediction and accuracy level of the KNN model used in this scenario for prediction. The accuracy score of the KNN model is 0.74. The precision of the k-nearest neighbour (KNN) algorithm is moderately inferior to other classification algorithms.

```
knn_y_pred = knn_model.predict(X_test)
```

```
accuracy_knn = accuracy_score(y_test, knn_y_pred)
print(f"Accuracy of KNN Model: {accuracy_knn:.2f}")
```

```
Accuracy of KNN Model: 0.74
```

Figure 4.22: Predict and accuracy of KNN.

Source: (Jupyter Notebook)

The below image depicts a KNN report regarding the information of various useful factors. Report of KNN equals classification report as well parameters to print the report of KNN. The precision value of the KNN model is 0.70 for the id 0 whereas the precision value is 0.80 for the id 1. The recall score for the 0 and 1 id is 0.82 and 0.67. The F1 score of the KNN model for the 0 and 1 id is 0.76 and 0.73. The support value of the KNN for 0 and 1 id is 4848 and 5152. The accuracy score of the KNN model is 0.74.

```
report_knn = classification_report(y_test, knn_y_pred)
print(report_knn)
```

```
              precision    recall  f1-score   support

           0       0.70      0.82      0.76      4848
           1       0.80      0.67      0.73      5152

    accuracy                           0.74     10000
   macro avg       0.75      0.74      0.74     10000
weighted avg       0.75      0.74      0.74     10000
```

Figure 4.23: Report of KNN

Source: (Jupyter Notebook)

The below image represents utilization of the decision tree algorithm in this process. Decision tree classifier has been imported from sklearn tree to use in this scenario. Dec tree variable is used to store the classifier on the other hand it uses fit method to pass x and y train data to obtain output of decision tree classifier model. A decision tree classifier algorithm is a machine learning algorithm that utilizes a tree-like model to construct forecasts *(Bansal et al., 2022)*. It operates by recursively diverging data into subsets based on the most substantial attribute at each node of the tree.

```
from sklearn.tree import DecisionTreeClassifier
```

```
dec_tree=DecisionTreeClassifier()
dec_tree.fit(X_train, y_train)
```

```
▾ DecisionTreeClassifier
DecisionTreeClassifier()
```

Figure 4.24: Decision Tree use

Source: (Jupyter Notebook)

The below image displays the prediction and accuracy score of the machine algorithm classifier decision tree. Dt_y_pred variable stores dec_tree with predict method and holds x_test parameter. The accuracy of the decision tree model is 0.81. The accuracy_score function holds two parameters such as y_test and dt_y_pred. The accuracy_dt variable holds this earlier value. This statement prints the accuracy of decision tree model while showing frequency of accuracy_dt value 2f.

```
dt_y_pred = dec_tree.predict(X_test)
```

```
accuracy_dt = accuracy_score(y_test, dt_y_pred)
print(f"Accuracy of Decision Tree Model: {accuracy_dt:.2f}")
```

```
Accuracy of Decision Tree Model: 0.81
```

Figure 4.25: Predict and accuracy of Decision Tree

Source: (Jupyter Notebook)

The below image displays the report of the decision tree classifier. The report_dt variable stores the value of classification_report. The passing parameters of the classification report are y_test and dt_y_pred. The above picture shows the print of the variable and output of useful factors. Useful factors of this decision tree classifier are precision, recall, f1 score, support and accuracy *(Tangirala et al., 2020)*. For 0 and 1 level id, the valuation of precision for the decision tree is 0.80 and 0.82. The recall score for 0 and 1 id of the classifier is 0.82 and 0.81. The f1-score value of the decision tree for level 0 and 1 is 0.81 and 0.82 and the support value is 4848 and 5152. The accuracy score of the decision tree classifier model is 0.81.

```
report_dt = classification_report(y_test, dt_y_pred)
print(report_dt)
```

```
              precision    recall  f1-score   support

           0       0.80      0.82      0.81      4848
           1       0.82      0.81      0.82      5152

    accuracy                           0.81     10000
   macro avg       0.81      0.81      0.81     10000
weighted avg       0.81      0.81      0.81     10000
```

Figure 4.26: Report of Decision Tree

Source: (Jupyter Notebook)

The below figure indicates utilization of logistic regression in this scenario. From the sklearn linear model, the logistic regression machine learning model is imported. Log_reg variable store logistics regression model while adding fit method as well as x and y train parameters with the variable. This is executed to obtain the result of logistic regression. Logistic regression is a supervised machine learning algorithm that achieves binary classification undertakings by envisioning the possibility of a consequence, possibility, or statement *(Nusinovici et al., 2020)*. The model produces a binary or dichotomous result defined as two probable consequences: yes/no, 0/1, or true/false.

```
from sklearn.linear_model import LogisticRegression
```

```
log_reg=LogisticRegression()
log_reg.fit(X_train, y_train)
```

```
▾ LogisticRegression
LogisticRegression()
```

Figure 4.27: Logistic Regression use

Source: (Jupyter Notebook)

The below image illustrates prediction and accuracy of logistic regression. The lr_y_pred variable stores log_reg with a predicts method. The passing parameter of this value is X_test and passing parameters of accuracy_score is y_test and lr_y_pred. The accuracy_lr variable stores the accuracy_score value. This prints the accuracy of the logistic regression model while showing frequency for accuracy_lr is 2f. The accuracy of the logistics regression model is 0.72.

```
lr_y_pred = log_reg.predict(X_test)
```

```
accuracy_lr = accuracy_score(y_test, lr_y_pred)
print(f"Accuracy of Logistic Regression Model: {accuracy_lr:.2f}")
```

```
Accuracy of Logistic Regression Model: 0.72
```

Figure 4.28: Predict and accuracy of Logistic Regression

Source: (Jupyter Notebook)

The below figure illustrates a report of the logistics regression model. The report_lr variable star classification report to print the value of the variable. Output of the logistic regression model is categorized into five segments such as recall, precision, f1 score, support, and accuracy. The precision value of the logistic regression model under the 0 and 1 id is 0.69 and 0.75. The recall value of 0 and 1 id is 0.75 and 0.69. The f1 score of the model for 0 and 1 id is 0.72 and 0.71. The accuracy of the logistic regression model is 0.72.

```
report_lr = classification_report(y_test, lr_y_pred)
print(report_lr)
```

```
              precision    recall  f1-score   support

           0       0.69      0.75      0.72      4848
           1       0.75      0.69      0.71      5152

    accuracy                           0.72     10000
   macro avg       0.72      0.72      0.72     10000
weighted avg       0.72      0.72      0.72     10000
```

Figure 4.29: Report of Logistic Regression

Source: (Jupyter Notebook)

4.2 Discussion

The evaluation portion provides the information data analysis technique. In this case, two data analysis techniques are used such as thematic analysis, and prediction model construction. The prediction model constriction method is used to predict the credit score for different customer types. The prediction model construction defines the involvement of the Python coding that is used to evaluate the prediction of the analytical functional areas of the section. The provided information defines the involvement of ML methods that are used to construct the prediction models such as Logistic Regression, KNN, Decision Tree, and so on. The models provide the predictive value for the evaluation process. The model also provides the information regarding the target variable which is used for the credit score prediction. In this case, the target variable is set to payment of minimum amount. This target variable provides the information about 0 and 1. Here 0 defines the ‘No’, and 1 defines ‘Yes’ option which is used to evaluate the data model.

The thematic evaluation provides information regarding multiple themes of the research that support the objectives of the research process. This provides the section of the evaluation that focuses on the main aim of the research. The evaluation of the themes also provides the functional areas of the data set which is used to evaluate the research process *(Anand et al., 2022)*. The constructive approach provides the defining parameters of the research such as credit score, data validation, and so on. All those factors are involved in evaluating the data objectives of the research process. The overview of the process supplies an announcement about the valuable elements of the research process that support the credit score evaluation process. It derives the necessary outcomes that are the key elements of the research factors and supports the objectives and the aim of the project.

4.3 Summary

Evaluations broken down by month and occupation are included in the analysis. The total amount of outstanding debts is computed after grouping occupations. The total number of bank accounts for each month is added together in a month-wise analysis, which offers insightful information on banking patterns. The dataset, which displays the total of monthly balances for various credit categories, is examined according to credit mix. Pie and doughnut plots are among the visualizations used to provide a clear image of how monthly balances and outstanding debts are distributed among various categories.

The investigation shows how credit-related data is handled in a reliable and organized manner. The data is reliable thanks to the cleaning and processing stages, and the descriptive and visual analytics offer insightful information on customer behaviour, delinquent accounts, and financial patterns. The analysis gains a predictive component with the use of KNN and Decision Tree machine learning models. In the financial sphere, these models can be useful instruments for risk assessment and credit grading. The data is easier to grasp when visualizations like pie plots, bar plots, and doughnut plots are used. The use of visualizations is essential for making difficult patterns and trends easier to understand.

CONCLUDING REMARKS

The credit scoring model has been found to be a mathematical model which is beneficial for the purpose of evaluating probability along with supporting customers with credit events. In this context, the current study has focused on analysing the significance associated with the credit scoring model to identify similar factors associated with the characteristics of loans and illustrate the benefits of effective consumer behaviour for economic cycles. For the purpose of development in business, the model can be helpful for determining investment as a trustworthy process along with identification of the length of the payment process in the business. The study has focused on demonstrating the application of an effective scoring model to help organisations in the process of lending loans as well as payments through online mode.

Linking with objectives

Linking with Objective 1: "To identify the concept of the credit scoring model".

The credit scoring model has been found to be a statistical model for analysis which is helpful for the estimation of probability in the lending process. Moreover, it has been identified that the credit scoring model is helpful in increasing transparency and fairness in financial transactions in which the devices are properly implemented, and companies gain profits *(Hohnen et al., 2021)*. Machine learning-based scoring models are increasingly popular in the current market. This has been found to be peer-to-peer lending that is recommended for business prediction regarding the risk and challenges for the purpose of taking loans. Network-based scoring model has been found to be helpful in analysing the existing risk factors within companies to manage their procedures of transaction and emphasise the capacity to estimate transaction decisions in social data.

Linking with Objective 2: "To evaluate the benefits of the credit scoring model for predicting the worthiness of individuals or businesses".

The incorporation of a credit scoring model in business operations has been found to be helpful for controlling the structure associated with risk allocation and cost functions in financial organisations, improving the trust of customers to manage their financial goals. In addition, in the context of individuals credit scoring model has been found to be beneficial and to allow them to access their personal loans according to their requirements from the market. It has been identified that machine learning tools are utilised for the purpose of estimating the automation process in transactions through a micro-credit scoring function to improve businesses. Behavioural and psychological traits of boulevards have been found to be helpful to minimise the complexity and issues during the process of credit scoring for individuals *(Pflügner et al., 2021).* Considering the social process and economic activities or functional progress associated with business in properly implementing the model of green credit scoring is viable in the transactional process.

Linking with Objective 3: "To evaluate the possible challenges of implementing a credit scoring model".

Various challenges have been found to be associated with properly operating the models of credit scoring in financial organisations. This model has been found to be used in innovative methods to predict financial losses along with emphasising the management of data history. Several barriers in credit scoring practices have been found to be the current environment of business issues associated with instability *(Khan et al., 2022).* It has been identified that in context of industry function financial data transfer customer engagement increased cyber-attacks and other aspects have increased that are based on real time estimation of information factor which is critical for organisations. Moreover, it has also been found that preservation of privacy through prediction of accurate cyber-attack area and random Central learning development concepts of business have been adversely affecting the structure associated with credit scoring model. In addition, the payment transaction area of individuals failure in the payment process and lack of history management have increased the risk associated with online transactions at any point of time.

Linking with Objective 4: "To identify the role of the credit scoring model in improving the financial landscape of an individual or in a business".

Implementation of strategic functions as well as effective goals of business are crucial to minimise potential challenges affecting the sustainability associated with credit scoring model in the current structure of business. Moreover, improvement in the structure of green credit scoring and evolution in traditional procedures of credit scoring sustainably in Companies it has been font to be necessary to apply appropriate strategic functions. Studies indicate that FICO scores are increasingly used in the developed countries to develop the quality associated with credit scoring in business along with emphasising the security of borrowing loans according to the structural process in a better manner *(Hohnen et al., 2021).* In addition, integration of hybrid credit scoring models can also be helpful in regression of logistics data management structure of organisations along with designing neural network structure in data collection *(Yang et al., 2021).* Determination of borrowing zones has been found to be crucial to secure the failure issues associated with individuals during conduction of copayments procedures.

Future Research Implication

The current research is based on the credit scoring model by developing a simple credit scoring model to predict the creditworthiness of individuals or businesses. It has several potential directions that can be focused on the future research and practical implications. Moreover, it has been identified that it has the implication to focus on investigating the ways for the purpose of increasing predictive power and accuracy associated with the models of credit scoring *(Djeundje et al., 2021).* It also includes incorporation of additional variables, refinement in existing components of the model and identifying different algorithms in the model. Future researchers can also focus on addressing the challenge associated with interpretability associated with credit scoring by developing methods to establish complicated models that are more understandable and have increased transparency along with allowing borrowing regulators and lenders to involve the factors affecting decisions related to credit. Future research can also investigate dynamic scoring in credit models by adapting to changing market friends, financial Behaviour of people and economic conditions with evolving time.

BIBLIOGRAPHY

Alazab, M., RM, S.P., Parimala, M., Maddikunta, P.K.R., Gadekallu, T.R. and Pham, Q.V., (2021). Federated learning for cybersecurity: Concepts, challenges, and future directions. *IEEE Transactions on Industrial Informatics*, *18*(5), pp.3501-3509.

Ampountolas, A., Nyarko Nde, T. and Constantinescu, C., (2021). A machine learning approach for micro-credit scoring. *Risks*, *9*(3), p.50.

Anarfo, E.B. and Abor, J.Y. (2020). Financial regulation and financial inclusion in Sub-Saharan Africa: Does financial stability play a moderating role? *Research in International Business and Finance*, *51*, p.101070.

Bazarbash, M., (2019). *Fintech in financial inclusion: machine learning applications in assessing credit risk*. International Monetary Fund.

Berg, T., Burg, V., Gombović, A. and Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, *33*(7), pp.2845-2897.

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.

Dastile, X., Celik, T. and Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, p.106263.

Giudici, P., Hadji-Misheva, B. and Spelta, A., (2019). Network-based scoring models to improve credit risk management in peer-to-peer lending platforms. *Frontiers in artificial intelligence*, *2*, p.3.

Goel, A. and Rastogi, S., (2023). Understanding the impact of borrowers' behavioural and psychological traits on credit default: review and conceptual model. *Review of Behavioral Finance*, *15*(2), pp.205-223.

Hasan, M.M., Popp, J. and Oláh, J., (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, *7*(1), pp.1-17.

Kozodoi, N., Jacob, J. and Lessmann, S., (2022). Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research*, *297*(3), pp.1083-1094.

Mhlanga, D., (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International journal of financial studies*, *9*(3), p.39.

Munkhdalai, L., Lee, J.Y. and Ryu, K.H., (2019), July. A hybrid credit scoring model using neural networks and logistic regression. In *Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceedings of the 15th International Conference on IIH-MSP in conjunction with the 12th International Conference on FITAT, July 18-20, Jilin, China, Volume 1* (pp. 251-258). Singapore: Springer Singapore.

Nabeeh, N.A., Abdel-Basset, M. and Soliman, G., (2021). A model for evaluating green credit rating and its impact on sustainability performance. *Journal of Cleaner Production*, *280*, p.124299.

Naranjo, P., Saavedra, D. and Verdi, R.S., (2022). The pecking order and financing decisions: Evidence from changes to financial-reporting regulation. *Journal of Accounting, Auditing & Finance*, *37*(4), pp.727-750.

Niu, B., Ren, J. and Li, X. (2019). Credit scoring using machine learning by combing social network information: Evidence from peer-to-peer lending. *Information*, *10*(12), p.397.

Niu, B., Ren, J. and Li, X., (2019). Credit scoring using machine learning by combing social network information: Evidence from peer-to-peer lending. *Information*, *10*(12), p.397.

Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J. and Baesens, B., (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, *74*, pp.26-39.

Paleyes, A., Urma, R.G. and Lawrence, N.D., (2022). Challenges in deploying machine learning: a survey of case studies. *ACM Computing Surveys*, *55*(6), pp.1-29.

Rahman, M.T., (2019). Testing trade-off and pecking order theories of capital structure: evidence and arguments. *International Journal of Economics and Financial Issues*, *9*(5), p.63.

Shen, F., Wang, R. and Shen, Y., (2020). A cost-sensitive logistic regression credit scoring model based on multi-objective optimization approach. *Technological and Economic Development of Economy*, *26*(2), pp.405-429.

Shen, F., Zhao, X. and Kou, G. (2020). Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory. *Decision Support Systems*, *137*, p.113366.

Siskin, B., Schmidt, N. and Stephens, B., (2021). Algorithmic Credit Scoring and FICO's Role in Developing Accurate, Unbiased, and *Fair Credit Scoring Models, p.1-12*.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, pp.333-339.

Suryono, R.R., Purwandari, B. and Budi, I., (2019). Peer to peer (P2P) lending problems and potential solutions: A systematic literature review. *Procedia Computer Science*, *161*, pp.204-214.

Wang, H., Kou, G. and Peng, Y., (2021). Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending. *Journal of the Operational Research Society*, *72*(4), pp.923-934.

Yang, Q., (2023), November. The applicability and rationality of "credit score based on digital footprint". In *2023 International Conference on Finance, Trade and Business Management (FTBM 2023)* (pp. 552-560). Atlantis Press.

Zhang, W., Yang, D. and Zhang, S. (2021). A new hybrid ensemble model with voting-based outlier detection and balanced sampling for credit scoring. *Expert Systems with Applications*, *174*, p.114744.

Abkenar, S.B., Kashani, M.H., Mahdipour, E. And Jameii, S.M., 2021. Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. Telematics and informatics, 57, p.101517.

Brandt, P. And Timmermans, S., 2021. Abductive logic of inquiry for quantitative research in the digital age. Sociological Science, 8, pp.191-210.

Chen, S., Liu, X., Yan, J., Hu, G. And Shi, Y., 2021. Processes, benefits, and challenges for adoption of blockchain technologies in food supply chains: a thematic analysis. Information Systems and e-Business Management, 19, pp.909-935.

Dawadi, S., Shrestha, S. And Giri, R.A., 2021. Mixed-methods research: A discussion on its types, challenges, and criticisms. Journal of Practical Studies in Education, 2(2), pp.25-36.

Djeundje, V.B., Crook, J., Calabrese, R. And Hamid, M., 2021. Enhancing credit scoring with alternative data. Expert Systems with Applications, 163, p.113766.

Hohnen, P., Ulfstjerne, M.A. and Krabbe, M.S., 2021. Assessing creditworthiness in the age of big data: a comparative study of credit score systems in Denmark and the US. Journal of Extreme Anthropology, 5(1), pp.29-55.

Hohnen, P., Ulfstjerne, M.A. and Krabbe, M.S., 2021. Assessing creditworthiness in the age of big data: a comparative study of credit score systems in Denmark and the US. Journal of Extreme Anthropology, 5(1), pp.29-55.

Hunziker, S. And Blankenagel, M., 2021. Research Design in Business and Management. Research Design in Business and Management.

KANG, E. And Hwang, H.J., 2021. Ethical conducts in qualitative research methodology: Participant observation and interview process. Journal of Research and Publication Ethics, 2(2), pp.5-10.

Kelly, L.M. and Cordeiro, M., 2020. Three principles of pragmatism for research on organizational processes. Methodological innovations, 13(2), p.2059799120937242.

Khan, K.I., Mata, M.N., Martins, J., Nasir, A., Dantas, R.M., Correia, A.B. and Saghir, U.S., 2022. Impediments of green finance adoption system: Linking economy and environment. Impediments of green finance adoption system: Linking economy and environment, (2), pp.217-237.

Pandey, P. And Pandey, M.M., 2021. Research methodology tools and techniques. Bridge Center.

Pflügner, K., Maier, C., Mattke, J. And Weitzel, T., 2021. Personality profiles that put users at risk of perceiving technostress: A qualitative comparative analysis with the big five personality traits. Business & Information Systems Engineering, 63, pp.389-402.

Pillai, A.A. and Kaushal, U., 2020. Research Methodology-An Introduction To Literary Studies. Central Asian Journal of Literature, Philosophy and Culture, 1(1), pp.1-11.

Raschka, S., Patterson, J. And Nolet, C., 2020. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. Information, 11(4), p.193.

Tomaszewski, L.E., Zarestky, J. And Gonzalez, E., 2020. Planning qualitative research: Design and decision making for new researchers. International Journal of Qualitative Methods, 19, p.1609406920967174.

Yang, F., Qiao, Y., Huang, C., Wang, S. And Wang, X., 2021. An automatic credit scoring strategy (ACSS) using memetic evolutionary algorithm and neural architecture search. Applied Soft Computing, 113, p.107871.

Chao X, Kou G, Peng Y and Viedma EH. (2021). Large-scale group decision-making with non-cooperative behaviors and heterogeneous preferences: an application in financial inclusion. Eur J Oper Res, 288(1):271–293. Https://doi.org/10.1016/j.ejor.2020.05.047.

Chi G and Zhang Z. (2017). Multi criteria credit rating model for small enterprise using a nonparametric method. Sustainability (switzerland). Https://doi.org/10.3390/su9101834.

Cucinelli D, Di Battista ML, Marchese M, and Nieri L. (2018). Credit risk in European banks: the bright side of the internal ratings-based approach. J Bank Finance, 93:213–229. Https://doi.org/10.1016/j.jbank fin.2018.06.014.

Cummings JR and Durrani KJ (2016). Effect of the Basel Accord capital requirements on the loan-loss provisioning practices of Australian banks. J Bank Finance, 67(December 2003):23–36. Https://doi.org/10.1016/j.jbankfin.2016.02.009.

Dastile X, Celik T, and Potsane M. (2020). Statistical and machine learning models in credit scoring: a systematic literature survey. Appl Soft Comput J, 91:106263. Https://doi.org/10.1016/j.asoc.2020.106263.

De Lima Silva DF, Ferreira L, and de Almeida-Filho AT. (2020). A new preference disaggregation TOPSIS approach applied to sort corporate bonds based on financial statements and expert's assessment. Expert Syst Appl, 152:113369. Https://doi.org/10.1016/j.eswa.2020.113369.

Doumpos M and Figueira JR. (2019). A multicriteria outranking approach for modeling corporate credit ratings: an application of the ELECTRE TRI-NC method. Omega (United Kingdom) 82:166–180. Https://doi.org/10.1016/j.omega.2018.01.003.

Dias Duarte F, Matias Gama AP and Paulo Esperanca J. (2017). Collateral-based in SME lending: the role of business collateral and personal collateral in less-developed countries. Res Int Bus Financ, 39:406–422. Https://doi.org/10.1016/j.ribaf.2016.07.005.

Dhliwayo, C. (2015). Conference and Exhibition for the Banking, Finance and Insurance Sectors held at the Harare International Conference Centre, Harare, Zimbabwe. Retrieved from https://www.rbz.co.zw/documents/publications/speeches/remarks-by-dr-dhliwayo-at-the-banking-finance-and-insurance-expo-29-july-2015(1).pdf.

Gastelum Chavira DA, Leyva Lopez JC, Solano Noriega JJ, Ahumada Valenzuela O and Alvarez Carrillo PA. (2017). A credit ranking model for a parafinancial company based on the ELECTRE-III method and a multiobjective evolutionary algorithm. Appl Soft Comput J, 60:190–201. Https://doi.org/10.1016/j.asoc.2017.06.021.

Hirsch B, Nitzl C and Schoen M (2018). Interorganizational trust and agency costs in credit relationships between savings banks and smes. J Bank Finance, 97:37–50. Https:// doi. Org/ 10. 1016/j. Jbank fin. 2018. 09. 017.

Ijadimaghsoodi A, Rasoulipanah H, Martínez López L, Liao H and Zavadskas EK. (2020). Integrating interval-valued multi-gran-ular 2-tuple linguistic BWM-CODAS approach with target-based attributes: site selection for a construction project. Comput Ind Eng, 139(November 2019):106147. Https://doi.org/10.1016/j.cie.2019.106147.

Ignatius J, Hatami-Marbini A, Rahman A, Dhamotharan L, and Khoshnevis P. (2018). A fuzzy decision support system for credit scoring. Neural Comput Appl, 29(10):921–937. Https://doi.org/10.1007/s00521- 016-2592-1.

Jackowicz K, and Kozłowski L. (2019). Social ties between SME managers and bank employees: financial consequences vs. SME managers' perceptions. Emerg Markets Rev. Https://doi.org/ 10.1016/j.ememar.2019.05.004.

Kahraman C, Onar SC, and Oztaysi B. (2015). Fuzzy multicriteria decision-making: a literature review. Int J computintellsyst, 8(4):637–666. Https://doi.org/10.1080/18756 891.2015. 1046325.

Kou G, Xu Y, Peng Y, Shen F, Chen Y, Chang K and Kou S. (2021). Bankruptcy prediction for smes using transactional data and two-stage multiobjective feature selection. Decis Support Syst, 140:113429. Https://doi.org/10. 1016/j.dss.2020.113429.

Shi B, Zhao X, Wu B and Dong Y. (2019). Credit rating and microfinance lending decisions based on loss given default (LGD). Financ Res Lett, 30(March):124–129. Https://doi.org/10.1016/j.frl.2019.03.033.

Zhang W, He H and Zhang S. (2019). A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: an application in credit scoring. Expert Syst Appl, 121:221–232. Https://doi.org/10.1016/j.eswa.2018.12.020.

Yang, F., Wang, X., Ma, H. And Li, J., 2021. Transformers-sklearn: a toolkit for medical language understanding with transformer-based models. BMC Medical Informatics and Decision Making, 21(2), pp.1-8.

Boateng, E.Y., Otoo, J. And Abaye, D.A., 2020. Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. Journal of Data Analysis and Information Processing, 8(4), pp.341-357.

Bansal, M., Goyal, A. And Choudhary, A., 2022. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short-term memory algorithms in machine learning. Decision Analytics Journal, 3, p.100071.

Tangirala, S., 2020. Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. International Journal of Advanced Computer Science and Applications, 11(2), pp.612-619.

Nusinovici, S., Tham, Y.C., Yan, M.Y.C., Ting, D.S.W., Li, J., Sabanayagam, C., Wong, T.Y. and Cheng, C.Y., 2020. Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of clinical epidemiology, 122, pp.56-69.

Ampountolas, A., Nyarko Nde, T., Date, P. And Constantinescu, C., 2021. A machine learning approach for micro-credit scoring. Risks, 9(3), p.50.

Ndayisenga, T., 2021. Bank loan approval prediction using machine learning techniques (Doctoral dissertation).

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.

Dastile, X. And Celik, T., 2021. Making deep learning-based predictions for credit scoring explainable. IEEE Access, 9, pp.50426-50440.

Bücker, M., Szepannek, G., Gosiewska, A. And Biecek, P., 2020. Transparency, auditability and explainability of machine learning models in credit scoring. Arxiv preprint arxiv:2009.13384.

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.

Li, Y. And Chen, W., 2020. A comparative performance assessment of ensemble learning for credit scoring. Mathematics, 8(10), p.1756.

Varun Kumar, K.S., Vijaya Kumar, V.G., Vijay Shankar, A. And Pratibha, K., 2020. Credit card fraud detection using machine learning algorithms. International Journal of Engineering Research & Technology (IJERT), 9(07), pp.5-8.

Xu, J., Lu, Z. And Xie, Y., 2021. Loan default prediction of Chinese P2P market: a machine learning methodology. Scientific Reports, 11(1), p.18759.

Anand, M., Velu, A. And Whig, P., 2022. Prediction of loan behaviour with machine learning models for secure banking. Journal of Computer Science and Engineering (JCSE), 3(1), pp.1-13.

APPENDIX (if necessary)