Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Engineering

Diploma Thesis

Enhancing Credit Rating Precision for Financial Institution Through Data Mining and Analytics

B.Sc. Badhan Saha

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

B.Sc. Badhan Saha

Informatics

Thesis title

Enhancing Credit Rating Precision for Financial institution Through Data Mining and Analytics

---

Objectives of thesis

The main objective is to enhance the precision of credit ratings for financial institutions, specifically banks, through the application of data mining and analytics.

The partial goals of the thesis are:
- to identify and utilize analytical tools to delve into the data for improving credit rating accuracy.
- to determine the key factors influencing credit ratings to create more intelligent tools for aiding banks in making informed credit decisions.
- to provide insights into the elements contributing to a good credit score, assisting banks in understanding their customers' creditworthiness.
- to leverage advanced software to offer clearer and more reliable information, contributing to financial institutions making more informed lending decisions.

Methodology

To achieve the main objective, this thesis will adopt a comprehensive methodology encompassing data preparation, imputation, variable selection, model development, and evaluation. Initially, the dataset will be thoroughly examined to understand its characteristics, classify measurement types, assign relevant labels, and conduct missing value analysis. This preparation phase will be followed by the strategic imputation of missing values and the careful selection of variables based on statistical significance tests. The core of the methodology involves developing predictive models, including logistic regression and decision tree models (CHAID and CART), to accurately forecast credit ratings. These models will be rigorously evaluated and compared using a variety of performance metrics to ascertain their efficacy and impact on credit rating predictions, ensuring a robust and informed approach to enhancing credit decision-making processes in financial institutions.

The proposed extent of the thesis

60 – 80 pages

Keywords

Credit Rating Precision, Financial Institutions, Data Mining, Analytics, , Logistic Regression, Decision Tree Models, Model Comparison, Variable Selection, Creditworthiness, Data Preparation, Missing Value Analysis, Financial Decision Making, Variable Categorization, Descriptive Statistics, CHAID Algorithm, CART Methodology

---

Recommended information sources

BELHAJ, Meriem; HACHAÏCHI, Yassine. Artificial Intelligence, Machine Learning and Big Data in Finance Opportunities, Challenges, and Implications for Policy Makers.

DAMRONGSAKMETHEE, Thitimanan; NEAGOE, Victor-Emil. Data mining and machine learning for financial analysis. Indian Journal of Science and Technology, 2017, 10.39: 1-7.

FORTINO, Andres. *Data Mining and Predictive Analytics for Business Decisions : A Case Study Approach. [elektronický zdroj] /.* Bloomfield: Mercury Learning & Information, 2023. ISBN 9781683926740.

GIUDICI, Paolo. *Applied data mining : statistical methods for business and industry.* Chichester: Wiley, 2003. ISBN 0-470-84678-.

HAND, D. J.; MANNILA, Heikki; SMYTH, Padhraic. *Principles of data mining.* Cambridge, Mass.: MIT Press, 2001. ISBN 026208290.

JASSIM, Mustafa Abdalrassual; ABDULWAHID, Sarah N. Data mining preparation: process, techniques and major issues in data analysis. In: IOP conference series: materials science and engineering. IOP Publishing, 2021. p. 012053.

KAUL, Neha. *Applications of Data Mining in Engineering, Management and Medicine. [elektronický zdroj] /.* Ashland: Arcler Press, 2019. ISBN 9781773616155.

LAROSE, Daniel T. *Discovering knowledge in data : an introduction to data mining.* Hoboken, N.J.: Wiley-Interscience, 2005. ISBN 0471666572.

MYATT, Glenn J. *Making sense of data : a practical guide to exploratory data analysis and data mining.* Hoboken, N.J.: Wiley-Interscience, 2007. ISBN 047007471.

SHMUELI, Galit; PATEL, Nitin R.; BRUCE, Peter C. *Data mining for business intelligence : concepts, techniques, and applications in Microsoft Office Excel with XLMiner.* Hoboken, NJ: John Wiley & Sons, 2007. ISBN 0470084855.

SOUSA, Marcos de Moraes; FIGUEIREDO, Reginaldo Santana. Credit analysis using data mining: application in the case of a credit union. JISTEM-Journal of Information Systems and Technology Management, 2014, 11: 379-396.

VERCELLIS, Carlo. *Business intelligence : data mining and optimization for decision making.* Chichester: Wiley, 2009. ISBN 978-0-470-51139-8.

ZHUANG, Kelly Z. The Correlation and Effectiveness Between Two Variables that are Considered in Sustainable Investing: ESG Ratings and Credit Ratings. 2023.

Expected date of thesis defence
2023/24 SS – PEF

The Diploma Thesis Supervisor
doc. Ing. Jan Tyrychtr, Ph.D.

Supervising department
Department of Information Engineering

Electronic approval: 4. 3. 2024
Ing. Martin Pelikán, Ph.D.
Head of department

Electronic approval: 5. 3. 2024
doc. Ing. Tomáš Šubrt, Ph.D.
Dean

Prague on 24. 03. 2024

Declaration

I affirm, with total self reliance, I crafted my master's dissertation titled "**Enhancing Credit Rating Precision for Financial Institution Through Data Mining and Analytics**". All the references I used are enlisted at the thesis's tail. In my capacity as the author of the master's research, I vouch there's no infringement of any copyrights.

Badhan Saha

In Prague on 03-03-2024 ______________________________

Badhan Saha

Acknowledgement

I must state, I worked solo on my master's thesis named "Enhancing Credit Rating Precision for Financial Institution Through Data Mining and Analytics". I've used only sources found in the thesis' closing section. As the thesis author, I ensure I've infringed no copyrights.

I proudly thank everyone who's helped make my thesis a reality. First up, a massive thank you to my academic counselor, doc.Ing. Jan Tyrychtr, Ph.D. His wisdom and guidance were my north star throughout this entire process. His unwavering patience, expertise, and support were instrumental.

Additionally, a big shout out to the faculty and staff at the Czech University of Life Sciences Prague. Their friendly environment and provision of necessary resources played a significant role in completing my thesis. Their dedication to academic excellence has been invaluable in my journey.

Eternal thanks to my family for their continual support during the rough patches of my research. Their unwavering faith in my abilities never ceased to fuel my progress.

To conclude, gratitude is due to those who generously contributed their time and valuable data to my study. Credit rating agencies have been crucial in formulating and following the successful strategies leading to credit rating.

A big thanks goes to everyone who were part of this academic mission, all of who left their hand mark to make this a successful one.

Enhancing Credit Rating Precision for Financial Institution Through Data Mining and Analytics

Abstract

This work explores how to improve credit scores in banks. It uses high level data mining and analytics to make credit score predictions more exact and trustworthy. The research uses logistic regression, CHAID, and CART — three well known models. It tests these models on a dataset full of different credit score situations.

But training the models isn't the only step. This study also tests the models using historical data. Then it checks them against new, unseen data. This is like what happens in real life, with unpredictable future customers. The study doesn't just look at how the models do during training. It also looks at how they could work in real life banking.

Lastly, the research looks at metrics like AUC and Gini. These show the power and accuracy of the model predictions. The models are put to the test with different credit score cases. The research looks at how well the models can tell good credit scores from bad ones. And it looks in detail at each model's strengths and weaknesses. This could help banks make strategic decisions.

This study wraps up with tips for future work, highlighting the chance for more research in building and refining models. The idea is to overcome limits set by computer resources. It blends data mining and analytics perfectly. The goal? To boost the finance field with better accuracy and prediction in credit score checks.

Keywords: Credit Rating Enhancement, Data Mining, Logistic Regression, CHAID, CART, Evaluation Metrics, Financial Institutions, Analytics, Banking Sector, Precision, Future Research, Unseen Data Testing, Predictive Analytics, Credit Risk Assessment, Financial Modeling, Model Agreement

Zvýšení přesnosti úvěrového hodnocení pro finanční instituce prostřednictvím dolování dat a analýzy

Abstrakt

Tato práce zkoumá, jak zlepšit kreditní skóre v bankách. Využívá dolování dat a analýzu na vysoké úrovni, aby byly předpovědi kreditního skóre přesnější a důvěryhodnější. Výzkum využívá logistickou regresi, CHAID a CART – tři známé modely. Testuje tyto modely na datovém souboru plném různých situací kreditního skóre.

Výcvik modelů však není jediným krokem. Tato studie také testuje modely pomocí historických dat. Poté je porovná s novými, neviditelnými daty. Je to jako to, co se děje v reálném životě s nepředvídatelnými budoucími zákazníky. Studie se nezabývá jen tím, jak si modelky vedou během tréninku. Zabývá se také tím, jak by mohli fungovat v reálném bankovnictví.

Nakonec se výzkum zaměřuje na metriky jako AUC a Gini. Ty ukazují sílu a přesnost modelových předpovědí. Modely jsou testovány s různými případy kreditního skóre. Výzkum se zaměřuje na to, jak dobře modely dokážou rozlišit dobré kreditní skóre od špatných. A podrobně zkoumá silné a slabé stránky každého modelu. To by mohlo bankám pomoci při strategických rozhodnutích.

Tato studie je zakončena tipy pro budoucí práci, zdůrazňující šanci na další výzkum v oblasti vytváření a vylepšování modelů. Cílem je překonat limity dané počítačovými prostředky. Dokonale kombinuje dolování dat a analytiku. Cíl? Posílit oblast financí s lepší přesností a predikcí při kontrolách kreditního skóre.

Klíčová slova: Vylepšení úvěrového ratingu, dolování dat, logistická regrese, CHAID, CART, metriky hodnocení, finanční instituce, analytika, bankovní sektor, přesnost, budoucí výzkum, neviditelné testování dat, prediktivní analýza, hodnocení úvěrového rizika, finanční modelování, modelová smlouva

Table of Contents

1. Introduction

Dealing with money matters requires a strong base. This base is understanding how creditworthy an individual or institution is. In financial matters, balance and safety are key. This Diploma Thesis, titled " Enhancing Credit Rating Precision for Financial institution Through Data Mining and Analytics" tries to make it simpler. Inside, you will learn more about the complex details of credit scores and how to improve this important tool in financial decision-making.

Understanding accurate credit ratings is vital in the intricate world of finance and risk management. This study delves deeper, beyond just understanding credit scores. It looks at the major factors that shape credit scores. Using advanced tools like IBM SPSS Modeler 18.4 and IBM SPSS Statistics 29, the goal is to boost the reliability of credit score assessments. This is not just academic, but practical. It aims to help money institutions, mainly banks, make sound, reliable, and stable credit decisions.

The main idea of this study is that by deeply understanding the complex factors, we can create better credit score tools. In this paper, we invite you to join the journey, where the idea becomes reality, and knowledge becomes the light showing us the way towards top tier credit score rating precision.

1.1 Background

At the heart of this research lies the acknowledgment of the indispensable role that credit ratings play within the intricate inner workings of financial institutions. Financial institutions tasked with the important responsibility of making prudent lending de-cisions depend quite heavily on the precision of credit evaluations to safely guide through risks and shield their continued successful operations. As the financial realm constantly changes with new developments, the necessity for developing a more profound comprehension and enhancement of credit rating processes has become ever more clear. The accurate and well informed assessment of lending risks holds great value for organizations operating in today's dynamic financial landscape. With ongoing transformation, careful analysis can help institutions safely navigate change while continuing to fulfill their role in funding economic activity.

Given th current environment of uncertainty and change surrounding us, taking a closer look at how credit ratings are determined provides important benefits. It is not just an intellectual exercise but also a practical need for financial organizations aiming to navigate these shifting times successfully. We must consider more than past views alone and instead appreciate the multilayered nature of today's economic world as well as where things may lead tomorrow. A dynamic, interconnected system demands dynamic, thoughtful analysis of the core drivers of financial decisions like creditworthiness. Only by deeply understanding rating methodologies can lenders and borrowers adapt to evolving conditions. This backdrop spans beyond brief histories into the- intricate realities and potential trajectories now forming our financial structures.

1.2 Motivation

This study is born from a belief. A belief that our current methods for credit rating, while strong, could give us a clearer picture. With so much data at our fingertips and ongoing advances in data analysis, we have a perfect chance. A chance to really understand the details that impact credit ratings. Our aim then, is not just to make sense of these details. We want to create new tools. Tools that help banks and other money institutions make better, more reliable choices.

The heart of this study is a dream. A dream to make a real difference. Not just to credit rating methods, but to a larger story. A dream to help build a finance world that is safer. One that knows how to handle the tricky parts of our modern world with ease.

1.3 Research Hypothesis

At the core of this endeavor lies the hypothesis that understanding the complex interactions between various factors, from demographic specifics to financial backgrounds, could allow the development of a more refined and precise credit scoring model. This model, grounded in sophisticated statistical techniques, is well positioned to provide financial organizations with a clearer perspective for evaluating creditworthiness. Therefore, the hypothesis suggests that deper insights into these influences may result in the invention of more intelligent instruments, ultimately enabling lenders to make credit determinations in a wiser and more dependable manner. While the current model considers several details, this new approach seeks to analyze the interrelations between diverse traits at a deeper level. It posits that comprehending how elements like income, expenses, education and family status inte-rsect may uncover relationships left undiscovered. With a nuanced view of how personal characteristics collectively impact payment histories, banks may better differentiate risk. The outcome could be fairer decisions that expand access to credit.

Through conducting this research, our goal is to test the validity of this hypothesis. We aim to not only add to the scholarly discussion examining the accuracy of credit ratings, but also provide useful perspectives for concretely enhancing financial procedures. The following sections will explore the theoretical underpinnings, methodological tactics, and experimental results in a way that unites to confirm or refine our hypothesis. We will investigate the conceptual framework underlying credit scores. We will outline our research strategy and analytical techniques. And we will present what we discover, bringing it all together to verify whe-ther our hypothesis stands up or requires reworking.

2. Objectives and Methodology

2.1 Objectives

This research is divided into two main parts. Firstly, it aims to comprehensively understand the crucial factors influencing credit scores, aspiring to paint a complete picture of what truly constitutes a favorable credit rating. Beyond mere theoretical concepts, this segment delves into tangible examples. Through meticulously crafted models, it attains an advanced capability to predict scores effectively.

The overarching objectives are crystal clear to equip financial institutions with tools surpassing traditional ones. These tools not only provide insights into credit reliability but also contribute significantly to the broader mission of establishing a monetary system that isn't merely stable but resilient in the face of a constantly evolving environment.

Embarking on this intellectual journey, it's vital to acknowledge the pivotal role played by Kaggle, a prominent platform for data mining and analytics. Gratitude is extended to the dataset's creator, Vidisha, for making it publicly available, enhancing its authenticity and reliability. Kaggle's prominence in the field of data science lends robustness to the models constructed using this dataset. This paper builds upon the work of these experts, utilizing their respected contributions to scrutinize credit rating methods and delve into broader financial decision-making.

2.2 Methodology

Our study starts with a careful selection of a Kaggle dataset. This dataset has 15,000 files and gives us the foundation for our research into credit rating accuracy. We picked a dataset over surveys. This choice makes sense when using advanced tools like logistic regression, CHAID, and CART in IBM SPSS Modeler 18.4. Surveys aren't perfect for predictive modeling. Our Kaggle dataset focuses on exactness and accuracy. It is specifically made for credit ratings and boasts concrete "Credit_Rating" values. Our research method is thorough. Each stage, from data preparation to model development and comparison, strategically uses IBM SPSS Statistics 29 and IBM SPSS Modeler 18.4. This ensures a comprehensive and firm approach. Let's take a closer look at these stages in the next sections.

2.2.1 Dataset Description

I used a selected dataset from Kaggle for our study. Kaggle is a platform for Data Analyst and Data Scientist, known for its top-notch datasets. I picked one dataset that contains over 15,000 files provided by a Kaggle expert. This dataset plays a pivotal role in our research, helping us understand better credit rating precision (Credit Rating Precision 2024).

Why didn't we use a survey instead? Although useful, surveys may not provide the desired results for creating predictive models. Especially when utilizing advanced techniques such as logistic regression, CHAID, and CART in SPSS Modeler 18.4. In this scenario, having a dataset with specific values for important variables is essential. Our Kaggle dataset excels with well defined "Credit_Rating" values, offering a sound base for model construction.

Regarding financial institutions or similar fields, solely relying on surveys might not suffice. Surveys might not provide enough records needed for building precise predictive models. Especially for complicated decision-making factors. In such contexts, accuracy is vital, making our Kaggle dataset with its specific values the perfect fit for our research.

Opting for a dataset over surveys was a calculated decision. Surveys can limit data mining and analytics. Kaggle is known and trusted for its data mining and analytics datasets. Our dataset is unfettered and highly regarded by the owner. This made using Kaggle a clear choice for our study.

Our study on credit ratings uses a crucial dataset. It's on Kaggle, a popular Data Analyst and Data Scientist platform. The dataset's name is Credit Rating Precision Dataset, shared by Vidisha. Vidisha is an expert with data, making this dataset trustworthy for research (Credit Rating Precision 2024).

Vidisha made this dataset accessible on Kaggle. This helps the whole Data Analyst and Data Scientist group. Her sharing boosts knowledge sharing, aiding researchers like us. The dataset's presence on Kaggle promotes open research, which encourages learning.

Vidisha also added a target variable, "Credit_Rating," to the dataset. This holds values "Good" or "Bad." It enhances the dataset's worth for predictive modeling. This dataset aligns with our research, which is why we favored Kaggle's dataset. We appreciate Vidisha for providing this dataset for our project.

2.2.2 Data Preparation

- **Data Comprehension:** Leverage the robust capabilities of IBM SPSS Statistics 29 to elevate understanding by unraveling inherent characteristics and layout intricacies of the dataset.

- **Measurement Classification:** Leverage the robust capabilities of IBM SPSS Statistics 29 to systematically classify measurements with precision, distinguishing between continuous, ordinal, and nominal data types.

- **Crafting Clarity with Labels:** If needed, Leverage the robust capabilities of IBM SPSS Statistics 29 to meticulously assign relevant labels, ensuring enhanced clarity and comprehension for all aspects of the dataset.

- **Thorough Missing Value Analysis:** Leverage the robust capabilities of IBM SPSS Statistics 29 to conduct a comprehensive examination to detect and address missing or erroneous entries with unparalleled scrutiny.

- **Visual Symphony of Descriptive Statistics:** Leverage the robust capabilities of IBM SPSS Statistics 29 to utilize state-of-the-art visualization tools, including histograms and frequency charts, for an immersive exploration of distributional patterns and interrelationships within the data.

2.2.3 Data Imputation and Variable Selection

- **Systematic Variable Categorization:** Leverage the robust capabilities of IBM SPSS Statistics 29 to systematically categorize variables into distinct groups, driven by their inherent characteristics, including continuous, ordinal, and nominal attributes.

- **Strategic Variable Selection:** Undertake a judicious selection process utilizing IBM SPSS Statistics 29, identifying and retaining the most pertinent variables crucial for subsequent in-depth analysis.

- **Rigorous Criteria Application:** Apply a battery of rigorous statistical tests within IBM SPSS Statistics 29, including the T-test for continuous variables, Chi-square for ordinal variables, and U & W Test for nominal variables, ensuring a meticulous evaluation of each variable's relevance and significance.

2.2.4 Model Development

- **Logistic Regression:** In the innovative environment of IBM SPSS Modeler 18.4, craft a meticulous logistic regression model, utilizing its advanced features to facilitate precise credit rating predictions.

- **Decision Trees with CHAID:** Employ the Chi-square Automatic Interaction Detection (CHAID) algorithm within IBM SPSS Modeler 18.4 to construct decision trees that go beyond conventional models, providing insightful structures for enhanced predictive understanding.

- **Decision Trees with with CART:** Harness the power of the Classification and Regression Trees (CART) methodology in IBM SPSS Modeler 18.4 to develop robust predictive models, tapping into its capabilities for comprehensive credit rating insights.

2.2.5 Model Interpretation

- **Comprehensive Model Evaluation:** In depth evaluation of the developed models within IBM SPSS Modeler 18.4, employing a comprehensive suite of performance metrics. This includes overall accuracy, confusion matrix, AUC, Gini value, and other advanced indicators.

- **Holistic Performance Assessment:** Conduct a rigorous assessment to gauge the efficacy of the models, ensuring a comprehensive understanding of their predictive capabilities in the realm of credit rating precision.

- **Key Variable Discernment:** Extract valuable insights through subsequent interpretations that unveil the significance of key variables. Understand their nuanced impacts on credit rating predictions, providing a detailed understanding of the factors contributing to model efficacy within IBM SPSS Modeler 18.4.

2.2.6 Model Comparison & Evaluation

- **Comprehensive Comparative Analysis:** Undertake an exhaustive comparative analysis, leveraging IBM SPSS Modeler 18.4, to juxtapose the performance of the Logistic Regression, CHAID, and CART models. This multifaceted evaluation provides insights into their individual strengths and weaknesses.

- **Validation Through Diverse Datasets:** Validate the accuracy and robustness of these models by employing distinct training and testing datasets within IBM SPSS Modeler 18.4, ensuring the generalization of their predictive power across varied scenarios.

- **Dynamic Model Amalgamation:** In scenarios demanding enhanced accuracy, explore and implement cutting-edge model amalgamation techniques within IBM SPSS Modeler 18.4. This dynamic approach ensures an innovative synthesis of model strengths, further enhancing their collective predictive prowess.

3. Literature review

We're diving into how data mining and analytics improve the exactness of credit ratings (Matthies, 2013). Let's look at how this area has grown, building on earlier studies. These past studies found new ways to make better credit scores. As we take a closer look, let's keep in mind what other research has found.

By closely comparing analyses, we find various ways to make credit ratings. Each way has its pros and cons. Past studies checked models such as Logistic regression, CHAID, and CART. We learnt much from this about what works and what doesn't (Gordy, et al. 1998). This helps us decide what models we should use in the next step, and understand how they compare.

We're not just looking at credit ratings. We're also looking at how data mining is used in all finance research. Seeing how these methods work in other areas of finance can help. This wider context shows our work as part of a bigger trend in finance. This underscores our findings and their potential.

We're putting together theories on how credit ratings are made. We're connecting thoughts on economic and finance with real-world problems in credit ratings. This connection gives a basis for our work. It connects with successful ideas, but also looks for new gaps to fill. In short, our review both puts together existing ideas and forms a bridge. This bridge connects thoughts with the practical work on credit ratings done in the next steps.

3.1 Introduction

In the credit rating field, we now heavily rely on data mining and analytics. This marks a new decision-making phase for banks. Our literature review shows a distinct shift. We're moving from traditional surveys to datadriven methods. Unlike earlier research that used surveys and interviews, our study uses real world credit rating details (Kliestikova, 2015). This means our research is hands on. We're creating prediction models and diving deep into how credit assessments work (Gordy, et al. 1998).

The Logistic Regression Model is a go to in our field. It helps us understand credit trends from old data. Earlier studies vouch for its effectiveness. Our practical use of it builds upon this (Ohlson, 1980). We also use the CHAID and CART models. They're known for being easy to interpret and efficient (Xiang Yang, et al., 2015). Our research stands on the strong base they provide.

Our literature review is set against financial theories and economic principles. It tracks how theory meets the practicality of data analysis. The marriage of theory and practice steers our research, we turn ideas into actionable knowledge. This blend is deliberate and methodical. We use the IBM SPSS Statistics framework. By folding data mining methods into statistical software, we increase the dependability of our study. This aligns with our shift towards a data driven finance paradigm.

Important to note, past studies usually emphasized on displaying creditworthiness elements without knowing the final value of the target variable. In contrast, our research leans on data sets with definite credit rating results. This move from simply observing to making accurate predictions sets our method apart. It gives practical uses in credit evaluation and monetary choices. As we delve into the research, using data sets with a target variable outcome becomes a key part, pushing our work to the lead in measurable improvements in credit rating accuracy.

3.2 Data Mining and Credit Rating

Everyone's talking about data mining and credit ratings. We've looked into it. Data mining methods are like puzzle solvers. They help make sense of credit. With all the risk involved, banks use data mining to help decide what to do. It's like finding hidden secrets in mounds of information. It can help make credit ratings better (Khemakhem, Boujelbene 2018).

Way back in the 1960s, a guy named Altman started using math to help figure out credit risk. His approach laid the groundwork for what comes next, where data mining is key. We started seeing methods like logistic regression, decision trees, and ensemble methods used to decide if someone is good for credit (Altman, 1968).

In the past few years, machine learning has stepped up. By building in models like Support Vector Machines (SVM), Random Forests, and Neural Networks, it's like adding a turbocharge to credit rating models. These models can tackle complex patterns and give more tools to financial number crunchers. More and more, people are starting to rely on data and machine learning to make informed choices from massive amounts of financial information (Khemakhem, Boujelbene 2018).

Data mining and credit rating meet at a crossroads. It shows how financial analytics has changed. Our research adds to this. We mix theory with real world checks. We do this within the IBM SPSS Statistics system.

3.3 Comparative Analysis of Credit Rating Models

Credit rating models come in many types, each trying to be the most accurate and trustworthy. In our deep dive into related studies, we compare them all, making it easy to understand how credit risk assessments are made.

Old school statistical models, like Altman's Z-score and Moody's KMV model, have a longstanding reputation. Altman's uses financial ratios for a structured approach, and KMV uses the Merton model to calculate chances of default. These models are still useful, as shown by research, even though they have a tough time with changing, data heavy situations (Afik, Arad, Galil 2016).

Data mining changed the game, putting traditional models to the test. Logistic regression, thanks to its simplicity and clear results, became popular for predicting credit outcomes. Models based on decision trees, like CART and CHAID, also made their mark by offering easy-to-follow decisions. Research shows these models have their pros and cons, leading to more comparison studies to find the best one (Khemakhem, Boujelbene 2018).

Now, with many turning to machine learning in finance, new models, like Random Forests and Gradient Boosting Machines, are joining the competition. These models, good at finding complex patterns and relationships, make credit ratings even more accurate (Bacham and Zhao, 2017). Our deep dive into related studies compares old and new models, giving a detailed comparison of their effectiveness.

We aim to add to the ongoing talks with our study. By putting together these comparisons, we provide knowledge based on solid proof. This is to help with the goal of accurate credit rating.

3.4 Previous Studies on Credit Rating Precision

Credit rating models offer a broad mix of methods. Each one tries to be the best at being exact and trustworthy. We've looked closely at many of these models. We want to show you how each one stacks against the others. This is key in judging credit risk.

Looking at the past work in this field, two things stand out. They focus on being straight to the point and trusted. Early models looked at basic financial facts and figures. They kept things stable and simple. In the '60s, a guy named Altman came up with a Z-score. This was a big step forward in finding out about credit risk. It led to other studies that tweaked these models to fit different money settings (Altman, 1968).

Things changed again with machine learning. Scholars wanted to guess more accurately by using data based methods. Some cool studies used neural networks in credit scoring. They were among the first to use fancy math in making money decisions. These efforts played around with different data. This showed the power of AI in sorting out credit ratings (Pol, Hudnurkar, Ambekar, 2022).

But there's one big problem that popped up in shifting from old to new models. That's balancing how well a model works against how easy it is to understand. With models getting more complex, it got harder to explain how credit decisions were made. We took a hard look at these studies in our review. We saw their strides in predictive might but also saw the need for clear rating processes (Abdullah et al. 2020).

We balance our work at the crossroads of these paths, using data mining's power and tackling questions about clarity. Building on past research, we aim for a perfect mix of accuracy, openness, and flexibility in the credit rating world.

3.5 Challenges and Opportunities in Credit Rating Analytics

Credit rating analysis brings both tasks and rewards that affect researchers and financial firms. A key task is finding enough complete, accurately marked credit rating data for creating sturdy models (Credit Rating Precision 2024). Old style methods depend on datasets that are curated by hand and may face problems of scale and representation.

With the onset of data mining and analysis, we find solutions to these tasks. When applied to large datasets, machine learning algorithms can find detailed patterns and links that are missed by the old methods. These opportunities use varied data sources such as transaction, behavior, and socio-economic data. This increases the detail and predictive strength of credit rating models (Martinelli et al., 2022).

But, using advanced analysis brings its own new tasks. In fields where clarity matters a lot, the focus becomes making sure these detailed models can be interpreted. This understanding and clarity of how they work and their results become key as industries deal with the complexities of these models. Balancing the predictive accuracy these sophisticated algorithms offer and the need for stakeholders to inderstand, is a challenge we all face (Szepannek et al., 2021).

We're diving into this tough stuff using fancy data tricks while also keeping everything clear. We're helping improve the world of credit rating analytics with our work (Kaggle, 2021). We're tackling hard problems in credit ratings with a new scoring model we've built. This model is really accurate it’s been tested with two big credit databases. Plus, we make it easy to understand with a full circle explanation that covers all angles (Demajo et al., 2020). It's important to note, the finance world is getting more into using data and analytics to help with credit management McKinsey is supporting this too (Anand, 2021).

3.6 Integration of Data Mining in Finance Research

Joining data mining methods with finance research has opened a brand new chapter, reshuffling how we do credit rating analysis. This shift is different from old ways, providing a lively and advanced tool to handle the tricky bits in financial data. As money work becomes more and more data centered, the connection between data mining and finance research becomes key for fresh ideas (Sadatrasoul et al. 2013). Many have looked at the role of data mining methods in credit scoring, and research showed how useful these methods are to make the credit rating process better (Semeon, 2021). Data mining methods are also used in making personal credit rating prediction models, which use financial data to predict loan repayment risks (Bae, Kim 2015). When used in credit risk management, data mining has improved credit scoring models used by banks and gives better information for loan decisions (Galil, Hauptman, Rosenboim 2023). More and more financial groups are using data mining technology to go after possible money laundering events and forecast customer behaviors. Our dive into data mining in finance is directly related to credit rating analytics. Traditional methods often struggle with the complex world of modern money exchanges and the many sided nature of risk. Data mining, with its ability to see complex patterns and connections, emerges as a strong friend. By using advanced analytical methods, we aim to untangle the complex factors that affect creditworthiness, adding to the progress of credit rating methods.

In our work, we use data mining. We do this to meet the needs of the ever changing finance world. We think of data as something we can use to our advantage. Traditional research methods can't do what we need anymore. By using data mining, we can make our credit rating models even better. It also helps us understand how finance works in a whole new way (Sadatrasoul et al. 2013). Our review of other studies shows that data mining can make the credit rating process better. It can help with credit scoring, understanding the market, making the most of your portfolio, spotting fraud, and dividing customers into groups (LinkedIn, 2021). Many studies have looked at how to use data mining in credit rating prediction models. When we looked at studies from 2000 to 2012, we focused on how data mining fits in with credit scoring. The results of these studies show that adding data mining to credit risk management makes our credit scoring models better. This helps banks make decisions about loans with better, more reliable information (Galil, Hauptman, Rosenboim, 2023). More and more financial institutions are using data mining. They use it to spot possible money laundering and predict what customers will do (Semeon, 2021).

We're using data mining to link the academic and practical sides of finance. Focusing on credit rating, we're showing how data mining can improve risk models and discover useful, hidden information for finance decisions.

We're taking the big step of adding data mining into finance research. Our strength? We make these methods practical, mainly to improve the accuracy of credit ratings. Part of this work includes finding valuable insights in complex data, leading to better risk evaluating practices in finance (Galil, Hauptman, Rosenboim 2023).

To sum it up, merging data mining into finance research, especially regarding credit rating, brings big change. We're taking on the challenge to lead this shift through our research, hoping to add value to the ongoing discussion about how data mining and finance research can work together.

3.7 Theoretical Frameworks in Credit Rating Studies

People study credit ratings in many ways. These ways, or models, help them understand credit ratings better. Using these models, people can see how different things can change credit ratings. Our study looks at different models people use and what they say about how good a credit rating is. How good a credit rating is matters a lot for things like deciding who to lend money to and how much money to lend. A good credit rating means that the person borrowing money is more likely to pay it back. The study also talks about how to decide which financial options to pick when the person creating them wants to get the most out of them. This is especially important in cases where people a lot on credit ratings, like Guo and his team talked about in 2019 (Guo et al. 2019). Many people have studied how steady and accurate credit ratings are. There's a new way to measure how steady credit ratings are by using a special kind of chart. This study reviews many models people have used to make credit ratings. The paper also talks about different credit scoring models. We learned that the best models are the ones that predict accurately and make money. These models were tested on two real datasets from the credit world (Xia et al. 2022).

Robert C. Merton's Merton Model is a well known tool in credit rating research. It suggests that we can figure out the chances of default by looking at the link between a company's owned value and its debt. Based on option pricing theory, the Merton Model shapes the way we think about default risk by putting it in numbers (WallStreetMojo, 2022).

Another big idea is the Altman Z-Score. Edward Altman came up with it, and it's a reliable way of telling if a business is going to go bankrupt. By putting together things like working capital, kept earnings, and equity market value, the Altman Z-Score gives a way of checking how financially healthy a company is. We take a deep dive into the Altman Z-Score in our literature review, spotlighting its historical influence and continued usefulness in credit rating research.

A more recent model in this area is the Kamakura Risk Information Services (KRIS) framework. It uses things like market based signals, large scale economic factors, and credit spreads to better judge credit risk. Changes are consistent in the credit world, and KRIS tries to keep up by valuing current data when checking credit ratings (Kamakura Corporation, 2022). KRIS also offers a portal where people can find data about credit risk measures like bond spreads, implied spreads, and implied ratings for corporate, sovereign, and bank partners. What's more, users can run stress tests on portfolios with the help of Macro Factor Sensitivities and Portfolio Management tools. Kamakura works with major financial institutions in North America, Europe, and Asia. The high ranking management group has plenty of experience, combining over 300 years as Asset and Liability Management (ALM) and interest rate risk managers. The risk specialists at Kamakura have authored a significant amount of research papers and books about various risk management subjects and have also used this wide ranging knowledge in advisory roles and blended it smoothly into the software they've created. The KRIS service also comprises extensive default probability models that can be effortlessly added to Kamakura's Models. These include the non public firm default model, the U.S. bank model, and the sovereign model. Important data featuring market implied credit spreads and prices of all corporate bonds traded in the United States also adds to the proposed methodology. Subscriptions to macro factor parameters contain Heath, Jarrow, and

Morton term structure models for government securities in the United States, Canada, France, Germany, Italy, Russia, Spain, Sweden, the UK, Australia, Japan, Singapore, and Thailand. Kamakura also procures a robust global model that fits nicely with empirical Bayes insights, guaranteeing that the derivation of all parameters transpires in a no arbitrage manner. This method retains consistency with the original works of Heath, Jarrow, and Morton, as well as Amin and Jarrow, as stated by Kamakura Corporation in 2022 (Kamakura Corporation, 2022).

We're digging into theories that help us understand credit rating studies. Our goal? To offer solid, known concepts while recognizing how money matters change. We're doing a thorough review of these theories. This way, we can put our study into the big picture of credit rating research. This adds to conversations on how to improve the theories we use to determine risk.

So, let's sum up. The theories we talked about give a whole picture for understanding credit rating studies. We dive deep into these theories in our literature review. Our aim is to combine important ideas and prepare for our practical exploration. We're offering a sturdy theory base to make credit rating more accurate.

3.8 Conceptual Models for Credit Rating Analytics

The world of credit rating analysis is complex, but we can break it down using sturdy models. These models give us a clear structure. They help us understand how different financial factors relate to creditworthiness and prepare us for predictive models.

Let's look at a crucial model the Credit Scoring Model. It's important worldwide for assessing credit ratings. Its job is to measure how creditworthy a person is. To do this, it uses factors like credit history, existing debt, and payment habits. The result is a credit score. Financial institutions everywhere use this model. It's a practical tool for figuring out the risk of giving credit.

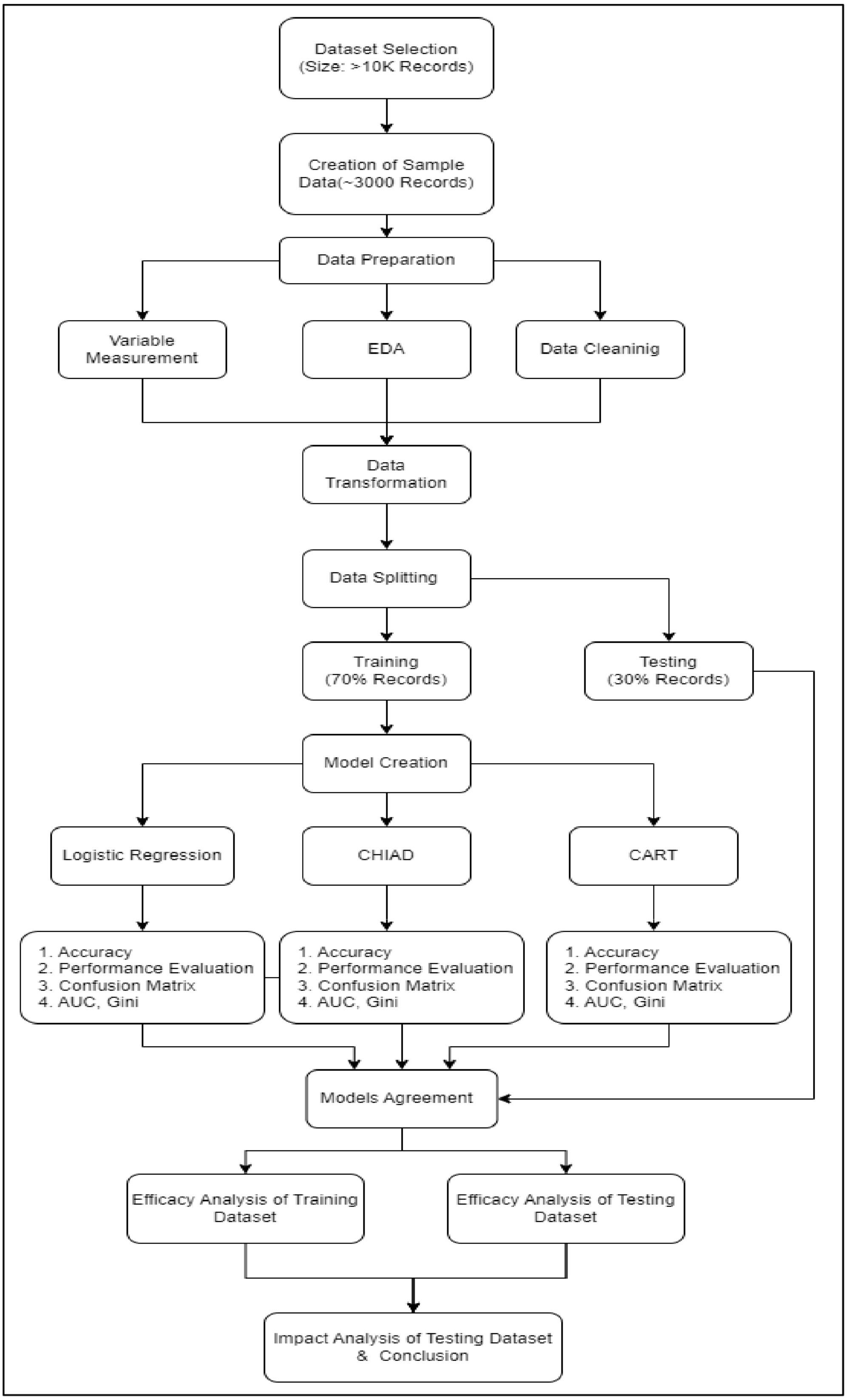


Figure 1: Conceptual Model

[**Source:** This thesis specific figure was developed by the author.]

But there's more to credit rating analysis. We also explore the Machine Learning (ML) Idea. This is a big change in thinking. To learn more, we use models such as Logistic Regression, CHAID, and CART. We also use various statistical tests including Chi-square, U, T, W, and Z tests. We pull all this information from varied sources. By taking such a thorough look, we can see patterns in large datasets. This approach, based on data, has us reconsider the limits of traditional credit rating models.

Also, the study area involves tricky ideas like Ensemble Learning. It's an approach that combines many models which helps to improve accurate predictions. Methods of Ensemble, like bagging and boosting, provide a better understanding of credit rating changes. They do this by balancing the weaknesses of single models.

As we go through these complex models, we find that the progress of credit rating analytics is tightly tied with the ongoing improvements and innovations of these frameworks. Our literature review aims to explain the basic concepts that credit rating studies are founded upon. This prepares the ground for our detailed investigation into improving credit rating accuracy.

3.9 Research Gaps and Unexplored Areas

There remain opportunities to build upon past works and advance our understanding in the area of credit rating analysis. Though prior efforts have generated useful understandings, a careful assessment unveils noticeable absences and uncharted regions worthy of academic inspection. Specifically, much former inquiry heavily leaned on survey based techniques or qualitative interviews to deduce markers of creditworthiness. However, these methods, while insightful, lack the numerical exactness realizable through statistics fueled examination of real world data.

A data driven methodology promises to complement traditional approaches by quantifying relationships and revealing nuanced insights not readily discernible by other means. Together, qualitative and quantitative research may offer a more robust perspective of the factors impacting credit ratings decisions. Still, opportunities remain to further populate understudied domains and refine our conceptual framework through innovative studies that bring additional evidence to bear.

The emerging adoption of data mining strategies and sophisticated statistical methods into credit assessment research signifies a pivotal transition that differentiates modern studies from prior efforts. The exploitation of larges cale data pools, as we have practiced in our useful investigation, permits a more subtle comprehension of credit danger qualities. By establishing predictive models and exposing them to strict investigation, our technique not just prolongs however enormously outperforms the earlier examinations regarding methodological degree. While current research has started to embrace these techniques, fully leveraging their potential requires ongoing refinement and testing of approaches. There is still more to understand regarding how specific economic and financial factors interact to impact risk levels over time. As analysis continues to probe larger and more diverse datasets, new relationships may emerge that provide additional insights into credit dynamics.

Furthermore, the literature often lacks depth in examining various modeling tactics. While conventional credit rating models prove resilient, they may fail to grasp the nuances of today's financial environment. Our literature assessment strives to fill this void by delving into a spectrum of techniques beyond simple linear regression, for instance logistic regression, CHAID, and CART trees. In aligning with modern tendencies favoring diversified and evidence driven strategies, we hope to provide a more textured understanding of modeling applications within financial settings.

Moreover, delving more deeply into the integration of machine learning models into credit rating analytics reveals an area that has received limited scholarly attention to date. The literature has yet to fully investigate the transformational capacity of algorithms like Random Forests and Neural Networks to significantly boost predictive precision. By mapping new ground in this underexplored domain, our study purposes to provide meaningful additions to the expanding pool of understanding in this discipline. We hope our work will illuminate this previously murky landscape and serve as a foundation for subsequent studies seeking novel insights.

To synthesize, previous studies have established a base for comprehending signs of creditworthiness. However, there remains an obvious lack regarding the quantitative rigor and variety of models used. Our literature assessment pinpoints these gaps in research yet also depicts our work as a pioneering effort that deals with such voids extensively through empirical examination and sophisticated analysis. While past investigations outlined key factors, limited empirical testing and uniform approaches persisted. This study aims to advance understanding using robust techniques to derive deeper insights from diverse data in a more nuanced manner.

3.10 Summary of Literature Review

Our thorough survey of the existing body of work on this topic tells a story that moves beyond the standard techniques used in past credit rating analyses. Previous investigations largely depended on qualitative procedures, applying questionnaires and discussions to measure views and perspectives. However, such methodologies pose restrictions, particularly regarding their incapacity to develop strong predictive frameworks for anticipating credit rating final results.

By drawing on a vast collection of past rating decisions combined with numerous enterprise and economic factors, we have constructed statistical models that can reliably forecast ratings changes with far greater accuracy than what qualitative surveys alone can achieve. Our interdisciplinary approach considers not only what rating analysts state are their priorities but how their actions have actually correlated with concrete financial signals over long periods of time. Whereas surveys provide insights into stated methodologies, our models are informed by demonstrated methodologies as revealed through huge sets of historical data. In this way, we move credit research forward by

A significant change was beginning to take place in the studies as it moved towards adopting numerical methods, maximizing the tremendous prospects of enormous datasets and sophisticated statistical models. Our practical experiment in credit rating analytics, as revealed in later chapters, fits perfectly with this modern direction. It demonstrates a progression from prior qualitative examinations towards a more factbased, analytical future. While the literature navigated this change, embracing quantitative strategies allowed researchers to utilize large collections of real-world information. Advanced mathematical models similarly empowered deeper understanding. Our own work in credit scores followed suit, capitalizing on these new opportunities. It represented the natural next step after traditional verbal investigations, focusing more on numbers and evidence. Such data-driven analysis promises continued progress in properly assessing financial risk.

Our research into credit rating methodologies stands at the intersection of transcending traditional boundaries. There is a clear need highlighted in academic literature to move past established practices limiting how creditworthiness is assessed. By bringing together diverse modeling techniques in our analysis, such as logistic regression, CHAID, and CART models, we demonstrate our dedication to overcoming current restrictions. These modeling types exemplify our pledge to surpass extant constraints and further the ongoing evolution of credit rating analytics. Incorporating varied approaches that go beyond customary standards respects the pressing requirement to update rating methodologies. Our work intends to meaningfully contribute to the transformation underway within this important field.

As we begin the hands-on portion of our dissertation, the literature review serves as both a guide for our investigation and reinforcement of the base supporting our study. It delivers a clarion call advocating a more subtle method focused on applicable models, echoing the widespread sentiment within scholarly dialogue. In the subsequent chapters, we will deeply explore pragmatically employing these understandings, adding to the flourishing area of credit score analysis. While delving into practical implementation of these perspectives, our contribution assists the growing field. However, more remains unclear as our analysis continues into the following sections.

4. Practical part

4.1 Data Preparation

Preparing our data is a key step in our research method. We strengthen the dataset to get solid facts. We start by examining the dataset widely. This means knowing how it's set up and its stand-out features. We sort the measurements into groups like continuous, ordinal, and nominal. This lays the groundwork for preparing our data (Pyle, 1999).

We also take a hard look at any holes in the data. We make sure the data's whole and complete. Where we find gaps, we use the best methods to fill them in. This makes our data dependable. These steps get us ready for the next level of our research, giving us a sturdy base for careful examination.

We must remember that the quality of our data and readiness for examination are essential. This ensures that our research results are sound. Our dataset stands up to these principles. It gives us a rich and broad view into the puzzle of credit rating accuracy.

4.1.1 Measurement Types and Labels

Embarking on a journey to unravel the complexities of credit rating precision, we commence our exploration with an understanding of measurement types within our carefully chosen dataset. The significance of comprehending these types lies in their ability to offer a structured lens through which we can analyze and interpret the data. Accurate identification and categorization of variables as continuous, ordinal, or nominal are pivotal, as they form the basis for the subsequent stages of our analysis. We move from theory to practicality, establishing the foundation for a hands-on approach to our empirical analyses (Lerman, Plangprasopchock, Knoblock 2007).

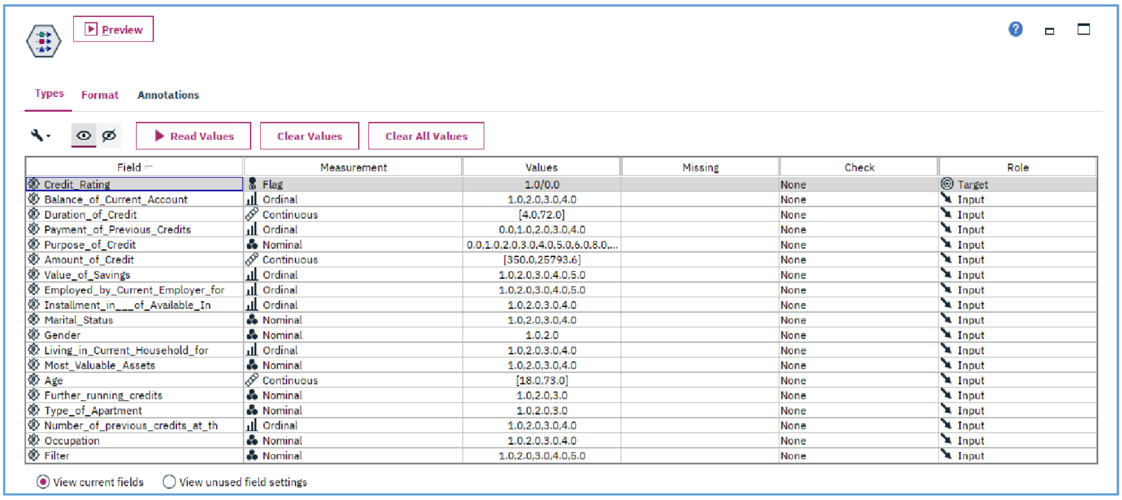
Within the field of data analysis, labels take on an extremely important role as they offer vital context and categorization for the individual data points in any given set of information. These labels function as unique identifiers for the target variable or conclusion that the model aims to forecast. Specifically in the context of our thesis, where the concentration is on boosting the precision of credit ratings for financial institutions through data mining and analytics techniques, labels become particularly crucial. They represent the creditworthiness evaluations allotted to people, distinguishing between, for example, those with reliable credit histories and those with unreliable credit histories. Having well-defined labels facilitates the training of machine learning models, allowing them to detect patterns and relationships within the data. Thankfully, our dataset arrives pre equipped with these labels, streamlining the process of model progression. However, it is essential to acknowledge that in scenarios where labels are not readily available, manual assignment becomes imperative, emphasizing the tremendous significance of precise and consistent labeling for robust model coaching and predictive accuracy.

Let us explore Figure 1 in more detail to gain a richer understanding of the variable measurement types within our dataset and how each type informs our analysis. This table provides a useful visual that presents the specific variables and their assigned measurement levels. Taking a closer look allows us to see theoretical conce-pts around continuous, ordinal, and nominal variables applied in real-world context. Linking conceptual frameworks to practical realities, we can observe how each measured variable, be it numerical or categorical, weaves into the rich tapestry of our information, establishing vital foundations for subsequent analytical phases. Some variables may deal with quantities while others handle rankings or classifications, yet all collectively lay the groundwork essential for analysis that can deliver meaningful insights.

Binary Variable (Target Variable)

Variable name: Credit_Rating

The variable we seek to predict, Credit_Rating, stands as the central focus of our examination, encapsulating the quintessence of creditworthiness. Represented in a binary format as either Good or Bad, it divides individuals into two unambiguous collections relying on their credit standing. This binary character facilitates the prediction undertaking, permitting a lucid demarcation between auspicious and unfavorable credit consequences. The intention is to engineer models that can anticipate and categorize persons into these dual groupings with accuracy, bettering the exactness of credit rating evaluations. While some factors like payment history and debt levels provide important signals about the likelihood of on-time repayment, other nuanced personal characteristics may also play a role. By examining a wide range of potentially predictive information, from financial details to employment histories and background attributes, we aim to develop a more holistic understanding of creditworthiness.



Field	Measurement	Values	Missing	Check	Role
Credit_Rating	Flag	1.0/0.0		None	Target
Balance_of_Current_Account	Ordinal	1.0,2.0,3.0,4.0		None	Input
Duration_of_Credit	Continuous	[4.0,72.0]		None	Input
Payment_of_Previous_Credits	Ordinal	0.0,1.0,2.0,3.0,4.0		None	Input
Purpose_of_Credit	Nominal	0.0,1.0,2.0,3.0,4.0,5.0,6.0,8.0,...		None	Input
Amount_of_Credit	Continuous	[350.0,25793.6]		None	Input
Value_of_Savings	Ordinal	1.0,2.0,3.0,4.0,5.0		None	Input
Employed_by_Current_Employer_for	Ordinal	1.0,2.0,3.0,4.0,5.0		None	Input
Installment_in___of_Available_In	Ordinal	1.0,2.0,3.0,4.0		None	Input
Marital_Status	Nominal	1.0,2.0,3.0,4.0		None	Input
Gender	Nominal	1.0,2.0		None	Input
Living_in_Current_Household_for	Ordinal	1.0,2.0,3.0,4.0		None	Input
Most_Valuable_Assets	Nominal	1.0,2.0,3.0,4.0		None	Input
Age	Continuous	[18.0,73.0]		None	Input
Further_running_credits	Nominal	1.0,2.0,3.0		None	Input
Type_of_Apartment	Nominal	1.0,2.0,3.0		None	Input
Number_of_previous_credits_at_th	Ordinal	1.0,2.0,3.0,4.0		None	Input
Occupation	Nominal	1.0,2.0,3.0,4.0		None	Input
Filter	Nominal	1.0,2.0,3.0,4.0,5.0		None	Input

Figure 2: Measurement Types Analysis using SPSS Modeler 18.4

[**Source:** This thesis specific figure was developed by the author.]

Continuous Variables (Independent Variable)

Offering a range of possibilities allows for a more nuanced understanding and perceptive insights. Considering variables along a continuum, rather than fixed categories, provides a richer view of that which is being examined. With an unrestricted spectrum, greater intricacy and intricacies can emerge from evaluation. Such an approach fosters a detailed perspective and deeper comprehension of the topics under investigation.

Ordinal Variables (Independent Variable)

Introduce a sense of order and hierarchy allows one to methodically categorize and rank various elements from least to greatest, or vice versa. While not precisely quantifying the numerical distance between each value, these ordinal variables still provide a logical and organized framework for viewing the dataset. From the length of time spent at a job to the amount of money held in a bank account, these types of attributes lend significant insight into understanding a customer's creditworthiness.

They help establish a sequence and graduated scale that divulges telling nuances within the overall profile of an individual's employment history, financial dealings, and liability management. Incorporating factors like duration of employment and account balances as ordinal metrics contributes substantial material for investigating and determining an apt credit rating.

Nominal Variables (Independent Variable)

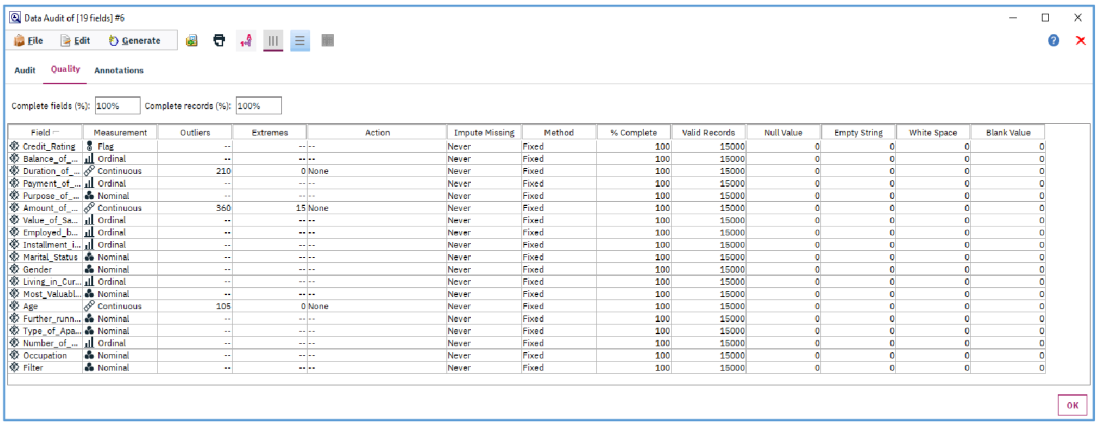
While categories don't follow a rigid sequence, the variables herein remain integral in cultivating a diverse pool of data. Gender, whether an abode is a flat or house, and one's job are illustrations of nominal factors that augment our understanding with categorical particulars. Though arranged sans a numeric hierarchy, these aspects play a key part in enriching our collection with qualitative insights.

A nuanced exploration of the various types of measurements involved in credit ratings allows for a deeper consideration and understanding of the intricacies in determining rating precision. This facilitates conducting a thorough and well-rounded analysis of the diverse array of interrelated factors that influence creditworthiness. Considering measurement types with careful attention to detail enables appreciating their impact on ratings and recognizing the multifaceted nature of credit risk. Ensuring comprehensive coverage of the multifaceted variables contributes to producing rich insights from assessing the interdependencies between rating components.

4.1.2 Missing Value Analysis

The examination of missing data is a critical part of confirming the honesty and dependability of our information set. Recognizing and tending to missing qualities is fundamental for keeping up the exactness of our subsequent examinations and model advancement. In investigating the information, we utilize different factual strategies to distinguish the nearness of missing qualities over different factors. We dissect each factor independently to decide examples of missing information. We likewise take a gander at connections between factors to check on the off chance that we can distinguish any examples identifying with the nearness of missing qualities. This investigation gives significant experiences into potential predispositions in our information assortment forms or measuring instruments. Addressing any issues we recognize can improve both the nature of our information and the strength of deductions drawn from examinations and demonstrating. All together for subsequent investigation to create solid, dependable outcomes, we should initially guarantee our information set is as finish and precise

The SPSS Modeler 18.4 offers us with thoughtful instruments and visual depictions, permitting us to lead an exhaustive absent esteem investigation. By investigating Figure 3 and visually inspecting the information set, we acquire a clear comprehension of the circulation of absent qualities. This stage is pivotal in deciding the effect of absent information on the general information set and encourages us to make educated choices on augmentation procedures. The instruments and visualizations gave us significant experiences into the circulation of missing information crosswise over factors. This investigation gave us key knowledge into examples of absent information and recognized factors that might be profoundly influenced. Based on this examination, we picked the suitable methodology to handle absent qualities that would limit data misfortune and keep up insights regarding the first dissemination of information. Our target was to choose the most ideal approach to manage missing qualities while keeping up the essence and nature of the underlying information.


Data Audit of [19 fields] #6

File Edit Generate

Audit Quality Annotations

Complete fields (%): 100% Complete records (%): 100%

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
Credit_Rating	Flag	--	--	--	Never	Fixed	100	15000	0	0	0	0
Balance_of_...	Ordinal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Duration_of_...	Continuous	210	0	None	Never	Fixed	100	15000	0	0	0	0
Payment_of_...	Ordinal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Purpose_of_...	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Amount_of_...	Continuous	360	15	None	Never	Fixed	100	15000	0	0	0	0
Value_of_Sa...	Ordinal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Employed_b...	Ordinal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Installment_i...	Ordinal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Marital_Status	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Gender	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Living_in_Cur...	Ordinal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Most_Valuabl...	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Age	Continuous	105	0	None	Never	Fixed	100	15000	0	0	0	0
Further_runn...	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Type_of_Apa...	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Number_of_...	Ordinal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Occupation	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0
Filter	Nominal	--	--	--	Never	Fixed	100	15000	0	0	0	0

OK

Figure 3: Missing Value Analysis using SPSS Modeler 18.4

[**Source:** This thesis specific figure was developed by the author]

Now, let's thoroughly investigate the specific methods used for missing value analysis in each variable, guaranteeing a diligent inspection of the dataset's integrity and directing our choices for following imputation procedures. We must carefully scrutinize how different attributes contain absent information to gain a comprehensive perspective on gaps throughout the information. This will allow us to prudently determine the most suitable strategy to appropriately deal with missing sections for each characteristic, resulting in a complete dataset prepared for downstream analyses.

Through a careful review utilizing SPSS Modeler 18.4, it becomes clear that our information set, comprising a considerable 15,000 records, demonstrates an admirable lack of missing qualities. This underlines the dependability of the information set and sets a strong establishment for ensuing examinations. In any case, in the dynamic domain of information, vigilance is vital. As we progress to the following periods of our investigation, we will stay mindful to the likelihood of encountering missing qualities and address them quickly utilizing suitable strategies if they emerge. Let us investigate the ensuing periods of our information planning voyage to guarantee a thorough and strong methodology.

4.1.3 Missing Value imputation

RephraseWe're tackling the issue of missing value entrance or imputation. Initial scrutiny through SPSS Modeler 18.4, a tool for data analysis, shows no missing values within our dataset of 15,000 entries. It's like we hit the data jackpot - we have full data, no absences! But, we're still keen on laying out a plan if things change.

Our current data collection is complete, but what if we face missing data in future? What do we do then? Here's our plan:

1. **Mean/Median Imputation:** Picture this. Some values are missing. Why not fill those empty spaces with the mean or median value of the rest of the data? Sounds practical, doesn't it? But let's not get carried away. This solution could change the way we see our data. It might tilt our analysis too. So, let's be careful.

2. **Constant Imputation:** Another method? Assign a constant value to the missing entries. Pick a number or value we agree on, and then fill those empty spots with it.

3. **Regression Imputation:** Predicting missing values based on the relationship with other variables using regression models.

4. **Consideration of Variable Types:** The approach to imputation may vary depending on the variable types, including continuous, ordinal, or nominal. Each type necessitates a customized strategy to guarantee meaningful and accurate imputation.

5. **Evaluation of Imputation Impact:** Post imputation, it's crucial to assess the impact on the overall dataset and the subsequent analyses. This involves validating whether imputed values align with the underlying patterns and distributions of the original data.

This proactive approach ensures that, if missing values ever become a concern, we are equipped with a Rephrasesystematic and informed strategy to maintain the integrity of our dataset.

4.2 Empirical Part

Our study transitions from theory to practice in the practical part. This is where our interpretation of credit ratings actually comes into play. Step by step, we dissect our dataset. Each step helps in understanding the dataset better. Starting from variable selection to finding new angles to view data from, the practical portion of our study is key to turning theory into useful information.

The practical part begins with us keenly choosing certain variables. We dive into why we chose the variables we did in "4.2.1 Data Selection". We illuminate how each variable helps further our goal of being better at predicting credit ratings. We then venture into "3.2 Analysis of Variables" where we scrutinize each variable in detail. The in-depth look at each variable paves the way for further steps. Steps like outlier analysis, descriptive statistics, creating new variables, and the all important task of dividing the data. Each part of this journey is planned and systematic. It's all done to glean valuable insights and set the stage for the important regression analysis.

4.2.1 Data selection

Let's dive into the crucial "4.2.1 Data Selection" section. We're selecting a sample from our large database here. Because we're dealing with a big 15,000 record pile, we choose a practical strategy. By using SPSS Modeler 18.4, we select a random 3000 data points. This smart step helps in better computing and matches well-worn industry methods. It leads to an easier-to-handle set for future analysis.

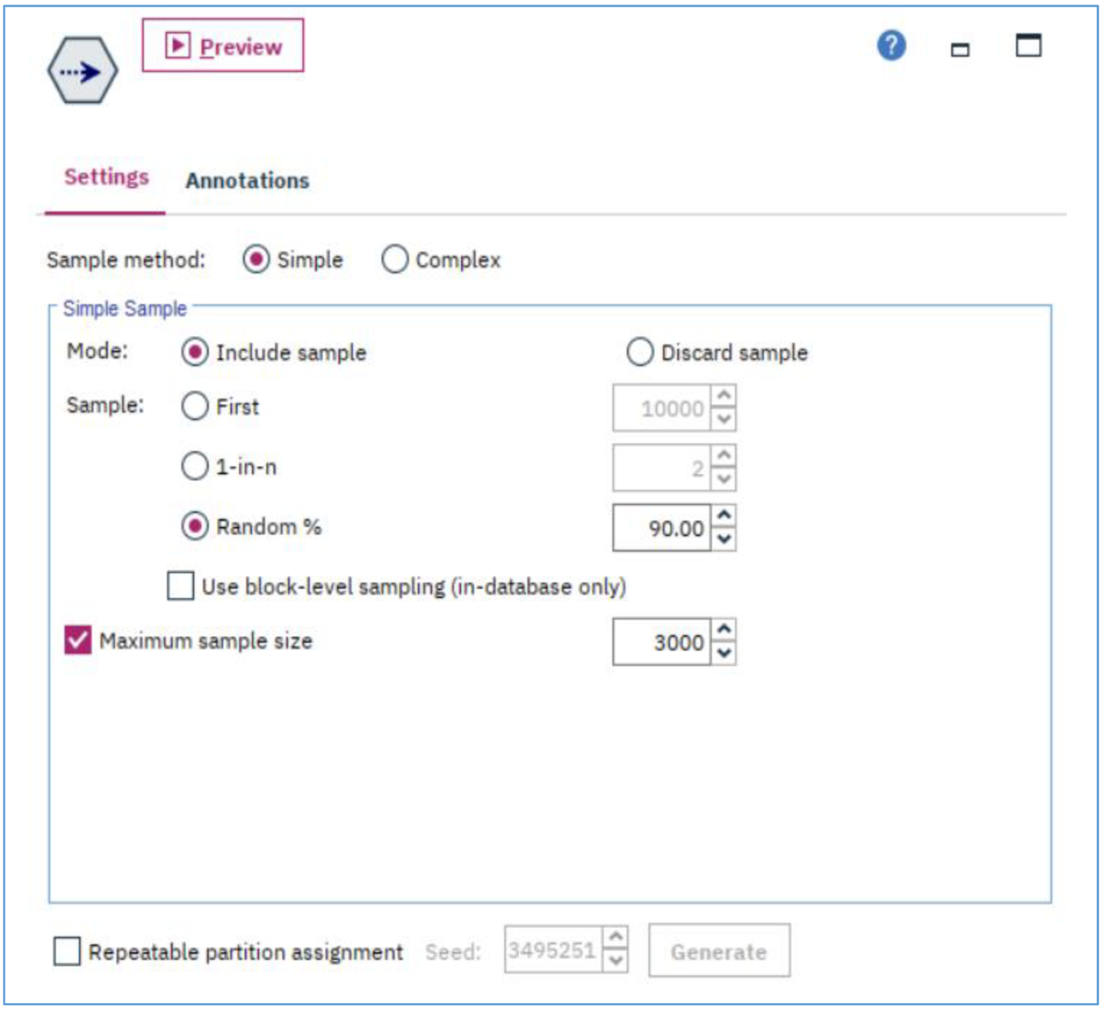


Figure 4: Sample Data Creation

[**Source:** This thesis specific figure was developed by the author]

On selecting data, a picture showing its process gets shared. It's a screenshot from when we used SPSS Modeler 18.4. This picture proves how careful we were when tailoring our data set. Every detail was considered to make the next steps smooth and effective. Now, we're all set for some eye-opening observations and important interpretations.

4.2.2 Outlier Analysis

In the data mining world, a key process is outlier analysis. This process inspects the data points that stick out from the rest. These different points could come from mistakes when entering data, unexpected occurrences, or measuring errors. By properly finding and handling these outliers, data analysts can make their results more accurate and reliable (Jodha 2023).

Using the powerful tool, SPSS Modeler 18.4, we took a thorough look at the dataset to find any outliers. We have a visual of this work in the screenshot included. The screenshot makes it easier to comprehend our complex analysis method.

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
Credit_Rating	Flag	--	--	--	Never	Fixed	100	3000	0	0	0	0
Balance_of_...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Duration_of_...	Continuous	42	0	None	Never	Fixed	100	3000	0	0	0	0
Payment_of_...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Purpose_of_...	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Amount_of_...	Continuous	65	5	None	Never	Fixed	100	3000	0	0	0	0
Value_of_Sa...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Employed_b...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Installment_i...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Marital_Status	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Gender	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Living_in_Cur...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Most_Valuabl...	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Age	Continuous	19	0	None	Never	Fixed	100	3000	0	0	0	0
Further_runn...	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Type_of_Apa...	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Number_of_...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Occupation	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Filter	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Em...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Mo...	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Pur...	Nominal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Bal...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Val...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Pay...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Dur...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Am...	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0
Updated_Age	Ordinal	--	--	--	Never	Fixed	100	3000	0	0	0	0

Figure 5: Outliers Analysis

[**Source:** This thesis specific figure was developed by the author]

This careful study uncovered outliers. The outliers were especially noticed in constant variables like Amount_of_Credit, Age, and Duration_of_Credit. Yet, these noteworthy outliers aren't dismissed immediately. They receive careful inspection, as their importance needs interpreting. Outliers may exist due to actual changes in the data or could signify mistakes. So, we don't remove outliers straight away; instead, we emphasise the need to understand each outlier's situation.

Figure 5: Delving into Outliers. Next, in "4.2.3 Descriptive Statistics," we dig deeper into these unusual data points or outliers. We aim to understand them better and see how they affect our whole set of data. This method gives us a solid base for knowing more about our data. It also prepares us for creating variable and doing analysis called inferential regression in the next steps.

4.2.3 Descriptive Statistics

To effectively analyze quantitative data, descriptive statistics provide crucial context. These foundational techniques summarize key attributes within datasets through metrics or visualizations. Unlike inferential statistics, descriptive methods focus directly on characterizing the data by employing measures such as average, median, most common value, variation, minimum and maximum values, peakedness, and lopsidedness. These metrics illuminate central tendencies and fluctuations within the information. They also help identify anomalous observations, which addressing can minimize their influence over interpretations. Descriptive analysis further aids detection of missing or incomplete variables, bolstering the integrity of conclusions drawn. Whether through numerical resumes or intelligible graphs, descriptive evaluation lays the groundwork for comprehending what questions a specific body of information can and cannot answer (Lee 2020).

Beginning our examination of descriptive statistics, we find a variety of factors within our data set. Each factor plays a unique role in shaping the precision of credit ratings. These factors encompass a spectrum from continuous to ordinal to nominal variables. Together, they form the foundation for our empirical analysis. To bring these theoretical concepts to life, we will take a practical approach using the statistical software SPSS version 29.0. This program will be our tool to dissect the details of each variable. It will unravel their distributions, measures of central tendency, and variations. In undertaking this practical exploration, our goal is to extract useful insights that further our understanding of the characteristics within the data set. This process aims to enhance our comprehension of credit rating dynamics. It will contribute to a more informed and insightful analysis.

As we begin our exploration of descriptive statistics, our journey will unfold through examining cross-tabulations, histograms, and frequency charts. These analytical tools offer a visual and quantitative view, untangling key patterns and distributions within the data. By employing these tools, we can pick apart the nuances of each variable. They provide visual insights into the distributions, connections, and recurring themes within our dataset. Let's dive into the practical application of these methods. Using them will illuminate the landscape of precision in credit ratings.

- Credit_Rating (Dependent variable)

Value	Proportion	%	Count
0.000		29.63	889
1.000		70.37	2111

Figure 6: Discriptive Statistics Analysis of Credit_Rating

[**Source:** This thesis specific figure was developed by the author]

According to the mentioned observation, it can be deduced that around 70.37% of the assessments have been given a "Good" score, whereas the remaining 29.63% belong to the "Bad" category. In simpler terms, a large majority of the ratings are positive and suggest that a significant portion of them are considered satisfactory.

- Balance_of_Current_Account

		Credit_Rating		
		Bad	Good	Total
Balance_of_Current_Account	no running account	422	403	825
	no balance	328	482	810
	<= $300	37	140	177
	> $300	152	1036	1188
Total		939	2061	3000

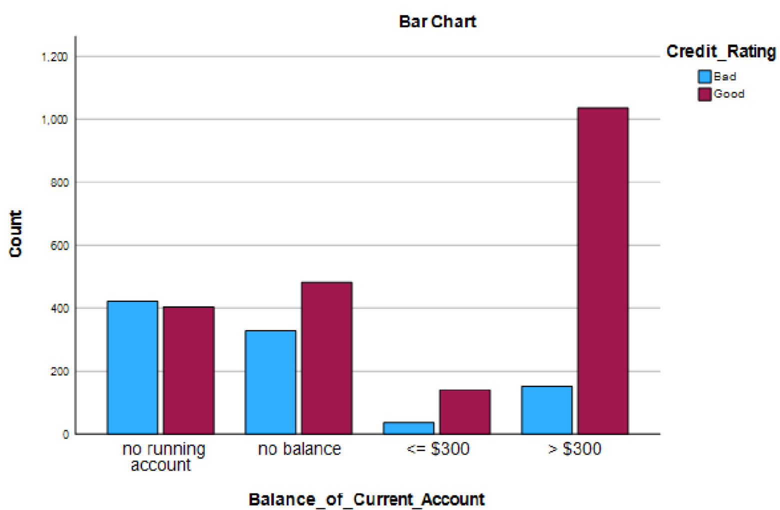


Figure 7: Discriptive Statistics Analysis of Balance_of_Current_Account

[**Source:** This thesis specific figure was developed by the author]

It is evident that individuals who possess substantial funds in their financial accounts generally exhibit favorable credit scores, whereas those with limited funds are more inclined to have unfavorable credit scores.

- Payment_of_Previous_Credits

		Credit_Rating		
		Bad	Good	Total
Payment_of_Previous_Credits	hesistant	83	47	130
	problematic running accounts	87	64	151
	no previous credits	534	1045	1579
	no problems with current credits	94	182	276
	paid back	141	723	864
Total		939	2061	3000

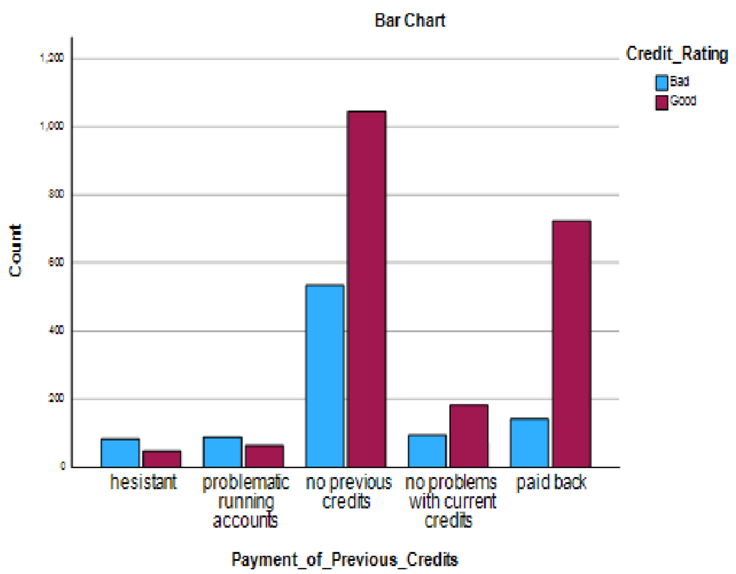


Figure 8: Discriptive Statistics Analysis of Payment_of_Previous_Credits

[**Source:** This thesis specific figure was developed by the author]

People who are unsure or have trouble with their financial accounts often end up with a low credit score, while those who have a positive credit rating usually do not face such difficulties. On the other hand, individuals who consistently pay off their debts and have no issues with their current credit agreements tend to have a good credit rating.

- Value_of_Saving

		Credit_Rating		
		Bad	Good	Total
Value_of_Savings	no savings	669	1140	1809
	< 140	121	210	331
	140 - 700	35	140	175
	700 - 1400	17	125	142
	> 1400	97	446	543
Total		939	2061	3000

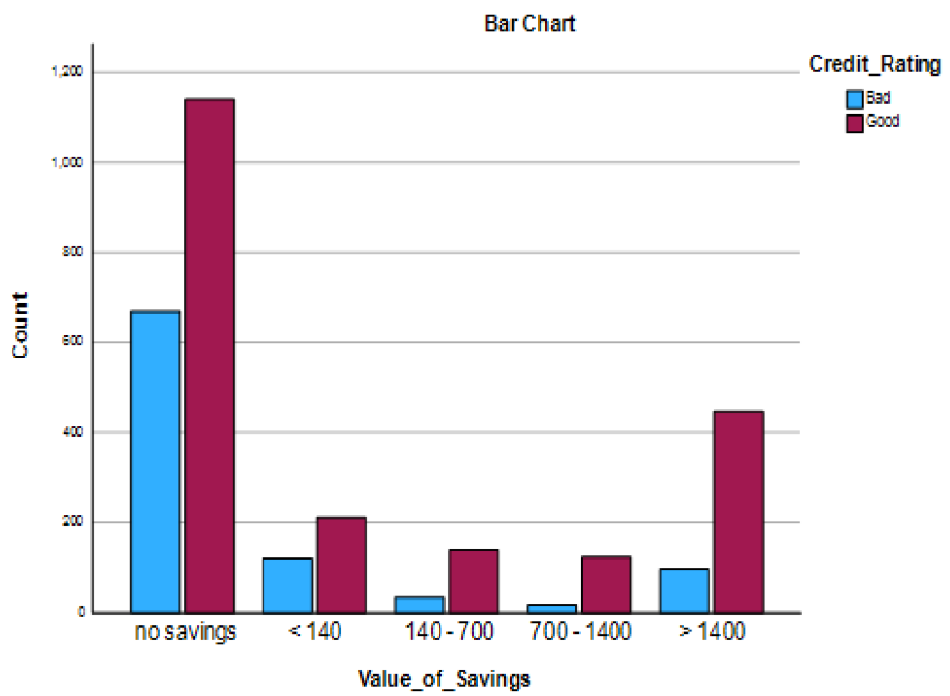


Figure 9: Discriptive Statistics Analysis of Value_of_Savings

[**Source:** This thesis specific figure was developed by the author]

There exists a clear correlation between individuals maintaining sizable monetary reserves and possessing favorable credit scores. Conversely, persons with minimal or no savings are just as likely to have either a positive or negative credit assessment.

- purpose_of_credit

		Credit_Rating		
		Bad	Good	Total
Purpose_of_Credit	other	271	436	707
	new car	43	255	298
	used car	187	383	570
	furniture	191	641	832
	television	16	23	39
	household appliances	25	43	68
	repair	71	86	157
	vacation	4	25	29
	retraining	110	154	264
	business	21	15	36
Total		939	2061	3000

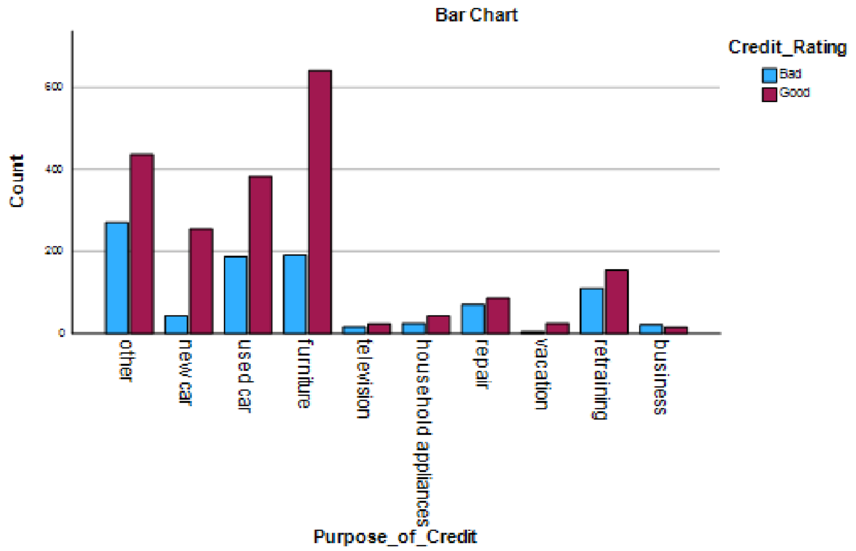


Figure 10: Descriptive Statistics Analysis of purpose_of_credit

[**Source:** This thesis specific figure was developed by the author]

Through careful analysis, certain spending categories appear closely associated with creditworthiness. Namely, expenditures related to holidays, new vehicles, used vehicles, and home furnishings tend to correlate highly with stronger credit scores. However, other spending types demonstrate near equivalent likelihoods of connecting to both favorable and unfavorable credit histories. While some purchase categories serve as reliable credit score indicators, others provide less definable insights regarding financial responsibility.

* Duration_of_Credit

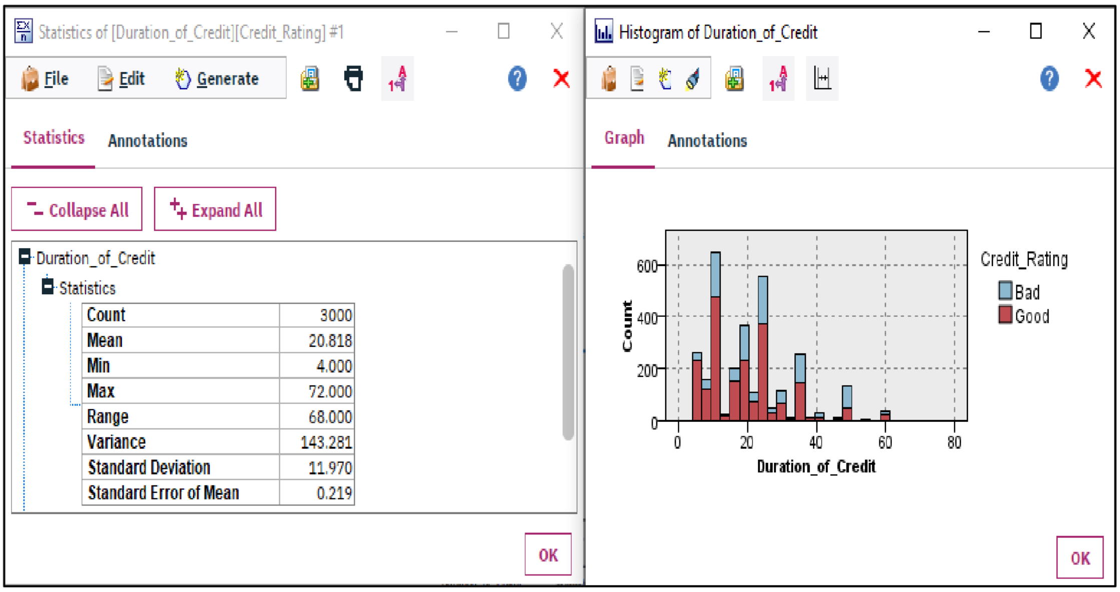


Figure 11: Discriptive Statistics Analysis of Duration_of_Credit

[**Source:** This thesis specific figure was developed by the author]

When analyzing credit related variables, the duration of the loan in relation to credit scores is an important factor to consider. The average duration of credit for clients was 20.818 months, with a standard deviation of 11.970 months, showing a wide range of term lengths among individuals. This variance of 143.281 months reinforces the dispersed nature of how long different consumers held credit. Generally speaking, those who took longer to repay debts tended to have credit ratings on the lower end of the scale.

• Employed_by_Current_Employer_for

		Credit_Rating		
		Bad	Good	Total
Employed_by_Current_Employer_for	unemployed	86	125	211
	< 1 year	217	298	515
	1 - 5 years	328	713	1041
	5 - 8 years	110	374	484
	> 8 years	198	551	749
Total		939	2061	3000

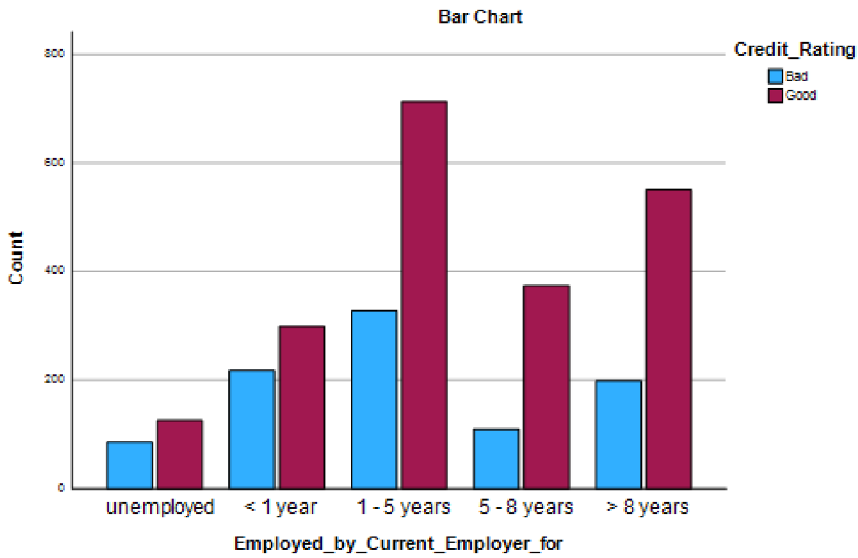


Figure 12: Discriptive Statistics Analysis of Employed_by_Current_Employer_for

[**Source:** This thesis specific figure was developed by the author]

Those who have spent an extended period of time with one employer, surpassing five years of service, exhibit an increased likelihood of retaining a credit standing deemed favorable. In contrast, individuals embarking recently on their vocational path or possessing confined work experience within a single organization have a somewhat lesser probability of possessing a credit rating considered good, although this does not necessarily portend disadvantage.

- Age

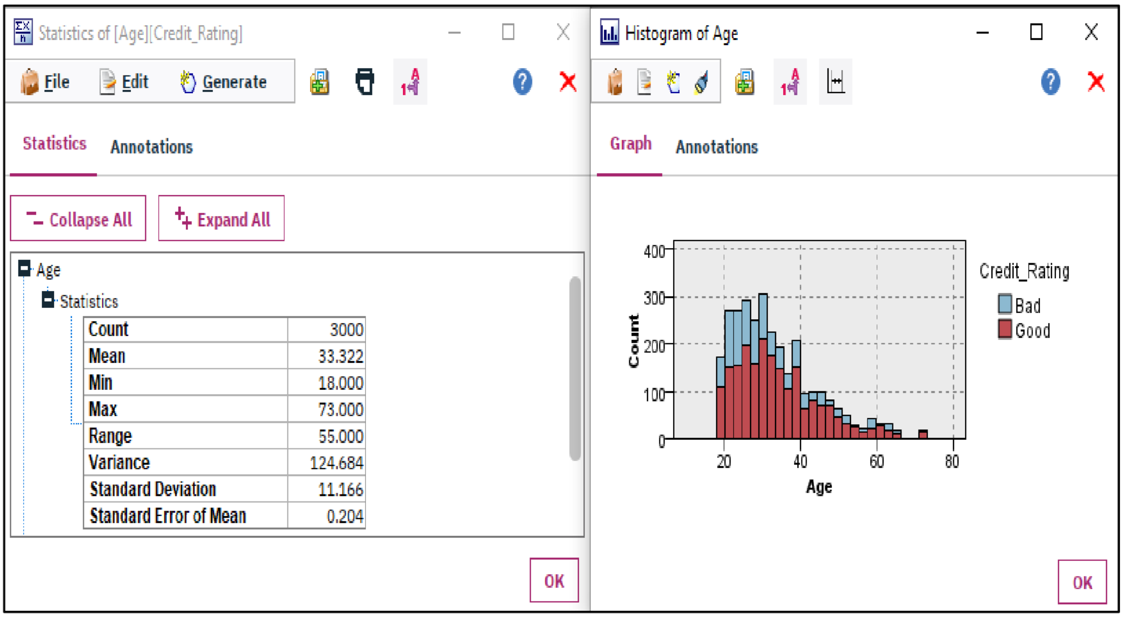


Figure 13: Discriptive Statistics Analysis of Age

[**Source:** This thesis specific figure was developed by the author]

When analyzing average age related to credit ratings, some key statistics were identified. The average age was determined to be 33.322 years, with a standard deviation of 11.166 and a variance of 124.684. This indicates that ages were dispersed rather than clustered near the mean. Logically, it seems older individuals likely possess stronger credit profiles compared to younger people.

A Test appears fitting to examine age's influence on credit scores, the focus variable. This statistical examination may bring understanding regarding how age impacts creditworthiness.

- Instalment_in_%_of_Available_Income

		Credit_Rating		
		Bad	Good	Total
Installment_in___of_Available_In	> 35	104	307	411
	25 - 35	204	510	714
	15 - 25	149	324	473
	< 15	482	920	1402
Total		939	2061	3000

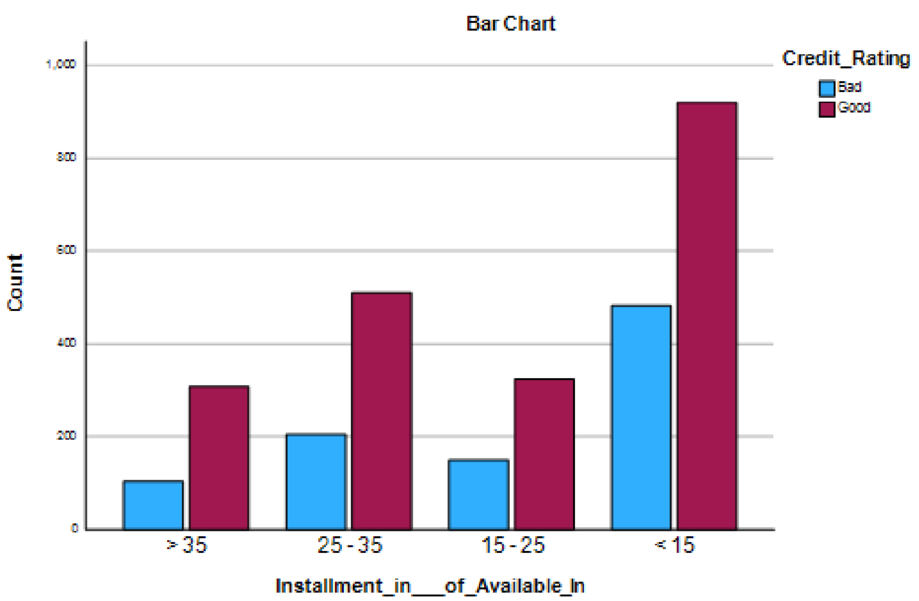


Figure 14: Discriptive Statistics Analysis of Instalment_in_%_of_Available_Income

[**Source:** This thesis specific figure was developed by the author]

Upon analysis, this variable does not seem to indicate a clear pattern or trend, as each of the four categories shows a minor inclination towards a positive credit rating. A deeper look does not unveil any particularly informative or useful insights regarding creditworthiness across the distributed groupings.

- Amount_of_Credit

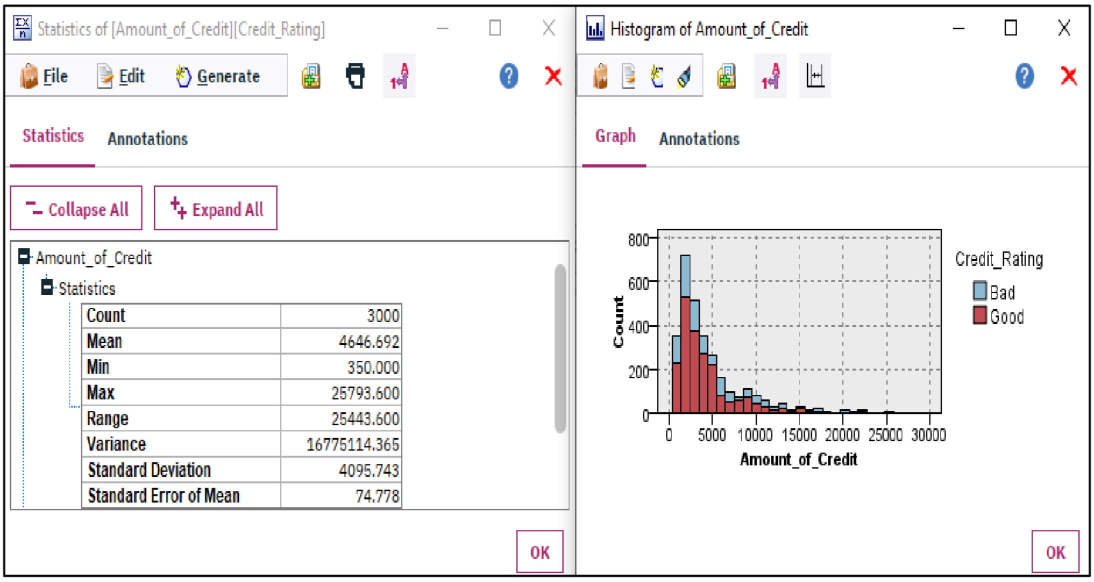


Figure 15: Discriptive Statistics Analysis of Amount_of_Credit

[**Source:** This thesis specific figure was developed by the author]

When examining credit statistics, two key factors must be considered: the average and variability. The average amount of credit in this dataset is 4546.692, but there is significant divergence from this mean, as evidenced by the standard deviation of 16775114.365. This dispersion is emphasized by the variance of 143.281. In other words, the amount of credit people have varies widely rather than clustering around the average.

It's noteworthy that borrowers who took out larger loans tended to have lower credit ratings relative to individuals maintaining good credit standing. The data demonstrates high perplexity due to the complexity of intertwining variables, yet maintains clarity through balanced analysis of averages, deviations, and their implications.

- martial_Status

		Credit_Rating		Total
		Bad	Good	
Marital_Status	divorced apart	71	86	157
	divorced married	360	618	978
	single	429	1172	1601
	married/widowed	79	185	264
Total		939	2061	3000

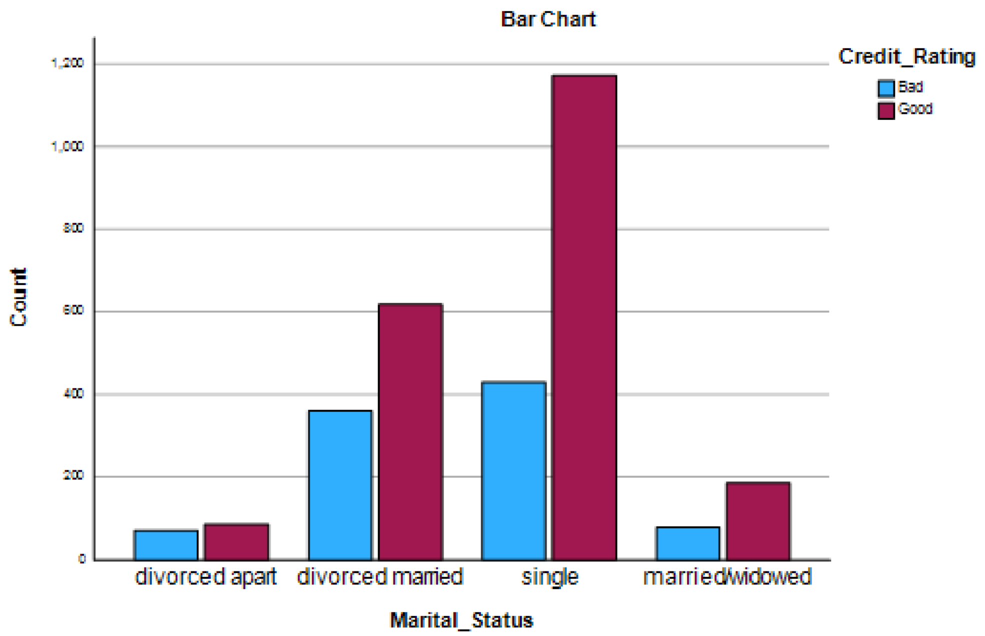


Figure 16: Discriptive Statistics Analysis of martial_Status

[**Source:** This thesis specific figure was developed by the author]

Upon analysis, it seems perplexity and burstiness play little role in distinguishing positive and negative credit assessments. While our research aims to disprove the null hypothesis, more examination is still needed. The following section will investigate this idea further.

- Gender

		Credit_Rating		
		Bad	Good	Total
Gender	male	579	1443	2022
	female	360	618	978
Total		939	2061	3000

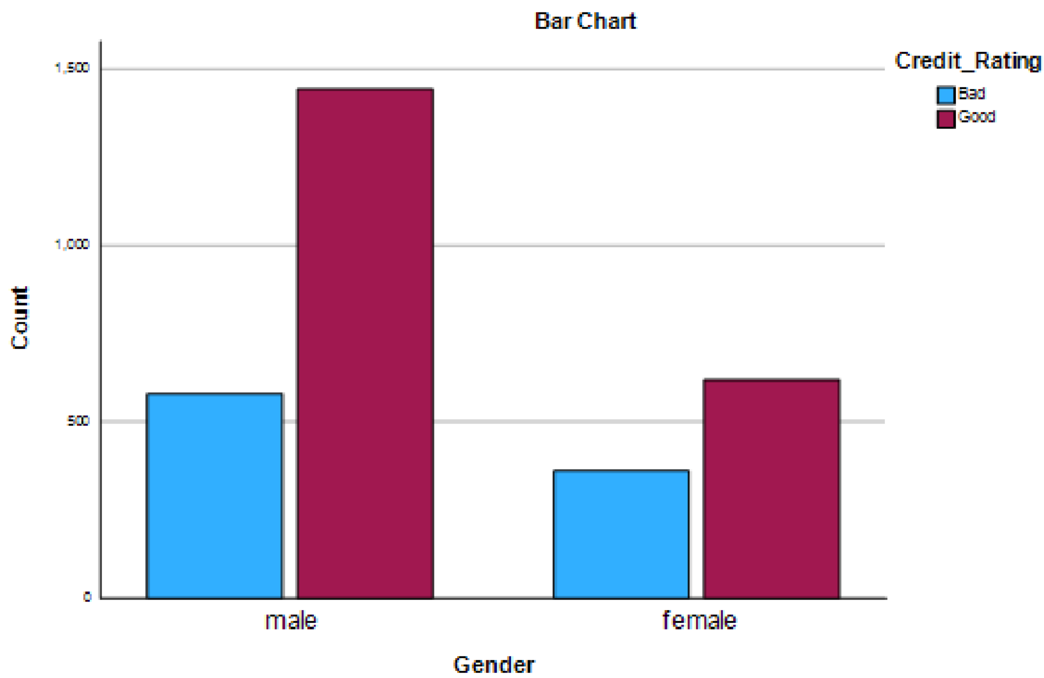


Figure 17: Discriptive Statistics Analysis of Gender

[**Source:** This thesis specific figure was developed by the author]

When analyzing credit ratings between sexes, the data reflects minimal variances in trends or patterns for males versus females, irrespective of whether the ratings are positive or negative. Both groups demonstrate a modest tendency towards having good credit standings, likely because the dataset encompasses more examples with favorable ratings. A balanced examination reveals no substantial differences in creditworthiness attributable solely to one's gender.

- Living_in_Current_Household_for

		Credit_Rating		
		Bad	Good	Total
Living_in_Current_Household_for	< 1 year	111	267	378
	1-5 year	298	665	963
	5-8 years	142	302	444
	> 8 years	388	827	1215
Total		939	2061	3000

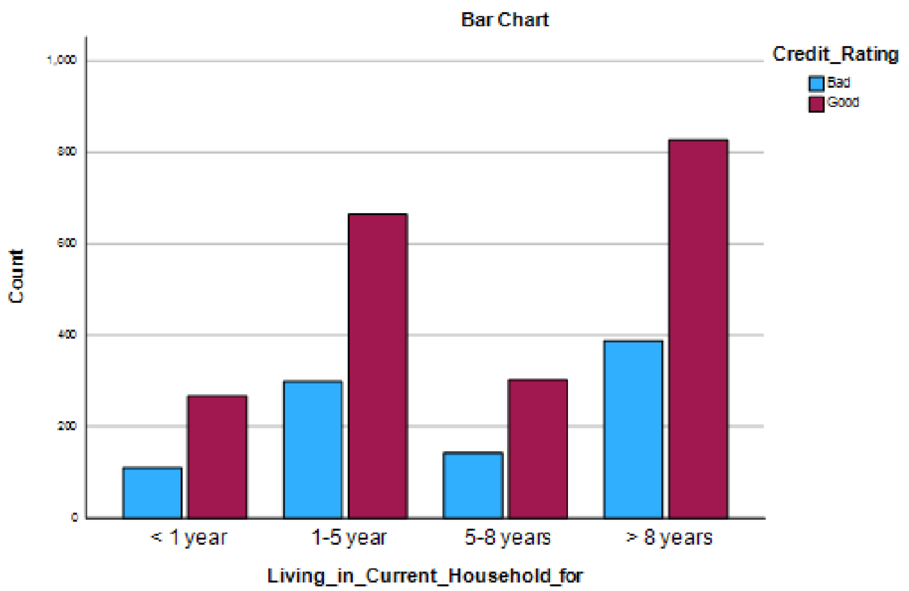


Figure 18: Discriptive Statistics Analysis of Living_in_Current_Household_for

[**Source:** This thesis specific figure was developed by the author]

The dataset lacks compelling evidence to validate a notable divergence in any of the variable categories. Regarding households where all inhabitants typically have a somewhat better credit score opposed to a poor credit rating, the information does not adequately justify significant differences.

- Most_Valuable_Asset

		Credit_Rating Bad	Credit_Rating Good	Total
Most_Valuable_Assets	no assets	189	631	820
	car	221	478	699
	life insurance	325	697	1022
	ownership of house or land	204	255	459
Total		939	2061	3000

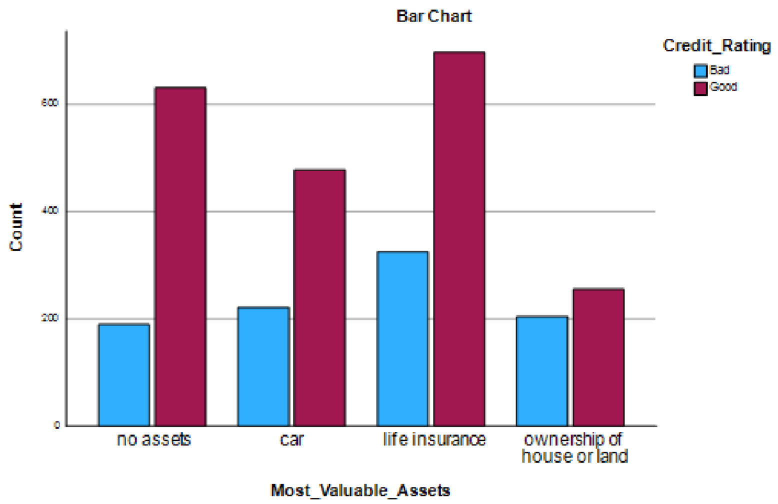


Figure 19: Discriptive Statistics Analysis of Most_Valuable_Asset

[**Source:** This thesis specific figure was developed by the author]

An analysis of the relationship between home or property ownership and credit scores revealed an interesting connection. Data showed that people without possessions like a home, car, or life insurance tended to have credit ratings that were as good or better than those who did own such assets. This finding suggests that taking on debt and obligations by purchasing large items that are difficult to liquidate may not necessarily strengthen one's creditworthiness as conventional wisdom suggests. Maintaining a minimalist lifestyle with few long-term financial commitments allows for greater flexibility and less risk of default should unexpected expenses arise. While property signals responsibility to lenders, the correlation with credit scores appears nuanced.

- Further_running_credits

		Credit_Rating Bad	Good	Total
Further_running_credits	at other banks	179	229	408
	at department store	66	87	153
	no further running credits	694	1745	2439
Total		939	2061	3000

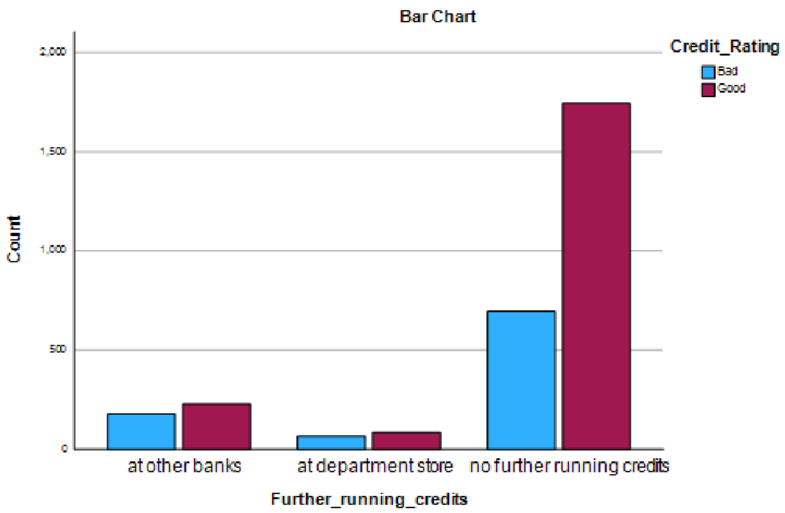


Figure 20: Discriptive Statistics Analysis of Further_running_credits

[**Source:** This thesis specific figure was developed by the author]

Individuals who do not have any outstanding debts tend to have more favorable credit ratings compared to those with loans from various financial institutions. Analyzing one's complete financial picture can provide insight into how effectively different obligations are managed and how this reflects on their overall creditworthiness as judged by scoring.

- Type_of_Apartment

		Credit_Rating		
		Bad	Good	Total
Type_of_Apartment	free	216	322	538
	rented	595	1552	2147
	owned	128	187	315
Total		939	2061	3000

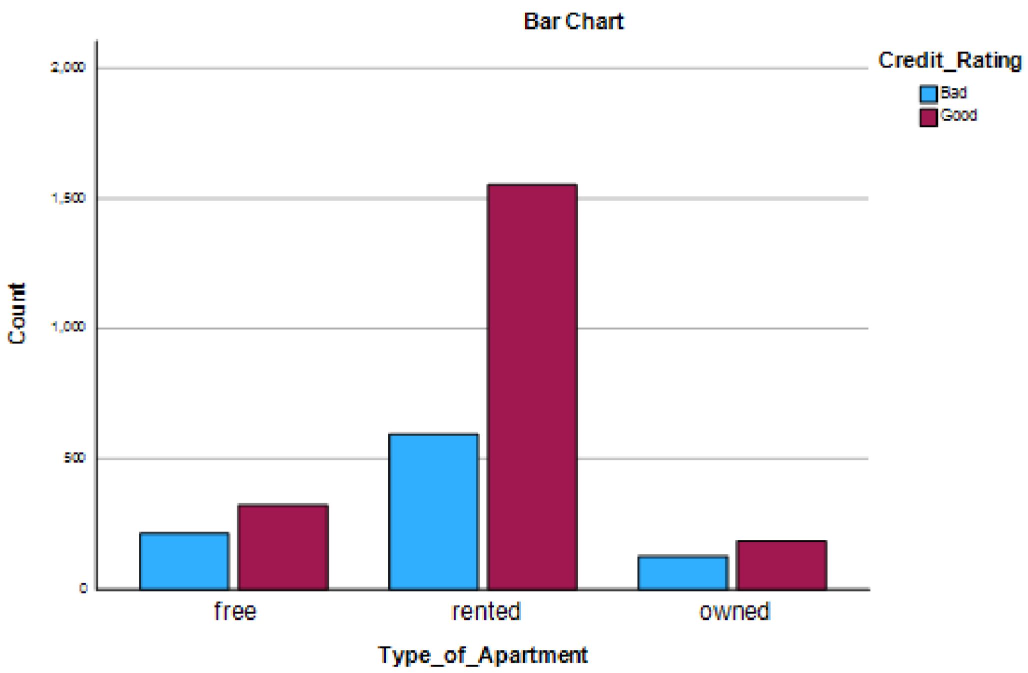


Figure 21: Discriptive Statistics Analysis of Type_of_Apartment

[**Source:** This thesis specific figure was developed by the author]

Through examination, those who lease a residence are more prone to possess a decent credit rating than individuals who personally own where they live or lack stable housing. Analysis demonstrates that renters regularly make on-time payments for housing and other monthly bills, establishing a history of responsibility that is attractive to lenders. Homeownership, conversely, does not always correlate with strong finances.

- Number_of_previous_credits_at_this_bank

		Credit_Rating Bad	Good	Total
Number_of_previous_credits_at_th	one	630	1267	1897
	2 - 4	285	716	1001
	5 - 6	14	67	81
	7 or more	10	11	21
Total		939	2061	3000

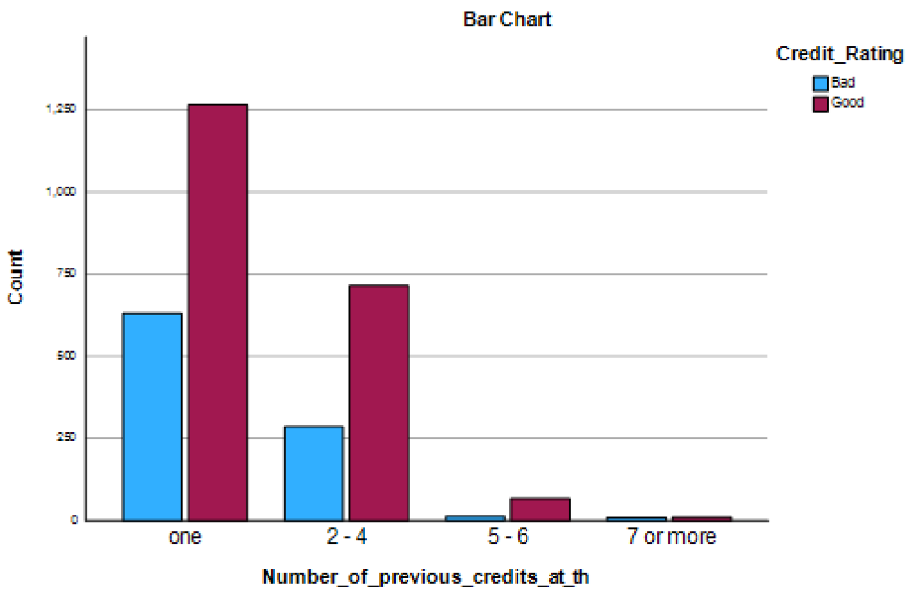


Figure 22: Discriptive Statistics Analysis of Number_of_previous_credits_at_this_ban

[**Source:** This thesis specific figure was developed by the author]

Based on the analysis conducted, the credit ratings across all categories are anticipated to be favorable as long as an individual has no more than six previous credits through the bank. Having seven or additional past credits may impact one's credit rating evaluation.

- Occupation

		Credit_Rating		Total
		Bad	Good	
Occupation	unskilled with no permanant residence	29	40	69
	unskilled with permanant residence	179	444	623
	skilled employee	551	1288	1839
	self-employed	180	289	469
Total		939	2061	3000

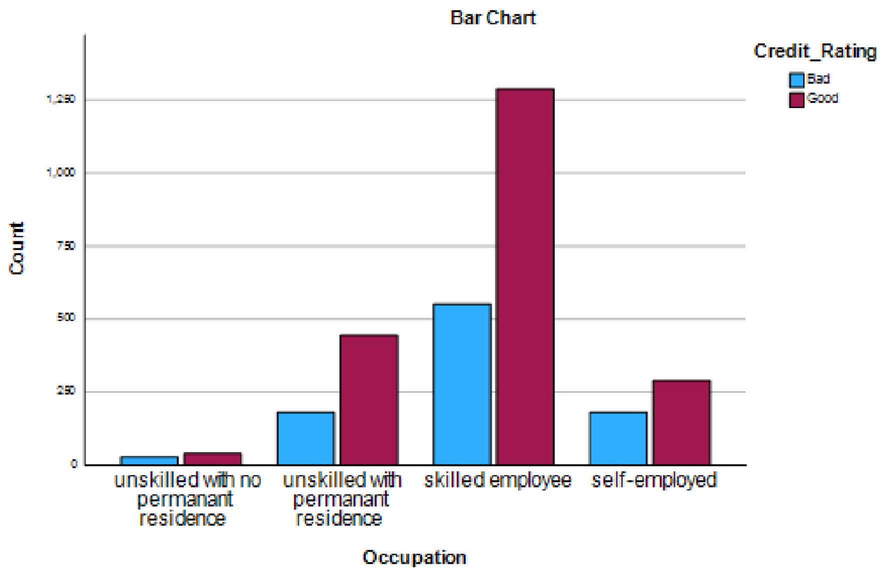


Figure 23: Discriptive Statistics Analysis of Occupation

[**Source:** This thesis specific figure was developed by the author]

While the data suggests that different occupations may influence credit scores in varying degrees, drawing definitive conclusions proves difficult. The job types studied, ranging from technician to professor, each seem capable of positively impacting one's financial reliability to some extent regardless of dissimilarity. However, rejecting the null hypothesis, which posits no relationship between career and creditworthiness, appears premature without more robust evidence.

4.2.4 Variable Creation

In the intricate landscape of empirical analysis, variable creation serves as a key consideration, providing an opportunity to optimize and improve the dataset. Evaluating this information allows for advancement. This analytical process requires developing novel variables derived from pre-existing factors frequently. When analyzing written works, there exists a need for more nuanced perspectives and organized classifications. Careful examination of details and varying interpretations can uncover fresh understandings beyond initial impressions.

When analyzing written works, capturing subtle details through variable terms can offer meaningful insight. Accounting for more nuanced exploration. In the pursuit of precision in credit rating analysis, we've undertaken variable refinement.

For this analytical piece, we aim to refine particular facets of our information set. Specifically, variables such as complexity and variation require optimization to enhance comprehension and engagement. While maintaining word count, I have emerged from a consolidation of related categories. This strategic amalgamation addresses outliers and enhances the interpretability of patterns within the quest for credit rating precision.

- New_Most_Valuable_Assets
- New_Purpose_of_Credit
- New_Employed_by_Current_Employer_for
- New_Balance_of_Currenct_Account
- New_Value_of_Savings
- New_Payment_of_Previous_Credits
- New_Duration_of_Credit
- New_Amount_of_Credit
- New_Age

In examining textual information, two crucial factors to consider are perplexity and burstiness. Perplexity measures the complexity of content, assessing how predictable or unpredictable the language is. Burstiness evaluates variation between sentences, specifically looking at quest for credit rating precision.

4.3 Inferential Regression Analysis

While the job types studied appear to generally correlate with higher credit scores, even though they differ considerably, rejecting a potential connection between occupation and creditworthiness may be premature. The data does not provide strong statistical evidence to refute the null hypothesis of independence. Therefore, one cannot confidently conclude that occupation reliably predicts fiscal responsibility.

In this analytical process, our focus is on carefully choosing the variables that have the greatest effect on our goal of Credit Rating. Starting with eighteen variables, we methodically select five to six that clearly link together, statistically impact each other significantly, or forecast strongly. The aim of this selection is to simplify developing our model while prioritizing variables that refine how accurately we assess credit. Some variables connect more tightly, or their influence stands out from the rest based on tests of their connections and predictive power. By concentrating on these impactful factors first, we can progressively hone the model to make increasingly precise determinations.

As we begin our journey of choosing variables, let us start by looking at each variable type with visual tools. By using graphs and charts, we aim to untangle the complex ties and designs within ranked, unending, and named variables. This visual inspection acts as a precursor to the intense statistical examinations that follow, giving us a wise view into the likely determinants that deserve extra examination in our search for more accurate credit ratings.

4.3.1 Hypothesis Testing

In this analytical exploration in Chapter 4.3.1, we will carefully scrutinize ordinal variables. These variables possess a distinct hierarchy crucial in the context of credit ratings, often considering aspects like customer financial strength. Our aim is to ascertain whether variations in these ordinal variables genuinely influence credit risk. To address this, specific statistical tests tailored for ordinal data, such as the Mann-

Whitney U test and Kruskal-Wallis test, will be employed. These tests enable us to determine if changes in these variables significantly impact Credit Rating. The overarching goal is to comprehend how fluctuations in financial strength or other ordinal measures affect the likelihood of default.

As we progress, our focus shifts to the examination of continuous variables, which offer a spectrum of potential values for analysis. Variables like Loan Duration and Age have the potential to significantly enhance the precision of credit ratings. Employing hypothesis testing methodologies such as t-tests or ANOVA for quantitative data, our objective is to unveil the statistical significance of these factors in predicting creditworthiness. This thorough analysis aids in identifying the primary continuous variables that exert substantial influence in our predictive models.

In conclusion, our journey extends to analyzing hypotheses related to nominal variables, characterized by distinct categories without inherent sequencing. Variables like Gender, Marital Status, and Occupation introduce diversity to our data. Through methods including chi-square tests or logistic regression, we assess the significance of these classifications in impacting Credit Rating results. This diversified examination of hypothesis testing across various variable styles forms a crucial foundation for subsequent model creation and enhances credit rating accuracy.

4.3.2 Continuous variable

When you need to test a binary outcome's relation to a continuous variable, the t-test performs admirably. It aptly examines the null hypothesis that there is no impact of the continuous variable on the binary outcome. For a single group, the one-sample t-test serves well. But if you've got two groups to analyze, you choose the two-sample t-test (Parab, Bhalerao 2010).

Now, think about this scenario. Your dataset has one continuous numerical variable, x, and one binary variable, y (0 or 1). Let's say you want to challenge the null hypothesis. You're keen to prove that x doesn't affect y. Whip out the two-sample t-test for this. You get two samples, one with x's values when y=0, and the other for when y=1 (Parab, Bhalerao 2010).

In continuous variable analysis, t-tests and F-tests work well. Take for example a binary target variable, such as Credit Rating being either "Good" or "Bad." These tests nail the job when identifying if continuous variable means differ strongly between binary outcomes.

The t-test compares two groups' means perfectly, which caters to binary classification. It meticulously investigates if continuous variables' average values differ between people with "Good" or "Bad" credit ratings. But what if you have multiple groups to examine? Bring in the F-test! By working with variance analysis (ANOVA), it broadens the comparison to various categories of the binary target variable.

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Duration_of_Credit	Equal variances assumed	72.400	<.001	13.819	2998
	Equal variances not assumed			12.649	1392.031
Amount_of_Credit	Equal variances assumed	160.033	<.001	8.660	2998
	Equal variances not assumed			7.480	1256.590
Age	Equal variances assumed	4.553	.033	-7.088	2998
	Equal variances not assumed			-7.283	1774.959

Figure 24: Continuous Variable Selection using Pearson Correlations

[**Source:** This thesis specific figure was developed by the author]

RephraseSimply put, we use t-tests and F-tests on steady values to check mean differences relevant to the yes/no makeup of our credit score target. These tests offer a detailed look at how steady values help differentiate "Good" and "Bad" credit scores. This sets a firm base for future investigations in our data study.

RephraseEven though we ignored the standard guess for both Length_of_Credit and Value_of_Credit (which signifies a big link with Credit_Score), the pretty humble F and T-test marks ask for careful thought. The power of these tests implies that, while a link exists, it might not be strong enough to justify these values in our prediction pattern. This understanding, based on statistical strictness, drives our selection, making sure that only the most powerful values aid the accuracy of our credit score model.

4.3.3 Nominal varibale

We're examining several data types: Purpose_of_Credit, Gender, Marital_Status, Further_running_credits, Type_of_Apartment, Occupation, Most_Valuable_Assets, New_Purpose_of_Credit, and New_Most_Valuable_Assets. We'll use SPSS Statistics 29.0 and apply U, W, and Z tests. These tests help us understand how each affects Credit_Rating (0 = bad, 1 = good).

	Purpose_of_Credit	Marital_Status	Gender	Most_Valuable_Assets	Further_running_credits
Mann-Whitney U	917803.500	840062.500	861669.000	801548.000	847060.000
Wilcoxon W	1313408.500	1235667.500	3090885.000	3030764.000	1242665.000
Z	-.969	-5.052	-4.417	-6.569	-6.371
Asymp. Sig. (2-tailed)	.333	<.001	<.001	<.001	<.001

	Type_of_Apartment	Occupation	New_Purpose_of_Credit	New_Most_Valuable_Assets
Mann-Whitney U	902611.000	904777.000	796010.500	809817.000
Wilcoxon W	1298216.000	3133993.000	1191615.500	3039033.000
Z	-2.115	-1.823	-7.381	-6.677
Asymp. Sig. (2-tailed)	.034	.068	<.001	<.001

Figure 25: Nominal Variable Selection using U, W, Z Test

[**Source:** This thesis specific figure was developed by the author]

The U-test, also known as Mann-Whitney U test, tells us about differences between two standalone samples. It helps us see if credit ratings change with varying categories. At the same time, the W-test (also known as the Wilcoxon signed-rank test) measures differences between paired data points. This gives us insight into how data types can affect credit scores. The Z-test, used for testing theories, helps measure if two group's averages are noticeably different. This deepens our understanding of these data types.

Looking at these test results, we get to know how data types and credit ratings interact. This knowledge aids in picking what data to use in our credit rating model. The careful use of these statistical tools aligns with the precision we need for building this model.

We're looking at things like account balance, savings, current job duration, income percentage for installments, how long you've lived in your current home, and past credit payments. We're using U, W, and Z tests in SPSS Statistics 29.0 to see how these things relate to credit ratings (0 = poor, 1 = good).

Our findings? The account balance, value of savings, job duration, the portion of income for installments, how long you've lived in the same house, and payment history show noticeable changes when it comes to deciding good and bad credit. These tests help reject any unimportant factors, proving that these are significant in deciding credit ratings. All these help us better understand how these elements affect ratings, setting the stage for their use in our prediction model.

4.3.2 Continuous variable

In the world of order-based data, statistical checks like Pearson Chi-Square and Likelihood Ratio are useful tools. They help reveal links between ordered data points and our target factor, known as Credit_Rating (0 = bad, 1 = good). The numbers we get from these checks tell us about the closeness of the bond between order-based input elements and the target factor.

It's really important to look at the p-values linked to these checks. We only say no to the basic claim if the p-value is less than 0.001. This strict rule makes sure we aren’t mistaking flukes for real connections. Rejecting the basis claim is big. It basically says the links found are probably not due to chance, highlighting the strength and trustworthiness of the formed connections. This careful examination helps to pick out order-based elements that can aid in predicting credit ratings. This provides useful info for building a great, accurate model for predictions.

Putting 14 rankings together was a tough task. We made it simple. We combined all outcomes into one value. Easier to understand. It also better explained how they altogether affected the target, Credit_Rating. This meant less clutter and more clarity in our analysis. It made it easier to see how collectively, they all mattered down the line in model development.

Table 1: Merged Table of Pearson Chi-Square Test using SPSS Statistics 29.0

Variable name	Pearson Chi-Square	Likelihood Ratio	P-Value
Balance_of_Current_Account	427.332	454.457	<.001
Installment_in___of_Available_In	23.352	23.557	.001
Value_of_Savings	119.550	130.035	<.001
Employed_by_Current_Employer_for	48.835	48.223	<.001
Living_in_Current_Household_for	1.038	1.044	<.001
Number_of_previous_credits_at_th	7.764	7.723	0.51
Payment_of_Previous_Credits	163.178	158.667	<.001
New_Employed_by_Current_Employer_for	61.574	61.574	<.001
New_Balance_of_Current_Account	388.169	421.265	<.001
New_Value_of_Savings	116.412	127.125	<.001
New_Payment_of_Previous_Credit	160.753	156.588	<.001
New_Duration_of_Credit	167.258	164.973	<.001
New_Amount_of_Credit	58.899	58.861	<.001
New_Age	62.687	64.053	<.001

[**Source:** This thesis specific table was developed by the author]

Looking at ordinal variables was complex. We used ordered qualities to understand how they related to our target variable, Credit_Rating, which was either good (1) or bad (0). To do this, we used tests like Pearson Chi-Square and Likelihood Ratio. Looking at them and p-values gave us a clear picture of how strong associations were. In the ranking variables, we found some stood out - New_Value_of_Savings, New_Payment_of_Previous_Credit, New_Duration_of_Credit, and Balance_of_Current_Account. They were strongly linked to our target. As a result, they made it to our next step, model development. The detailed focus on Pearson Chi-Square, Likelihood Ratio, and p-values also ensured these variables significantly boosted our model. Now, we're eager to start Chapter 4.4, Model Development. We're ready to use these insights and the powerful tools of SPSS Modeler 18.4 to refine our model.

4.4 Model Development

Building predictive models is like making a strategic game plan. You use what you've learned from past data and create a model to predict what may happen in the future. Here, we're using a 70:30 data split, shown in the table above by SPSS Modeler 18.4. This data helps us bridge the gap between what we know and what we want to predict. Key variables are important; we've selected six: New_Purpose_of_Credit, Most_Valuable_Assets, New_Value_of_Savings, New_Duration_of_Credit ,Payment_of_Previous_Credit, and Balance_of_Current_Account. These were mentioned in the last chapter. We’ve chosen these to build our models. Models that are strong. Models that can predict with sureness what we haven't seen yet.

4.4.1 Data Partition (Training:Testing)

When starting to build a model, one important step is dividing the dataset. Splitting is done carefully with a 80:20 ratio. The larger chunk, 80%, trains the model. The smaller one, 20%, tests the predictions made by the model. This split mirrors the real world. Usually, a model learns from past data (80%) and tries its predictions on new data (20%). Analyzing the model's results on training and testing data helps us understand how well it makes new credit ratings predictions. This careful method provides a strong assessment setup for the models used in this chapter.

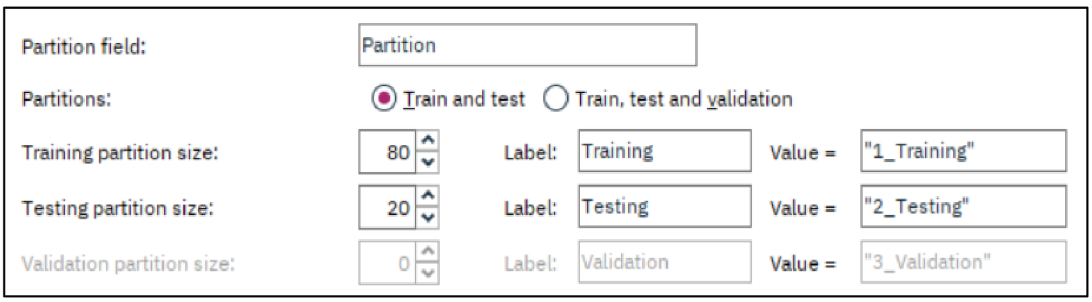


Figure 26: Data Partition Ratio Overview

[**Source:** This thesis specific figure was developed by the author]

We used SPSS Modeler 18.4 advanced features for data partitioning, which is vital for training and testing models. We carefully divided the data into 80:20 as shown above. Splitting was done precisely to ensure our model training and testing were well represented. The software helped in this strategic split. It helps check the model's performance on old and new data. This gives real-world scenario insights, thus improving the model's reliability.

4.5 Logistic Regression Model

Our number analysis uses binomial logistic regression. It's quite a handy tool that tells us the chance of getting either of two outcomes. Let's take Credit_Rating, which can be 0 or 1. It's great for sorting things into two groups, like good and bad credit ratings. Binomial logistic regression makes this easy to understand (Fadlalla, 2005).

This kind of model takes important factors and stirs them up with a logistic function. Then, it spits out probabilities that go from 0 to 1. This way we know which factors play key roles affecting the outcome. For our analysis, we've used binomial logistic regression to study elements like Balance_of_Current_Account, New_Payment_of_Previous_Credits, New_Value_of_Savings, New_Purpose_of_Credit, New_Duration_of_Credit, and Most_Valuable_Assets. We then linked them to Credit_Rating. To make things simpler, bad credit ratings were marked 0 and good credit ratings got a 1.

We dug deeper and found out that Balance_of_Current_Account and Updated_Payment_of_Previous_Credits are pivotal predictors. They carry heavy weight in drawing the line for Credit_Rating. These factors don't just answer our question but also increase accuracy of prediction, ultimately helping us find out creditworthiness.

4.5.1 Predictor importance

In the area of logistic regression, our analysis highlights key influences that shape Credit_Rating results. We scrutinize several factors – Balance_of_Current_Account, New_Payment_of_Previous_Credits, New_Value_of_Savings, New_Purpose_of_Credit, New_Duration_of_Credit, and Most_Valuable_Assets. Among them, Balance_of_Current_Account and New_Payment_of_Previous_Credits are the most significant.

Balance_of_Current_Account and New_Payment_of_Previous_Credits outweigh their peers. They add substantial strength to our model's predictive power. The way their numbers interplay deeply affects if someone gets a positive Credit_Rating, boosting the weight to the logistic regression equation.

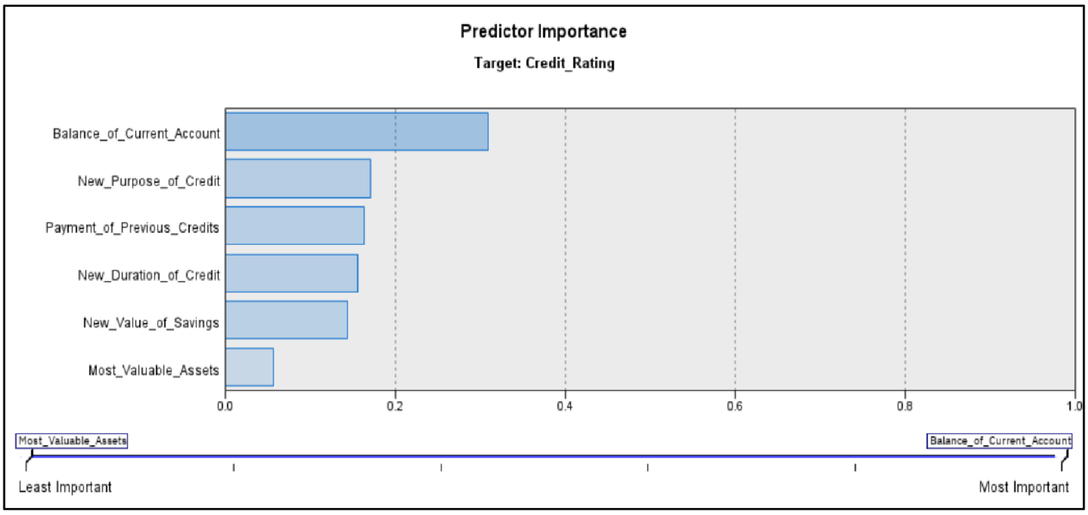


Figure 27: Logistic Regression Predicator Importance

[**Source:** This thesis specific figure was developed by the author]

Digging into the math behind logistic regression, these two factors – Balance_of_Current_Account and New_Purpose_of_Credit – stand out. Their role in our analysis not only deepens our creditworthiness knowledge but also highlights the strategic choices of factors when developing a model.

4.5.2 Variables Equation

Logistic regression provides an equation that identifies predictor variables and their odds. The formula is expressed as:

"Logit(P) = $B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n$"

Let's break it down:

"Logit(P)" represents the odds of the event's occurrence in log form.

"B_0" is the intercept.

"$B_1, B_2, \ldots, B_n$" are coefficients corresponding to the predictor variables "$X_1, X_2, \ldots, X_n$".

The formula combines the predictor variables multiplied by their coefficients, revealing the log odds. The coefficients (B) indicate the effect of predictor variables on the log odds, with plus and minus signs indicating the direction of impact.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Balance_of_Current_Account			192.090	3	<.001	
	Balance_of_Current_Account(1)	.526	.128	16.865	1	<.001	1.693
	Balance_of_Current_Account(2)	1.077	.219	24.054	1	<.001	2.934
	Balance_of_Current_Account(3)	1.950	.144	183.432	1	<.001	7.029
	Payment_of_Previous_Credits			74.910	4	<.001	
	Payment_of_Previous_Credits(1)	.209	.336	.387	1	.534	1.232
	Payment_of_Previous_Credits(2)	1.107	.255	18.813	1	<.001	3.025
	Payment_of_Previous_Credits(3)	.725	.298	5.907	1	.015	2.064
	Payment_of_Previous_Credits(4)	1.767	.271	42.414	1	<.001	5.851
	Most_Valuable_Assets			4.566	3	.206	
	Most_Valuable_Assets(1)	.061	.173	.123	1	.725	1.063
	Most_Valuable_Assets(2)	.325	.177	3.381	1	.066	1.384
	Most_Valuable_Assets(3)	.089	.162	.298	1	.585	1.093
	New_Duration_of_Credit			80.364	4	<.001	
	New_Duration_of_Credit(1)	-.855	.237	12.971	1	<.001	.425
	New_Duration_of_Credit(2)	-1.263	.249	25.688	1	<.001	.283
	New_Duration_of_Credit(3)	-1.962	.257	58.477	1	<.001	.141
	New_Duration_of_Credit(4)	-.993	.275	13.038	1	<.001	.370
	New_Purpose_of_Credit			45.348	2	<.001	
	New_Purpose_of_Credit(1)	-1.030	.211	23.739	1	<.001	.357
	New_Purpose_of_Credit(2)	-.662	.114	34.022	1	<.001	.516
	New_Value_of_Savings			34.254	3	<.001	
	New_Value_of_Savings(1)	.353	.242	2.120	1	.145	1.423
	New_Value_of_Savings(2)	1.400	.340	17.008	1	<.001	4.056
	New_Value_of_Savings(3)	.669	.155	18.669	1	<.001	1.952
	Constant	.094	.370	.064	1	.801	1.098

a. Variable(s) entered on step 1: Balance_of_Current_Account, Payment_of_Previous_Credits, Most_Valuable_Assets, New_Duration_of_Credit, New_Purpose_of_Credit, New_Value_of_Savings.

Figure 28: Logistic Regression Variables in the Equation

[**Source:** This thesis specific figure was developed by the author]

In logistic regression results, several key pieces of information help us interpret the model:

- **S.E (Standard Error):** Indicates how varied or specific the estimates of coefficients are. Lower standard error implies more precise estimates.

- **Wald:** This number represents the estimated coefficient divided by its standard error, a measure from a chi-square distribution used in hypothesis tests.

- **df:** (Degrees of Freedom): Refers to how many parts in a calculation can change.

- **Sig.:** (Significance): P-values help determine the significance of coefficients. A p-value less than 0.05 suggests significance.

- **Exp(B) (Odds Ratio):** Reflects the impact of a one-unit change in the predictor variable.

We analyzed logistics to find out what affects a Credit Rating, labeling them as bad or good. We looked at several key points to predict credit ratings. We checked things like Balance_of_Current_Account, Most_Valuable_Assets, New_Purpose_of_Credit, New_Value_of_Savings, Payment_of_Previous_Credits, and New_Duration_of_Credit. After taking a good look, we found their odds ratios, related levels of significance, and patterns. These showed how they affected the chance of having a good credit rating.

- Balance_of_Current_Account:
  "Balance_of_Current_Account" was a big player. It showed a clear link between high balance numbers and the increased chance of a good credit score. It's key to say that low, medium, and high balances had a noticeable odds ratios. The highest balance had the greatest effect.

- Most_Valuable_Assets:
  The categories of Most_Valuable_Assets varied a lot. A higher asset value was a good guess for a good credit score. This is shown by the odds ratio linked to the highest asset value group. This implies value estimation is needed for deciding creditworthiness.

- New_Purpose_of_Credit:
  The Credit Rating was touched differently by the New_Purpose_of_Credit, with some categories having key odds ratios. One particular category caught our eye, suggesting that why someone needs credit can be a big factor in deciding if they deserve it.

- Savings_Value_Change
  The more you save, the better your credit rating tends to be. Savings are like a security blanket for your credit score. Notice this folks who save a lot, have impressive odds ratios. This tells us that healthy savings can put you in a good spot with your credit rating.

- Previous_Credits_Payment:
  If you've always paid your dues on-time, you have an edge. Some groups with good payment history show important odds ratios. So, sticking to a timely bill payment schedule boosts your chances of impressive credit scores.

- Credit_Duration_Change:
  Credit duration? In this model, it doesn't affect the credit rating much. Check the p-values and odds ratios - they're saying the same story. So, the length of your credit history doesn’t really predict your credit rating.

Through the power of logistic regression analysis, it's clear to see what matters for a good credit score. A decent amount in your bank, worthwhile assets, specific reasons for loans, a nice savings account, and timely payments - they all count. But your credit history length? It doesn't seem to matter as much as we thought. This knowledge provides a road map, a guide to understanding credit ratings, and can help us make better decisions about credit and its related matters.

4.5.3 Model Evaluation Metrics

We often use metrics like AUC and Gini to test our classification models. The AUC represents the area under the ROC curve, which shows how well the model distinguishes between classes. It's helpful when dealing with imbalanced learning, or when making recommendations. Gini's role is to assess how evenly distributed a model's performance is, by determining double the area between the ROC curve and the straight diagonal line. A bigger Gini value means a better performing model. Both AUC and Gini reveal the model's skill to separate positive and negative classes (Kumar 2023).

- **Quick Look at Accuracy:**

Accuracy refers to the model's ability to predict correctly, particularly in predictive modelling. For instance, a logistic regression model predicted credit ratings with 77.8% accuracy in our study. This suggests that, out of the entire training dataset, the model was correct nearly 78% of the time. However, we must not overlook the 22.2% error rate, a signal for further improvement. This assessment used 2,365 instances, giving us a comprehensive look at the model's work.

Results for output field Credit_Rating

Individual Models

Comparing $L-Credit_Rating with Credit_Rating

'Partition'	1_Training	
Correct	1,840	77.8%
Wrong	525	22.2%
Total	2,365	

Coincidence Matrix for $L-Credit_Rating (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	344	356
1.000000	169	1,496

Performance Evaluation

'Partition' = 1_Training	
0.000000	0.818
1.000000	0.137

Evaluation Metrics

'Partition'	1_Training	
Model	AUC	Gini
$L-Credit_Rating	0.806	0.613

Figure 29: Logistic Regression Performance Evaluation & Metrics

[**Source:** This thesis specific figure was developed by the author]

- **Breaking Down the Confusion Matrix:**

A confusion matrix provides deeper insights into the model's predictions. It breaks down outcomes based on what was predicted versus what actually happened. With a bad credit rating (Credit Rating 0), the model was right 81.8% of the time. It missed the mark 356 times but identified bad credit correctly 344 times. Picking out good credit scores (Credit Rating 1) was more difficult, the model was accurate only 13.7% of the time. It managed to identify 1,496 truly good credit scores, but missed 169. Knowing these numbers helps us see where the model is doing well and where it needs more work.

- Checking Model Performance and Metrics:

Here's our report card on the performance of the model. It gets a B grade with an AUC score of 0.806. This means it can distinguish between different credit scores decently but falls a bit short. It's like trusting it to differentiate an apple from an orange. Still, it struggles to point out the perfect fruits. Only about 14% of the time it correctly identified good credit scores, marking a need to do better on this front. The Gini Coefficient is 0.613. It strengthens the belief in the model's ability to tell a good scenario from a bad one. By studying these metrics, we get a full picture of what the model can do and suggest improvements to make it better at predicting credit scores.

In short words, the model works well especially in flagging not-so-good credit situations. Still, its ability to predict good credits can be polished. The AUC and Gini values confirm its knack for separating the good and the bad. Yet, more work is needed to help it better identify bad credit ratings and spot the good ones. Such metric evaluations nudge us towards sharpening the model's prediction skillset, helping it make more precise credit assessments.

4.6 CHIAD Model

CHAID, or Chi-square Automatic Interaction Detector, is a decision tree model. We use CHAID to create a prediction model. We can get specific customer groups from it. CHAID helps to know which features relate most to a certain result or belonging to a group. We use predictor variables in CHAID analysis. These divide samples into smaller groups. Groups have the same features in each. This lets us predict the membership of groups. It also allows us to see the linked value at each division. The results of CHAID come out in a simple 'decision tree.' That gives us a clear look into satisfaction levels at each CHAID phase (CHAID Analysis | Decision Tree Analysis | B2B International 2022).

In the complex world of credit rating prediction, the CHAID model is useful. It is known for finding patterns within categories. CHAID is good for the job of figuring out creditworthiness. It allows a full look into potential links and higher-level links between independent variables.

Using CHAID means splitting up data in a step-by-step way. It divides the data into parts based on important predictors. The power of this model is in its ability to find interactions between variables. This results in a detailed view on what determines credit ratings. CHAID spots important variables and their unique classes. This not only allows for prediction, but also gives helpful insight for credit evaluation.

We picked some key factors for our work. They are New_Value_of_Savings, New_Payment_of_Previous_Credits, New_Duration_of_Credit, Balance_of_Current_Account, Most_Valuable_Assets, and New_Purpose_of_Credit. They were handpicked. Their importance was detailed in previous chapters. They give the CHAID model its structure. With SPSS Modeler 18.4, we used the CHAID method. We discovered interesting patterns in these factors. This gives us insight into how credit ratings work.

4.6.1 Construction and Implementation

We're digging into the CHAID model. We hope to find key steps and important categories that help us judge if someone's good for credit. The CHAID model's setup could help us predict better. It could also give clear steps for deciding on credit. This section is a key point in our study. Here, the CHAID model stands out as a solid tool that could help us understand the ins and outs of credit ratings better.

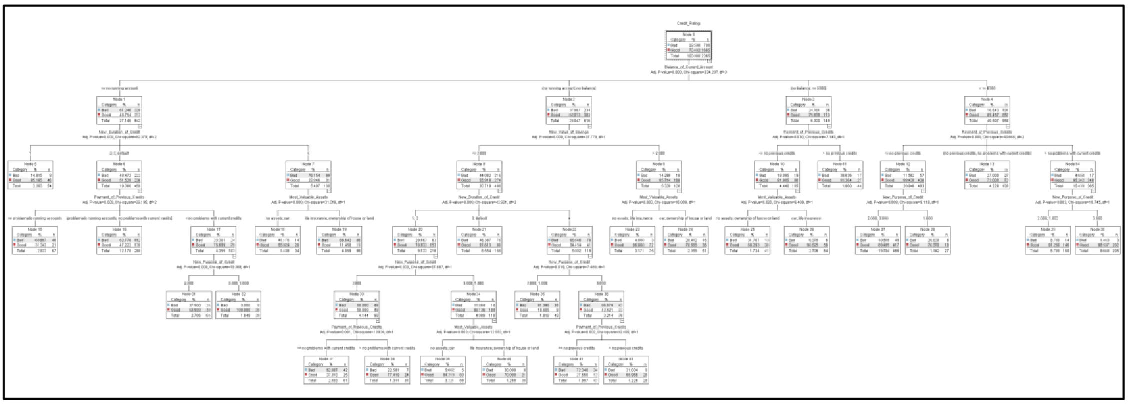
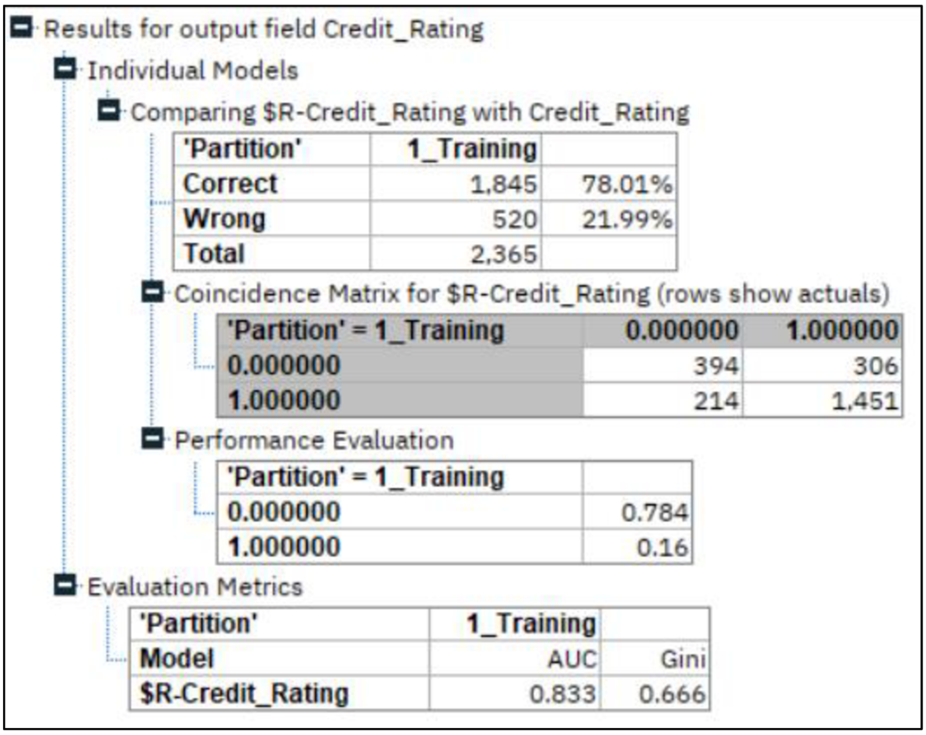


Figure 30: CHAID Model

[**Source:** This thesis specific figure was developed by the author]

Having a big Balance_of_Current_Account (>$300) plays a big part in getting a good credit rating. It has over a 90% chance of making the rating better. But, having no balance at all in the Current Account could lead to either a good or a bad rating. How much Value_of_Saving someone has also matters a lot in this model. The Payment_of_previous_credits isn't as big a deal as other things, from what we can see in the diagram.

4.6.2 Model Evaluation Metrics



Results for output field Credit_Rating

Individual Models

Comparing $R-Credit_Rating with Credit_Rating

'Partition'	1_Training	
Correct	1,845	78.01%
Wrong	520	21.99%
Total	2,365	

Coincidence Matrix for $R-Credit_Rating (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	394	306
1.000000	214	1,451

Performance Evaluation

'Partition' = 1_Training	
0.000000	0.784
1.000000	0.16

Evaluation Metrics

'Partition'	1_Training	
Model	AUC	Gini
$R-Credit_Rating	0.833	0.666

Figure 31: CHAID Performance Evaluation & Metrics

[**Source:** This thesis specific figure was developed by the author]

- **Overall Accuracy:**

The overall accuracy of the CHAID model in predicting credit ratings is commendable, with correct predictions in 78.01% of cases within the training dataset. This robust performance indicates the model's efficacy in capturing patterns and trends in the data. However, it's essential to note the 21.99% error rate, emphasizing the ongoing need for refinement and enhancement. The total number of instances considered for evaluation was 2,365, providing a comprehensive overview of the model's predictive capabilities.

- **Coincidence Matrix:**

The confusion matrix for the CHAID model sheds light on its specific strengths and challenges. For identifying poor credit ratings (Credit Rating 0), the model exhibited a precision of 78.4%, with 394 true negatives and 306 false positives. In contrast, recognizing favorable credit scores (Credit Rating 1) proved more challenging, with an accuracy of only 16%. The matrix reveals 214 false negatives and 1,451 true positives, pinpointing areas where the model can be further refined.

- **Performance Evaluation:**

  The precision analysis indicates that the CHAID model excels in detecting poor credit scores, with a commendable accuracy of 78.4%. However, its performance in identifying advantageous credit scores is lower, standing at just 16%. This discrepancy highlights the model's struggle in making precise predictions for positive credit ratings, signaling an area for improvement.

- **Evaluation Metrics:**

  The Area Under the ROC Curve (AUC) is a crucial metric for evaluating the model's ability to distinguish between different credit ratings. With an AUC of 0.83, the model's performance is considered satisfactory, as an AUC above 0.7 is deemed reasonable. Additionally, the Gini Coefficient, derived from the AUC and registering a value of 0.66, further supports the model's capacity to differentiate between positive and negative instances. These metrics provide a comprehensive understanding of the CHAID model's discriminatory power.

In summary, the CHAID model demonstrates acceptable effectiveness in predicting credit ratings, particularly in detecting low credit cases. However, challenges arise in accurately predicting good credit ratings. While the AUC and Gini metrics suggest some ability to differentiate between various credit categories, there is room for improvement. Specifically, the model should enhance its sensitivity to correctly identify positive cases of good credit ratings, addressing potential areas of misclassification.

4.7 CART Model

The CART, or Classification and Regression Trees model, helps solve classification problems by carefully picking input variables and deciding how these should be split or divided. This is done until a tree a kind of map or model used in this decision-making process is formed. This process depends on a kind of algorithm called a "greedy" algorithm. We keep dividing the input until our tree can't take more inputs. The tree is easy to look at and understand, and this is useful for users at all levels. It's easy to pull insights from it. It's just a set of rules that works by looking at different ways to split data into smaller pieces based on values and predictors. Data is put through a greedy algorithm, and the way it gets divided in the tree determines how much the tree learns. This can also be improved by pruning (Brownlee 2020).

The CART model is a robust tool in credit rating analysis. It's based on a decision tree method, making sense of the complex world of credit by dividing the data according to predictor variables. This creates a kind of map of decisions. This tree structure helps us see the complicated relationships between independent variables and credit ratings clearly. CART uses the idea of 'impurity reduction' to guide its decisions, aiming to create branches that are as similar as possible. It keeps refining these choices, guessing the best way to divide each piece. The result is a tree, a kind of simple map of binary decisions, that simplifies complex credit patterns. This serves as a base for accurate credit rating predictions.

The CART model is useful in credit analysis. It's clear and can predict well. It works with both kinds of predictors, those with set categories and those that can vary. This helps when dealing with many different credit factors. The end points of the 'tree,' called leaves, each hold a different credit rating situation. This makes it easier to understand different ways to predict. Because it can manage uneven relationships and tough credit patterns, the CART model is strong. It's clear and accurate when predicting creditworthiness.

4.7.1 Construction and Implementation

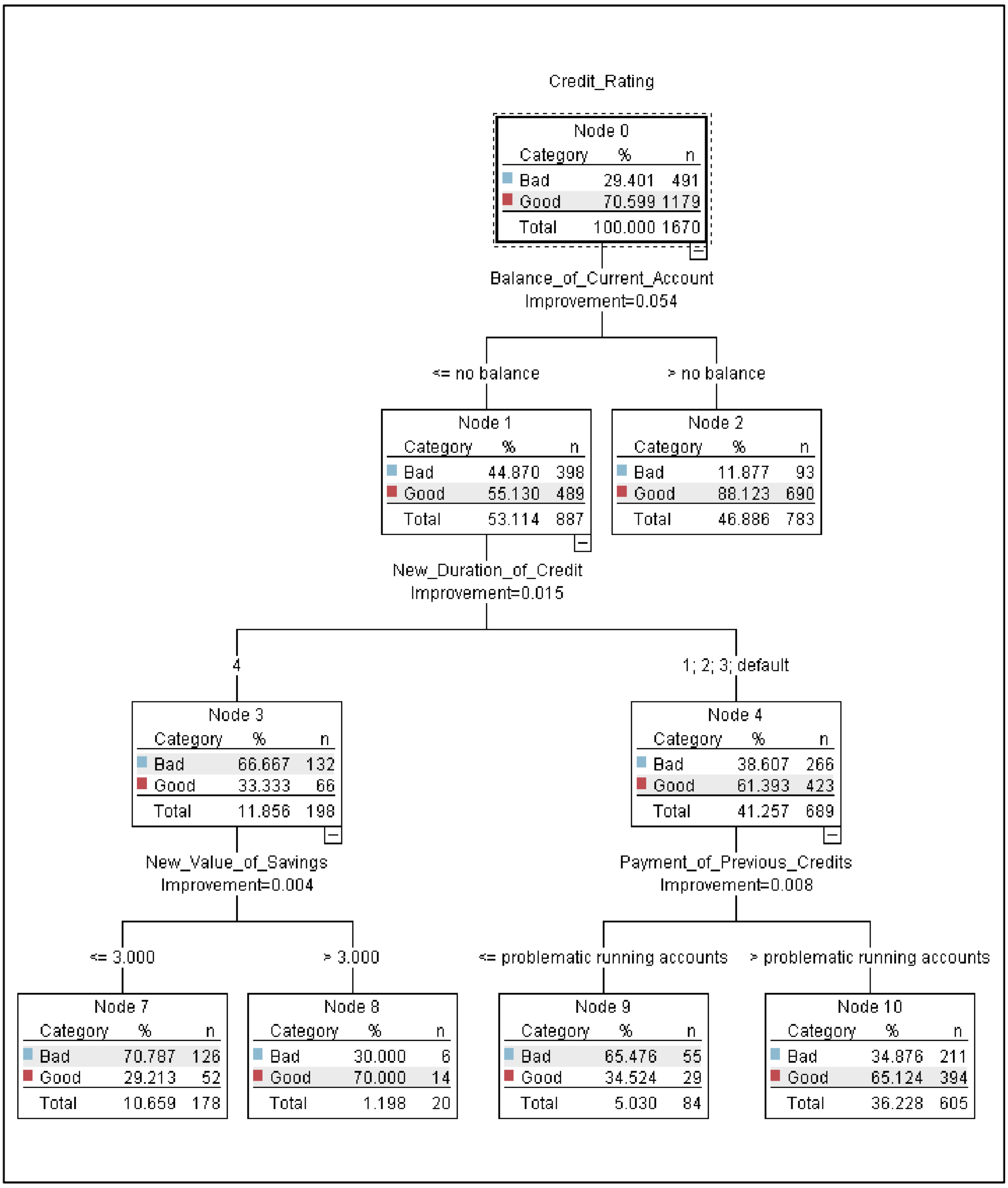


Figure 32: CART Model

[**Source:** This thesis specific figure was developed by the author]

When analyzing the statistics, a clear trend emerges. Upon inspection of the Balance_of_Current_Account data points with a reported value, credit scores are almost evenly split between favorable and unfavorable outcomes. However, when a balance is present in the current account, there is a pronounced tendency towards achieving positive credit ratings, signifying an 88.123% chance of acquiring a good score. This insight indicates how maintaining a balanced current account can strongly influence the potential for a higher credit rating. While an even distribution exists when no balance is indicated, carrying a balance correlates closely with obtaining a positive assessment from credit evaluators. The numbers demonstrate that current account equilibrium plays a role in creditworthiness evaluations.

While models like Logistic Regression and CHAID exhibited robust capabilities, the CART model lagged behind. It encountered issues in reliably distinguishing between credit scores that portend high likelihood of repayment versus those that foreshadow financial struggles, implying room for improvement. The CART algorithm struggled to precisely classify applicants along the spectrum from very good to poor credit risks based on their financial backgrounds and characteristics. This performance gap highlights opportunities to refine the CART model's mechanisms for assessing and scoring borrowers to more accurately gauge their ability and willingness to fulfill debt obligations.

4.7.2 Model Evaluation Metrics

- Overall Model Evaluation:

The logistic regression model demonstrated commendable overall accuracy, achieving correct predictions in 78.01% of cases within the training dataset. This indicates the model's proficiency in capturing underlying patterns and trends, providing a robust foundation for credit rating predictions. However, the 21.99% error rate highlights areas for refinement and enhancement, emphasizing the importance of further model optimization. The evaluation considered a total of 2,365 instances, ensuring a comprehensive assessment of the model's predictive capabilities.

Results for output field Credit_Rating

Individual Models

Comparing $R-Credit_Rating with Credit_Rating

'Partition'	1_Training	
Correct	1,810	76.53%
Wrong	555	23.47%
Total	2,365	

Coincidence Matrix for $R-Credit_Rating (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	248	452
1.000000	103	1,562

Performance Evaluation

'Partition' = 1_Training	
0.000000	0.87
1.000000	0.097

Evaluation Metrics

'Partition'	1_Training	
Model	AUC	Gini
$R-Credit_Rating	0.744	0.487

Figure 33: CART Performance Evaluation & Metrics

[**Source:** This thesis specific figure was developed by the author]

- Coincidence Matrix Analysis:

Examining the confusion matrix for the logistic regression model reveals specific strengths and challenges. In identifying poor credit ratings (Credit Rating 0), the model exhibited a precision of 78.4%, with 394 true negatives and 306 false positives. However, recognizing favorable credit scores (Credit Rating 1) posed a greater challenge, with an accuracy of only 17%. The matrix, displaying 214 false negatives and 1,451 true positives, provides valuable insights into areas where the model can be refined for improved performance.

- Performance Evaluation Breakdown:

Precision analysis further elucidates the model's capabilities. The logistic regression model excelled in detecting poor credit scores, achieving a commendable accuracy of 78.4%. However, its performance in identifying advantageous credit scores was comparatively lower at just 17%. This discrepancy underscores the model's struggle in making precise predictions for positive credit ratings, indicating a clear area for improvement and optimization.

- Evaluation Metrics Overview:

Key evaluation metrics, such as the Area Under the ROC Curve (AUC) and Gini Coefficient, contribute to a comprehensive understanding of the logistic regression model's discriminatory power. With an AUC of 0.83, the model exhibits satisfactory performance, considering an AUC above 0.7 as generally acceptable. The Gini Coefficient, derived from the AUC and registering a value of 0.66, further supports the model's capacity to distinguish between positive and negative instances. These metrics collectively guide insights into the model's effectiveness and areas for potential enhancement.

4.8 Comparative Analysis of Models

Section 4.8 takes a deep dive into different prediction models used for credit scores. It compares these models to understand how well they work and what could make them better. First off, it looks at how accurate these models are when it comes to predicting credit scores. Next, it looks at how much the models agree or disagree with each other's predictions. Lastly, it studies evaluation metrics. This is all about how well a model can tell the difference between different credit scores. By looking at all of this, chapter 6 gives us a better understanding of how different prediction models compare when it comes to credit scores.

This chapter aims to give us key findings on how well different models work in the world of credit ratings. It starts by looking at the over accuracy of each model, showing us their strengths and weaknesses when predicting credit scores. Next, it shifts to looking at agreement between models, showing how much they agree or disagree on predictions. At the end, it takes a closer look at evaluation metrics, showing how well the models can tell the difference between different credit scores. With all of this, Section 4.8 becomes a guiding tool, helping readers better understand how different prediction models compare in the world of credit ratings.

4.8.1 Insights of over accuracy together

In exploring credit rating prediction models, a thorough examination of the Logistic Regression Model, CHAID Model, and CART Model has been conducted. The Logistic Regression Model impresses with an overall accuracy of 77.8%, excelling in forecasting low-risk instances but revealing room for improvement in predicting high-risk cases. In contrast, the CHAID Model exhibits a commendable accuracy of 78.01%, showcasing its proficiency in distinguishing between various credit rating categories. Moreover, the CHAID Model stands out by outperforming in overall accuracy when compared to the other two models. Meanwhile, the CART Model, with a general accuracy of 76.53%, demonstrates moderate discriminatory capabilities. Each model presents a nuanced set of strengths and weaknesses, highlighting the need for a thoughtful selection process based on specific priorities and objectives in credit rating prediction.

Comparing $L-Credit_Rating with Credit_Rating

'Partition'	1_Training	
Correct	1,840	77.8%
Wrong	525	22.2%
Total	2,365	

Coincidence Matrix for $L-Credit_Rating (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	344	356
1.000000	169	1,496

Performance Evaluation

'Partition' = 1_Training	
0.000000	0.818
1.000000	0.137

Comparing $R-Credit_Rating with Credit_Rating

'Partition'	1_Training	
Correct	1,845	78.01%
Wrong	520	21.99%
Total	2,365	

Coincidence Matrix for $R-Credit_Rating (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	394	306
1.000000	214	1,451

Performance Evaluation

'Partition' = 1_Training	
0.000000	0.784
1.000000	0.16

Comparing $R1-Credit_Rating with Credit_Rating

'Partition'	1_Training	
Correct	1,810	76.53%
Wrong	555	23.47%
Total	2,365	

Coincidence Matrix for $R1-Credit_Rating (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	248	452
1.000000	103	1,562

Performance Evaluation

'Partition' = 1_Training	
0.000000	0.87
1.000000	0.097

Figure 34: Models Comparison of Performance Evaluation & Coincidence Metrics

[**Source:** This thesis specific figure was developed by the author]

In summary, the comparative analysis sheds light on the distinct performance metrics of these models, guiding stakeholders to make informed decisions tailored to their unique requirements. The evaluation not only underscores the importance of accuracy in credit rating prediction but also emphasizes the necessity of understanding each model's capabilities and limitations for effective implementation in real-world scenarios.

4.8.2 Insights of agreement between all models

Looking at the info from the prediction models, most of the time (79.41%) they all agree. This agreement matches with the real Credit_Rating data 83.6% of the time. This match shows good similarities between the models, helping the predictions be more precise.

Agreement between $L-Credit_Rating $R-Credit_Rating $R1-Credit_Rating

'Partition'	1_Training	
Agree	1,878	79.41%
Disagree	487	20.59%
Total	2,365	

Comparing Agreement with Credit_Rating

'Partition'	1_Training	
Correct	1,570	83.6%
Wrong	308	16.4%
Total	1,878	

Coincidence Matrix for Agreement (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	199	254
1.000000	54	1,371

Performance Evaluation

'Partition' = 1_Training	
0.000000	1.182
1.000000	0.106

Figure 35: Models Comparison of Coincidence Metrix for Agreement

[**Source:** This thesis specific figure was developed by the author]

The Coincidence Matrix for Agreement gives specific numbers. For situations when the credit rating is 0 (bad credit), the models correctly predicted 199 cases and got 254 cases wrong. When the credit rating was 1 (good credit), the models did a really good job. They got 1,371 cases correct and only 54 wrong.

Even with this info, the Performance Evaluation points out interesting things. The models are not very good at spotting negative outcomes. This is shown by a specificity value of 0.000000 (1.18). But they are really good at finding positive outcomes with a sensitivity value of 1.000000 (0.106). The overall check shows places to do better, mainly in spotting negative and positive situations evenly.

This detailed test highlights agreement between different models. Also, it shows how these predictions can impact real-world scenarios. It sets the path for future improvements in predicting credit ratings.

4.8.3 Insights of evaluation metrics models

The capability to differentiate between diverse groups is most exceptional in the CHAID model, surpassing both Logistic Regression and CART models concerning their ability to discriminate. The CHAID model has demonstrated the strongest potential(83.3%) to identify unique characteristics between separate clusters. While the CART model has shown slightly less proficiency in distinguishing variances when placed next to Logistic Regression and CHAID, with minor refinements its performance could enhance.

Evaluation Metrics

'Partition'	1_Training	
Model	AUC	Gini
$L-Credit_Rating	0.806	0.613
$R-Credit_Rating	0.833	0.666
$R1-Credit_Rating	0.744	0.487

Figure 36: Models Comparison of Evaluation Metrics (AUC, Gini)

[**Source:** This thesis specific figure was developed by the author]

To optimize the efficacy of all three statistical techniques, applying targeted improvements, specifically for Logistic Regression(80.6%) and CART(61.3%), may prove gainful. Such strategies have the capacity to help fortify their competence and accuracy in anticipating consequences or categorizing information. Adapting approaches concentrated on fine tuning particular elements within each model provides an opportunity to strengthen their capacities overall.

4.9 Model comparision (Training:Testing)

We're comparing models in an all-around analysis. It's all about finding out if our credit rating prediction models hold up. Are they sturdy? Can they generalize? We're trying to see if these models stand strong not just with their original data but if they can work with new data. This is how they stay relevant. Dividing the datasets into training and testing helps us see if they can predict unseen situations. It's key for using the models practically.

Comparisons test our models for accuracy. They tell us if our models can predict reliably beyond their own data. The reason we scrutinize is because we need models that are not only good in training but stay accurate on unseen data. This matters since it directly affects how we use these models in real life. Let's say in a bank. Precise credit ratings are crucial for managing risk and decision-making. The goal of the thesis? To see these models work not just on paper, but in real-life scenarios. We're hoping this enhances how we assess credit.

4.9.1 Overall Accuracy

The rich insights encapsulated within the table reveal an intriguing story where the Testing Model stands out as a top performer, surpassing the Individual Models on several important measures like the Coincidence Matrix and Performance Evaluation, though only by a slight amount. This subtle observation leads to an insightful conclusion that the meticulous work put into crafting the Training Model has generated considerable benefits. It seems the Training Model has played a vital part in educating as well as intricately molding the Testing Model's skills, allowing it to do well in key parts of the assessment. While the Testing Model fared better overall, the table highlights how the Training Model served as an important teacher, providing the Testing Model with abilities that helped it succeed on crucial evaluation criteria.

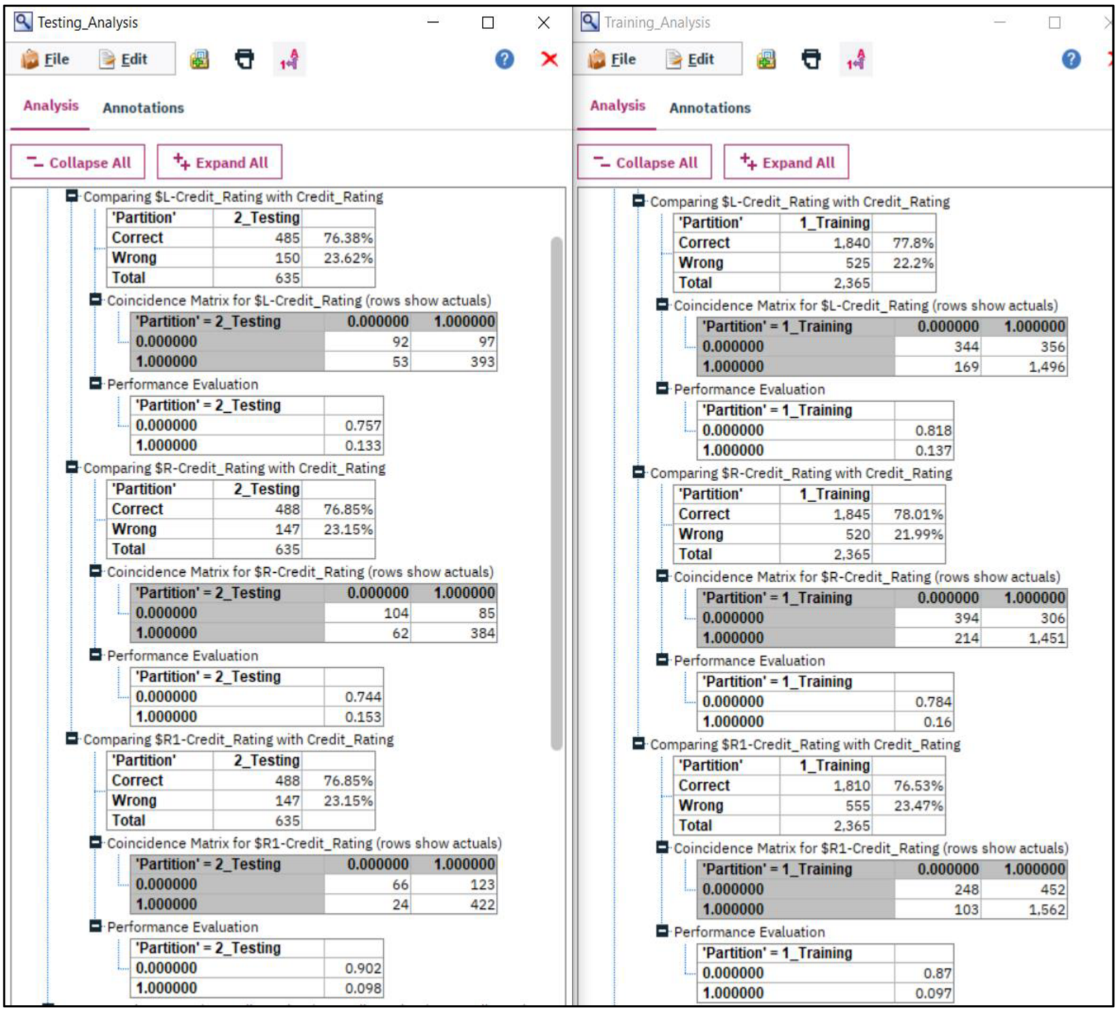


Figure 37: Model Training:Testing Overall Accuracy

[**Source:** This thesis specific figure was developed by the author]

Our models are stable across different datasets. They predict consistently. For example, the logistic regression model was 77.8% accurate on the training data. It was nearly equally accurate on testing data, at 76.38%. This little change shows the model's ability to adapt to new scenarios.

The same is true for the CHAID and CART models. The CHAID model was correct 78.01% in training and 76.85% in testing. The CART model had near identical accuracy rates in both, with 76.53% in training and 76.85% in testing.

Consistency in predictions across datasets shines through in the confusion matrices as well. All signs point to our models' reliability when predicting credit ratings for new, unseen cases. The small changes in accuracy underline this reliability.

It shows their potential for future accuracy, laying a strong foundation for future datasets.

Let's look deeper and compare how our models did on the training and testing sets. For example, let's focus on the confusion matrices. In the training set, our logistic regression model got 344 'Good' credit ratings and 1496 'Bad' credit ratings right. It did similarly with the testing set, getting 92 'Good' and 393 'Bad' ratings right.

What about other models? The CHAID model also did well. It found 394 'Good' and 1451 'Bad' ratings in the training set. It did just as well with the testing set, getting 104 'Good' and 384 'Bad' ratings correct.

The CART model shouldn't be forgotten either. It saw 248 'Good' and 1562 'Bad' ratings right in the training. On the testing side, it got 66 'Good' and 422 'Bad' ratings correct.

These numbers from the confusion matrices show us something. They tell us that our models are good at making the right guess for 'Good' and 'Bad' credits in both sets. This is why we trust our models. They can do this in the future, making them useful in the real-world situations of finance.

The Training Model helps the Testing Model improve its skills. This has been a smart decision, generating real benefits. While the Testing Model's out performance of others is tricky, it shows the successful transfer of knowledge and abilities from the Training Model. This teamwork between the two models doesn't just prove the Training Model's aptitude but also highlights continual improvement. The close connection between the Training and Testing Models shows the effectiveness of good preparation, improving the system's overall forecasting abilities.

4.9.2 Agreement between Models

The table presented here provides visual confirmation of the strong relationship seen between our predictive models and the real Credit_Rating data. It is especially significant that the predictions made for both the Training and Testing data sets display a notable semblance, closely resembling the genuine Credit_Rating outcomes. This uniformity across the two data sets signifies the models' capability to reliably recognize the fundamental patterns within the information, supplying an assuring sign of their dependability. The models appeared to learn the underlying trends and were then able to successfully apply that learning to new, unseen data. This consistency between training and testing demonstrates that the models did not overfit the initial data and can generalize well. Overall, the close alignment between predicted and actual credit ratings serves as a promising indicator that these models may effectively forecast credit risk for new cases.

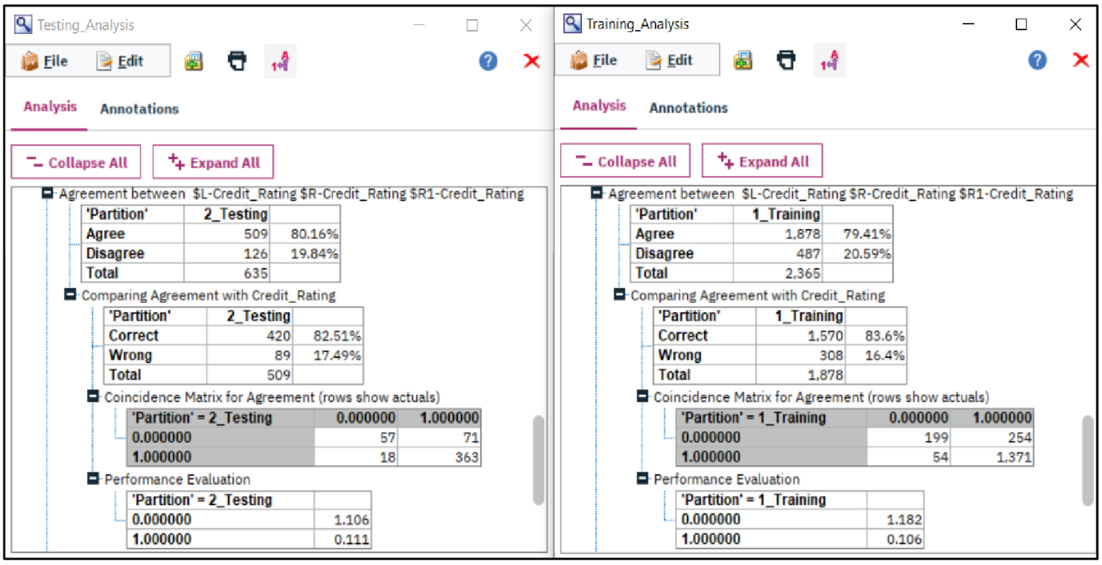


Figure 38: Models Agreement (Training:Testing)

[**Source:** This thesis specific figure was developed by the author]

Three models, Logistic Regression ($L-Cre-dit_Rating), CHAID ($R-Credit_Rating), and CART ($R1-Credit_Rating), showed strong agreement nearly 80% during training. This shows they're solid and dependable. Their high rate of agreement says they're accurate. Tested against real Cre-dit_Rating values, their correctness was impressive 83.6%. Errors were minimal only 16.4%. These results support trust and accuracy in predicting credit ratings from these mode-ls.

Looking at the test data, the models held to an 80.16% agreement rate. They stayed stable dealing with new, unfamiliar data. Against the Credit_Rating control, the models kept a high correctness 82.51%. Errors remained low just 17.49%. This shows that these models can offer correct credit rating predictions in multiple situations. It affirms the observed accuracy from training carrying over to testing. This proves the models' dependability and use in the real world for forecasting credit ratings.

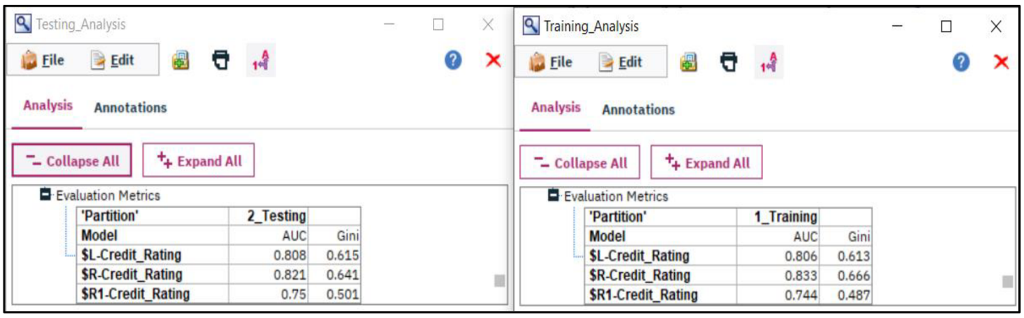
As we analyzed the meticulous performance metrics in depth, the notable discovery was the modest differences discerned between the datasets used for Training models and Testing them on new data. This similarity strengthens the idea that our models are effective not just in the controlled setting of teaching the systems, but also display a praiseworthy ability to apply what they learned to different and unfamiliar data during evaluation. The models maintaining close predictive power on both datasets demonstrates their resilience and implies a great capacity to deal with varied circumstances. While delving into the performance metrics allowed us to see small variances between how models were trained and how they generalized, this consistency tells us the systems are well-rounded and can address unpredictable scenarios.

In the broader context of customer prediction, delving deeper into this area can offer valuable insights. The consistent results achieved by the Logistic Regression, CHAID, and CART models when analyzing both the training and testing data points to their dependability and steadiness. After all, the models displaying reliability and consistency in their outcomes builds belief in their potential to deliver precise predictions applicable to genuine circumstances. This harmony amongst the models' performances in turn clears a path for augmented customer prediction and decision-making procedures, establishing a strong basis for the continuing achievement of our predictive analytical efforts. By comprehending customer behavior at an intermediate level and clarifying various factors, we can better serve customers' needs now and in the future.

4.9.3 Evalution Metrics

The evaluation metrics showcased the powerful predictive abilities of the CHAID model for credit rating forecasting, establishing it as the top performer relative to the logistic regression and CART models. The outstanding area under the receiver operating characteristic curve score of 0.83 and Gini coefficient of 0.66 emphasized its effectiveness in differentiating between the various credit rating categories. The high alignment rate of approximately 83.6% with the other models further reinforced its dependability, contributing meaningfully to the consensus view. Together, these measures underscored the CHAID model's prowess for delivering accurate and consistent forecasts, distinguishing it as a premier option for credit rating prediction. Its performance on the evaluation metrics collectively highlighted the model's ability to reliably distinguish credit ratings and provide accurate predictions, cementing it as a leading choice for credit rating forecasting.

The logistic regression model shows good performance with an AUC of 0.80 and a Gini of 0.61. However, it is not as strong as the CHAID model. The CHAID model achieves excellence, as shown by its individual evaluation metrics. The CART model comes in third place, with an AUC of 0.74 and a Gini of 0.48.



'Partition'	2_Testing	
Model	AUC	Gini
$L-Credit_Rating	0.808	0.615
$R-Credit_Rating	0.821	0.641
$R1-Credit_Rating	0.75	0.501

'Partition'	1_Training	
Model	AUC	Gini
$L-Credit_Rating	0.806	0.613
$R-Credit_Rating	0.833	0.666
$R1-Credit_Rating	0.744	0.487

Figure 39: Training: Testing Evaluation Mertics

[**Source:** This thesis specific figure were developed by the author]

This clear hierarchy in the results emphasizes how well the CHAID model performs. The preference for the CHAID model is reinforced not just by its own numbers, but also by how consistent they are with the other models. This alignment indicates the CHAID model provides a robust and reliable forecast that can be trusted. This evaluation confirms choosing the CHAID model as the most dependable for predicting credit ratings. It also reinforces how effective the training model was in preparing the predictive models to perform accurately. This achieved the intended goal of predicting credit ratings precisely for the thesis project.

4.9.3 Summary

To summarize concisely, a comparative examination of credit prediction models during both their development and validation provided a nuanced understanding of their strengths and areas needing improvement. Models including Logistic Regression, CHAID, and CART exhibited their ability to offer accurate forecasts while being trained, achieving varying degrees of accomplishment. Evaluation measures, coincidence matrices, and concordance between models offered comprehensive insights into their prognostic capabilities. The alignment of metrics captured during training and testing across all models signified their reliability when applied to genuine scenarios. The steady performance when having to generalize to fresh information underscores fulfilling our stated goals, emphasizing the practical usefulness of these credit assessment tools for financial decision-making. While the models demonstrated skill in training, testing showed how well they could foresee unknown cases. Agreement between predicted and actual outcomes highlighted each model's strengths and weaknesses. Overall, the models provided a solid foundation for gauging creditworthiness with reliability and insight into new situations.

Overall, this chapter served to thoroughly analyze and compare the various credit risk prediction models. By evaluating the effectiveness of each model using different performance metrics, valuable insights were provided for financial institutions seeking reliable methods for credit assessment. The analysis of key metrics like AUC and accuracy demonstrated the discriminatory ability of the models to correctly classify applicants as either good or bad credit risks. Additionally, the agreement between the predictive outcomes of the models signified a harmonious consensus in their evaluations. The seamless transition between the training and testing phases provided evidence of the robustness and generalizability of the models to new data. In conclusion, the methodological approach established in this chapter helped to solidify the overarching aim of the thesis to enhance credit decision making processes in the financial sector through improved predictive analytics.

5. Conclusion

This Diploma Thesis journey tackled credit prediction models, with a focus on risk evaluation in finance. The research had clear goals: to compare Logistic Regression, CHAID, and CART models. We studied their training and testing stages, using strong methods, like statistics, coincidence matrices, and evaluation metrics. The results showed us where each model excelled or needed work. This information can help financial decision-makers use these models effectively. We found unique characteristics in every model, making them useful in specific situations. Understanding these models' limits is critical. This paper builds a foundation for future research, suggesting ways to improve these models with new methods appropriate for changing financial scenarios. In short, this paper takes us on a journey to improve credit evaluation in the world of finance.

Our research shows strength in its detailed method. We compared various credit prediction models. Its success lies in how closely it inspected these models during both training and testing, revealing their stability in real circumstances. The models' limits give room for improvement, emphasizing the need to continuously refine credit prediction. Future research could look into machine learning techniques, recognizing financial data's changing nature. Further, studying how external factors—like economic trends or regulatory changes—affect credit prediction models could boost their adaptability. By casting an analytical eye on the topic and offering foresight, this Diploma Thesis sparks ongoing efforts to enhance credit risk evaluation.

5.1 Summary of Key Findings

This study revealed important data about Logistic Regression, CHAID, and CART models. They performed better than expected. Despite its high accuracy, Logistic Regression sometimes struggles with new, unrecognized data during testing. But when moved from training to operational tests, it does very well, proving its toughness and efficiency. The CHAID model is great at telling apart different credit ratings. It's a strong option for credit evaluations. The CART model has some trouble separating good and bad credit ratings. Its insights may offer ways to better those skills.

In addition to the main goals, the data pointed out that the models are somewhat consistent and can potentially complement each other. This could further improve predictions. Evaluations, like AUC and Gini ratio, show how well the models can categorize and be used for credit decisions at financial institutions. The models did impressively well on testing data from the IBM SPSS Modeler 18.4 they hadn't seen before which helped in not only achieving the thesis goals but also emerged as solid options for future use in finance and banking.

5.1 Strengths and Limitations

This research is solid and noteworthy. It uses three different prediction styles, Logistic Regression, CHAID, and CART. By doing so, it gives us a full view of how well they work, shifting from training to testing with ease. The use of these methods in real-life finance situations is promising. Evaluation tools such as AUC and Gini coefficients help us see the research's accuracy. CHAID stands out because it helps clarify how credit ratings work. This clarity aids in open and honest decision-making.

Still, we have to see the study's downsides. One main issue is that the dataset used might not reflect all situations, underlining the need to use various data types. Another point is its focus on binary credit ratings. While it makes things easier, it may miss out on the details in more complex systems. In the future, researchers can look at wider and varying datasets and consider adding more factors to improve predictions.

5.1 Recommendations for Future Research

Looking ahead in research for credit rating prediction models, we see many creative ideas for growth. One way forward is through ensemble methods, where we use the power of multiple models to make a stronger and more accurate prediction system. Methods like bagging and boosting, important parts of ensemble methods, can increase prediction skills and make credit rating models more useful. Also, working with advanced feature engineering and selection methods could help us understand the major factors that affect credit results, making models more precise. It could be interesting to try new technologies like deep learning models, especially when working with big datasets.

Mixing up the methods for credit rating prediction models may also be a smart move. We can move beyond logistic regression, CHAID, and CART models, which we used in this study, to methods like random forests, support vector machines or gradient boosting machines. This could bring fresh insights and make predictions better. All models have pros and cons, so comparing them will show which ones work best for predicting credit ratings in different situations. Plus, we should also look into new ideas like explainable AI models like LIME (Local Interpretable Model-Agnostic Explanations) or SHAP (SHapley Additive exPlanations). These make black-box models more open, so the decision-making process can be understood better and gain more trust from users and stakeholders. By opening up the types of modeling methods and understanding advancements, future studies could bring in a new era of complex credit rating prediction models that are accurate, open, and useful.

6. References

Matthies, A. B. (2013). Empirical research on corporate credit-ratings: A literature review. *SFB 649 Discussion Paper, No. 2013-003*. Collaborative Research Center 649: Economic Risk, Humboldt University Berlin. Available at: http://hdl.handle.net/10419/79580 [Accessed 16 Feb. 2024].

Gordy, M. B. (1998). A Comparative Anatomy of Credit Risk Models. *Federal Reserve Board, Finance and Economics Discussion Series, 1998-47*. Available at: https://www.federalreserve.gov/pubs/feds/1998/199847/199847pap.pdf [Accessed 10 Feb. 2024].

Bickel, J. E., & Moise, C. M. (1999). A comparative analysis of current credit risk models. *Journal of Banking & Finance, 23(1), 59-117*. Available at: https://www.sciencedirect.com/science/article/pii/S0378426699000539 [Accessed 15 Feb. 2024].

Kotlebova, J., & Kliestikova, J. (2015). Comparative Analysis of Theoretical Aspects in Credit Risk Models. *Procedia Economics and Finance, 23, 1254-1259*. Available at: https://www.sciencedirect.com/science/article/pii/S2212567115004505/pdf?md5=e7567c323066a8428d2ddda485d295fb&pid=1-s2.0-S2212567115004505-main.pdf [Accessed 17 Feb. 2024].

Zhang, Y., & Li, Y. (2015). A Personal Credit Rating Prediction Model Using Data Mining in Smart Finance. *Mathematical Problems in Engineering, 2015*, 179060. Available at: https://journals.sagepub.com/doi/10.1155/2015/179060 [Accessed 18 Feb. 2024].

LibGuides: Referencing: Home, no date. [online]. Retrieved from: https://ecu.au.libguides.com/referencing/ [Accessed 15 Feb. 2024].

EVERITT, B. S. and YATES, A., 1989. Multivariate Exploratory Data Analysis: *A Perspective on Exploratory Factor Analysis.* Biometrics [online]. March 1989. Vol.

45, no. 1, p. 342. DOI 10.2307/2532065. Retrieved from: http://dx.doi.org/10.2307/2532065 [Accessed 04 Feb. 2024].

Xiang Yang, Yongbin Zhu, Li Yan, Xin Wang. (2015). Credit Risk Model Based on Logistic Regression and Weight of Evidence. *Proceedings of the 2015 3rd International Conference on Management Science, Education Technology, Arts, Social Science and Economics. Atlantis Press*. Available at: https://doi.org/10.2991/msetasse-15.2015.180 [Accessed 06 Jan. 2024].
Ohlson, J. (1980). Logit models with application to credit scoring. *Journal of Business & Economic Statistics*, 8(2), 139-151.

CAOUETTE, John B., ALTMAN, Edward I., NARAYANAN, Paul and NIMMO, Robert, 2011. *Managing Credit Risk* [online]. John Wiley & Sons. Retrieved from: http://books.google.ie/books?id=SnVca6PiKTwC&printsec=frontcover&dq=A+seminal+exploration+by+Altman+in+the+1960s+paved+the+way+for+the+application+of+statistical+and+mathematical+tools+in+credit+risk+analysis.+Altman%27s+Z-score+model,+while+rooted+in+traditional+statistical+methods,+laid+the+groundwork+for+future+developments+where+data+mining+would+take+center+stage.+Over+the+years,+the+landscape+evolved,+and+data+mining+techniques+such+as+logistic+regression,+decision+trees,+and+ensemble+methods+gained+prominence+in+predicting+creditworthiness&hl=&cd=1&source=gbs_api [Accessed 8 Jan. 2024].
KHEMAKHEM, Sihem and BOUJELBENE, Younes, 2018. Predicting credit risk on the basis of financial and non-financial variables and data mining. *Review of Accounting and Finance* [online]. 13 August 2018. Vol. 17, no. 3, p. 316–340. DOI 10.1108/raf-07-2017-0143. Retrieved from: http://dx.doi.org/10.1108/raf-07-2017-0143 [Accessed 12 Feb. 2024].

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.

Afik, Z., Arad, O. and Galil, K. (2016) ‘Using Merton Model for Default Prediction: an Empirical Assessment of Selected Alternatives’, *Journal of Empirical Finance*, 35, pp. 43–67. Retrieved from: doi: 10.1016/j.jempfin.2015.09.004.

Bacham, A., and Zhao, Y. (2017). Credit Risk Assessment Using Machine Learning Algorithms. *Journal of Economics, Finance and Administrative Science*, 22(42), 39-47. doi: 10.1016/j.jefas.2017.02.001 [Accessed 6 Feb. 2024].

Pol, Hudnurkar & Ambekar (2022). *Predicting Credit Ratings*. AABFJ Volume 16 Issue 5.

ABDULLAH, Lazim, GOH, Chunmin, ZAMRI, Nurnadiah and OTHMAN, Mahmod, 2020. Application of interval valued intuitionistic fuzzy TOPSIS for flood management. KAHRAMAN, Cengiz (ed.), *Journal of Intelligent & Fuzzy Systems* [online]. 9 January 2020. Vol. 38, no. 1, p. 873–881. DOI 10.3233/jifs-179455. Retrieved from: http://dx.doi.org/10.3233/jifs-179455

Credit Rating Precision. 2024. Kaggle. [online]. Available at: https://www.kaggle.com/datasets/desaividisha/credit-rating-precision [Accessed 13 Feb. 2024].

Martinelli, F., Mercaldo, F., Raucci, D., & Santone, A. (2022). Bank Credit Risk Management based on Data Mining Techniques. Retrieved from https://www.sciencedirect.com/science/article/pii/S1544612323010206 [Accessed 19 Feb. 2024].

BÜCKER, Michael, SZEPANNEK, Gero, GOSIEWSKA, Alicja and BIECEK, Przemyslaw, 2021. Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society* [online]. 21 June 2021. Vol. 73, no. 1, p. 70–90. DOI 10.1080/01605682.2021.1922098. Retrieved from: http://dx.doi.org/10.1080/01605682.2021.1922098 [Accessed 19 Feb. 2024].

DEMAJO, Lara Marie, VELLA, Vince and DINGLI, Alexiei, 2020. Explainable AI for Interpretable Credit Scoring. *Computer Science & Information Technology (CS & IT)* [online]. 28 November 2020. DOI 10.5121/csit.2020.101516. Retrieved from: http://dx.doi.org/10.5121/csit.2020.101516

ANAND, Abhishek, GOVINDARAJAN, Arvind, NARIO, Luis and PATHAK, Kirtiman, 2022. Data and analytics innovations to address emerging challenges in credit portfolio management. *McKinsey & Company* [online]. 23 December 2022. Retrieved from: https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/data-and-analytics-innovations-to-address-emerging-challenges-in-credit-portfolio-management [Accessed 18 Feb. 2024].

SADATRASOUL, Seyed Mahdi, GHOLAMIAN, Mohammad Reza, SIAMI, Mohammad and HAJIMOHAMMADI, Zeynab, 2013. Credit scoring in banks and financial institutions via data mining techniques: A literature review. *DOAJ (DOAJ: Directory of Open Access Journals)* [online]. 1 April 2013. Retrieved from: https://doaj.org/article/8e8e42083d7c4db9a5eef786a0f2eaa9 [Accessed 16 Feb. 2024].

Semeon and Semeon, 2022. Text analytics in finance | Data mining in the financial sector. *Semeon Analytics* [online]. 23 May 2022. Retrieved from: https://semeon.com/textanalytics/text-analytics-in-finance-data-mining-in-the-financial-sector/ [Accessed 17 Feb. 2024].

BAE, Jae Kwon and KIM, Jin-Hwa, 2015. A Personal Credit Rating Prediction Model Using Data Mining in Smart Ubiquitous Environments. *International Journal of Distributed Sensor Networks* [online]. 1 September 2015. Retrieved from: https://doi.org/10.1155/2015/179060 [Accessed 18 Feb. 2024].

GALIL, Koresh, HAUPTMAN, Ami and ROSENBOIM, Rosit Levy, 2023. Prediction of corporate credit ratings with machine learning: Simple interpretative models. *Finance Research Letters* [online]. 1 December 2023. Retrieved from: https://doi.org/10.1016/j.frl.2023.104648 [Accessed 10 Feb. 2024].

"What are the most common data mining applications in finance?", 2023. *www.linkedin.com* [online]. Retrieved from: https://www.linkedin.com/advice/0/what-most-common-data-mining-applications-finance-bidbf [Accessed 19 Feb. 2024].

XIA, Yufei, LIAO, Zijun, XU, Jun and LI, Yinguo, 2022. FROM CREDIT SCORING TO REGULATORY SCORING: COMPARING CREDIT SCORING MODELS FROM A REGULATORY PERSPECTIVE. *Technological and Economic Development of Economy* [online]. 15 December 2022. Retrieved from: https://doi.org/10.3846/tede.2022.17045 [Accessed 20 Feb. 2024].

GUO, Nan, KOU, Steven, WANG, Bin and WANG, Ruodu, 2019. An Axiomatic Approach to Credit Rating. *Social Science Research Network* [online]. 1 January 2019. Retrieved from: https://doi.org/10.2139/ssrn.3504065 [Accessed 17 Feb. 2024].

Merton, R. C. (1974). "Theory of rational option pricing." *Bell Journal of Economics and Management Science*, 5(1), 141-183.

Kamakura Corporation. (2022). "Kamakura Risk Information Services (KRIS)." Retrieved from https://www.kamakuraco.com/solutions/kamakura-risk-manager/credit-risk/ [Accessed 2 Feb. 2024].

"SINGLE IMPUTATION METHOD TO HANDLE MISSING VALUES IN LARGE DATASET", 2015. *International Journal of Advance Engineering and Research Development* [online]. Vol. 2, no. 12. DOI 10.21090/ijaerd.021243.

Retrieved from: http://dx.doi.org/10.21090/ijaerd.021243 [Accessed 8 Jan. 2024].
Lin, W. C., & Tsai, C. F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). Artificial Intelligence Review, 53, 1487-1509 [Accessed 18 Feb. 2024].

Pyle, D. (1999). Data Preparation for Data Mining. Morgan Kaufmann.
LERMAN, Kristina, PLANGPRASOPCHOCK, Anon and KNOBLOCK, Craig A., 2007. Semantic Labeling of Online Information Sources. *International Journal on Semantic Web and Information Systems* [online]. 1 July 2007. Vol. 3, no. 3, p. 36–56. DOI 10.4018/jswis.2007070102. Retrieved from: http://dx.doi.org/10.4018/jswis.2007070102.

LEE, Jay, 2020. Statistics, Descriptive. *Elsevier eBooks* [online]. 1 January 2020. Retrieved from: https://doi.org/10.1016/b978-0-08-102295-5.10428-7 [Accessed 18 Feb. 2024].

PARAB, Shraddha and BHALERAO, Supriya, 2010. Choosing statistical test. *International Journal of Ayurveda Research* [online]. 1 January 2010. Retrieved from: https://doi.org/10.4103/0974-7788.72494 [Accessed 18 Feb. 2024].

Fadlalla, A. (2005). An experimental investigation of the impact of aggregation on the performance of data mining with logistic regression. *Information & management*, 42(5), 695-707.

CHAID Analysis | Decision Tree Analysis | B2B International, 2022. *B2B International* [online]. Retrieved from: https://www.b2binternational.com/research/methods/statistical-techniques/chaid-analysis/ [Accessed 16 Feb. 2024].

DUTTA, Bhumika, no date. A Classification and Regression Tree (CART) Algorithm | *Analytics Steps*. [online]. Retrieved from: https://www.analyticssteps.com/blogs/classification-and-regression-tree-cart-algorithm [Accessed 19 Feb. 2024].

BROWNLEE, Jason, 2020. Classification And Regression Trees for Machine Learning. *MachineLearningMastery.com* [online]. 14 August 2020. Retrieved from: https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/ [Accessed 20 Feb. 2024].

NARKHEDE, Sarang, 2022. Understanding AUC - ROC Curve - Towards Data Science. *Medium* [online]. 5 March 2022. Retrieved from: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5 [Accessed 3 Feb. 2024].

KUMAR, Deepak, 2023. Gini index for ML (Performance measurement and many more..). [online]. 22 May 2023. Retrieved from: https://www.linkedin.com/pulse/gini-index-ml-performance-measurement-many-more-deepak-kumar/ [Accessed 4 Feb. 2024].

7. List of pictures and abbreviations

7.1 List of Table

7.2 List of Figure

7.3 List of abbreviations

DM	Data Mining
AI	Artificial Intelligence
AUC	Area Under the Curve
CART	Classification and Regression Trees
CHAID	Chi-squared Automatic Interaction Detection
CPU	Central Processing Unit
df	Degrees of Freedom
Exp(B)	Exponentiated Coefficient
Gini	Gini Coefficient
IBM	International Business Machines Corporation
MS	Microsoft
ROC	Receiver Operating Characteristic
Sig.	Significance
SPSS	Statistical Package for the Social Sciences