# Abstract

This dissertation explores the growing role of alternative credit scoring systems and their impact on lending practices within the fintech loan application landscape. By utilizing advanced machine learning techniques, such as Logit and Probit Regression (LR), Support Vector Machine (SVM), and Deep Neural Networks (DNNs), this research aims to replicate lender loan acceptance decisions and predict default risks. The proposed two-phase model differentiates between predicting loan rejections in the first phase and assessing default risks for approved loans in the second phase, providing insights into customer churn.

A holistic analysis of the credit scoring model of Paytm, a prominent lender, reveals the significant influence of alternative data, particularly the quality of digital transactions, on loan outcomes. This raises concerns about the equitable implementation of opaque models that reshape risk evaluation. Expanding the investigation, the study examines three fintech-bank partnerships and evaluates the capital allocation of banks participating in instant loan platforms. This assessment highlights the growing focus on loans facilitated through collaborations between fintech startups and banks, indicating a notable shift in the allocation of bank balance sheets.

The research demonstrates that alternative credit scoring systems play a crucial role in the calculative infrastructure, enabling specific institutions to overcome obstacles associated with risk-based pricing. These systems also serve as strategic collaboration points between technology startups and financial entities, capitalizing on new revenue streams.

Furthermore, the study focuses on loans utilized for personal and entrepreneurial purposes, revealing an interesting dichotomy in model performance. The first phase demonstrates superior performance when trained on the entire dataset, while the second phase excels when focused on the small business subset. This discovery emphasizes potential variations in the screening and default prediction processes for small business loans.

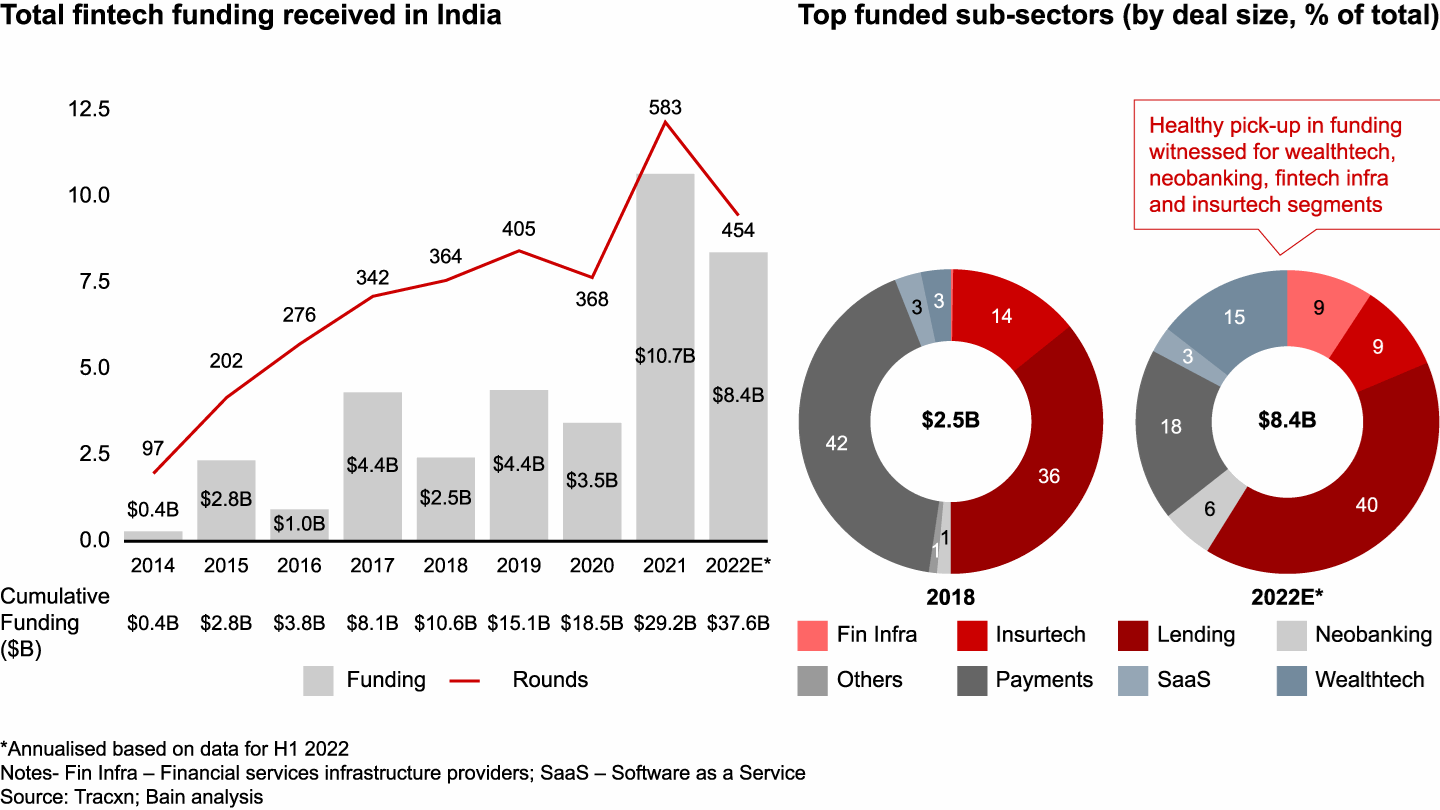
This research establishes the potential of machine learning algorithms in refining credit risk assessment, enhancing informed lending decisions, and effectively forecasting loan defaults. By harnessing artificial intelligence, financial institutions can improve risk management frameworks and strengthen the soundness of lending practices.

# Index

|  |  |  |
| --- | --- | --- |
| **SL. NO.** | **TOPIC** | **PAGE NO.** |
| 1 | **Chapter 1 – Introduction** | 1-5 |
| 2 | **Chapter 2 – Organizational Profile** | 6-11 |
| 3 | **Chapter 3 – Review of Literature and Research Design** | 12-16 |
| 4 | **Chapter 4 Data Analysis and Interpretation** | 17-22 |
| 5 | **Chapter 5 – Summary of Findings, Conclusion and Suggestions** | 23-28 |
| 6 | **Bibliography and Appendices**  **25.3.1 References**  **25.3.2 Questionnaire**  **25.3.3 Synopsis Registration Form**  **25.3.4 Progress Report**  **25.3.5 Plagiarism Report**  **25.3.6 Any other relevant documents** | 29-34 |

# Chapter 1 Introduction

In today's fast-changing world, lending and borrowing money has become more intricate than ever. Especially in places like Bangalore, where technology is booming and people have diverse financial needs, figuring out who might not pay back their loans on time is a big challenge. The rapid strides made in the expanses of machine learning (ML) and artificial intelligence (AI) have spawned the emergence of alternative credit scoring systems, fundamentally transforming how lenders gauge risk and make pivotal investment choices within the sphere of consumer debt. These novel models, leveraging unconventional data sources and sophisticated processing methods, herald a crucial departure from the norms of risk-based pricing - a cornerstone for financial institutions. Risk-based pricing delineates interest rates based on the anticipated odds of loan defaults. In this context, the rise of alternative credit scoring, supported by algorithmic models, surpasses prevailing obstacles and ushers in a comprehensive re-evaluation of risk assessment.



**Growth of Indian Fintech Industry**

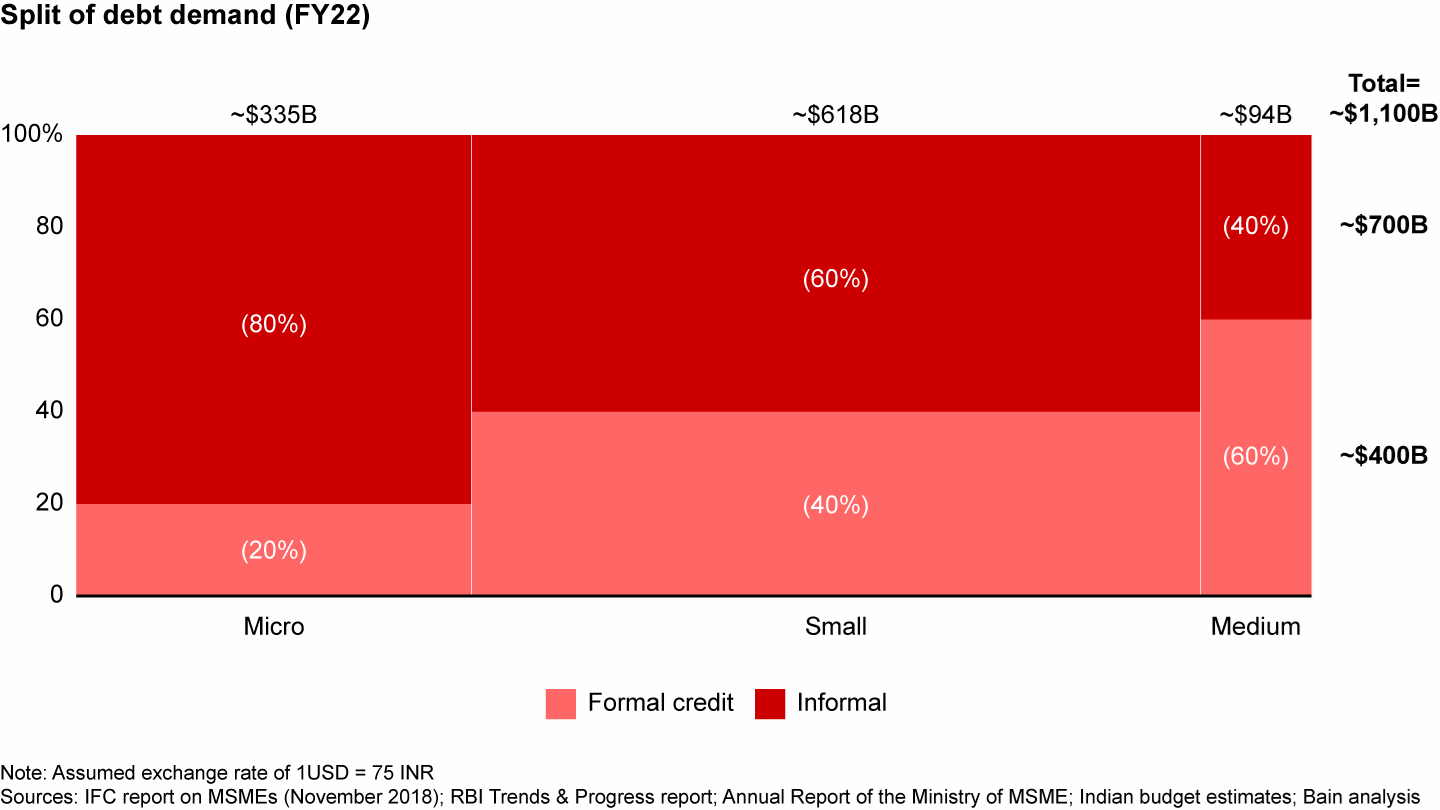
**Source: Bain & Company** [**[1]**](https://www.bain.com/insights/india-fintech-report-2022-sailing-through-turbulent-tides/)

In the world of consumer finance, when individuals apply for credit cards or auto loans, lending institutions traditionally rely on credit scores as a crucial factor in determining approvals. Credit scoring involves a statistical assessment conducted by financial institutions or credit bureaus to evaluate a borrower's creditworthiness. This assessment takes into account various factors such as credit history, demographic information, and credit behavior. Generally, borrowers with higher credit scores have a better chance of securing loans at lower interest rates. The importance of credit scoring for financial institutions goes beyond assessing the risk of default; even slight improvements in credit scoring mechanisms can lead to significant financial gains.

This dissertation is all about finding a smarter way to predict if someone might not be able to pay back the money they borrowed.We're going to look at two important things: how people decide to stop borrowing money (we call this "customer churn"), and how they behave with money they've borrowed (which we call "credit behavior"). These things can tell us a lot about whether they'll be able to pay back their loans.

Coming from a detailed study of how Paytm, a major lending company, assesses credit using advanced computer methods, this research brings attention to the big role those different kinds of information, like having various assets as collateral, past financial behaviours and financial literacy, play in deciding if someone will be able to pay back a loan. This discovery gleams a light on the small but important details that go into deciding the risks of lending and raises important questions about fairness when using complicated credit scoring algorithms to make these decisions. This study doesn't just stick to one company; it also looks at how tech companies and banks work together. By closely looking at how banks are doing financially and how they partner with platforms like CIBIL, this research shows that there's a clear change in how money is being given out as loans.

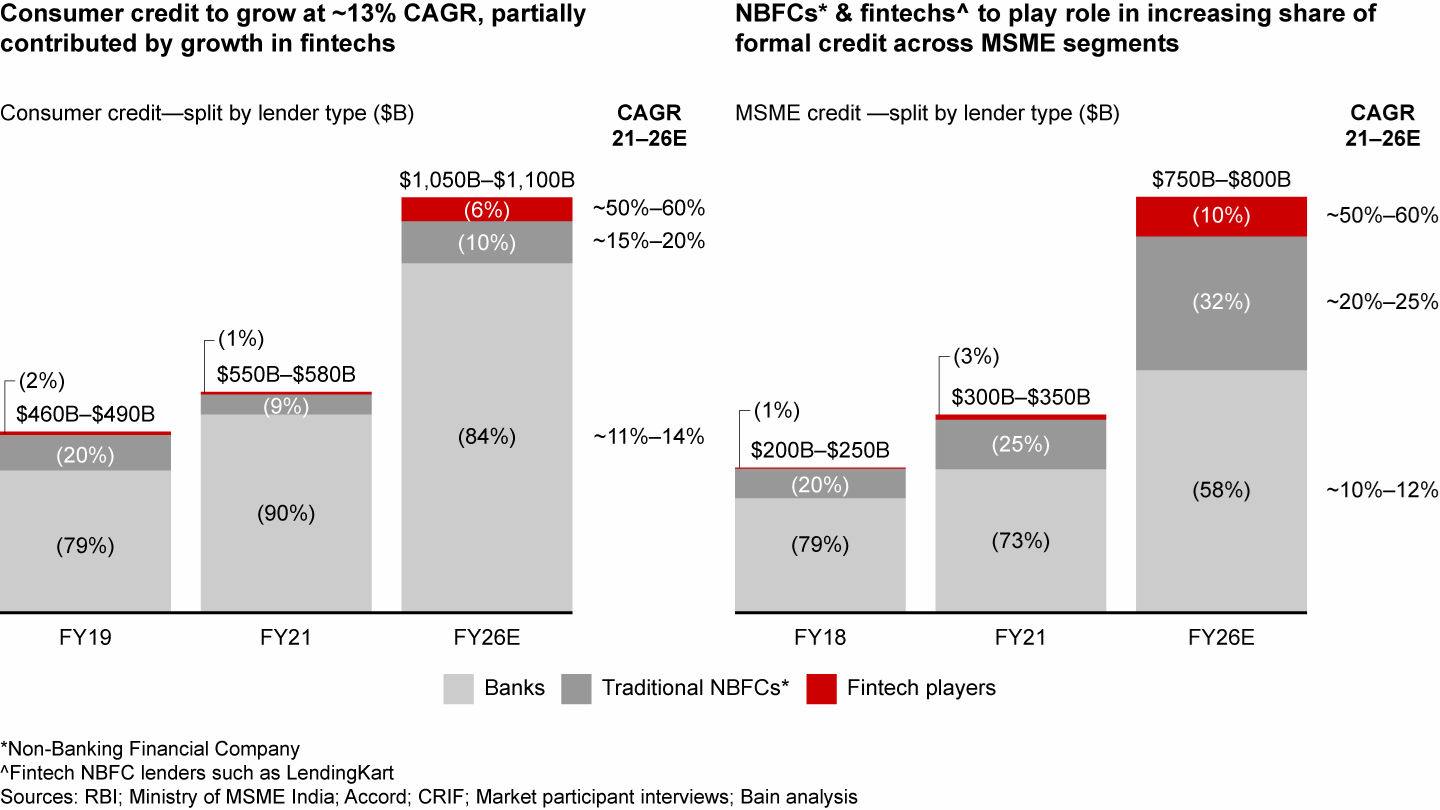
Bangalore is a special place for this study because it's a mix of different businesses and a growing technology industry. This means lots of people have different kinds of jobs and financial situations. We'll use information from the past to make a smart system that can predict if someone might not pay back their loan in the future. Moreover, the study directs its focus onto loans sought for personal and entrepreneurial aspirations as an encapsulated domain of inquiry. Intriguingly, the performance of computational models manifests divergence when encompassing the entire dataset vis-à-vis concentrating on loans for small-scale business ventures. This anomaly underscores potential discrepancies in the assessment processes and mechanisms for projecting loan defaults in this specific loan category.



**Indian Debt Demand**

**Source: Bain & Company (Oct 2022)** [**[2]**](https://www.bain.com/insights/india-fintech-report-2022-sailing-through-turbulent-tides/)

However, it is important to note that a significant portion of the population lacks an established credit history, particularly in emerging economies where credit reporting systems may be underdeveloped. This lack of credit information hinders financial institutions' ability to establish a robust credit scoring framework that can differentiate high-risk borrowers from the larger pool of applicants, which becomes even more challenging in the context of peer-to-peer (P2P) lending platforms. Furthermore, the absence of credit data can lead to deserving candidates being denied access to credit. With the increasing prominence of P2P lending, payday loans, and online microlending markets in developing economies, it is crucial for financial institutions to explore more sophisticated methods of assessing borrowers' likelihood of default.



**Indian Consumer Credit Market Growth**

**Source: Bain & Company** [**[3]**](https://www.bain.com/insights/india-fintech-report-2022-sailing-through-turbulent-tides/)

Currently accounting for about 7% of India’s $1.4 trillion FS EV, the fintech sector is expected to grow to $350 billion in EV by 2026, representing nearly 15% of FS market cap. Covid-19 led to transformational shifts in consumer behaviour and accelerated digital adoption:

1. Non-cash payments soared, with more than 75% year-over-year (YoY) growth in UPI transactions between FY20–21.
2. Digital lending apps (DLAs) accounted for more than 60% of loans disbursed by nonbank financial companies (NBFCs) in FY21.
3. Over 35 million demat accounts were added in FY22 (till Nov’21), thereby increasing the tally of demat accounts by 63%, from 55 million in FY21 to nearly 90 million in FY22.

To make this smart system, we're going to use a type of machine learning technique that's really good at understanding patterns called "Ayuda neural network algorithms” which is supplemented with Various types of regression models. These will help us analyse all the information we gather from people and predict whether they might have trouble paying back the money they borrowed.

Even during the Covid-19 pandemic, loans for shopping without any security have been steadily increasing, growing by about 25% in the last three years (from FY19 to FY22):

* Credit cards have shown strength with a growth of about 19% every year.
* Personal loans have grown really well, increasing by around 29% every year.
* Loans for buying things like appliances have also bounced back to the levels before the pandemic and have grown by about 13% every year.

The places that are not big cities have been the main reason for this growth in loans without security. In the smaller towns (called Tier 4), these loans have grown by about 32% every year, while in the big cities (called Tier 1), they have grown by about 18% every year over the last three years.

The amount of money people are borrowing for different kinds of loans, like personal loans and loans for appliances, is getting smaller. This is because more companies that lend money (like fintech and NBFC lenders) are focusing on giving out smaller loans. The average amount of money people borrow from NBFCs for personal loans has gone down by 70% in the last two years.

For credit card loans, the smaller loans (less than $650) have been growing really well, about 12% every year for the last three years. This is happening mostly in the smaller towns. Personal loans for smaller amounts (less than $650) have been growing really fast too, about 120% every year, and most of these loans (85%) are given to people younger than 35. Just like credit card loans, this growth is also happening more in the smaller towns.

Loans for buying appliances in smaller amounts have also bounced back, growing by about 11% every year in the last three years. More than 70% of these loans have been given to people under 40, with about 36% going to people aged 30 to 40, and 37% to people under 30. This shows that many younger people, like millennials and Gen Z, want loans for buying things. People who are getting a loan for the first time are really good at repaying their credit card loans, but they are having some trouble repaying personal loans and loans for appliances.

All of this shows that there is a clear trend towards smaller loans for shopping, especially among younger people. The growth in the smaller towns and among new customers also tells us that companies are using the internet more to give out loans and using different kinds of information to decide who should get a loan.

A big part of this research is understanding how people who might not have a good credit history – like those who borrow money instantly without much checking – can still be understood and predicted. These risky loans are a big challenge, and we want to figure out how to make lending money to these people safer for everyone.

In a nutshell, this dissertation is all about using new technology to understand people's money habits better and predict if they might not pay back their loans. By doing this, we hope to make lending and borrowing money more secure and reliable. This research will help us to understand the tricky ways we figure out who gets loans and how it's changing. We'll look at how computer stuff works to learn about money, check how it affects real life, and see how tech companies and banks team up. This will help us see how new fancy technologies are changing how we decide if giving out loans is safe or not.

Keywords: predictive model, loan defaults, customer churn, credit behaviour, technology, financial habits, instant loans, credit profile.

# Chapter 2 – Organizational Profile

***About Flipcarbon Integrated Solutions Pvt Ltd***

Flipcarbon Group was set up in 2014 to provide complete strategy, execution design, and

deployment solutions to growth-oriented organizations. It works closely with startups, SMEs,

MSME Organizations, Large Multinationals, and family-run businesses to unlock the full

potential of the business.

The following tenets comprise the core of what is done at Flipcarbon:

• A commitment to rapid growth.

• A culture that enables rapid growth through autonomy but controls costs through discipline.

• Capability practices that not only understand the dynamics of doing new things but also the

people implications of attempting to do new things.

• Developing and deploying more competitive people.

Flipcarbon: Virtual CFO & Accounting Services is the wing of operations where I operated.

Flipcarbon’s ‘specific-solution-to-specific-issues-of-specific-client’ orientation comes from

our inspiration, Carbon, which forms unique bonds for unique circumstances with unique

elements to give us unique compounds

***Background and Current Description***

Flipcarbon Integrated Solutions Private Limited is an unlisted private company incorporated on 04 December 2014. It is classified as a private limited company and is located in Bangalore, Karnataka. Its authorized share capital is INR 1.00 Cr and the total paid-up capital is INR 91.50 lac.

Description: The company provides HR & Financial Consulting, end-to-end HR & Financial Outsourcing, Modular HR Services, and Financial Services

***Products & Services of CFO Vertical:***

CFO Advisory, Accounting, taxation and MIS services, Payroll processing, Handling PF and ESI Compliances, Company secretary functions for Indian Private Limited Companies. Virtual CFO, Accounting & Bookkeeping Services for Growing Startups & Businesses. Cashflow, Cost & Working Capital Management, Business Plans, Valuation Modelling & Financial Modelling, Board & Management Reporting, Investment Advisory, Transaction Support, Fundraising, Financial Strategy & Planning

***Our Team***

The company has 3 directors and no reported key management personnel. The longest-serving director currently on board is Prabhash Nirbhay was appointed on 04 December 2014. Interestingly he is also the founder of this Organisation, Prabhash Nirbhay has been on the board for more than 8 years and is also the founder of the organization. Additionally, Lokesh Mehta was appointed as a director on 10th August 2015. The most recently appointed director is Alok Ranjan, who was appointed on 01 March 2018 and is also the current acting CEO of the organization. Prabhash Nirbhay has the largest number of other directorships with a seat at a total of 6

companies. In total, the company is connected to 5 other companies through its directors.

The Bangalore branch has over 13 Employees while the Flipcarbon family has over 30+ experts in 3 branches.

The CFO vertical where I interned is headed by Mr. Deepak Kewalramani, who is the founding partner for CFO services and brings almost 3 decades of expertise in financial analytics and strategic advisory. I reported to Ms. Pramila Lakra who has almost half a decade’s experience in Financial Due Diligence and Forecasting and another strategic advisory in the field of Finance relating to Mergers and acquisitions, funding advisory, etc.

The organization was in need of a business intelligence expert who could automate the process of preparing their financial statements and prepare eye catchy and informative visualizations using Microsoft Suite and Power BI, I fit into the role without any friction as I had a background in both finance and business analytics. I was able to leverage my experience as a business analyst and identify interoperability problems between different verticals of the firm and suggested a framework for the integration of solutions which was a tech-stack (SaaS).

With more than 400 years of combined management experience, eight years of consulting

experience, and an ever-growing list of more than 130+ absolutely ecstatic clients, and 20+

Experts, we are poised to be the gold standard in management consulting.

# Chapter 3 – Review of Literature and Research Design

# Review of Literature

Customer Churn Prediction for Fintech Companies Using Artificial Neural Networks (Pooja Malhotra , Punit Patel , Neel Shah) (2020)

The study's findings highlight the pressing need for advanced techniques in predicting customer churn. By leveraging the potential of Artificial Neural Networks and data mining, businesses can devise strategies to retain customers and counteract revenue loss. The future direction for this field lies in refining and enhancing the accuracy of predictive models, thus enabling businesses to stay ahead in the competitive landscape. The study underscores the significance of understanding customer behavior, utilizing cutting-edge technologies, and crafting proactive strategies to ensure enduring success in the face of churn challenges.

In the modern business landscape, the immense surge in data production across industries has been pivotal. This data, when harnessed effectively, holds the key to a company's success. Extracting meaningful insights from raw data falls under the umbrella of Data Mining, which involves uncovering concealed knowledge through advanced technologies. Customer-generated data constitutes a significant portion of this information, rendering customers a valuable asset. Beyond acquiring new customers, companies now strive to retain existing ones. Customer churn signifies the cessation of business between a company and its customers, underscoring the importance of accurate customer behavior prediction and relationship-building.

In industries like banking, customer relationships are paramount, especially for long-term commitments like home loans. The sophisticated nature of modern customers demands personalized offers and incentives to maintain their loyalty. Evidently, customer churn poses a major hindrance to organizational growth. As such, efficient churn calculation methods are essential. This enables organizations to closely monitor customer behavior, anticipate churn in its early stages, and enact preventive strategies.

**Conclusion:** This study underscores the significance of accurately predicting customer churn using advanced techniques. While traditional algorithms such as SVM, Random Forest, and Linear Regression fall short of desired accuracy levels, the application of Artificial Neural Networks emerges as a promising solution. This technology-driven approach leverages the power of data mining to unearth valuable insights from customer-generated data, enabling companies to proactively counter churn by tailoring strategies to retain customers.

By forecasting customer behavior accurately, businesses, particularly those in the fintech sector, can anticipate and mitigate churn. This aids in nurturing lasting customer relationships and promoting sustained growth. As industries continue to grapple with the challenges posed by churn, the integration of Artificial Neural Networks into churn prediction models presents a pathway to maintaining customer loyalty and maximizing business success.

**Keywords:** Customer Churn, Data Mining, Artificial Neural Networks, Linear Regression, Support Vector Machine, Random Forest.

An Empirical Study on Loan Default Prediction Models (Uzair Aslam, Hafiz Ilyas Tariq Aziz, Asim Sohail, and Nowshath Kadhar Batcha) (2019)

This article's findings suggest that the integration of machine learning algorithms and neural networks holds promise for improving credit risk assessment and loan approval processes. It encourages further research into refining these predictive models to enhance their accuracy and applicability in real-world scenarios. Ultimately, the study contributes to the evolving landscape of credit risk assessment, offering insights that can guide both researchers and practitioners in the finance industry.

The article delves into the significance of loan lending in the financial world and acknowledges its role in overcoming financial constraints for individuals and businesses. However, the concept of credit risk, associated with the possibility of borrowers failing to repay loans on time, is highlighted as a major concern. This risk, termed credit default, can lead to severe consequences for both parties involved. Despite the risks, loan lending remains a lucrative endeavour for financial institutions, contributing to profit-making and business sustainability.

Traditionally, creditworthiness assessment relied on credit scores assigned to individuals based on their historical data. However, modern advancements in machine learning and neural networks have revolutionized the approach to credit risk assessment. The article underscores how machine learning algorithms can autonomously predict credit scores by analysing individuals' historical data, enabling lenders to identify potential defaulters and manage risks more effectively. The practice of loan lending, while advantageous for both lenders and borrowers, comes with inherent risks, notably credit risk. This study explores existing literature that focuses on predictive models for credit risk assessment, utilizing machine learning algorithms. The objective of the paper is to present a comprehensive understanding of loan default prediction models and their implications.

Machine Learning Based Customer Churn Prediction In Banking (Manas Rahman, V Kumar) (2020)

This study presents a method that employs machine learning techniques, a subset of artificial intelligence, to predict customer churn in the banking industry. The research explores the potential of identifying churn by analysing customer behaviour. The study employs classifiers like KNN, SVM, Decision Tree, and Random Forest, along with feature selection methods to enhance system performance. The research is conducted on a churn modelling dataset from Kaggle, aiming to identify an optimal model with higher precision and predictability.

Recognizing and understanding consumers is essential for corporations. This concept is reinforced by Liu and Shih, who stress the need for innovative marketing strategies to meet consumer expectations and enhance loyalty. The competitive market necessitates a more effective utilization of marketing resources, and technology, including data mining techniques, serves as a tool to extract marketing insights and guide business decisions. The paper underscores the critical importance of predicting customer churn at early stages. The proposed framework leverages machine learning techniques to predict customer churn in the banking sector. This approach holds potential for helping organizations retain customers by identifying potential churn cases. The study emphasizes the need for predictive models that aid in retaining customer loyalty and boosting organizational success.

The study's findings hold implications for the banking sector, emphasizing the value of early-stage churn prediction. The integration of machine learning techniques offers a novel avenue for addressing customer churn. Future research could delve into refining and optimizing the proposed framework, potentially incorporating more advanced machine learning algorithms. Additionally, exploring how this framework can be adapted to other industries could lead to broader applications and insights.

The study, situated in the banking sector, unveils the potential of machine learning techniques in predicting customer churn. The dynamic nature of the market necessitates innovative approaches to customer retention. As corporations strive to balance customer acquisition and retention, predictive models fueled by machine learning offer a promising solution. The research contributes to the evolving landscape of customer churn prediction, guiding future research and aiding the banking industry in tackling this pivotal challenge.

Keywords: Customer Churn, Machine Learning, Banking, Predictive Models, Data Mining.

Sample selection in credit-scoring models (William Greene) (1998)

The paper "Sample Selection in Credit-Scoring Models" delves into the intricate realm of credit scoring, a vital process for financial institutions like credit-card vendors. Credit scoring involves evaluating loan applicants based on statistical models. However, a significant challenge arises due to the sample selection process, as models are constructed using historical data from individuals who have already received loans. This introduces a potential bias in evaluating new applicants randomly drawn from the entire population. The paper scrutinizes this sample selection issue across three distinct applications, each requiring a different statistical model and estimation technique.

The study examines three applications impacted by sample selection, shedding light on how this phenomenon influences the measurement of variables pertinent to credit-card vendors. The applications entail predicting loan default, modeling expenditure, and assessing the number of derogatory reports in credit histories. Each application demands a specific statistical model and estimation technique, accounting for the sample selection issue.

In credit assessment scenarios, loan officers meticulously evaluate large loan applications, whereas credit-card vendors, dealing with millions of applications, rely on statistical models to assign scores. The models sort applications based on statistical profiles of successful borrowers, offering a statistical measure to gauge applicants. However, a fundamental problem emerges due to the data used to build these models. Models are built using data from individuals whose applications were accepted, while evaluations must be made for randomly arriving applications from the entire population.

The paper's findings have crucial implications for credit-card vendors and financial institutions, emphasizing the potential bias introduced by sample selection. This has repercussions on predictive accuracy and business decisions. The study encourages the development of models that address the sample selection problem to yield more accurate predictions.

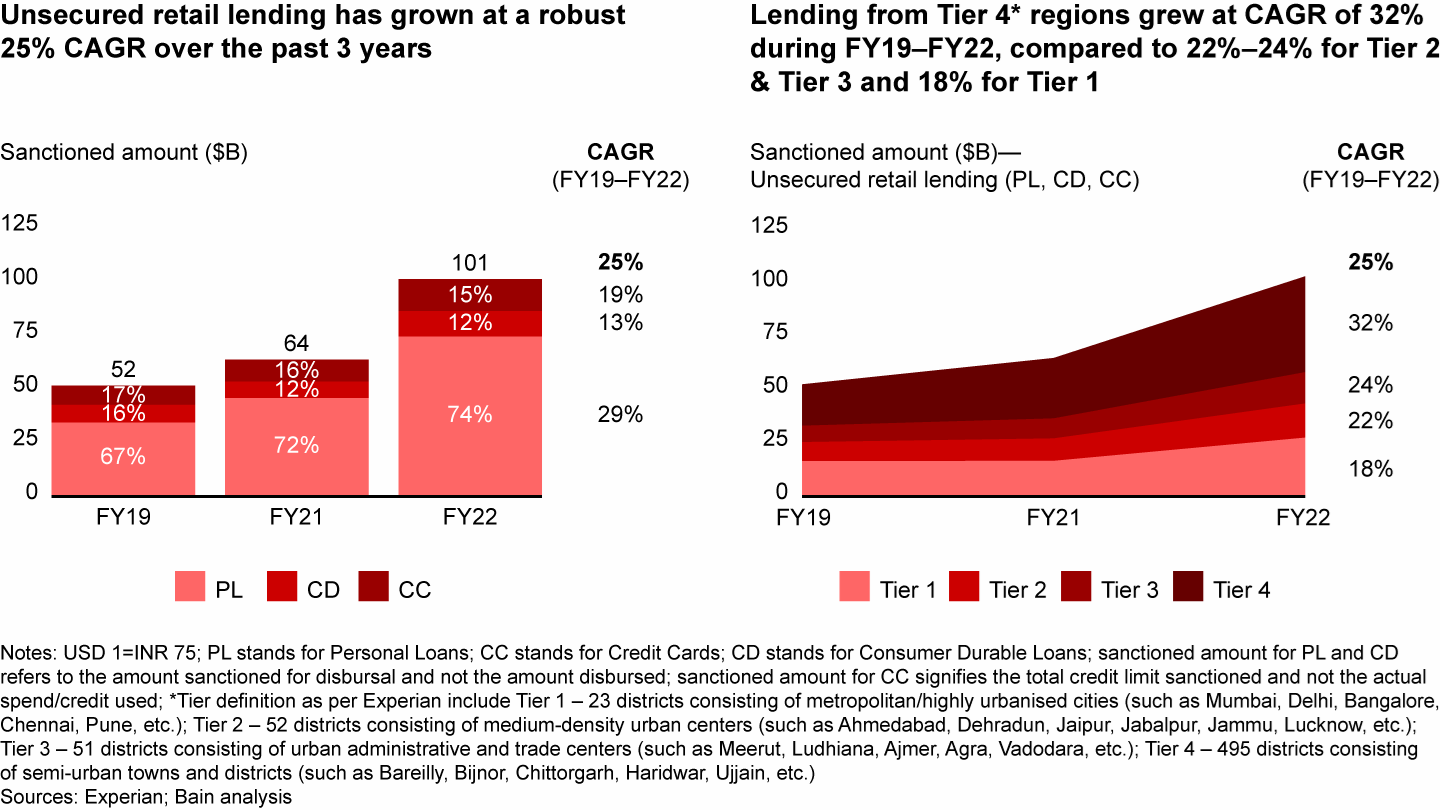
The study illuminates the necessity of constructing models that account for sample selection biases, enhancing the precision of credit-scoring predictions and contributing to a more accurate evaluation of loan applicants.

**Keywords:** Sample Selection, Credit-Scoring Models, Loan Default Prediction, Statistical Models, Financial Institutions

# Problem Statement

Credit risk assessment is of paramount importance for fintech companies, particularly those operating in the online lending market, where peer-to-peer lending, RBI Approved loan apps which provide unsecured loans to their customers and institutions who deal with other vulnerable derivatives which are prone to credit defaults and customer churns. Traditional credit scoring methods, such as credit bureau scores, have limitations in capturing the complex and dynamic factors that influence borrowers' repayment behavior. Moreover, they heavily rely on historical data, which might not accurately reflect the current economic and social conditions of borrowers. In the context of Bangalore, the Silicon Valley of India, the challenges posed by an unregulated credit market add to the urgency of developing a more advanced and robust credit scoring model.

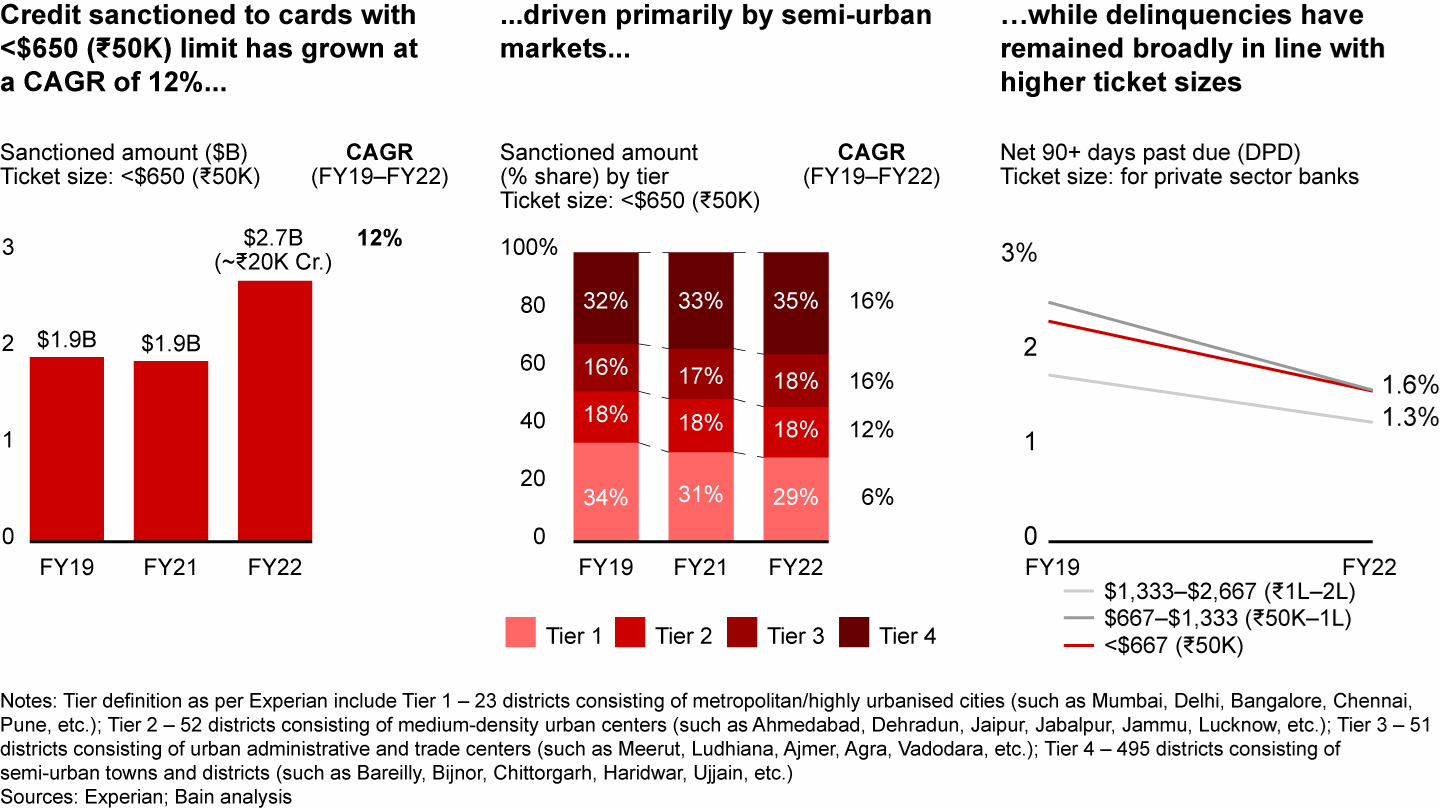
Despite the rapid growth of the online lending market in Bangalore, very little research has been conducted on credit risk assessment, specifically concerning the incorporation of various types of data such as demographic, financial, behavioral, and social media data. The lack of comprehensive studies leaves fintech companies with insufficient tools to accurately predict the probability of loan default and churn for loan applicants in this region.



**Unsecured Retail Lending Market Growth**

**Source: Bain & Company** [**[4]**](https://www.bain.com/insights/india-fintech-report-2022-sailing-through-turbulent-tides/)

Furthermore, the unregulated nature of the credit market in this spectrum of unsecured instant loan providers adds another layer of complexity. The absence of stringent regulations and monitoring mechanisms could lead to an increased risk of potential borrowers with poor creditworthiness slipping through the cracks. This further emphasizes the urgency for a reliable and transparent credit scoring model, with special reference to the creation of a stable, reliable and dynamic credit profile that can mitigate credit risks and empower fintech companies to make well-informed loan approval and risk management decisions.



**Credit Profiling In Semi Urban Market**

**Source: Bain & Company** [**[5]**](https://www.bain.com/insights/india-fintech-report-2022-sailing-through-turbulent-tides/)

Thus, this dissertation aims to fill the research gap by developing an innovative credit scoring model that harnesses the power of machine learning techniques. By leveraging diverse data sources collected through a comprehensive questionnaire from a target population of working professionals aged 25 and above in Bangalore, the model will create customer profiles based on a 4x4 matrix. Moreover, the model will be designed to handle challenges such as missing values, outliers, imbalanced classes, and nonlinear relationships in the data.

The expected outcome of this research is to provide fintech companies with a robust credit scoring tool, which is not singular rule based, that can accurately predict the probability of loan default and churn, thereby improving the assessment of creditworthiness for loan applicants in Bangalore's unregulated online lending market. Additionally, this study will contribute significantly to the existing literature on credit risk assessment in the online lending market, with specific reference to Bangalore. By shedding light on the factors that influence borrowers' repayment behavior and loyalty, this research aims to offer valuable insights for informed decision-making in this fast-evolving and dynamic industry.

# Objectives of The Study

1. Identifying the various contributing factors to the risks in digital lending and mitigating them.
2. Accurately predict loan default and churn likelihood in Bangalore's online lending platforms.
3. Develop an advanced credit scoring model using machine learning by incorporating diverse variables such as demography, social, financial behaviors and historical information about the respondent

# Scope of The Study

This study explores the fields of designing an innovative credit scoring system strengthened by machine learning methodologies. The system's focus is specifically on the growing online lending ecosystem in Bangalore. This dynamic landscape involves a combination of various factors, including demographics, financial tendencies, and social interactions, which collectively enhance the predictive capabilities of the developed model.

However, the scope of this project goes beyond predictive ability. It aims to address a prevalent concern in the credit industry, especially in the online arena, which often struggles with inadequate regulatory oversight. By harnessing the power of intelligent technology, this study aims to empower fintech entities to make wise and well-informed decisions. These decisions are strategically aimed at avoiding the risks associated with loans, thereby promoting a more secure and sustainable lending ecosystem.

# Limitations of The Study

1. Focusing solely on a particular demographic narrows the scope of the findings, thereby jeopardizing the model's broader relevance to a diverse range of borrower profiles.
2. The seamless implementation of the credit scoring model into fintech operations may encounter difficulties attributable to intricacies and the potential requirement for recalibration, thereby affecting its practical integration.
3. Evolving borrower behaviors, economic shifts, and the emergence of new platforms extend beyond the confines of historical data, exerting an influence on the predictive capacity of the model and underscoring the imperative of consistent evaluation and adjustment.

# Research Gap

In the realm of financial services, the process of deciding who gets loans and who doesn't is not as simple as it may seem. Traditionally, banks and lenders use certain methods to determine whether someone is eligible for a loan or credit. These methods, known as credit scoring, play a pivotal role in shaping lending decisions. However, there's a catch – these methods have limitations, and they might not always take into account all the factors that influence a person's ability to repay a loan. This is where the research gap comes into play.

Imagine a scenario where someone wants to get a credit card or a loan to buy a car. The bank or lender would usually check their credit score to decide if they're a good fit for the loan. A credit score is like a grade that tells lenders how likely someone is to pay back the money they borrow. The higher the credit score, the better the chances of getting the loan, and often at a lower interest rate. But here's the thing: these credit scoring methods have their limitations. They might not consider all the different things that affect someone's ability to repay the loan. This becomes even more important when we look at places like Bangalore, where the rules and regulations for lending money might not be as strict as in other places.

So, what's the big deal? Why does this matter? Well, it matters because lending is a serious business. Banks and lenders want to make sure they're giving loans to people who will actually pay them back. And if they can predict whether someone will repay the loan, they can make better decisions about lending money. This is not just about the bank's profits – it's also about the people who are borrowing the money. If they get a loan that they can't repay, it could lead to financial difficulties and problems.

This is where our research comes in. We're aiming to bridge this gap in our understanding of credit scoring. We want to find a better way to predict who will repay loans and who might struggle. But it's not just about prediction – it's about making lending fairer and more accurate. And we're doing this by using a powerful tool called machine learning.

In our case, we're going to use machine learning to create a new and improved credit scoring model. This model will be designed to take into account a bunch of different things that traditional methods might miss out on. We're talking about things like a person's background, how they behave financially, and even their interactions on social platforms.

Now, why is Bangalore so important in all of this? Well, Bangalore is a hub of innovation, especially in the financial technology sector. This city is witnessing a rapid growth in online lending platforms, which makes it a perfect playground for our research. With more and more people turning to online platforms for loans, it's becoming crucial to have a reliable and effective way to assess their creditworthiness. And that's where our advanced credit scoring model comes in – it's tailor-made for the unique dynamics of Bangalore's online lending scene.

But there's more to it. The semi-regulated nature of the credit market in Bangalore adds an element of urgency to our research. In places where the rules for lending money are not as strict, there's a higher chance of risky loans being given out. This is a problem for both the lenders and the borrowers. Lenders risk losing money, and borrowers could end up with loans they can't afford to repay. Our research aims to tackle this challenge by providing a smarter way for lenders to make decisions. By using machine learning, we want to help fintech companies in Bangalore and beyond to make better choices and avoid these risky loans.

To sum it all up, our research is like a puzzle – we're trying to fit all the missing pieces together. We're developing a credit scoring model that's not just smarter but also fairer. By using machine learning, we want to revolutionize how we decide who gets loans and who doesn't. With a focus on Bangalore's online lending market, we're stepping into a world where innovation meets finance, making lending more reliable, more accurate, and ultimately, more beneficial for everyone involved.

# Research Design

The research methodology initiates with the meticulous collection and pre-processing of a comprehensive dataset sourced from the working professionals in the dynamic realm of Bangalore. This dataset encompasses a rich array of variables spanning demographics, financial metrics, behavioural patterns, and social attributes. The data undergoes a rigorous cleansing and refinement process to ensure its quality and relevance, thus paving the way for subsequent rigorous analysis.

With a robust dataset in place, the construction of clear and testable research hypotheses ensues, encompassing both null and alternative stances. These hypotheses stand as the pivotal launchpad for hypothesis testing, which is at the heart of empirical scrutiny.

The investigative prowess is further amplified through the employment of Probit and Logit Regression methodologies. These techniques unfurl the intricacies of causal interplay among variables, particularly attuned to the domain of loan default propensities and churn dynamics. By unearthing these causal threads, the research aims to illuminate the driving forces shaping the lending landscape. The first objective of the study is to predict customer churn by analyzing customer behavior and preferences. A well-structured questionnaire will be designed to gather relevant information from a sample population of 250 individuals. The questionnaire will cover demographics, financial details, past loan history (if applicable), and banking behavior. The collected data will be preprocessed to handle any missing values and ensure data quality. Machine learning classifiers, including KNN, Support Vector Machine (SVM), Decision Tree, and Random Forest, will be employed to explore the likelihood of churn. The study seeks to identify key factors contributing to customer churn, empowering banks to develop proactive strategies for customer retention and engagement.

|  |  |  |
| --- | --- | --- |
| Steps | Tools | Impact on Research |
| Data Collection and Pre-processing | Google Forms, Excel | 34% |
| Hypothesis Formulation | N/A | 14% |
| Building Model Using Alyuda Neural Network | Python - Tensorflow, Pytorch | 28% |
| Probit and logit Regression to identify the causation effect | Python - Statsmodel | 12% |
| Two Way MANOVA For acceptance or failure to accept Null Hypothesis | Python –Statsmodel | 12% |

The second objective is to assess the creditworthiness of loan applicants in the instant loan app sectors employed through fintech platforms. The same questionnaire data will be used to build a comprehensive credit scoring model. Machine learning techniques will be applied to handle the data, including handling missing values, encoding categorical variables, and scaling numerical features. By training various models, such as Logistic Regression, Decision Trees, Random Forest, or Gradient Boosting, the study will determine the most accurate model for predicting the probability of loan default and churn. The credit scoring model will serve as a transparent and reliable tool for fintech companies and banks to make informed decisions on loan approval and risk management.

The research will evaluate the performance of both the churn prediction and credit scoring models using appropriate evaluation metrics, such as accuracy, precision, recall, F1-score, and ROC curves. The results of this study will offer valuable insights into customer behavior, loan risk assessment, and the factors influencing creditworthiness and customer churn in the banking sector.

The crux of the research lies within the Model Building phase, where a sophisticated Machine learning technique, Ayuda Neural Network is deployed to actualize an intricate credit scoring model. This neural architecture is primed through meticulous training, leveraging the meticulously preprocessed dataset. The network's architecture optimization process is geared towards harnessing optimal predictive efficacy, thereby enhancing its proficiency in extrapolating meaningful credit assessment insights.

The spotlight then shifts to hypothesis validation, orchestrated through the Two-Way Multivariate Analysis of Variance (MANOVA) technique. This analytical behemoth elucidates latent differentiations in means across various groupings and delves into the convoluted interactions interwoven within variables. MANOVA's analytical prowess equips the research with incisive insights to validate or invalidate formulated hypotheses, thereby underscoring the empirical foundation of the study.

As a lodestar of ethical tenets, data privacy is vigilantly guarded, underpinning the research's commitment to preserving the sanctity of participant confidentiality and research integrity. This ethical compass safeguards the rights and trust of the participants whose data contributes to this scholarly exploration.

The crescendo of the research journey culminates in the interpretation of results. This stage demands a meticulous dissection of the gleaned insights and emergent relationships. The goal is to orchestrate an astute analysis that facilitates the articulation of definitive assertions regarding causal influences, hypothesis validation, and the nuanced dominion exerted by distinct variables on the intricate domain of credit evaluation.

# Sample Design

