Міністерство освіти і науки України

Національний університет «Києво-Могилянська академія»

Факультет економічних наук

Кафедра фінансів

Магістерська робота

освітній ступінь – магістр

на тему: «**Застосування моделей машинного навчання в оцінці кредитоспроможності» / «Machine learning utilization in solvency evaluation**»

Виконав: студент 2-го року навчання, спеціальність 072
«Фінанси, банківська справа та страхування»

Недождій Михайло Андрійович

Керівник: Буй Тетяна Григорівна
кандидат економічних наук, доцент

Рецензент ____________________
(прізвище та ініціали)

Магістерська робота захищена

з оцінкою «______________________»

Секретар ЕК __________Донкоглова Н.А.

«____»______________ 2024 р.

Київ – 2024

Table of Contents

INTRODUCTION

In recent years, there has been a significant surge in credit scoring utilization due to factors like wide accessibility of data, increase in raw computation power amongst financial institutions and a growing demand for efficient solvency evaluation practices [1]. On top of that, credit scoring functionality has been expanding beyond binary decision making (e.g. accepting or rejecting a loan application) to other parts of the credit process such as setting credit limits and pricing financial services depending on the risk profile of the customer.

As a result, there has been a surge of sophisticated and accurate credit scoring models on the market that have been developed both for public companies and their stocks (depending on investment viability) and for individuals (depending on their likelihood of paying back a loan). While the market is moving away from proprietary in-house credit scoring solutions towards branded publicly available ones, there has been a significant increase in demand for these solutions in Ukraine.

Ukraine's economy was significantly impacted by the ongoing war, which has caused a severe negative impact on the volume and total amount of non-performing loans within the country. While loan portfolio health for state-owned and private banks in Ukraine is closely monitored by the National Bank of Ukraine, there are other types of financial institutions in the country that give out loans and could benefit from efficient credit scoring solutions. These are the so-called microfinancial institutions, non-banking financial companies that give out short-term payday loans to individuals, with traditionally high interest rates accruing daily.

For both these types of financial companies, we want to propose multiple credit scoring methods, including more traditional modeling algorithms (e.g. logistic regression) and data mining methods (e.g. decision tree, random forest, and gradient boosted decision trees). All these models will be tested on a publicly available borrower dataset to measure their individual performance. Also, we will be comparing these algorithms in terms of complexity of deployment and implementation, for both Ukrainian banks and microfinancial institutions.

The contribution that this paper makes to pre-existing literature on the topic is the utilization of the abovementioned models on real borrower data and provided suggestions on how these models can be implemented in the Ukrainian financial market. Also this paper features up-to-date statuses of the non-performing loan ratios in Ukrainian banks, and makes predictions on what NPL ratios could be for Ukrainian microfinancial institutions, both of which are highlighted in Chapter 1.

Relevance of the topic. Considering the status of financial markets in Ukraine and the NPL ratios for state-owned banks (which will be highlighted in Chapter 1), we assume that there is significant demand for new, efficient credit scoring algorithms on the market that could help Ukrainian banks and non-banking financial institutions balance out their loan portfolios and adhere to National Bank of Ukraine's increasing requirements regarding loan portfolio health.

The purpose and main objectives of the paper. The purpose of this paper is to provide an overview of the current state of Ukraine's personal loan market and propose multiple viable modeling solutions that can be applied in Ukrainian banks and non-banking financial institutions. The main objectives of this paper are the following:

- To provide historical background on credit scoring;
- To review the most widespread credit scoring solutions for individual borrowers;
- To evaluate the current state of the Ukrainian personal loan market, both from a banking perspective and from a perspective of microfinancial companies;
- To review up-to-date data mining methodology for businesses;
- To highlight the fundamental parts of logistic regression and decision tree-based classification algorithms;

- To create an overview of metrics that could be used for multi-class classification model performance evaluation;
- To compare performance of a logistic regression model to data mining decision tree-based modeling methods;
- To create recommendations based on the modeling results for Ukrainian banks and non-banking financial institutions.

The **object** of this paper is the Ukrainian consumer credit market and its unique characteristics.

The **subject** of this paper is machine learning-enabled credit scoring methodology and utilization of ML algorithms for solvency evaluation.

Structure of the paper. This paper features an introduction, 3 chapters (Chapters 1-2 are primarily theoretical, Chapter 3 highlights our application of multiple modeling algorithms on a borrower dataset) and conclusions.

Chapter 1 reviews the historical background of credit scoring and the currently available branded solutions for both scoring stocks depending on their investment viability and scoring individuals based on their solvency. The 2nd part of this chapter is dedicated to an overview of the personal loan market in Ukraine. It features a comprehensive outlook on the current structure of bank loans in Ukraine and the non-performing loans ratio in the country. It also features a review of microfinancial institutions' activity in Ukraine, an overview of the current state of the payday loan market and information on the regulatory efforts made by the National Bank of Ukraine to stabilize that market.

Chapter 2 starts with a brief overview of different data mining modeling algorithms that are primarily used in businesses and potential vectors of their application. It also goes into the base terminology for machine learning-based modeling algorithms, like supervised/unsupervised learning. The 2nd part of this chapter is a more detailed outlook on the methodology of logistic regression and 3 specific data mining algorithms: decision trees, random forest, and gradient boosted decision trees (XGBoost). It features a comprehensive overview of each of the modeling methods and graphical interpretations of their target outputs.

Chapter 3 features an overview of individual borrower data that exists in the public domain. It has a brief overview of the metrics we used to measure model performance and the reasoning behind choosing these specific metrics. Moving on, it highlights the results of our modeling efforts, both as metrics and as a graphical representation of the results in the form of confusion matrices. The chapter wraps up with a conclusion regarding modeling results and recommendations about specific modeling algorithms and their applicability for Ukrainian financial institutions.

CHAPTER 1 CREDIT SCORING AND ITS ROLE IN SOLVENCY EVALUATION

1.1 The role of credit scoring in consumer lending

In the early days of modern credit reporting none of the evaluation processes associated with loans were automated. Instead, most loan applicants in larger banks had to get interviewed by a loan officer. Upon examining the applicant's credit history and verifying with the credit bureau for any previous bankruptcies, the bank will decide whether to grant a loan to the applicant [2].

As the demand for automation grew in the mid-20th century, there have been developments in the field, utilizing linear programming methods for credit scoring. While these earlier models had judgmental variables and scores, they have laid a foundation for organized and transparent solvency evaluation for the future [3,4].

In the US, most of the credit reporting at the time was done by CRSPs (Credit Reporting Service Providers) that over time developed into bigger organizations that provided more accurate credit scoring data. In 1970, the FCRA (Fair Credit Reporting Act) was passed in the US, requiring all the CRSPs to publicize their records and establish a discrimination-free environment when making credit decisions. The FCRA, on top of technological advancements motivated CRSPs to grow and merge to large-scale CRSPs like the ones that exist today [3].

As the volume of loan applications grew over time, with it came the global increase in computing power available to financial institutions. This led to increased demand for automated and statistically based credit scoring solutions that had improved predictive capabilities compared to their predecessors.

As a result, nowadays there are a multitude of credit scoring methods that are used by CSPs. Historically, credit scores were based on the loan applicant’s payment history, and confirming or denying their solvency utilizing that data has been the foundation of most modern credit scoring models [5]. Over time, these have developed into modern day credit scoring methods like linear regression, decision trees etc., as well as various machine learning-enabled black box methods that are utilized primarily by non-banking institutions.

There is a growing trend among Credit Service Providers (CSPs) and Credit Reporting Service Providers (CRSPs) to explore other data sources than the applicant's credit history. These sources could include open banking transactions, data derived from mobile phone usage and other digital channels. In markets relying on traditional credit scoring models, potential borrowers must possess a sufficient history of credit information to be considered scorable. Without this historical data, generating a credit score becomes impossible, often leaving a potentially creditworthy borrower unable to secure lending.

The more common term for credit scoring institutions is Consumer Reporting Agencies, or CRAs. These are organizations that generate consumer reports based on individuals' historical financial transactions. The three largest CRAs in the US are Equifax, Experian, and TransUnion [6]. These credit bureaus typically compile consumer credit reports that encompass historical data related to repayment on various credit products, including mortgages, student loans, credit cards, and car loans. The reports could also cover credit applications, bankruptcies, and debts in collection. CRAs can also offer specialized consumer reporting services, such as employment screening and tenant screening. These services may include information on credit repayment history, utility, telecom, and rent payments, employment records, insurance claims, arrests, bankruptcies, and check writing and account management. These consumer files usually lack details about consumer income or assets, as well as personal information such as race or ethnicity, religious or political preferences, and medical history [7].

Credit scoring is the process of using statistical methods to predict the probability of default/delinquency for a loan applicant. The result of credit scoring is an estimate of the loan applicant's probability of default, or in other words, their probability of not paying back the loan they're applying for. This probability is typically represented as a numeric representation, indicating the likelihood of a consumer or business consistently and fully repaying credit obligations, encompassing additional costs like interest and fees [4]. Typically, the higher this number is for an individual, the higher their probability of paying back the loan, and the more opportunities they get when applying for a loan.

For large corporations, businesses etc. creditworthiness is determined through a credit rating. The majority of credit ratings for public companies are publicly disclosed and are used by debt investors in the investment appraisal process. Credit ratings are typically not numerical, but are rather represented by letters, ranging from AAA (highest rating) to C/D (lowest rating). There are three main credit rating agencies: Standard & Poor's (S&P), Moody's Investor Services (Moody's) and Fitch IBCA (Fitch) [8]. While they use different scales for rating, they are comparative between one another and there is equivalence across the credit ratings (Table 1.1).

The significance of credit ratings is in their role in determining a business' credit access, shaping the terms and conditions of credit facilities, including interest rates set by CSPs. Also, as mentioned above, credit ratings play a large role in investment assessment for debt investors, e.g. venture debt companies.

Table 1.1 Structure of credit ratings across Moody's, S&P and Fitch

<table>
<tr><th rowspan="2"></th><th colspan="2">Moody's</th><th colspan="2">S&P</th><th colspan="2">Fitch</th></tr>
<tr><th>Long-Term</th><th>Short-Term</th><th>Long-Term</th><th>Short-Term</th><th>Long-Term</th><th>Short-Term</th></tr>
<tr><td rowspan="7">Investment grade</td><td>Aaa</td><td rowspan="6">Prime 1</td><td></td><td rowspan="6">A-1+</td><td></td><td rowspan="6">F1+</td></tr>
<tr><td>Aa1</td><td>AAA</td><td>AAA</td></tr>
<tr><td>Aa2</td><td>AA+</td><td>AA+</td></tr>
<tr><td>Aa3</td><td>AA</td><td>AA</td></tr>
<tr><td>A1</td><td>AA-</td><td>AA-</td></tr>
<tr><td>A2</td><td></td><td></td></tr>
<tr><td>A3</td><td>Prime 2</td><td>A+</td><td>A-1</td><td>A+</td><td>F1</td></tr>
</table>

Continuing Table 1.1

<table>
<tr><th rowspan="2"></th><th colspan="2">Moody's</th><th colspan="2">S&P</th><th colspan="2">Fitch</th></tr>
<tr><th>Long-Term</th><th>Short-Term</th><th>Long-Term</th><th>Short-Term</th><th>Long-Term</th><th>Short-Term</th></tr>
<tr><td rowspan="5">Investment grade</td><td>Baa1</td><td></td><td>A</td><td></td><td>A</td><td></td></tr>
<tr><td></td><td rowspan="4">Prime 3</td><td>A-</td><td rowspan="2">A-2</td><td>A-</td><td rowspan="2">F2</td></tr>
<tr><td>Baa2</td><td>BBB+</td><td>BBB+</td></tr>
<tr><td>Baa3</td><td>BBB</td><td rowspan="2">A-3</td><td>BBB</td><td rowspan="2">F3</td></tr>
<tr><td></td><td>BBB-</td><td>BBB-</td></tr>
<tr><td rowspan="18">Non-investment grade</td><td rowspan="6">Ba1</td><td rowspan="18">Not Prime</td><td>BB+</td><td rowspan="6">B</td><td>BB+</td><td rowspan="6">B</td></tr>
<tr><td>BB</td><td>BB</td></tr>
<tr><td>BB-</td><td>BB-</td></tr>
<tr><td>B+</td><td>B+</td></tr>
<tr><td>B</td><td>B</td></tr>
<tr><td>B-</td><td>B-</td></tr>
<tr><td rowspan="5">Ba2</td><td>CCC+</td><td rowspan="5">C</td><td>CCC+</td><td rowspan="5">C</td></tr>
<tr><td>CCC</td><td>CCC</td></tr>
<tr><td>CCC-</td><td>CCC-</td></tr>
<tr><td>CC</td><td>CC</td></tr>
<tr><td>C</td><td>C</td></tr>
<tr><td>Ba3</td><td rowspan="7">D</td><td rowspan="7">/</td><td rowspan="7">D</td><td rowspan="7">/</td></tr>
<tr><td>B1</td></tr>
<tr><td>B2</td></tr>
<tr><td>B3</td></tr>
<tr><td>Caa</td></tr>
<tr><td>Ca</td></tr>
<tr><td>C</td></tr>
</table>

Reference: Formed by the author based on [8]

For individual borrowers, credit score is traditionally a numeric representation of a person's creditworthiness. These are usually given out by risk assessment agencies, like the Fair Isaac Corporation (FICO) in the US. The implementation of credit scoring technology in small-medium sized enterprises has shown to result in a significant rise in profitability. In the automotive industry, this results in a reduction of financing for high-risk car loan applicants, while lending to less risky borrowers grew, fueled by increased financing options for higher-quality, pricier vehicles [9].

In the modern economic climate though, there is little need for SMEs to create their own proprietary credit scoring models to evaluate creditworthiness of their customers, when branded, publicly available credit scores are available, like FICO. While the FICO Score is only one of the credit score brands available on the market, we want to use it as an example as to how consumer credit scores are defined.

FICO scores, created by the Fair Isaac Corporation (FICO), are the primary credit scoring system utilized by lenders throughout the United States. These scores offer a uniform numerical evaluation of a person's creditworthiness, impacting determinations on loan approvals, interest rates, and credit limits. FICO scores can range from 300 to 850 and are split into multiple ratings:

- Exceptional – 800-850 score range;
- Very good – 740-799 score range;
- Good – 670-739 score range;
- Fair – 580-669 score range;
- Poor – 300-579 score range [10].

The exact methodology of calculating a FICO score is proprietary and not available to the public. However, FICO have published the weighting of the scores on their website. Payment history, reflecting the borrower's timely payments, is the most significant factor, contributing 35% to their score. Amounts owed, which refers to the ratio of a borrower's debt to their credit limits, constitute another 30%. The length of credit history contributes 15% to a borrower's score, with established accounts positively impacting their FICO score. Credit mix represents 10%, with FICO favoring borrowers who demonstrate proficiency in managing various types of debt like mortgages, auto loans, and revolving credit. New credit makes up the remaining 10%, with FICO prioritizing borrowers who have not recently opened multiple credit accounts (Figure 1.1) [11].

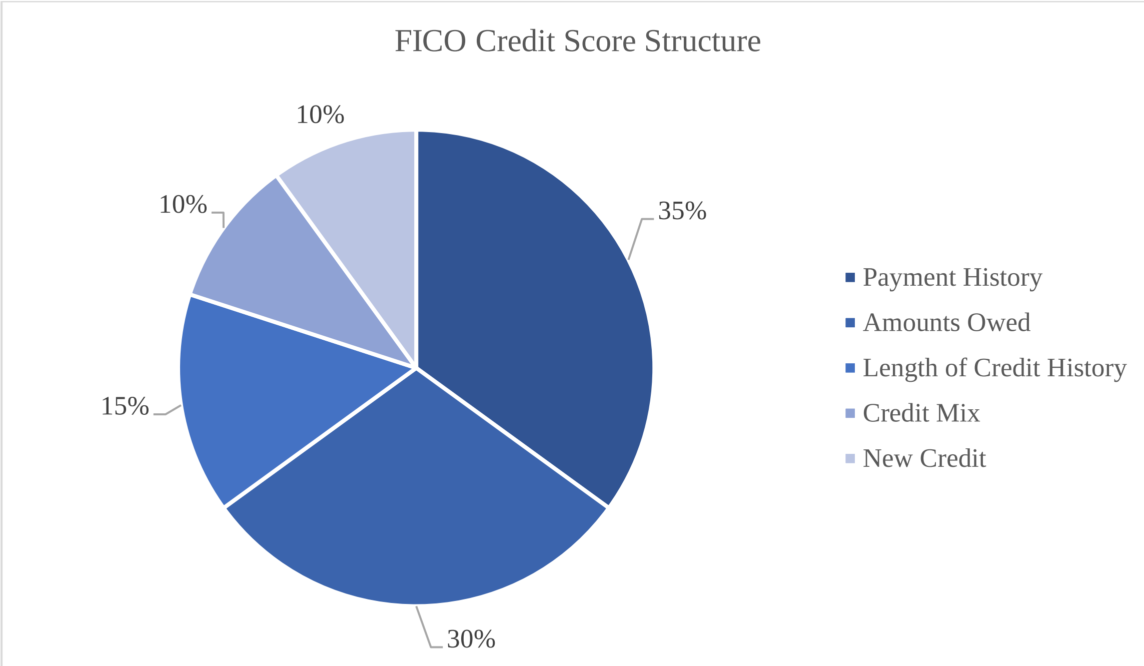

Figure 1.1 FICO Credit Score Structure

Reference: [11]

1.2 The importance of solvency evaluation and the economic impact of NPLs for different types of financial institutions

Credit scoring is an essential part of any bank’s vetting process for potential borrowers, as consumer credit (depending on the bank’s business model) can be a crucial part of the institution’s income. As of January 1st, 2024, in Ukraine customer credit on average contributes to 20.03% of the bank’s assets, most of which consists of loans given out to businesses [12]. There are certain strategic advantages for a banking institution to give out consumer loans, such as:

- Revenue generation through interest and fees;
- Portfolio diversification;
- Building customer relationships with businesses with the goal of cross-selling.

Ever since the beginning of the full-scale war in Ukraine in February 2022, there has been a negative trend of the increase of volume of non-performing loans (NPLs) in banks' portfolios. Prior to the russian invasion of Ukraine, there has been a positive trend in NPL volumes in Ukraine's financial system, which resulted in an almost UAH 300 billion NPL volume reduction [13].

One of the very important figures to consider when looking at the Ukrainian banking system is the NPL ratio, which the share of credits in a banks' portfolio that have defaulted. As of March 2024, the NPL ratio for Ukrainian banks (including state-owned banks) is 36.6%. State-owned banks hold ≈75% of those non-performing loans with PrivatBank holding ≈40% of them (Figure 1.2). There is a large imbalance between NPL ratios for private/foreign banks in Ukraine and state-owned banks, as the nationalization of PrivatBank in 2016 recognized a large portion of its credit portfolio as NPLs [14].

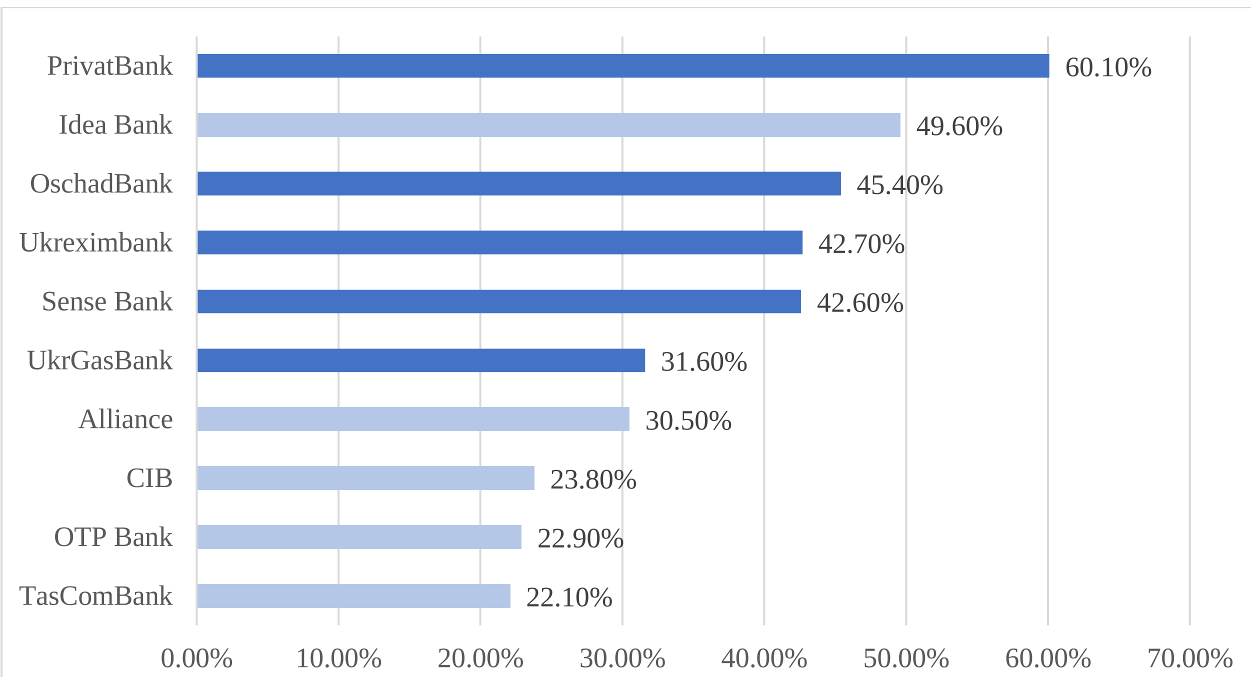

Figure 1.2 Top 10 banks in Ukraine by NPL ratios (government banks highlighted in a darker hue)

Reference: [13]

The full-scale war resulted in damage to assets and collateral, income loss, and decreased borrower solvency. Consequently, borrowers' capacity to repay loans diminished, degrading the quality of banks' loan portfolios, and requiring increased provisioning. Significant loan losses have already been acknowledged, with provisioning exceeding UAH 100 billion since the onset of the full-scale war, amounting to over 12% of banks' loan portfolios by the end of February 2022. Due to the war, the economic crisis, and energy-related disruptions, potential losses in the loan portfolio could reach 30% [13].

Compared to other countries in Europe, Ukraine’s NPL ratio stands amongst the highest in the region. As of Q2, 2023 the NPL ratio in the Euro economic area was 1.85%, compared to Ukraine’s 36.6% [15]. This high of an NPL ratio in Ukrainian banks (especially state-owned banks) has been caused by a multitude of factors, such as:

1. An Asset Quality Review (AQR) conducted by the National Bank of Ukraine, which forced a lot of domestic banks to recognize a portion of their credit portfolios as non-performing loans [16];
2. Introduction of a stricter definition of non-performing loans in accordance with international banking standards as per NBU Board Resolution №351 [17];
3. Recognition of non-performing loans given out by PrivatBank after it’s nationalization in 2016 [14].

Prior to the beginning of the full-scale war in Ukraine, there were plans in motion for state-owned banks to sell a large portion of their NPLs over 2022-2023 below their nominal value. In April 2020, the Ukrainian Cabinet of Ministers authorized state-owned banks to sell non-performing loans (NPLs) at prices below their face value. Prior to this decision, there were no established legal provisions allowing state-owned banks to engage in such transactions. Consequently, these banks had previously relied on alternative strategies like restructuring or foreclosures rather than selling their rights to claim debts [18]. Unfortunately, due to the ongoing war these plans were never realized, and state-owned banks' credit portfolios in Ukraine are still overwhelmed with non-performing loans.

When it comes to non-banking financial institutions that give out loans, in Ukraine there exists a large prevalence of so-called "microfinancial institutions". These are licensed financial institutions that specialize in giving out payday loans. Payday loans are short-term loans that are usually characterized by high fees and high interest rates, accruing daily [19]. These companies mostly don't have a banking license and only give loans out to individuals. As of Q2 2023, there have been 629 registered microfinancial organizations in Ukraine that have given out UAH 40.62b of loans in 2023 (Figure 1.3) [20].

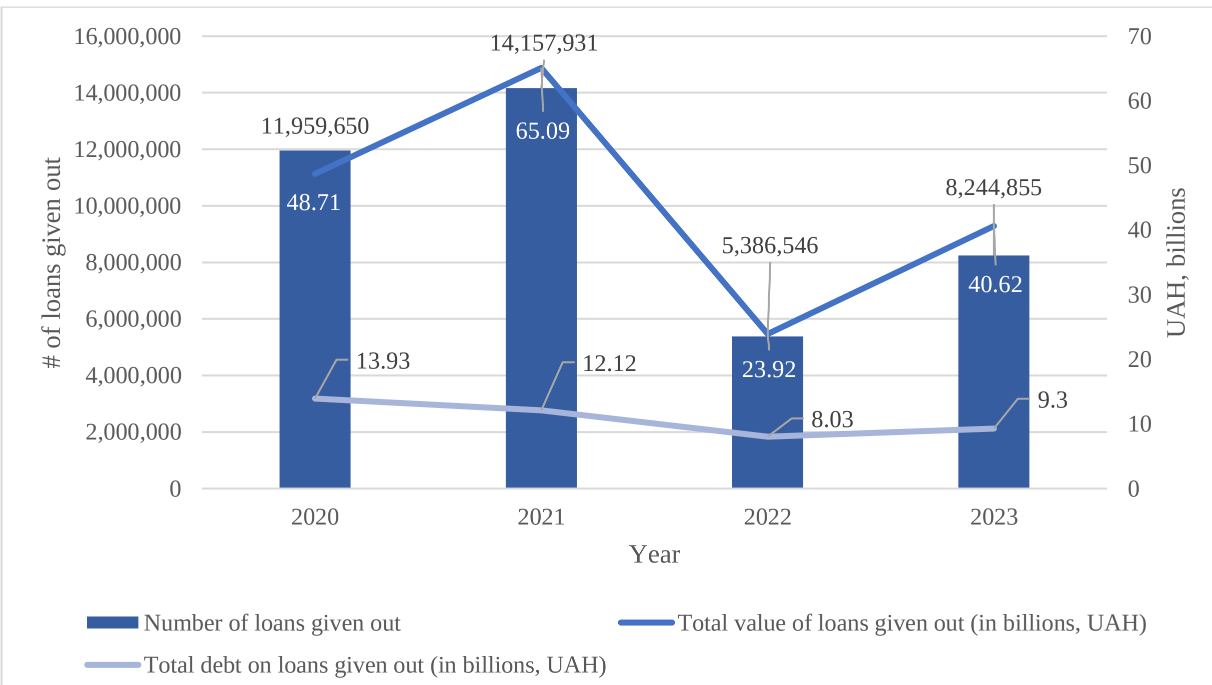

Figure 1.3 Overview of the Ukrainian payday loan market

Reference: [21]

Where microfinancial organizations differ from banks is in their loan application process. The customer base for payday loans generally comes from lower income households that either don't have sufficient credit history to apply for a bank loan or can't afford to pay interest and bank fees associated with long-term loans [22]. Also worth mentioning is that microfinancial organizations in Ukraine have very low requirements for an individual to apply for a loan, which is combined with short review periods. This, combined with a standard practice to offer the first loan to a customer with a heavily discounted rate (≈0.01% per day) results in most microfinancial organizations in Ukraine having high NPL ratios of up to 45% [23].

Payday loans also stand out due to their convenient online application process, which allows potential borrowers to assess various options available in the market and select the institution offering the largest loan with minimal waiting time. This emphasis on quick approval has led microfinancial institutions to employ automated credit application approval systems, a departure from the traditional technologies utilized by Ukrainian banks, which prioritize speed of approval over credit scoring accuracy.

This deliberate neglect of solvency evaluation by microfinancial companies is more apparent when reviewing the data from their borrowers. We’ve done research in the past, based on an anonymous dataset from a Ukrainian microfinancial company, where out of 34895 individual borrowers that were approved for a loan, 42.125% (or 14704 individuals) of them defaulted. This data was used to plot a whale curve and split all borrowers into 4 classes, depending on their profitability for the company (Figure 1.4) [23].

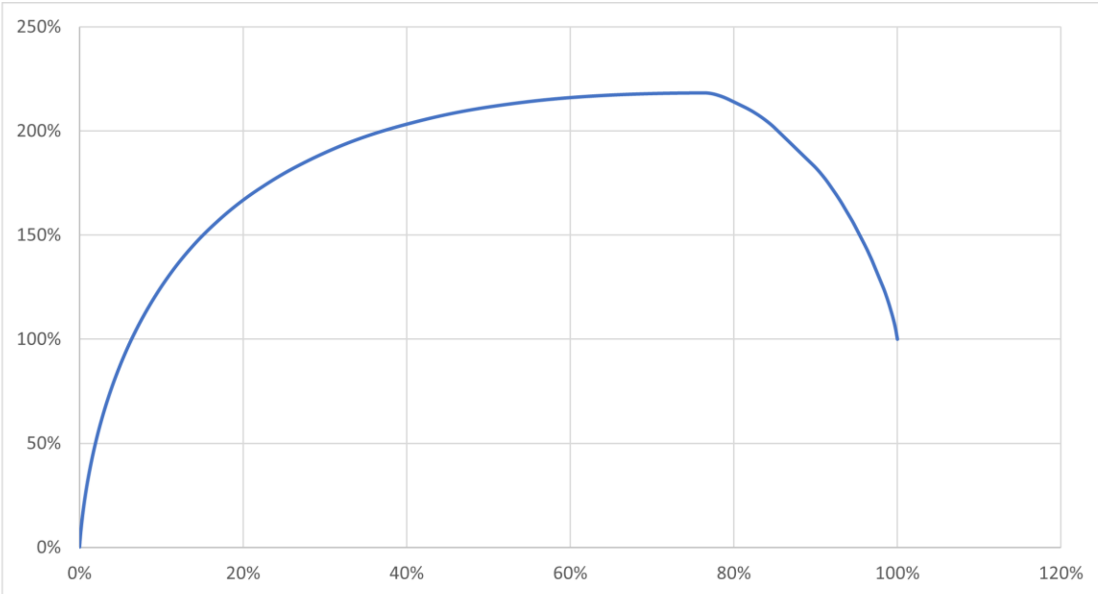

Figure 1.4 Whale curve graph for a sample Ukrainian microfinancial company

Reference: [23]

Based on their profitability, all borrowers were split into 4 classes:

1. Class D (non-profitable) – customers who defaulted on the loan, they have not paid off the loan debt and made no interest payments, represented by 80-100% on the X axis of the whale curve;
2. Class C (non-profitable) – customers who are overdue and are in grace period, they have partially paid out the loan debt and have not made any interest payments, represented by 57.875-80% on the X axis of the whale curve;
3. Class B (profitable) – customers who have paid out their loan debt and have made minimal interest payments, represented by 20-57.875% on the X axis of the whale curve;
4. Class A (profitable) – customers who have paid out their loan debt and made significant interest payments, represented by 0-20% of the X axis of the whale curve.

While this dataset does not represent the entirety of the microfinancial organizations market in Ukraine, it is representative of what their business structure typically is. In this specific case, the company heavily relies on the top 20% of their borrowers to pay out significant interest fees. Simultaneously, they are neglecting their credit portfolios' health by having 42.125% of their borrowers be a net loss, and either classified as defaulters or being in grace period (Table 1.2). This is a strong indication of lack of desire amongst microfinancial companies to attempt to give out loans to individuals who can afford to pay them back.

Table 1.2 Average datapoints for each borrower class present in the dataset.

Class	Number of borrowers	Profit (in UAH)	Age	Total paid (UAH)	Loan term (days)	Loan number
A	6982	4698.51	32	16973.22	24.71	3
B	13209	618.12	31	4270.75	23.88	1.53
C	5997	-3180.54	32	2396.54	24.61	1.8
D	8707	-3358.69	32	2255.34	24.74	1.85

Reference: [23]

In December 2023, a law was passed in Ukraine, proposed by the National Bank of Ukraine, that sets stricter regulations for microfinancial organizations in the company to protect the most economically vulnerable segments of the population. This law has the goal of improving upon the existing legal framework for payday loans and introducing several fundamental changes to the market in order to attempt to balance out microfinancial companies' credit portfolios [24].

As per the National Bank of Ukraine, "microfinancial companies deliberately neglect to check borrower's solvency before providing them with a loan, simultaneously covering the risk of non-performing loans with incredibly high interest rates. According to the Office for the Protection of the Rights of Consumers of Financial Services of the National Bank, hundreds of thousands of borrowers of microfinancial companies are in a state of chronic debt, which requires for them to constantly take out new loans to repay previously taken ones and cannot independently break this vicious circle.

The new law brings change to the market and introduces responsible lending by financial companies, when the interests of the borrower are taken into account and consumer rights are protected [25].”

The law, amongst other restrictions, included:

- Introduction of a strict interest rate limit for payday loans, set at 1% per day (currently payday loan rates can reach up to 1800% yearly), implemented over the next 8 months;
- A requirement to include information on the daily interest rate in the loan agreement;
- The right for the National Bank of Ukraine to set minimum requirements for a borrower’s creditworthiness in order to reduce loan pressure for individual borrowers;
- A requirement to exclusively extend the loan agreement upon the agreement of both sides;
- Additional sanctions that apply if a microfinancial company does not abide with these restrictions [24].

To add upon the National Bank of Ukraine’s statements, a reduction in maximum interest rates financial companies can set on payday loans would most likely cause them to implement proper credit scoring procedures, which as a result would reduce the share of NPLs in their credit portfolios and help protect the most economically vulnerable portion of the population who use these payday loans the most [19].

Conclusions to Chapter 1

Credit scoring is an integral part of each financial institution’s day-to-day operations. Throughout history, credit scoring has evolved from manual labor, performed by each individual bank or institution to automated algorithms, allowing for more businesses and individuals to evaluate their solvency through these tools. As the demand for debt worldwide grew, these credit scoring algorithms became more sophisticated and accurate, and led to certain companies (e.g. Fair Isaac Corporation) to create branded solutions that are widely accepted on the market and are applicable to both public companies and individuals.

As Ukraine's economy is heavily impacted by the ongoing war, there is an increased demand for accurate credit scoring solutions for banks and non-banking financial institutions. The non-performing loan (NPL) ratio for Ukrainian banks is approaching 36.6% as of March, 2024, which is marginally higher than the NPL ratios in the neighboring European Union, where they sit at 1.85% as of Q2 2023. Main holders of these NPLs are the state-owned banks, which hold ≈75% of the outstanding NPLs in the country. The issue of NPLs in Ukraine is as relevant as ever, which is also recognized by the National Bank of Ukraine through their continuous efforts to enforce good portfolio health amongst banks and non-banking financial institutions.

Another widespread type of financial institution that struggle with non-performing loans in Ukraine are microfinancial institutions, which provide payday loans to individual borrowers. As of Q3, 2023 there have been 629 registered microfinancial institutions in Ukraine that have given out UAH 40.62b of loans throughout 2023. The particular feature of this market is that these companies do not attempt to maintain good loan portfolio health, as they are receiving most of their income from the top 20% of their borrower base. As loan health is neglected in these institutions, the National Banks of Ukraine has been enforcing restrictions for microfinancial institutions in regards to transparency and maximum interest rates, which could lead to a surge in demand amongst these companies for a proper credit scoring solution that would fit their needs.

CHAPTER 2 MACHINE LEARNING METHODOLOGY AND IT'S UTILIZATION FOR CONSUMER CREDIT SCORING

2.1 Overview of machine learning methodology for businesses

Credit scoring can be performed with one of the two targets: binary or multinomial. The abovementioned FICO score is a numeric representation of credit scoring with a binary target, as the output for each individual can be interpreted as a value ranging from 0 to 1. This is the most popular form of credit scoring, placing each ranked individual on a 0-100% scale, depending on their creditworthiness and probability of default.

On the opposite end of the spectrum, there exists credit scoring with multinomial targets, e.g. classification models. These models can include multiple proprietary targets and be used, for example to attempt to predict not only whether the potential borrower will default, but also how much interest they will pay if they are provided with a loan. As part of this work, we want to focus on both statistical and data mining modeling methods and how these methods can be effectively utilized for solvency evaluation. More specifically, we want to review the methodology of machine learning and how machine learning-enabled modeling methods can be used for credit scoring.

In regard to machine learning methods, these hardware-enabled modeling methods appeared relatively recently due to the increase of computational power and developments in the field of AI. The concept of machine learning was first coined in mid-20th century by Arthur Lee Samuel in an article "Some Studies in Machine Learning Using the Game of Checkers" [26]. Early adoptions of machine learning have heavily relied on proprietary hardware, and budget-wise were only available to large corporations.

In recent years, there has been a surge in usage of machine learning-enabled credit scoring models. Recent developments in the field suggest that for banks, utilization of ML can reduce expenses on overdue customers by 25% [27]. While ML-enabled credit scoring models are effective, they have similar drawbacks to traditional modeling methods when it comes to using them on low-quality, or incomplete data. Also, as a lot of ML algorithms that are utilized for prediction are “black box” solutions, they are challenging to scale for larger financial institutions e.g. banks.

Machine learning is a science that studies and creates algorithms for learning and prediction from data inputs. Where it differs from more traditional modeling methods is that machine learning-enabled algorithms adapt and make decisions and predictions based on the data, compared to more conservative modeling methods, which have the disadvantage of following static program instructions [28]. Over the years, a multitude of data mining methods were implemented to address different business tasks, some of which being:

1. **Classification**, or class probability estimation, is a method used for predicting which class an individual customer belongs to. Usually, classes proprietary for each model and are mutually exclusive. In a classification context, a data mining process generates a model that, when given a new individual, determines the individual's class membership. Instead of predicting a class, a scoring model assigns a score representing the likelihood (or another measure of probability) that the individual belongs to each class. Classification and scoring are closely related, and typically, a model capable of one task can be adapted to perform the other;
2. **Regression** is a modeling method that predicts or estimates the numerical value of a variable for a specific individual. For our credit scoring case, this numerical value would be the probability of default for said individual. As mentioned above, regression and classification are closely related, and yet have different goals. Classification aims to predict whether an event will occur, while regression measure the frequency with which that event will (or will not) happen;
3. **Similarity matching** is a data mining method that attempts to identify similar customers, based on the data available about them. These methods are most

frequently used for product recommendations in B2C markets. Also, similarity matching can be utilized to find companies that fit your ICP (Ideal Customer Profile) to prospect potential clients that are similar to your existing ones, based on the input data;

4. **Clustering** is a graphical modeling method, similar to classification, but with undefined classes. It allows to group customers or individuals by their common features to identify whether they have any, and whether your customer base can be split into segments. Clustering is valuable for initial data exploration within a domain to identify existing natural groups, which may then suggest the utilization of other data mining methods;
5. **Co-occurrence grouping** is a modeling method that attempts to group together entities, based on transactions involving them. It is most commonly used in B2C purchase suggestion algorithms, like "these items are most commonly bought together" etc. Where this method differs from clustering, is that clustering involves examining the similarity between objects based on their attributes, whereas co-occurrence grouping assesses the similarity of objects based on their frequent appearance together in transactions;
6. **Profiling** (also known as behavior description) is a modeling method that attempt to identify a typical behavior amongst a group of individuals/clients. Behavior can be analyzed across an entire population or at more granular levels, such as small groups or individuals. Profiling is commonly employed to establish typical behavioral patterns amongst clients, which is valuable for detecting anomalies in applications like fraud detection and monitoring for computer system hacks and intrusions;
7. **Link Prediction** is a data mining modeling method that attempts to predict links or connections between items by suggesting that a link should exist and estimating the strength of said link. It is most commonly used for B2C

products, like friend suggestions on social media and recommendation engines for movies etc.;

8. **Data Reduction** is a modeling method that attempts to reduce a large dataset to a smaller one that contains the most important features of the original dataset. This algorithm is mostly used internally by businesses as part of their data science efforts, as subsampling a large dataset while preserving feature importance can reveal patterns in the data and create additional insights into its structure;
9. **Causal Modeling** is a data mining modeling method that attempts to predict causality between events, understand what actions influence others. This causality identification is especially important to test the performance of marketing campaigns, sales etc. Techniques for creating causal models range from things such as randomized controlled experiments (also known as A/B tests) to more sophisticated observational methods [29].

All machine learning-enabled data mining methods can be broadly split into two types: supervised learning algorithms and unsupervised learning algorithms, depending on how limited their respective output is. Supervised learning algorithms are created with a specific target result in mind. While all data mining models require input data (also known as a training dataset), supervised machine learning algorithms also require an example of output data that the model will use as an example. A dataset used for a supervised machine learning algorithm could be, for example, a list of individual borrowers, each having their credit score as part of the dataset. That credit score, when used as a dependent variable, would be an example of a target output for the model [30].

On the other hand, unsupervised learning methods do not have target output data. This results in a relatively unstructured output of the model, usually represented graphically as clustering of unlabeled data. While the lack of a target output is greatly limiting when it comes to modeling results interpretation, this also eliminates the biggest data constraints of supervised learning, which is that there must be data on the target [29, 30].

Out of the abovementioned data mining methods, classification, regression and causal modeling are traditionally solved with supervised methods; similarity matching, link prediction and data reduction could be either and clustering, co-occurrence grouping and profiling are generally solved with unsupervised methods [29].

One thing to consider when implementing data science methodology in businesses, is that most (if not all) of these methods require pre-existing IT infrastructure in the company, and often a data science team to maintain and improve the model over time. This results in data science modeling being more applicable for businesses in industries that are technologically advanced and where process optimization can lead to significant cost savings. This primarily includes different verticals within manufacturing (e.g. automotive, semiconductors, pharmaceuticals) and professional services, like finance and business consulting (Table 2.1).

Table 2.1 Current status of machine learning implementation in different industries

Sector	In evaluation/ planning	Initial prototypes	Being used in selected sectors	In productive operation in broad sections of the company
Automotive	60%	20%	0%	20%
Machine and system engineering	52.9%	17.7%	29.4%	0%
Chemical manufacturing	30%	40%	30%	0%
Metalworking	38.5%	46.2%	15.4%	0%
Consumer goods	18.8%	43.8%	25%	12.5%
Logistics and transportation	16.7%	41.7%	41.7%	0%
Banking and insurance	28.6%	33.3%	28.6%	9.5%

Continuing table 2.1

Sector	In evaluation/ planning	Initial prototypes	Being used in selected sectors	In productive operation in broad sections of the company
Professional services	38.5%	23.1%	30.8%	7.7%
Pharmaceuticals and healthcare	22.2%	55.6%	22.2%	0%
IT, telecommunications, and media	11.8%	47.1%	26.5%	14.7%
Public sector and education	27.8%	44.4%	27.8%	0%

Reference: [31]

Table 2.1 is based on a B2B report from 2016, and we can assume that in recent years data mining and machine learning methods have become a lot more widespread in manufacturing industries. On top of this, there have been a lot of recent developments in B2C large language models (or LLMs) developed by the likes of OpenAI, Google and Microsoft. Presented in the form of chat bots, these LLMs have seen widespread usage in B2C markets, and could potentially be implemented in businesses in the future.

Circling back to data mining methods that can be utilized for credit scoring, as part of our modeling efforts in Chapter 3 of this paper we will be creating multiple credit scoring models and testing them on a dataset adapted for supervised learning algorithms. Three modeling methods were selected for this task, for each of which we want to go into more detail about how they are built and how they differ from one another:

- Logistic regression (with a multinomial target);
- Decision tree-based models, including:
    - Random forest;

- Boosted decision trees.

2.2 Multinomial logistic regression and decision tree methodology

Logistic regression is a type of a regression model, utilized primarily for classification tasks. This classification algorithm can be used when the dependent variables are categories and all of them can be split into discrete outputs [32]. These outputs are dichotomous in nature and are usually coded 1 (indicating success) or 0 (indicating failure). In terms of input data, a logistic regression

For a graphical interpretation of what the modeling results of logistic regression look like, we can use a binary target model to output the classification results on a graph. As logistic regression is a linear model, its output is the best hyperplane (i.e. line) that separates the two classes based on the input data (Figure 2.1). In a binary classification, the decision function is a line on a 2-dimensional plane, splitting all individuals into 2 classes (green and red).

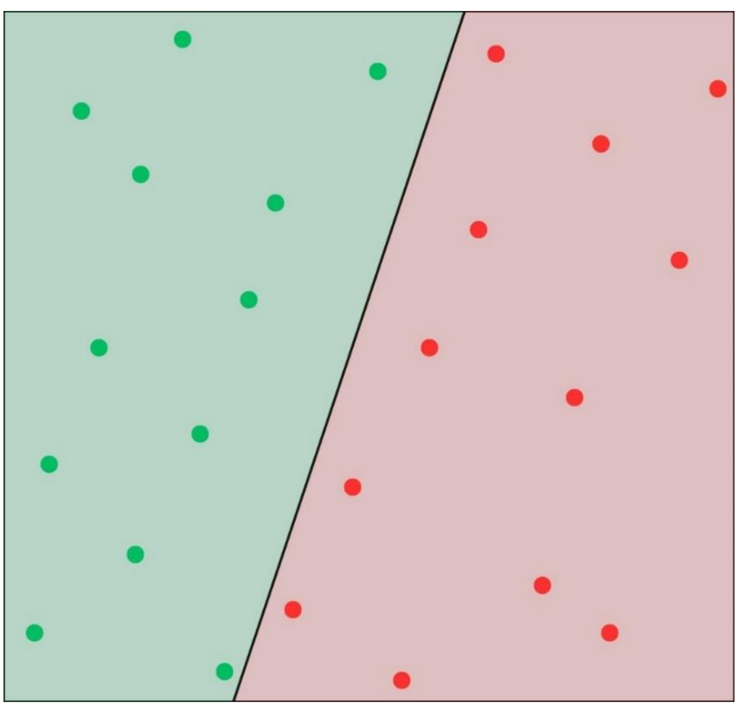

Figure 2.1 Decision function of a binary logistic regression model.
Reference: Built by the author based on [31]

When it comes to logistic regression with a non-binary target, the target can be either ordinal, or multinomial (which is the method we've used for our modeling efforts in Chapter 3). Both ordinal and multinomial logistic regression methods have multi-class targets. Where they differ from one another is that in ordinal logistic regression, classes are ordered (or scaled), usually from "best" to "worst", while multinomial algorithms have their classes unranked [33].

Moving onto decision trees, these are relatively common machine learning algorithms that are often used because of their ease of interpretation, compared to other ML-based algorithms. It is a non-parametric supervised learning algorithm that can be utilized for both regression and classification tasks [34]. The result is a flowchart-like tree structure that consists of:

1. Root node – starting point of the tree, representing the entire dataset;
2. Internal nodes – secondary nodes, representing testing on features or attributes of the initial dataset;
3. Branches – lines representing possible outcomes of the test at an internal node;
4. Leaf nodes – nodes at the bottom of the decision tree, indicating either a predicted probability (for regression) or a class indication (for classification) [35].

One of the main advantages of decision trees is its ease of interpretation, both in methodology and graphical interpretation of the model. Decision tree learning uses a "divide and conquer" approach by conducting a search to find the best split points within a tree. This splitting process continues recursively from the top down until most or all records are classified under specific class labels. Whether all data points are classified into homogeneous sets largely depends on the complexity of the decision tree.

Smaller trees can more easily achieve pure leaf nodes where data points belong to a single class. However, larger trees struggle to maintain this purity as they often end up with too little data in each subtree, a phenomenon known as data fragmentation, which can lead to overfitting [36]. An example of a graphical representation of a decision tree model can be found in Figure 2.2 and as part of our modeling efforts in Chapter 3 of this paper.

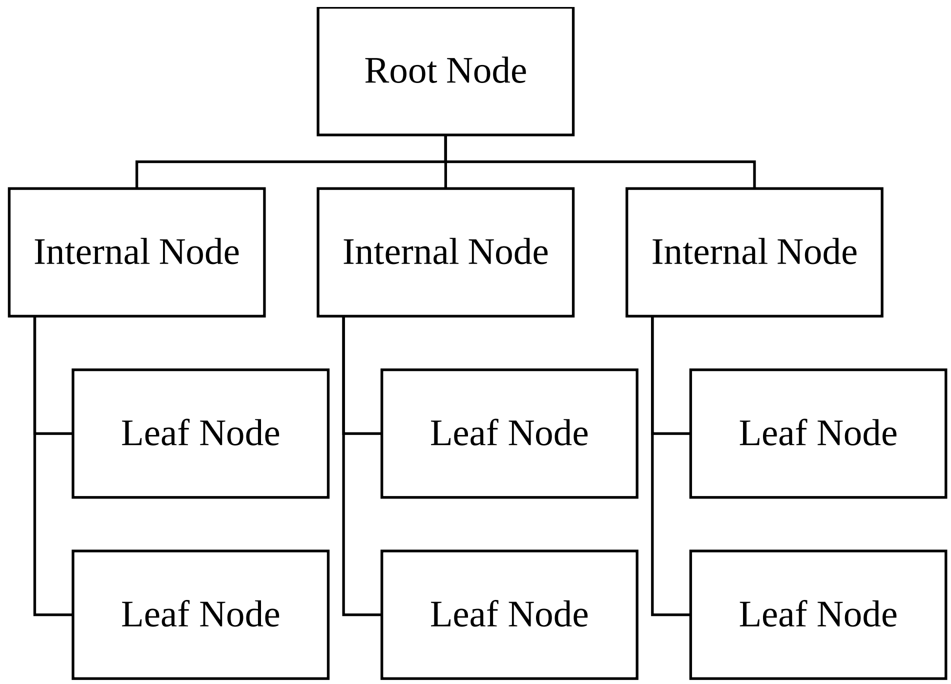

Figure 2.2 A graphical representation of a decision tree model
Reference: Built by the author based on [35]

Decision tree models can also be grouped into ensembles, or decision forests, which is a model consistent of multiple decision trees. The output of this modeling method differs depending on the modeling target. In this paper we want to primarily focus on randomization-based ensemble techniques with multi-class classification, specifically random forest and boosted decision tree algorithms.

Random forest in a randomization-based ensemble technique, which includes a multitude of decision tree algorithms ran simultaneously. The algorithm creates an ensemble (also called a forest) of decision trees and lets them vote on the most popular class. While growing these forests, random vectors are generated that govern the growth of each decision tree in the ensemble. After a large (often pre-selected) number of trees are generated, they vote for the most popular class amongst the decision trees [36]. A graphical representation of this decision-making process is shown on Figure 2.3.

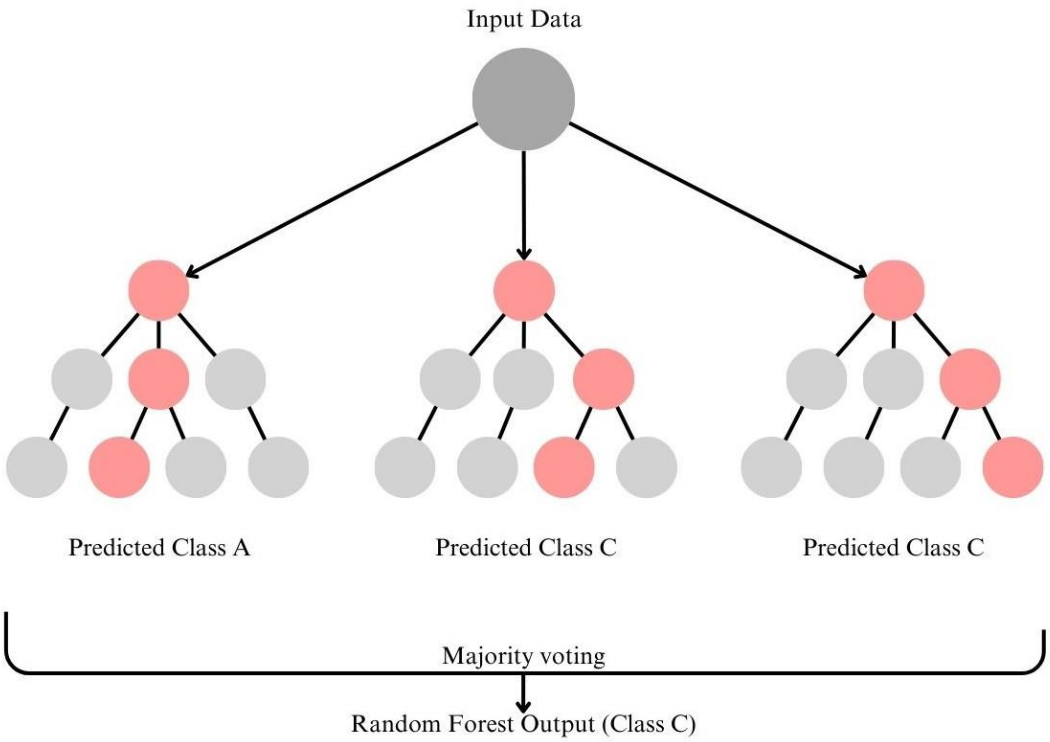

Figure 2.3 A graphical representation of the random forest ensemble algorithm for multi-class prediction

Reference: Built by the author based on [37]

Boosted decision tree algorithms are more hardware limited and complex, compared to the abovementioned methods. In general, boosting is a method in which many "weak" decision trees are grouped together to form a strong classifier. Each tree in the model is created iteratively and its output is given a weight relative to its accuracy [38]. The ensemble output is a weighted sum and is calculated as follows:

$$\hat{y}(x) = \sum_t w_t h_t(x), \text{ (2.1)}$$

where h(x) is the tree's output,

t – number of the tree in order;

w – decision tree weight, depending on its accuracy (higher accuracy=higher value) [38].

The objective function of a boosted decision tree is the following:

$$obj = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \omega(f_i), \text{(2.2)}$$

where $l(y_i, \hat{y}_i)$ is the training loss function, the distance between the actual value and the value predicted by the *i*th decision tree;

and $\omega(f_i)$ is the regularization function, which penalizes the complexity of *i*th decision tree [38, 39].

There are a multitude of decision tree boosting algorithms available on the market, that primarily differ in how they approach minimization of the loss function by iteratively adding learners [38]. In this paper, we want to focus on one of the more popular boosting algorithms available, which is extreme gradient boosting, also known as XGBoost.

XGBoost is a scalable machine learning library for parallel decision tree boosting. It is distributed as an open-source package and is widely used for a multitude of regression and classification tasks [40]. The impact of XGBoost has been widely acknowledged in various machine learning and data mining challenges, notably demonstrated in competitions hosted by Kaggle. In 2015, among 29 winning solutions featured on Kaggle's blog, 17 utilized XGBoost. Eight of these solutions relied solely on XGBoost for model training, while others combined XGBoost with neural networks in ensemble approaches. By comparison, deep neural networks were used in 11 solutions, making XGBoost the most prevalent method [41].

XGBoost's efficiency stems from its scalability in all scenarios, exhibiting over tenfold speed enhancements compared to prevailing solutions on a single machine and accommodating billions of examples in distributed or memory-constrained settings.

Several key optimizations contribute to this scalability, including a novel tree learning algorithm tailored for sparse data and a weighted quantile sketch procedure that enables instance-weighted approximate tree learning [41].

For our specific use case, we have utilized the XGBoost library for a classification task, more specifically to split borrowers into pre-defined classes, depending on whether or not they can pay back the loan, and how long they would be overdue for. One of the specifics of utilizing XGBoost for a classification task is the existence of hyperparameters in the library that help tune the model to control overfitting and/or handle an imbalanced dataset.

Before we move on to hyperparameters, it is also worth discussing the main 2 problems that can be solved with hyperparameter tuning in XGBoost: overfitting and class imbalance. Overfitting is an issue specific to supervised machine learning algorithms, that occurs when the model is overly trained on the training dataset and starts to underperform when used on other datasets, that can differ from the training one [42]. There are a multitude of statistical methods that allow for overfitting prevention, like shrinking the training dataset or "early stopping", which is a process of reducing model complexity, so it doesn't overfit (Figure 2.4). There are multiple ways to prevent overfitting using hyperparameters as part of the XGBoost library, which we will be going over later in this chapter.

Class imbalance is an issue that is not unique to supervised machine learning algorithms and is relatively common in data science. It means that the dataset used for modeling does not have equal amounts of representatives for each class. In non-simulated datasets, class equality is uncommon, which results in the model being more accurate at predicting the most represented class, and less accurate at predicting others. This can happen both in datasets with binary and multi-class outputs. XGBoost has multiple hyperparameters that can help negate class imbalance for binary classification, but for multi-class it is more challenging as it requires more intricate transformations done to the original dataset.

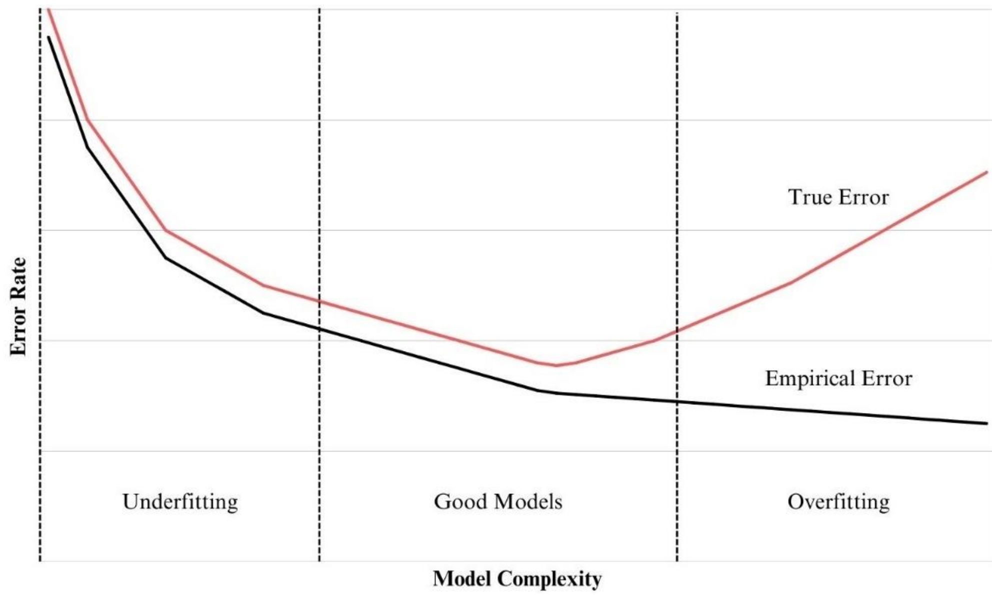

Figure 2.4 A relationship between training error and test error as a function of model complexity.

Reference: Built by the author based on [43]

As we will be using the Python package of XGBoost for our modeling efforts, the primary hyperparameters we want to highlight are the following:

- *max_depth* – a hyperparameter that controls the maximum depth of each decision tree, can range from 0 to ∞; this parameter can help reduce overfitting, as reducing this value makes the model less complex, and therefore less likely to overfit;
- *min_child_weight* – a hyperparameter responsible for the minimum sum of instance weight needed in a child. If the tree partition step results in a leaf node with the sum of instance weight less than *min_child_weight*, then the building process will give up further partitioning. Decreasing this value decreases model complexity and can help prevent overfitting, can range from 0 to ∞;

- *eta* or *learning_rate* – a hyperparameter that increases/decreases the step size shrinkage used in each update. Same as with the other 2 hyperparameters used to control overfitting, a reduction in *learning_rate* reduces model complexity and make the algorithm less conservative, can range from 0 to 1;
- *scale_pos_weight* – a hyperparameter used for handling imbalanced classes. It controls the balance of positive and negative weights in the dataset, and should be set to the sum of negative instances divided by sum of positive instances. By default it is set to 1, but this hyperparameter can't be used for multi-class classification;
- *max_delta_step* – a hyperparameter that sets what maximum delta step each leaf output is allowed to be. In logistic regression, an increase of this hyperparameter can help with extremely unbalanced classes, can be set from 0 to ∞ [44].

While not all hyperparameters mentioned here are applicable for a multi-class classification algorithm, they provide a better understanding of how the XGBoost library functions and how highly customizable its prediction capabilities are. Because of its customizability and scalability, over the years XGBoost has been used for a multitude of different classification and prediction models, such as: web text classification, customer behavior prediction, hazard risk prediction and product categorization [41]. As part of our modeling efforts, we will utilize the abovementioned logistic regression and decision tree-based algorithms to create classification models, compare their performance and evaluate their usability for the Ukrainian financial market.

Conclusions to Chapter 2

As business processes in financial companies become more challenging and difficult to address over time, there is an increased demand on the market for reliable solutions for not only credit scoring, but other business processes that can be automated. As a result, there exists a plethora of data mining solutions that can address these needs. These methods can be broadly split into supervised and unsupervised algorithms, depending on whether the output target of the model is specified withing the input data. There are certain advantages and disadvantages to each of those types of solutions, but as part of this paper we want to primarily focus on supervised deep learning algorithms that can perform classification tasks and can be utilized for solvency evaluation.

As part of this paper, we are primarily focusing on 4 algorithms: logistic regression, decision trees, random forest, and gradient boosted decision trees (the XGBoost algorithm). All these methods can be adapted for both regression and classification tasks, and are widely used in different industries, including financial services.

CHAPTER 3 SOLVENCY EVALUATION UTILIZING MACHINE LEARNING

3.1Credit Data Overview

Considering the problem of NPLs in Ukrainian banks and non-banking financial institutions we want to test multiple modeling methods on the same customer dataset. Our assumption is that with stricter credit portfolio limitations coming from The National Bank of Ukraine, financial institutions in Ukraine will have to reconsider their existing loan applications methods. This is especially relevant for microfinancial institutions that provide payday loans, as they have the lowest standards when it comes to borrowers and target the most economically vulnerable households in the country [19]. Our goal is to attempt to improve upon the more traditional multinomial credit scoring methods (e.g. logistic regression) and weigh the advantages and disadvantages of each approach.

Due to the nature of credit application data, there is a general scarcity for datasets that feature non-simulated data. For the purposes of this thesis, we have tested multiple datasets to find out the one that would fit our research best. In our research, we have found the following publicly available datasets of real borrowers [45]:

- Country data from the UCI repository, featuring data from countries such as Germany, Australia, Taiwan, Japan and Poland [46,47,48,49,50];
- Datasets from a data science competition platform Kaggle, some of which feature non-simulated data e.g.:
    - "Give me some credit", a dataset created for a competition back in 2011, data source is unknown, but it is not classified as a simulated dataset [51];
    - "Home credit default risk", a dataset provided by Home Credit, a Dutch financial institution from their internal databases [52];

- Lending Club individual borrower data, which was public until 2021 and has been preserved in public repositories [53].

Out of the datasets available to us, we have disqualified all public data except the Lending Club borrower dataset for multiple reasons. Most of the UCI data is scarce and features a small number of observations ranging from 125 for Japan up to 30000 for Taiwan. Regarding Kaggle datasets, we have already had some experience working with the “Give me some credit” dataset, while the “Home credit default risk” dataset lacked a clear independent variable and is structured in a way to make it more challenging for people who participated in the competition [54]. So, for our research, we have decided to use a dataset from Lending Club to test the effectiveness of different classification algorithms and compare them to machine learning-based solutions.

Lending Club is a US-based financial services company that provides personal and business loans, as well as financial consulting and investment advisory services [55]. The overview of the dataset we used is the following:

- Time frame: 2007 Q1 – 2018 Q4, each row has a separate timestamp;
- Number of observations/individual borrowers: 1,048,563;
- Total number of features: 151;
- Number of features used for modeling: 14.

During the modeling process, we have shrunk the number of features to avoid overfitting the model and speed up the modeling process. Target variable is **loan_status**, split into 3 classes depending on the status of the loan. Originally the dataset featured 7 classes, and as part of our feature engineering process we have grouped them into 3 (Table 3.1). Another reason for grouping our target classes was a large class imbalance that was present in the original dataset. As we can observe in Table 3.1, 4 out of 7 classes represented 1.04% or less of the borrowers, which would create additional challenges in the modeling process. As a result, we have grouped all borrowers into 3 classes:

- Class 0 – borrowers who have fully paid out their loan;

- Class 1 – borrowers who have not yet paid out their loan, but haven't defaulted;
- Class 2 – borrowers who defaulted and were unable to pay out the loan.

Table 3.1 Class distribution in the Lending Club dataset

Name of the class in the original dataset	Number of observations per class	Percentage of the borrowers represented by the class	Grouping classes as part of the feature engineering process
Fully Paid	482595	46.02%	0
Current	419550	40.01%	1
In Grace Period	4227	0.4%	2
Late (16-30 days)	2104	0.2%	2
Late (31-120 days)	10879	1.04%	2
Default	21	0.00%	2
Charged Off	129187	12.32%	2

Reference: Built by the author based on [53]

When selecting variables for our model, a conscious decision was made to reduce the number of features to prevent overfitting and make the model more flexible for applications outside of our use case. In the end, 13 independent variables were selected:

1. **Loan_amnt** – the listed amount of the loan applied for by the borrower; If at some point in time, the credit department reduced the loan amount, then it was reflected in this value, represented in absolute values in USD;
2. **Funded_amnt** – The total amount committed to that loan at that point in time, represented in absolute values in USD;

3. **Funded_amnt_inv** – The total amount committed by investors for that loan at that point in time, represented in absolute values in USD;
4. **Term** – The number of payments on the loan. Values are in months and can be either 36 or 60;
5. **Int_rate** – interest rate on the loan;
6. **Installment** – the monthly payment owed by the borrower if the loan originates;
7. **Annual_inc** – the self-reported annual income provided by the borrower during registration, represented in absolute values in USD;
8. **Dti** – a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income;
9. **Delinq_2yrs** – the number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years;
10. **Open_acc** – The number of open credit lines in the borrower's credit file;
11. **Revol_util** – Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit;
12. **Last_fico_range_high** – The upper boundary range the borrower's last FICO pulled belongs to;
13. **Last_fico_range_low** – The lower boundary range the borrower's last FICO pulled belongs to.

Prior to selecting these independent variables, we have done extensive research into feature importance for all the 81 numeric variables in the dataset (out of 151 total). These specific features were selected to balance out the metric output of the model and present variables that can be realistically obtained from an individual customer in a Ukrainian financial institution. Outside of the FICO score range (which could be replaced with an equivalent rating provided based on credit history), all these variables can be obtained from an individual Ukrainian borrower. A more detailed overview of feature importances in the dataset can be viewed as part of Appendix A.

3.2 Application of data mining methods for credit scoring in Ukrainian financial markets

For this classification task, we have used 4 modeling methods to compare between one another: logistic regression, decision trees, random forest, and gradient boosted decision trees (XGBoost). All three models were created in Python using the Jupyter Navigator GUI. First, let's start with the overview of our logistic regression model.

To evaluate effectiveness of the 3 selected modeling methods, 4 metrics were used: **accuracy** of the model, **precision**, **recall**, and the **F1 score**. For graphical interpretation, we have utilized accuracy confusion matrices, which represent the accuracy of the model for multi-class prediction. An example of a multi-class confusion matrix can be found in Table 3.2. Rows represent actual (or true) classification, while columns represent classification that the model predicted. Classes are listed in the same order in rows and columns, so correctly classified datapoints are located diagonally from top left to bottom right [56].

Table 3.2 A sample multi-class confusion matrix

True labels	a	150	2	21
	b	20	46	42
	c	12	34	78
		a	b	c
		Predicted labels		

Reference: Built by the author based on [56]

To further go into detail regarding the modeling metrics we've used, we also need to introduce the concept of true/false positives and true/false negatives in classification models. True Positive values in a binary classification model are the ones that have been labeled as positive by the model, and are actually positive, while True Negative values are the ones that have been labeled as negative by the model and are actually negative. This can be represented on a confusion matrix like Table 3.3.

Table 3.3 A binary confusion matrix

<table>
<tr><td rowspan="2">True labels</td><td>Positive</td><td>True Positives (TP)</td><td>False Negatives (FN)</td></tr>
<tr><td>Negative</td><td>False Positives (FP)</td><td>True Negatives (TN)</td></tr>
<tr><td colspan="2" rowspan="2"></td><td>Positive</td><td>Negative</td></tr>
<tr><td colspan="2">Predicted labels</td></tr>
</table>

Reference: Built by the author based on [56]

Accuracy is a metric that was used to calculate the fraction of correct predictions our model made and is calculated as follows [57]:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}, (3.1)$$

Accuracy can also be calculated from the previously mentioned confusion matrices, and can be calculated as follows for a binary classification model:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, (3.2)$$

where TP – total number of True Positive values;

TN – total number of True Negative values;

FP – total number of False Positive values;

FN – total number of False Negative values.

So, for our sample confusion matrix (Table 3.2), the accuracy calculation would be as follows:

$$Accuracy = \frac{150+46+78}{405} = \frac{274}{405} = \mathbf{0.6765}, \quad (3.3)$$

As we've mentioned previously, the Lending Club dataset we've used for our model is class-imbalanced and there is significant disparity between class representation, so accuracy can't be used as the only evaluation metric. That is why we introduced Precision, Recall, and the F1 score to further evaluate model effectiveness.

Precision is defined as the fraction of True Positive (TP) values divided by the total amount of positively predicted values [56, 58]:

$$Precision = \frac{TP}{TP+FP}, \quad (3.4)$$

Recall is defined as the fraction of TP values divided by the total amount of values that were actually positive [56, 58]:

$$Recall = \frac{TP}{TP+FN}, \quad (3.5)$$

Utilizing precision and recall, we have also implemented the weighted F1 score as the main performance metric for all 4 of our models. The F1 score is defined as a harmonic mean between precision and recall [59]. The logistic regression model with a multi-class target was created by utilizing the scikit-learn library for Python [60]. The modeling process produced the following metrics:

- **Accuracy** – 0.5452;
- **Precision** – 0.5667;
- **Recall** – 0.5637;
- **F1 score (weighted)** – 0.5452.

As logistic regression is one of the most widely used and accepted methods of credit scoring, we will use these metrics as a baseline for our next 3 modeling methods. The accuracy confusion matrix for our logistic regression model can be found in Figure 3.1.

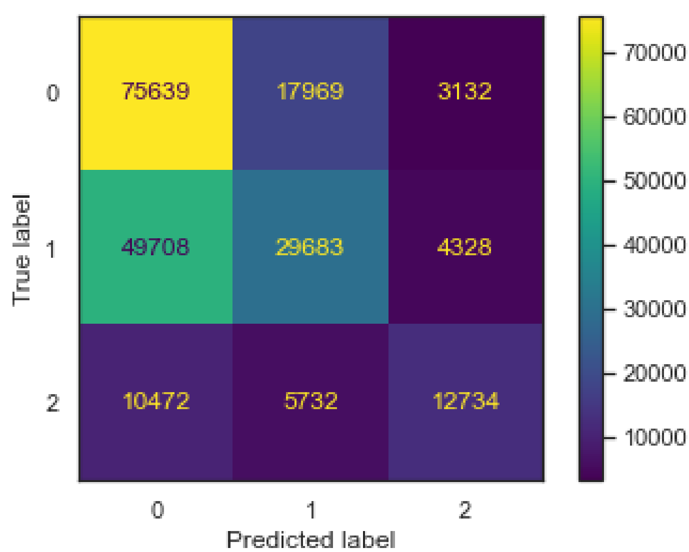

Figure 3.1 Accuracy confusion matrix of the logistic regression model built on Lending Club data.

Reference: Author's developments

The 2nd model we created was a basic decision tree model, built by utilizing the scikit-learn library for Python. The modeling process for a decision tree model produced the following metrics:

- **Accuracy** – 0.5985;

- **Precision** – 0.5997;
- **Recall** – 0.5985;
- **F1 score (weighted)** – 0.5817.

The decision tree model produced minor improvements over the logistic regression model performance-wise, or about a 6.7% improvement in the weighted F1 score and a 9.77% improvement in accuracy. The graphical representation of our decision tree model can be found in Figure 3.2.

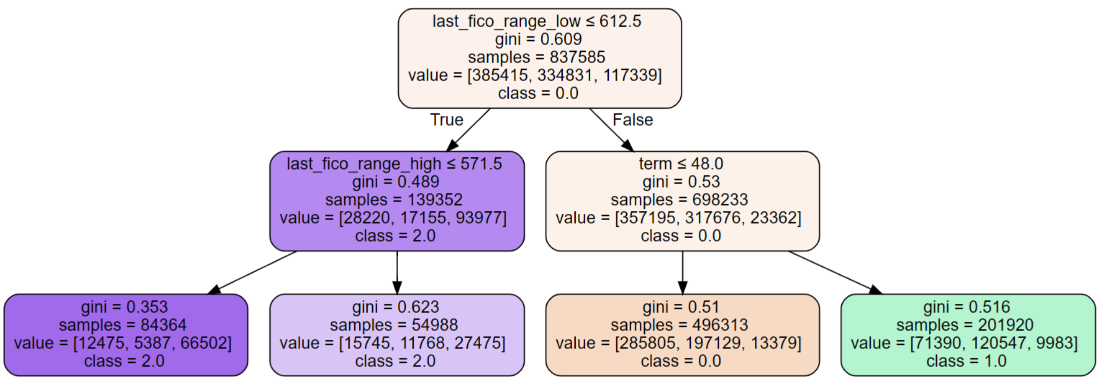

Figure 3.2 Graphical representation of our decision tree model built on Lending Club data.

Reference: Author's developments

For this specific task, a decision tree model with a depth of 2 was chosen for ease of graphical interpretation. A further depth increase to 3 and 4 improves the weighted F1 score of the model by 1.13% and 3.14% respectively, compared to the baseline model. For decision tree formation, Gini impurity was used to measure the quality of each split. As evident from the graphical interpretation of the model, the FICO score of the borrower and the loan term are the most important features in the dataset. We will touch on feature importance in further versions of the model. The accuracy confusion matrix for our decision tree model can be found below (Figure 3.3).

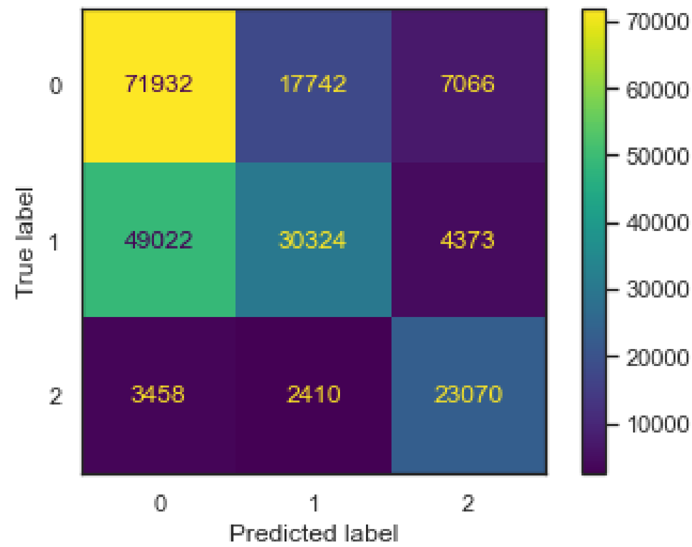

Figure 3.3 Accuracy confusion matrix of the decision tree model built on Lending Club data.

Reference: Author's developments

To compare with our baseline results, we have decided to create a random forest model on the same dataset. Random forest is a machine learning-enabled algorithm, that creates a multitude of decision trees, each of which are built for a subsample of the entire dataset. For a classification task (like in our case), the algorithm decides on the most frequent categorical variable amongst the created decision trees and outputs that as a result [36, 61]. Our modeling process for the random forest model produced the following metrics:

- **Accuracy** – 0.6977 to 0.7023;
- **Precision** – 0.6973 to 0.702;
- **Recall** – 0.6977 to 0.702;
- **F1 score (weighted)** – 0.6973 to 0.702.

As is evident from the modeling results, there is a certain degree of randomness when using the random forest algorithm for classification or regression tasks, which results in boundaries between which all metrics fall. However, both the highest and lowest metrics output by the random forest algorithm are a significant improvement over our logistic regression algorithm. Metrics wise, the random forest algorithm brings a ≈28% improvement to both accuracy and the weighted F1 score, compared to logistic regression (Figure 3.4).

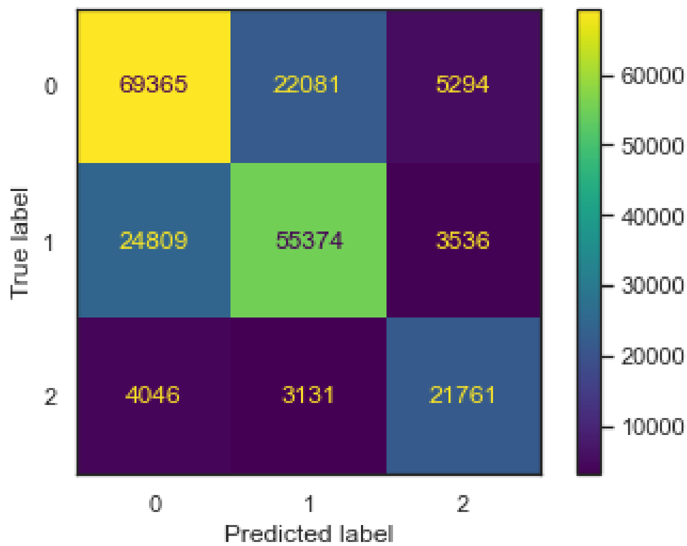

Figure 3.4 Accuracy confusion matrix of the random forest model built on Lending Club data.

Reference: Author's developments

Moving on, we utilized a black box machine learning algorithm XGBoost to attempt to improve upon our modeling results. XGBoost (stands for Extreme Gradient Boosting) is a gradient boosted decision tree machine learning algorithm and is one of the leading ML algorithms for regression and classification tasks [62, 63]. One of the main specifics of using this algorithm is that the model is highly customizable through a set of hyperparameters, which were tuned as follows:

- *gamma* – The minimum reduction in loss necessary to warrant a further partition on a leaf node within the tree; larger values increase conservativeness of the algorithm, set to **0.1** (default=0);

- *learning_rate* – The reduction in step size applied during updates to prevent overfitting; larger values increase conservativeness of the algorithm, set to **0.38** (default=0.3);
- *max_delta_step* – Maximum delta step that each leaf's output is allowed to be; positive values of this parameter make the algorithm more conservative, set to **1** (default=0, no limit);
- *min_child_weight* – The minimum sum of instance weights required in a child node; if the partitioning step of the tree results in a leaf node with a sum of instance weights lower than min_child_weight, the building process will halt further partitioning; increase of this parameter will make the algorithm more conservative, set to **5** (default=1);

Below you can find the accuracy confusion matrix for our XGBoost model, prior to hyperparameter tuning (Figure 3.5). As we are working with a multi-class classification task, the extent towards which we can utilize hyperparameter tuning is limited, as we can't directly address class imbalance in a multi-class dataset.

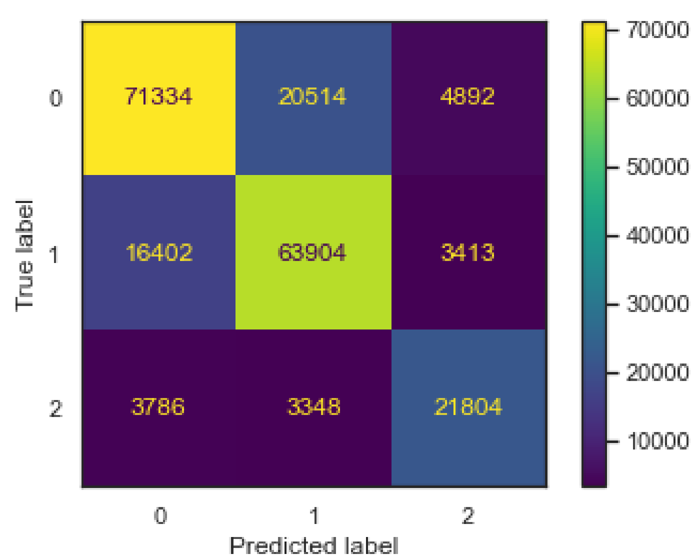

Figure 3.5 Accuracy confusion matrix of the XGBoost model built on Lending Club data (pre-hyperparameter tuning).

Reference: Author's developments

Also important to mention here are the feature importances for each of the datapoints we've used for this model, which are the following:

1. **Loan_amnt** – 0.0235;
2. **Funded_amnt** – 0.0844;
3. **Funded_amnt_inv** – 0.0473;
4. **Term** – 0.2167;
5. **Int_rate** – 0.1777;
6. **Installment** – 0.0088;
7. **Annual_inc** – 0.0086;
8. **Dti** – 0.0116;
9. **Delinq_2yrs** – 0.0083;
10. **Open_acc** – 0.0081;
11. **Revol_util** – 0.0099;
12. **Last_fico_range_high** – 0.3946.

Remaining hyperparameters were left at their default values. The hyperparameter turning, combined with the XGBoost algorithm produced the following metrics:

- **Accuracy** – 0.7505;
- **Precision** – 0.7518;
- **Recall** – 0.7505;
- **F1 score (weighted)** – 0.7507.

The metrics produced by the XGBoost algorithm are the highest amongst the 4 models built, outperforming logistic regression and random forest by 37.7% and 6.9% in the weighted F1 score respectively (Figure 3.6). The accuracy confusion matrix for our XGBoost model can be found below.

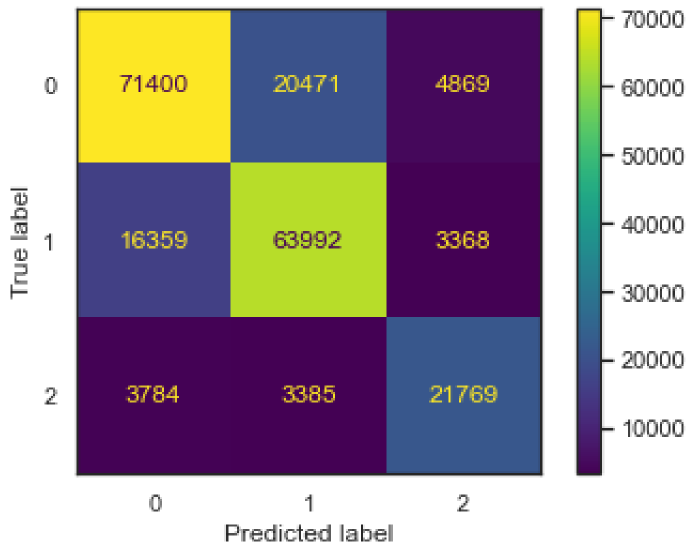

Figure 3.6 Accuracy confusion matrix of the XGBoost model built on Lending Club data (post-hyperparameter tuning).

Reference: Author's developments

While out of the models we've tested, ML-based modeling methods outperform logistic regression by metrics, there are also additional considerations when it comes to selecting the right modeling method for a financial institution. As previously mentioned, the main two types of loan institutions in Ukraine we want to highlight are banks and microfinancial organizations, and for each of them there could be a different modeling solution they'd want to implement to increase solvency of their clients.

The main drawback of using ML-based models is the difficulty in interpreting the results. While solutions like XGBoost provide feature importance for each of the input datapoints used in the modeling process, they are greatly limited when it comes to variable interpretation. Also, since it is a relatively new methodology for the market, it would need a proper implementation environment and development power in order to perform adequately, compared to the institution's pre-existing credit scoring solution [64].

Logistic regression, however, also has certain drawbacks when applied for credit scoring in financial institutions. Same as with machine learning-enabled methods, it requires development efforts and a separate (though, smaller, and not as proprietary) team to maintain the predictive algorithms. Also, there is an argument of iterations of the same model and their manual implementation, one the new iteration has been proven more accurate than the last (Table 3.4).

Table 3.4 Pros and Cons between Logistic Regression and Gradient Boosted Decision Trees

	Logistic Regression	**Gradient Boosting**
Pros	Direct parameters interpretation	Fast development
	Easy deployment	Stress of parameters
	Greater stability over time	Flexibility (Target: Binary/Multinomial/Interval)
	Business view	
	Market confidence	
Cons	Development effort	Needs an implementation environment
	Manual fit (iterations)	New methodology for the market
		Low variable interpretation

Reference: Built by the author based on [64]

For banks with established credit portfolios and long-standing corporate clients, it is challenging to introduce new methodology on top of their pre-existing credit scoring processes as they tend to be conservative. Ukrainian state-owned banks are holding ≈75% of non-performing loans in the country, however, that is also influenced by other factors besides outdated credit scoring methodology [13]. Considering this, we believe that for banking institutions logistic regression modeling of borrower’s credit scores can be a good fit, if that is the method they used before. Parameters of logistic regression models can be more easily explained to the end customer, while its ease of deployment allows the analytics team to deploy newer, better iterations of the model relatively fast.

Finding a good credit scoring solution for microfinancial organizations in Ukraine is challenging, as they currently prioritize speed of loan application and approval over balancing out their credit portfolio. With the latest regulatory developments in the market, the tendency could shift for payday loan companies to attempt to reduce the share of non-performing loans in their portfolios, simultaneously with a significant reduction in interest rates [24]. For these types of companies, and for other financial companies operating on the consumer loan market in Ukraine, utilizing machine learning-enabled modeling methods could be a viable alternative to logistic regression. It is especially relevant for newly founded companies, as they can create the infrastructure and the implementation environment for deploying such solutions, while not reducing customer transparency by switching to a different credit scoring method.

Conclusions to Chapter 3

The main goal of our modeling efforts was to empirically compare the 4 selected modeling methods by attempting to predict solvency evaluation on a publicly available dataset of individuals that applied for a loan. Out of the multiple datasets we've found, we have decided to use a dataset from Lending Club, a US-based financial services company providing loans to individuals because of its data quality, lack of severe class imbalance and a large number of instances in the database. All the models were measured based on the following metrics: accuracy, precision, recall and the F1 score.

Results of our modeling efforts placed these algorithms in the following order best-worst, based on the F1 score: XGBoost, Random Forest, Decision Tree, Logistic Regression. However, performance metrics are not the only measure for these models, and whether or not they can be effectively used in Ukrainian financial institutions is the deciding factor.

Based on the advantages and disadvantages of the presented methods, we have concluded that while the 3 data mining algorithms we tested provided better metrics compared to logistic regression, however for banks and their more traditionally conservative business structures, switching to data mining algorithms would require a significant initial investment and a considerable change in the existing infrastructure. We would delve deeper into these issue in our conclusions and provide analysis on which data mining methods are more efficient for each type of financial institution in Ukraine.

CONCLUSIONS

As it has been 10 years since the start of the war in Ukraine and over 2 years since the full-scale invasion, the Ukrainian economy has been in a constant struggle to balance out military expenses and creating a stable economy. One of the main indicators of a stable financial system we highlighted as part of this paper is the non-performing loans ratio, or the percentage of loans in a banks' portfolio that the borrower has defaulted on. Over the years, there have been a number of events that led to the increase of the NPL ratio in Ukraine:

- An Asset Quality Review conducted by the National Bank of Ukraine since 2018, which as a result recognized a lot of outstanding loans as non-performing [16];
- Introduction of a stricter definition of non-performing loans by the NBU in accordance with international banking standards since 2016 [17];
- The nationalization of certain banks like PrivatBank in 2016 and Sense Bank in 2023, which led to a recognition of a significant portion of their loan portfolios as non-performing [14, 65].

These, and other events led to a state of the financial market where the NPL ratio in Ukraine is at 36.6%, with ≈75% of those loans help by state-owned banks (including PrivatBank and Sense Bank) [13]. Simultaneously, there has been a significant surge in recent years in the amount of non-banking financial institutions in Ukraine that provide short-term payday loans that are infamous for their low loan portfolio health and disregard towards solvency evaluation processes [25].

With this in mind, we want to propose an alternative to existing credit scoring methods on the market, by applying data mining methodology on an individual borrower dataset and comparing its performance to a "traditional" credit scoring method, logistic regression. To our knowledge, there is no existing publicly available dataset of Ukrainian individual borrowers, so the dataset we selected was taken from a US-based financial institution called Lending Club.

Four models were created on the same Lending Club dataset as part of research conducted for this paper: a logistic regression model, a decision tree model, a random forest model and a model utilizing the XGBoost library. All models were ranked based on accuracy and their weighted F1 score and produced the following results:

- Logistic regression
  - Accuracy – 0.5452;
  - F1 Score – 0.5452.
- Decision tree
  - Accuracy – 0.5985;
  - F1 Score – 0.5817.
- Random forest (best estimator)
  - Accuracy – 0.7023;
  - F1 Score – 0.7020.
- XGBoost
  - Accuracy – 0.7505;
  - F1 Score – 0.7507.

According to the algorithm's comparative performance, they can be ordered in the following manner based on their weighted F1 score: XGBoost, Random Forest, Decision Tree, Logistic Regression. As we highlighted in Chapter 3 however, performance metrics are not the only meaningful measurements of the model when it comes to real world applications. When comparing XGBoost to Logistic Regression, there is a series of compromises a financial services company must make when selecting one over another for their credit scoring efforts. The issue inherent to XGBoost and other data mining algorithms is their severe lack of interpretability for the end user, who is ultimately going to be the individual borrower coming to your institution for a loan.

While there is a plethora of “black box” credit scoring solutions on the market, going from logistic regression-based modeling to data mining methods would be challenging for large-scale financial institutions due to the cost of developing a custom data mining solution and creating an infrastructure for its utilization. When it comes to application of these modeling methods for the Ukrainian financial market, these are the conclusions we want to draw:

1. While data mining methods outperform logistic regression by the metrics we selected, the infrastructural cost of deployment of these solutions would mean that, most likely, only small-scale institutions would be able to realize their full potential. As mentioned in Chapter 3, we believe that the continuous efforts by the National Bank of Ukraine to improve loan portfolio health amongst microfinancial companies could lead to an increased demand among them for an accurate credit scoring solution. And for companies that do not have pre-existing proprietary solvency evaluation techniques implemented, the cost of deployment would be comparable to logistic regression.
2. For banks, a deeper analysis of the Ukrainian loan market is required to draw solid conclusions. If one of the state-owned banks would publish an anonymized dataset of their loan applicants, we could train models on that data and directly compare our performance to that of the banks’ model. As of now, we believe that if major Ukrainian banks are using logistic regression-based models for solvency evaluation, these models can be vastly improved upon over iterations and lead to a significant reduction of NPLs in the banks’ portfolios.

BIBLIOGRAPHY

1. Asli Demirguc-Kunt, Leora Klapper, Dorothe Singer "Financial Inclusion and Inclusive Growth: A Review of Recent Empirical Evidence". URL: https://documents1.worldbank.org/curated/en/403611493134249446/pdf/WPS8040.pdf
2. Aire "Credit where credit is due: scoring the right balance in today’s economy". URL: https://www.credit-connect.co.uk/wp-content/uploads/2016/07/Aire_ResearchPublication_UNDER-STRICT-EMBRAGO-UNTIL-THURSDAY-28-SEPTEMBER-2.pdf
3. William E. Hardy Jr., John L. Adrian Jr. "A linear programming alternative to discriminant analysis in credit scoring".
4. The World Bank Group "Credit Scoring Approaches Guidelines". URL: https://thedocs.worldbank.org/en/doc/935891585869698451-0130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf
5. The United States Federal Register "Fair Credit Reporting (Regulation V)". URL: https://www.federalregister.gov/documents/2011/12/21/2011-31728/fair-credit-reporting-regulation-v
6. Congressional Research Service "Consumer Credit Reporting, Credit Bureaus, Credit Scoring, and Related Policy Issues". URL: https://crsreports.congress.gov/product/pdf/R/R44125#
7. Federal Reserve "Federal Fair Lending Regulations and Statutes, Equal Credit Opportunity (Regulation B)". URL: https://www.federalreserve.gov/boarddocs/supmanual/cch/fair_lend_reg_b.pdf
8. Krista Santos "Corporate credit ratings: a quick guide". URL: https://www.treasurers.org/ACTmedia/ITCCMFcorpcreditguide.pdf

9. Liran Einav, Mark Jenkins, Jonathan Levin "The impact of credit scoring on consumer lending". URL: https://www.jstor.org/stable/43186418
10. The Fair Isaac Corporation "What is a FICO® Score?". URL: https://www.myfico.com/credit-education/what-is-a-fico-score
11. The Fair Isaac Corporation "What's in my FICO® Scores?". URL: https://www.myfico.com/credit-education/whats-in-your-credit-score
12. National Bank of Ukraine, Supervisory data "Aggregated outstanding amounts on balance sheet accounts of the Ukrainian banks (breakdown by banks)". URL: https://bank.gov.ua/files/stat/aggregation_2024-02-01_eng.xlsx
13. National Bank of Ukraine "Loan Portfolio Quality (NPLs)". URL: https://bank.gov.ua/en/stability/npl
14. PrivatBank "PrivatBank under State Protection". URL: https://pb.bank.gov.ua/en.html
15. Statista "Non-performing loans and advances (NPL) as a share of gross loans in the euro area from 2nd quarter 2015 to 2nd quarter 2023". URL: https://www.statista.com/statistics/1122984/non-performing-bank-loans-in-europe/
16. National Bank of Ukraine "Resilience Assessment of Ukraine's Banks". URL: https://bank.gov.ua/en/supervision/about/diagnostics-of-banks#
17. National Bank of Ukraine "Board Resolution №351 "On the Regulation for Measuring Credit Risk Generated by Banks’ Asset Operations". URL: https://zakon.rada.gov.ua/laws/show/v0351500-16#Text
18. Yuriy Fedoriv, KPMG "Large NPL sale in Ukraine: why now and entry considerations". URL: https://kpmg.com/ua/en/home/media/press-releases/2022/02/large-npl-sale-in-ukraine-why-now-and-entry-considerations.html

19. The Pew Charitable Trusts "Payday Lending in America: How Borrowers Choose and Repay Payday Loans". URL: https://www.pewtrusts.org/~/media/assets/2013/02/payday_loans_report2overview_finalforweb.pdf
20. Opendatabot "Понад 650 тисяч мікрокредитів беруть українці щомісячно". URL: https://opendatabot.ua/analytics/mfo-2023-2
21. Opendatabot "Кількість мікрокредитів зросла у 1,5 рази за 2023 рік". URL: https://opendatabot.ua/analytics/mfo-2023
22. Silvia Szilagyiova "The Effect of Payday Loans on Financial Distress in the UK". URL: https://www.sciencedirect.com/science/article/pii/S2212567115013349
23. Mykhailo Nedozhdii "Застосування скорингових моделей у небанківських фінансових установах" / "Scoring models applications in non-bank financial institutions (NBFIs)".
24. Закон України "Про внесення змін до деяких законів України щодо удосконалення державного регулювання ринків фінансових послуг". URL: https://zakon.rada.gov.ua/laws/show/3498-20#Text
25. National Bank of Ukraine "Національний банк вітає ухвалення закону, який запроваджує відповідальне кредитування в Україні". URL: https://bank.gov.ua/ua/news/all/natsionalniy-bank-vitaye-uhvalennya-zakonu-yakiy-zaprovadjuye-vidpovidalne-kredituvannya-v-ukrayini
26. Artur Lee Samuel "Some Studies in Machine Learning Using the Game of Checkers". URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5392560
27. Kavyashree GN, Prof. Arvind L.N., International Journal of Research Publication and Reviews "Artificial Intelligence and Machine Learning in Financial Sector". URL: https://ijrpr.com/uploads/V5ISSUE1/IJRPR22064.pdf

28. Ziyu Shang, Zhongyuan Wang "Machine Learning in Finance: A Brief Review". URL: https://www.atlantis-press.com/article/125977091.pdf
29. Foster Provost, Tom Fawcett "Data Science for Business".
30. Johann Faouzi, Olivier Colliot "Machine Learning for Brain Disorders, Classic machine learning algorithms". URL: https://hal.science/hal-03830094v1/file/Chapter%2002%20-%20Final.pdf
31. Jacqueline Fauland, SSI Schäfer "Artificial Intelligence in Logistics". URL: https://www.ssi-schaefer.com/resource/blob/1080670/06d87a3eff1abfbdd7af3875404b724a/white-paper-artificial-intelligence-in-logistics-dam-download-en-16558--data.pdf
32. Daniel Jurafsky, James H. Martin "Speech and Language Processing". URL: https://web.stanford.edu/~jurafsky/slp3/5.pdf
33. Oscar Torres-Reyna "Logit, Probit, and Multinomial Logit models in R". URL: https://www.princeton.edu/~otorres/LogitR101.pdf
34. Mengye Ren, Matthew MacKay "CSC 411: Introduction to Machine Learning". URL: https://www.cs.toronto.edu/~mren/teach/csc411_19s/lec/lec03.pdf
35. International Business Machines "What is a decision tree?". URL: https://www.ibm.com/topics/decision-trees
36. Leo Breiman "Random Forests". URL: https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf
37. Tony Yiu, Towards Data Science "Understanding Random Forest". URL: https://towardsdatascience.com/understanding-random-forest-58381e0602d2
38. Katherine Woodruff "Introduction to boosted decision trees". URL: https://indico.fnal.gov/event/15356/contributions/31377/attachments/19671/24560/DecisionTrees.pdf
39. XGBoost "Introduction to Boosted Trees". URL: https://xgboost.readthedocs.io/en/stable/tutorials/model.html

40. GitHub "XGBoost". URL: https://github.com/dmlc/xgboost
41. Tianqi Chen, Carlos Guestrin "XGBoost: A Scalable Tree Boosting System". URL: https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf
42. Xue Ying "An Overview of Overfitting and its Solutions". URL: https://iopscience.iop.org/article/10.1088/1742-6596/1168/2/022022/pdf
43. Theja Tulabandhula, Cynthia Rudin "Machine Learning with Operational Costs". URL: https://arxiv.org/pdf/1112.0698.pdf
44. XGBoost "XGBoost Parameters". URL: https://xgboost.readthedocs.io/en/stable/parameter.html
45. GitHub "Credit Scoring Datasets". URL: https://github.com/JLZml/Credit-Scoring-Data-Sets?tab=readme-ov-file
46. UC Irvine Machine Learning Repository "Statlog (German Credit Data)". URL: https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data
47. UC Irvine Machine Learning Repository "Statlog (Australian Credit Approval)". URL: http://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29
48. UC Irvine Machine Learning Repository "Default of credit card clients". URL: https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients
49. UC Irvine Machine Learning Repository "Japanese Credit Screening". URL: https://archive.ics.uci.edu/dataset/28/japanese+credit+screening
50. UC Irvine Machine Learning Repository "Polish Companies Bankruptcy". URL: https://archive.ics.uci.edu/dataset/365/polish+companies+bankruptcy+data
51. Kaggle "Give Me Some Credit". URL: https://www.kaggle.com/competitions/GiveMeSomeCredit/overview
52. Kaggle, Home Credit Group "Home Credit Default Risk". URL: https://www.kaggle.com/c/home-credit-default-risk/overview

53.Kaggle, Nathan George "All Lending Club loan data". URL: https://www.kaggle.com/datasets/wordsforthewise/lending-club

54.Mykhailo Nedozhdii "Scoring models in solvency evaluation". URL: https://ekmair.ukma.edu.ua/items/acfecc5b-a257-470c-a5c5-5e447c7a4670

55.LendingClub "About Us". URL: https://www.lendingclub.com/company/about-us

56.Margherita Grandini, Enrico Bagli, Giorgio Visani "Metrics for Multi-class Classification: An Overview". URL: https://arxiv.org/pdf/2008.05756.pdf

57.Google for Developers "Classification: Accuracy". URL: https://developers.google.com/machine-learning/crash-course/classification/accuracy

58.Google for Developers "Classification: Precision and Recall". URL: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

59.Yutaka Sasaki "The truth of the F-measure". URL: https://www.researchgate.net/publication/268185911_The_truth_of_the_F-measure

60.scikit-learn Website. URL: https://scikit-learn.org/stable/

61.IBM "What is random forest?". URL: https://www.ibm.com/topics/random-forest

62.NVIDIA "XGBoost". URL: https://www.nvidia.com/en-us/glossary/xgboost/

63.XGBoost Documentation. URL: https://xgboost.readthedocs.io/en/stable/

64.Paulo Celio Di Cellio Dias, Serasa Experian; Melissa Forti, Bradesco; Marc Witarsa, Serasa Experian "A comparison of Gradient Boosting with Logistic Regression in Practical Cases". URL: https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1857-2018.pdf

65.The Ministry of Finance of Ukraine "Уряд прийняв рішення про участь держави у виведенні з ринку АТ «СЕНС БАНК» для забезпечення стабільності фінансової системи України та захисту інтересів вкладників й інших кредиторів банку". URL: https://www.mof.gov.ua/uk/news/uriad_priiniav_rishennia_pro_uchast_derzhavi_u_vivedenni_z_rinku_at_sens_bank_dlia_zabezpechennia_stabilnosti_finansovoi_sistemi_ukraini_ta_zakhistu_interesiv_vkladnikiv_i_inshikh_kreditoriv_banku-4116

APPENDICES

Appendix A

Variable Name	Feature Importance
open_il_12m	4142.196777
open_rv_12m	1542.601929
term	1466.71521
last_pymnt_amnt	733.4886475
last_fico_range_high	615.616333
open_il_24m	254.6094971
open_rv_24m	212.7248993
loan_amnt	162.6050568
total_rec_prncp	118.7225571
funded_amnt_inv	96.75402069
installment	43.25679398
funded_amnt	40.52630615
total_cu_tl	30.1675148
total_acc	28.19219208
total_rec_int	25.10643959
pub_rec	24.87986755
open_acc_6m	22.25640488
total_pymnt	18.50349426
fico_range_low	17.83610535
num_rev_accts	17.26969719
pub_rec_bankruptcies	16.57479668
int_rate	15.3834753
total_pymnt_inv	14.69959641
max_bal_bc	14.51488781
inq_fi	13.28515148
all_util	11.18903351
open_act_il	9.626737595
pct_tl_nvr_dlq	9.454991341
mo_sin_old_rev_tl_op	9.399244308
total_bal_il	8.076700211
annual_inc	7.443414688
num_rev_tl_bal_gt_0	7.229462624
inq_last_6mths	6.92124939
inq_last_12m	6.853680134
num_actv_rev_tl	6.300444603
tot_cur_bal	5.448857307
mths_since_rcnt_il	5.411586285
num_actv_bc_tl	5.122211456

tax_liens	5.08290863
dti	4.770335197
acc_now_delinq	4.636672974
num_bc_tl	4.611124516
num_sats	4.559018612
revol_util	4.481777191
il_util	4.318377495
mort_acc	4.23985815
tot_hi_cred_lim	3.761975288
total_bal_ex_mort	3.722453117
mths_since_recent_revol_delinq	3.707401276
avg_cur_bal	3.69135499
bc_open_to_buy	3.655034065
mths_since_recent_inq	3.550478697
bc_util	3.545904398
acc_open_past_24mths	3.525101662
delinq_amnt	3.390398741
num_bc_sats	3.378497124
open_acc	3.367823839
total_bc_limit	3.286400318
mths_since_recent_bc	3.280458927
total_il_high_credit_limit	3.274564028
tot_coll_amt	3.256260633
revol_bal	3.222468376
delinq_2yrs	3.211682796
num_op_rev_tl	3.189350128
total_rev_hi_lim	3.158354044
percent_bc_gt_75	3.126595497
mo_sin_rcnt_tl	3.02512455
mo_sin_old_il_acct	3.02234149
num_tl_90g_dpd_24m	2.967826366
num_il_tl	2.947360277
mo_sin_rcnt_rev_tl_op	2.87351799
mths_since_recent_bc_dlq	2.767369032
num_accts_ever_120_pd	2.594109058
num_tl_120dpd_2m	2.508686781
num_tl_op_past_12m	2.105986357
chargeoff_within_12_mths	1.289459944