TUNIS BUSINESS SCHOOL
UNIVERSITY OF TUNIS

Republic Of Tunisia
University Of Tunis
Tunis Business School

END OF STUDIES INTERNSHIP REPORT IN BUSINESS ANALYTICS & IT

as a partial requirement for the degree of Bachelor in Business Administration

TITLE

Leveraging Machine Learning Techniques to Enhance Credit Risk Assessement and Provisioning Decisions in The Banking Sector

REALISED BY
Zeineb HAJJI

ACADEMIC ADVISOR
Dr. Montassar BEN MESSAOUD

PROFESSIONAL SUPERVISOR
Mr Wissem KRIMI

alBaraka

Academic Year
2024-2025

APPROVAL

APPROVED BY

ACADEMIC ADVISOR

Name	Signature	Date

PROFESSIONAL SUPERVISOR

Name	Signature	Date

ACADEMIC EVALUATOR

Name	Signature	Date

Declaration

I certify that I am the author of this project and that any assistance I received in its preparation is fully acknowledged and disclosed in this project. I have also cited any source from which I used data, ideas, or words, either quoted or paraphrased. Further, this report meets all of the rules of quotation and referencing in use at TBS, as well as adheres to the fraud policies listed in the TBS honor code.

No portion of the work referred to in this study has been submitted in support of an application for another degree or qualification to this or any other university or institution of learning.

Student Name................................ Signature .... Date........................

Acknowledgements

I thank above all God
for the guidance, strength, faith, and blessings bestowed upon me throughout the realization of my end-of-studies project.

I would like to express my sincere gratitude to everyone who supported me throughout the completion of this end of studies project.

To my academic supervisor

I am deeply thankful to Dr. Montassar Ben Messaoud, my academic advisor at Tunis Business School, for his ongoing guidance, constructive feedback, and encouragement. His expertise and insightful advice played a crucial role in shaping the direction and quality of this project.

To my professional supervisor

I wish to extend my thanks to Mr. Wissem Krimi, my professional supervisor at Al Baraka Bank, for giving me the opportunity to carry out this project in a real-world context and for his constant support and availability during my internship.

To the Risk Direction team at Al Baraka Bank

My appreciation goes as well to the entire Risk Direction team at Al Baraka Bank for their collaboration, trust, and valuable assistance. Working alongside such a dedicated and knowledgeable team has been a truly enriching and formative experience.

To my family

I am deeply grateful to my family for their unconditional love, patience, and constant support throughout my academic journey. Their encouragement during challenging moments and belief in my potential have been my greatest source of strength. To my parents, especially thank you for your sacrifices, values, and the trust you placed in me. This achievement would not have been possible without your unwavering presence by my side.

To my friends

I would like to sincerely thank my friends for always being there with kind words, motivation, and shared moments of joy and relief. Your support, whether emotional or practical, helped me stay balanced and focused during this journey. I'm lucky to have such caring and understanding people around me.

Abstract

This project explores the use of machine learning techniques to classify credit risk levels of clients within a financial institution. Accurate credit risk classification is essential for minimizing financial losses and improving loan management. The primary goal of this research is to predict the future credit risk category of each client by analyzing their past financial behavior. The dataset includes financial and time-based information, enabling both static and dynamic modeling approaches.
We have developed four machine learning models and evaluated: Bayesian Network, Hidden Markov Model, Random Forest, and XGBoost. Clients with only one year of data were analyzed using a Static Bayesian Network, while Dynamic Bayesian Networks were applied to those with multi-year records to capture time-related patterns. Hidden Markov Models were used to identify hidden transitions in risk behavior. Ensemble methods like Random Forest and XGBoost were employed to improve prediction accuracy and handle class imbalance.
We assessed each model using accuracy, precision, recall, F1-score, and Quadratic Weighted Kappa (QWK), which is especially useful for ordinal classification tasks. Among the tested models, XGBoost demonstrated the best performance, achieving a QWK score of 0.91 and an accuracy of 0.93. These results indicate its strong ability to classify clients across multiple risk categories.
This study demonstrates the practical value of advanced machine learning models in enhancing credit risk assessment. The findings support their integration into financial decision making processes and risk management strategies.

Keywords: Credit Risk, Risk Classification, Machine Learning, Dynamic Modeling, XGBoost.

TABLE OF CONTENTS

List Of Figures

List Of Tables

Abbreviations

NPL	Non Performing Loan
ML	Machine Learning
IFRS	International Financial Reporting Standards
BN	Bayesian Network
DAG	Directed Acyclic Graph
CPT	Conditional Probability Table
DBN	Dynamic Bayesian Network
HMM	Hidden Markov Model
QWK	Quadratic Weighted Kappa
CRISP-DM	Cross Industry Standard Process for Data Mining
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
API	Application Programming Interface
RESTful API	REpresentational State Transfer Application Programming Interface

Introduction

According to a recent article published by (World Bank, 2024), credit risk remains one of the most significant challenges for financial institutions, especially during times of economic uncertainty. Banks need to find better ways to manage this risk, as relying only on traditional tools is no longer enough. With increasing global financial instability and stricter international regulations, such as Basel III and IFRS 9, banks must adopt more advanced methods to evaluate and manage credit risk.
Traditional credit risk assessment techniques mostly depend on financial ratios and expert opinion.While these approaches have been used for decades, they are limited. They usually provide a static view of a client's financial situation at a specific point in time and often fail to detect how a client's behavior changes over time.This becomes a serious issue when trying to identify early warning signs that a client might become risky in the future.Moreover, another limitation of traditional methods is the class imbalance problem. In most loan portfolios, the number of risky clients is much smaller than the number of good clients.This imbalance can cause ignoring the high-risk clients, making it harder to detect and respond to potential defaults. In developing countries like Tunisia, banks often face additional difficulties.These include less access to high quality data, fewer resources to adopt new technologies, and rapidly changing local market conditions. All these factors make credit risk evaluation even more complex.
In addition, Islamic banks, such as **Al Baraka Bank**, face further challenges due to the nature of their financial products. Since Islamic banking follows Sharia principles, it avoids interest-based lending and uses profit and loss sharing contracts. This unique approach creates a different risk profile, which makes it harder to apply global standards like IFRS 9 directly without adjustments. The main goal of this project is to build and test machine learning models that can better predict a client's future credit risk level using their historical financial and behavioral data.The project uses internal banking data collected over four years. Several advanced models are tested, including Bayesian Networks,Hidden Markov Models, Random Forest, and XGBoost. Each model is designed to capture different aspects of client behavior. These models are then evaluated using a comprehensive set of performance metrics, with particular emphasis on their ability to correctly classify clients across different risk categories with particular attention to the Quadratic Weighted Kappa (QWK) that account for ordinal risk classes.
To make the results useful for the bank, the project also includes the development of a user friendly tool.This tool is designed to help risk managers explore client risk profiles and apply the model's predictions in their decision making, without requiring any understanding of the underlying models or technical implementation.
To better manage the project and respond effectively to the main problem, a study plan has been developed.

This report is divided into five main chapters. Chapter 1 provides a general presentation of Al Baraka Bank and the project context. Chapter 2 presents a literature review covering credit risk concepts, machine learning applications in finance, and relevant modeling approaches. Chapter 3 details the methodology, including the CRISP-DM framework, dataset characteristics, and model implementations. Chapter 4 describes the technical implementation aspects, including data preprocessing and model development. Chapter 5 presents the results and findings, comparing model performance and discussing business implications.

Chapter 1

General Presentation

This chapter introduces Al Baraka Bank Tunisia, a pioneer in Islamic finance, by presenting an overview of its history, vision, and strategic positioning through a SWOT analysis. It then describes the context of my internship within the bank's Risk Management directory . The final section sets the foundation for the core problem addressed in this project: the need for a forward looking approach to credit risk assessment.

1.1 Introduction To The Bank

1.1.1 History Of The Bank

Islamic finance is an ever-evolving financial sector, and Al Baraka Bank Tunisia has played a pioneering role in its development in Tunisia and the Maghreb region. It was created by the Tunisian State and Sheikh Salah Abdallah KAMEN, founder of the DALLAH AL BARAKA group, on June 15, 1983, under the name: BEIT ETTAMWIL TOUNSI SAOUDI-BEST BANK. As part of the unification of the group's business identity, BEST BANK has changed its name in 2009 to become ”Al Baraka bank Tunisia” as part of Al Baraka Banking group-ABG.In 2013, Al Baraka bank Tunisia obtained approval to transform an off-bank shore into a Universal Resident Bank, enabling it to meet the needs of the financial services of all categories of customers: individuals, professionals, and companies in the means of a trusting and lasting partnership.

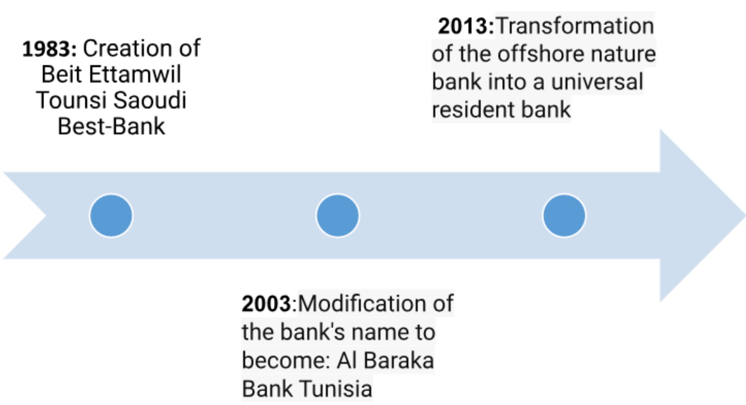

Figure 1.1: History of The Bank

1.1.2 Vision of Al Baraka Bank

To be a global leader in innovative crowdfunding, offering an ethical and agile financial system designed for the digital age.(Al Baraka Bank Tunisia, 2025)

1.1.3 Mission of Al Baraka Bank

To meet the financial needs of communities around the world by conducting business in a customer-centric, digital-first ethical approach, based on our core beliefs to share our business successes with our business partners: our customers, our employees, our shareholders, and our community at large.(Al Baraka Bank Tunisia, 2025)

1.1.4 Ownership and Shareholding structure

The capital of Al Baraka Bank Tunisia is set at 120,000,000 Tunisian dinars, with ownership distributed as follows: 78.4% is held by Al Baraka Group BSC (c), the parent company, while 10% is owned by the Tunisian State and 10% by the CNSS. The remaining 1.6% is held by other smaller shareholders. This ownership structure reflects the bank's strong ties to both local and international entities, reinforcing its position in the Tunisian banking sector.

Table 1.1: Bank Profile

Company Name	Al Baraka Bank Tunisia
Creation	1983
Localisation	Av.Cheikh Mohamed Fadhel Ben Achour Tunis
Legal Form	Public Limited Company
Sector of Activity	Financial Sector
Capitalization	120 million dinars
Branch Network	38 and 2 boxes

1.1.5 SWOT Analysis

We use SWOT analysis to build an understanding of the strategic framework of AL Barka Bank.

Strengths

- **42 years of experience in islamic finance:** Since its creation on June 15, 1983, Al Baraka Bank was the leader in Islamic finance in Tunisia.It was the only Islamic bank in Tunisia operating in this domain until 2009 when its first competitor Zitouna Bank joined the industry.
- **Rapid Growth:** By obtaining the approval for the transformation from a non-resident bank to a resident universal bank,the number of agencies increased from 8 agencies in 2013 to 38 agencies.This shows the rapid growth and expansion policy adopted by top management.
- **Wide Range of Services:** The bank offers a variety of services, including retail banking, corporate banking, and investment management.

Weaknesses

- **Over-Reliance on Traditional Banking:** Despite digital efforts, the bank still relies heavily on its physical branch network, which increases operational costs.
- **Limited International Presence:** The bank is less competitive in international markets compared to some global financial institutions.

Opportunities

- **Low Competition:** In Tunisian banking sector , the market is saturated. However this is not a disadvantage for the Islamic Banks due to the limited number of direct competitors.
- **Foreign Investment Prospects:** Al Baraka Banking Group is leader in Islamic Finance , its presence in 15 Arab Countries can easily facilitate foreign investment while strengthening the bank's position locally.

Threats

- **Ambiguous market conditions:** The Tunisian banking sector's environment is unstable and complex.Furthermore , the last couple of years, The Central Bank has introduced new drastic measures and published new structures has mainly affected Islamic banks.This has increased several costs and reducing at the same time initial margins.
- **Cybersecurity Risks:** With the rise in digital banking, the risk of cyberattacks and data breaches is becoming a growing concern.

1.2 Internship Description

My internship took place in the headquarter of Al Baraka Bank in Tunis.I was assigned to work in the Risk management department.

1.2.1 Central Directorate Of Risk Management And Internal Control

The central directorate for risk management and internal control, which was recently established in March 2008, reports directly to general management. The purpose of this directorate is to ensure that the bank's risks are in accordance with its risk policies, management objectives, and profitability. It exercises continuous control over the risks associated with financing, profit rates, and liquidity, as well as operational risks, and informs the Board of Directors of the risk situation to which the bank is exposed. Below, we present the organizational chart of the central management of risk management and internal control.

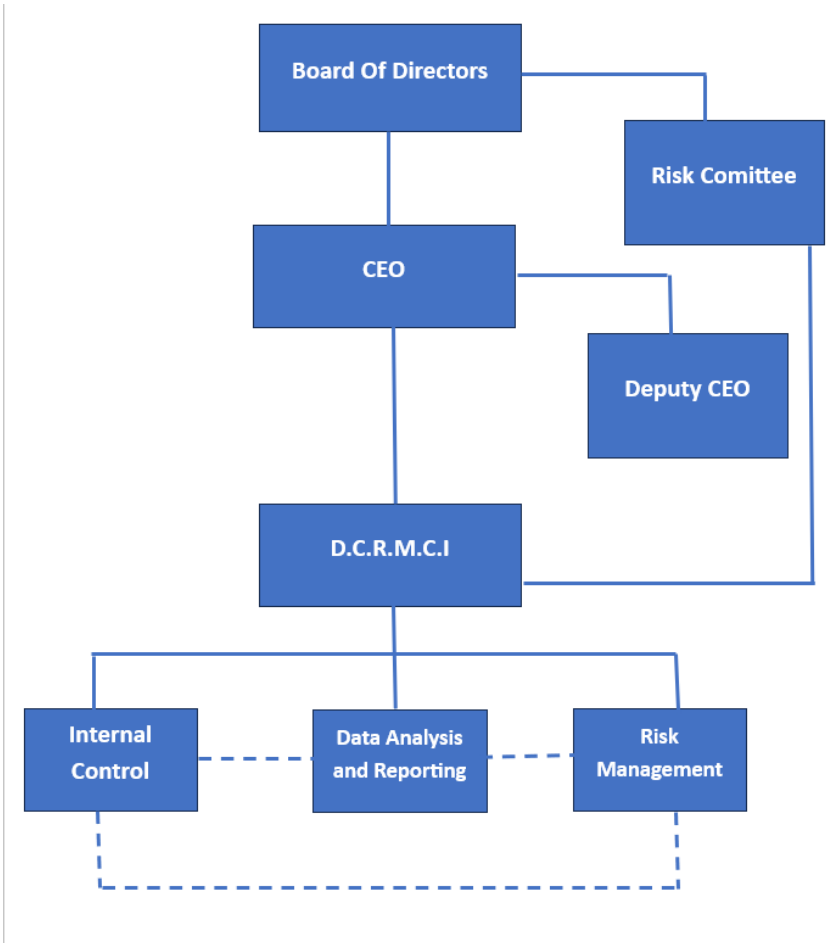

Figure 1.2: Hiearchy Of The Central Directorate Of Risk Management And Internal Control

1.2.2 Objectives Of The Internship

The objectives of the internship have been divided in two levels: strategical level and operational level.

Strategic Objectives: At the strategic level, the main goal is to gain an understanding of how the bank manages risk on a global scale. This includes studying the bank's internal risk management policies, the procedures followed, and the mechanisms in place for continuous monitoring and control. Additionally, the different types of risks faced by the bank and the key indicators used to track them were examined.

Operational Objectives: This internship aims to enhance both technical and professional skills. On the technical side, real data was used to analyze credit risk through tools such as Python and Excel for data processing and visualization. Machine learning models were applied to predict changes in credit risk. Moreover, the application of Bayesian Networks was explored to better understand relationships between various financial indicators. On the professional side, the internship provided an opportunity to strengthen critical thinking, problem solving and clear communication of results.

Each objective is designed to ensure progressive learning and create real added value for Al-Baraka Bank, both in terms of operational efficiency and compliance with regulatory requirements.

1.2.3 Challenges Of The Internship

During the internship, several challenges emerged that required adaptability and rapid learning. One of the initial difficulties involved understanding the terminology and concepts specific to credit risk management. With a background rooted in technical and academic training, it was initially difficult to align business related terminology from the banking sector with the data analysis methods previously studied. Gaining insight into how banks evaluate and manage credit risk in practice involved frequent discussions with team members and independent research. Another key challenge was related to data confidentiality. Due to the sensitive and regulated nature of the financial environment, access to client data was limited. Strict internal policies had to be followed to ensure data protection, and only anonymized or partially accessible data was available. These constraints occasionally impacted the ability to train and validate models effectively. Additionally, communicating model results in a way that could support business decision making proved to be complex. Machine learning outputs needed to be presented in a clear and accessible format for risk managers, without omitting critical technical details. Achieving this balance between model performance and interpretability was especially important in the context of a highly regulated sector like banking.
Despite these difficulties, each challenge helped me grow professionally. I gained a deeper understanding of the banking world and how data science can be applied to real business problems.

1.3 Project Context

In today's uncertain and fast-changing economic world, banks need to do more than just assess the current financial risk of their clients they also need to predict how this risk might change in the future. Being able to detect early signs that a client's financial situation might get worse is very important. It helps banks take action early and avoid situations where clients can no longer pay back their loans, which could seriously hurt the bank's financial health.
This project focuses on solving a major problem: traditional credit risk models are not very effective at looking ahead. Most current methods only focus on where the client stands financially right now. These include internal rating systems that often don't give a full picture and Stress Testing models that have failed to predict new and growing risks. Because of this, banks struggle to prepare for future risks that could affect their financial strength, performance, and ability to stay within regulatory requirements.
Our project offers a more advanced solution to this problem. We use machine learning to predict credit risk and measure how it could affect key financial indicators like solvency and capital. By analyzing past data and spotting risk patterns, our models can warn banks in advance about clients who are likely to become riskier. This forward-looking approach helps banks make better decisions, manage risk more effectively, and stay strong in a challenging economic environment.

1.4 Conclusion

The first chapter highlighted the strategic importance of Al Baraka Bank's mission and operations, revealing both its strengths and the challenges it faces in a competitive and evolving financial landscape. My integration into the Risk Management department clarified the growing demand for analytical tools that support faster, data informed decisions. This context emphasized the relevance of developing predictive models for credit risk, aligning both academic goals and business needs, and laying the foundation for the technical approach presented in the next chapters.

Chapter 2

Background

This chapter looks at how credit risk is defined and why it matters in finance. It also reviews the different models used to predict credit risk, from older statistical methods to newer machine learning techniques. We also introduce models like Bayesian Networks and Hidden Markov Models, which help understand how risk changes over time. The goal is to show what has already been done in this field and what is still missing. This helps explain why our study is needed and how it adds something new.

2.1 Definition and Importance of Credit Risk

Credit risk is a fundamental concept in finance, particularly within the banking sector, as it reflects the likelihood that a borrower or counterparty will fail to fulfill their financial obligations as agreed. This risk extends beyond simple loan defaults and can affect various financial instruments, such as bonds, interbank lending, derivatives, and trade finance transactions. Effective management of credit risk is essential to maintaining financial stability and ensuring the long-term viability of financial institutions. According to the Basel Committee on Banking Supervision (1999), credit risk is defined as the potential for a loss resulting from a borrower's failure to repay a loan or meet other financial commitments (Basel Committee on Banking Supervision, 1999). This definition forms the basis of global regulatory standards that emphasize the need for prudent risk management practices. The Basel II (Basel Committee on Banking Supervision, 2006) and Basel III (Basel Committee on Banking Supervision, 2011) accords, in particular, introduced a more risk-sensitive approach to capital requirements, requiring banks to align their capital buffers with the credit quality of their assets (Investopedia, nd). These frameworks also encouraged the adoption of more advanced credit assessment tools and internal risk rating systems, especially in response to the lessons learned from the 2007–2008 financial crisis. The financial consequences of poorly managed credit risk can be severe. Banks that fail to adequately assess borrower risk may experience rising levels of non-performing loans (NPLs), which can erode their capital base and reduce profitability. In more severe cases, unchecked credit risk can lead to insolvency and contribute to systemic financial instability. This has been historically evident, as major banking failures are often linked to inadequate credit evaluation practices, weak monitoring systems, and insufficient responsiveness to changing economic conditions. Beyond compliance with regulatory standards, the strategic management of credit risk is central to effective decision-making in banking. It supports the evaluation of borrowers' creditworthiness, informs loan pricing, guides capital allocation, and helps institutions maintain a balanced and diversified credit portfolio. By accurately measuring and managing credit risk, banks can optimize their risk-return profile and enhance the resilience of their operations.

In recent years, the adoption of advanced analytical methods such as machine learning and probabilistic modeling has further strengthened the ability of financial institutions to model and predict credit risk. These technologies enable a more nuanced analysis of borrower behavior, facilitate early detection of default signals, and improve overall portfolio risk management.

2.2 Machine Learning in Finance

In the past few years, machine learning (ML) has emerged as a powerful tool in the financial industry, offering significant advantages in terms of predictive accuracy, automation, and data-driven decision-making. Financial institutions increasingly rely on ML algorithms to analyze vast and complex datasets, uncover patterns, and make forecasts that were previously unattainable with traditional statistical methods. This technological shift reflects the broader trend of digital transformation within the finance sector (Chen and Zhao, 2021). Machine learning refers to a subset of artificial intelligence that enables systems to learn from data without being explicitly programmed. In finance, its applications are diverse and expanding rapidly, covering areas such as fraud detection, algorithmic trading, portfolio optimization, customer segmentation, and, notably, credit risk assessment (Brynjolfsson and McAfee, 2017). By learning from historical data, ML models can detect subtle, non-linear relationships between variables and adapt to changing patterns in real time capabilities that are particularly valuable in dynamic and uncertain financial environments. One of the most transformative uses of ML in banking is in the domain of credit scoring and risk classification. Traditional credit risk models often rely on linear assumptions and expert-defined scoring systems, which may fail to capture the complex interactions among borrower characteristics, macroeconomic conditions, and behavioral trends. In contrast, machine learning algorithms such as decision trees, random forests, support vector machines, and neural networks can process high-dimensional data and identify intricate patterns that enhance predictive performance. Furthermore, ML enables continuous learning, allowing models to improve as new data becomes available. This adaptability is critical in financial applications, where market conditions, borrower behavior, and regulatory requirements evolve over time. The integration of ML techniques into financial risk management has led to more granular risk assessments, better early warning systems, and more efficient allocation of capital. Despite its benefits, the use of machine learning in finance also presents challenges. Model interpretability, data quality, and regulatory compliance are key concerns. In particular, the "black-box" nature of some advanced models, such as deep neural networks, raises questions about transparency and accountability especially in high-stakes contexts like credit decisions. As a result, there is growing interest in explainable AI (XAI) methods that aim to make ML outputs more understandable to human users and regulators (Brynjolfsson and McAfee, 2017). Overall, the incorporation of machine learning into financial practices represents a paradigm shift, enhancing the analytical capabilities of institutions while also demanding new standards in model governance, ethical use, and regulatory oversight.

2.3 Credit Risk Modeling: Traditional Approaches vs. Machine Learning

Credit risk modeling is a central task in financial risk management. It involves estimating the likelihood that a borrower will fail to meet their financial obligations. Traditionally, financial institutions have relied on statistical models to assess credit risk. However, the rise of machine learning (ML) has introduced more powerful and flexible tools, leading to a growing interest in comparing the two approaches in terms of performance, interpretability, and practical application (Rogojan and Badea, 2023).
The table 2.1 compares traditional statistical models with machine learning methods in the context of credit risk prediction.

Table 2.1: Comparison of Traditional and Machine Learning Models for Credit Risk

Traditional Statistical Models	**Machine Learning Models**
Rely on assumptions such as linearity, normality, and independence	Do not require strong assumptions about data distribution
Examples: Logistic Regression, Discriminant Analysis,Stress Test	**Examples:** Decision Trees, Random Forest, XGBoost, Neural Networks, k-NN
Provide good interpretability and are easy to implement. Focus mainly on the client's current financial state. Limited ability to capture emerging risks or temporal patterns.	Can model complex, non-linear relationships but are often less interpretable. Enable forward-looking analysis using large volumes of historical data. Capable of identifying hidden patterns and early risk migration signals.

Comparison and Challenges : Several studies have confirmed that ML models particularly ensemble methods and deep learning consistently outperform traditional statistical models in credit risk prediction (Bitetto et al., 2023). However, this increased accuracy often comes with a cost: reduced interpretability. Many ML models, especially complex ones, act as "black boxes making it difficult to explain why a certain prediction was made (Wang et al., 2024). This lack of transparency can be a barrier to adoption, especially in highly regulated sectors like banking. Furthermore, ML models require more computational resources and careful tuning of hyperparameters. They are also more prone to overfitting if not properly validated (Noriega et al., 2023). As a result, current research is focusing on the development of explainable artificial intelligence (XAI) techniques, which aim to make ML models more transparent and trustworthy (Bussmann et al., 2023). In conclusion, traditional statistical models offer simplicity, interpretability, and regulatory familiarity, while machine learning provides superior predictive performance and flexibility. The choice between the two depends on the context and the specific goals of the risk assessment process. Increasingly, hybrid approaches and explainable ML techniques are being explored to combine the strengths of both worlds.

2.4 Bayesian Networks in Risk Analysis

Bayesian Networks (BNs), also known as belief networks or probabilistic directed acyclic graphical models, represent a powerful framework for reasoning under uncertainty, making them particularly well suited for applications in financial risk analysis, including credit risk modeling (Pavlenko and Chernyak, 2010; Sun and Shenoy, 2015). BNs provide a graphical representation of probabilistic relationships among a set of variables, offering a unique blend of statistical rigor and intuitive visualization. **Core Concepts:**

A Bayesian Network consists of two main components:

- **Structure (Qualitative Part):** A Directed Acyclic Graph (DAG) where nodes represent random variables (e.g., financial ratios, macroeconomic indicators, default status) and directed edges represent conditional dependencies. The absence of an edge implies conditional independence under certain conditions.
- **Parameters (Quantitative Part):** A set of Conditional Probability Tables (CPTs) associated with each node, quantifying the probability distribution of a node given its parent nodes. For root nodes, the CPT specifies prior probabilities.

The strength of BNs lies in their ability to combine prior knowledge (including expert opinions or domain knowledge) with observed data through Bayesian inference. They allow for probabilistic reasoning in multiple directions: predicting outcomes (e.g., probability of default), diagnosing causes given effects, and explaining away alternative causes.

Applications in Credit Risk: BNs have been applied to various areas of credit risk:

- **Credit Scoring:** Estimating default probability for borrowers based on financial and demographic variables (Malagrino et al., 2017; Khashman, 2010).
- **Concentration Risk:** Assessing exposure concentration across counterparties or sectors (Pavlenko and Chernyak, 2010).
- **Systemic Risk:** Modeling risk transmission across institutions and markets (Ballester et al., 2023).

Table 2.2: Comparison between Static and Dynamic Bayesian Networks

Feature	**Static Bayesian Networks (BNs)**	**Dynamic Bayesian Networks (DBNs)**
Temporal Modeling	Do not account for temporal evolution	Explicitly model temporal dependencies across time steps
Structure	Single-layer graph	Multi-slice graph (e.g., 2-slice temporal Bayesian network - 2TBN)
Variables	Represent a snapshot at one time point	Include time-indexed variables (e.g., X_t, X_{t+1})
Applications	Good for point-in-time risk assessment	Ideal for tracking borrower behavior over time and credit migration
Complexity	Less computationally intensive	More complex due to temporal links and parameter replication

In conclusion, Bayesian Networks provide a flexible and powerful probabilistic framework for modeling credit risk, offering advantages in handling uncertainty, integrating prior knowledge, and modeling complex dependencies, including temporal dynamics through DBNs.

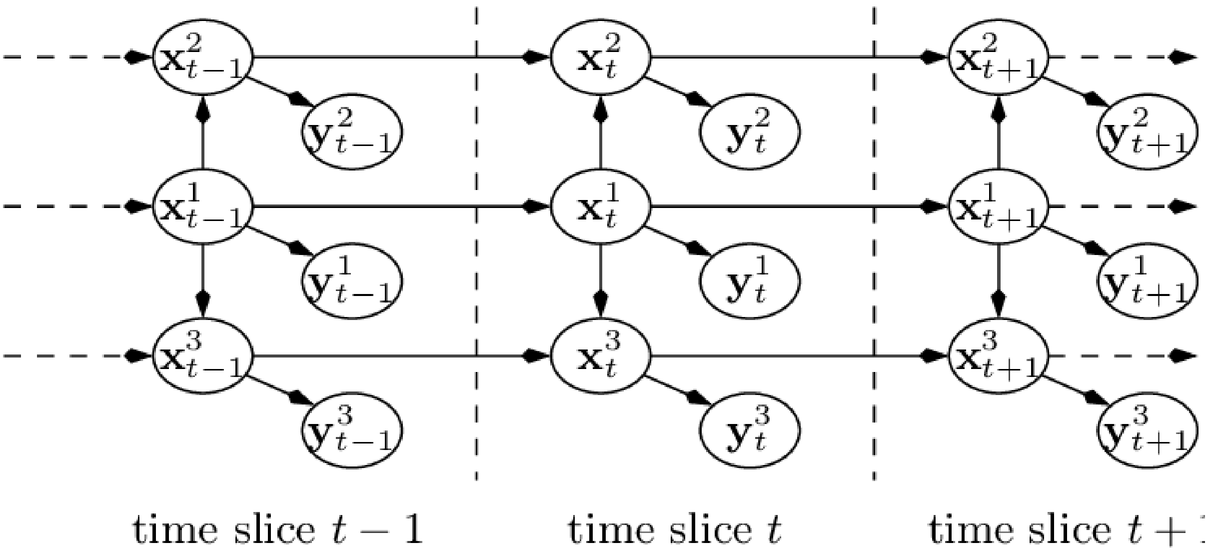

Figure 2.1: Dynamic Bayesian Network

2.5 Ensemble Learning Methods

Ensemble learning refers to a class of techniques in machine learning that combine the predictions of multiple base models to achieve improved accuracy, robustness, and generalization. Instead of relying on a single predictive model, these methods aggregate the outputs of several learners, harnessing the collective performance to outperform individual models (Opitz and Maclin, 1999; Rokach, 2010). This approach is particularly effective in complex tasks like credit risk assessment, where ensemble models often yield superior performance due to their ability to manage non-linearity, high dimensionality, and noisy data (Shi et al., 2022; Martin et al., 2024).

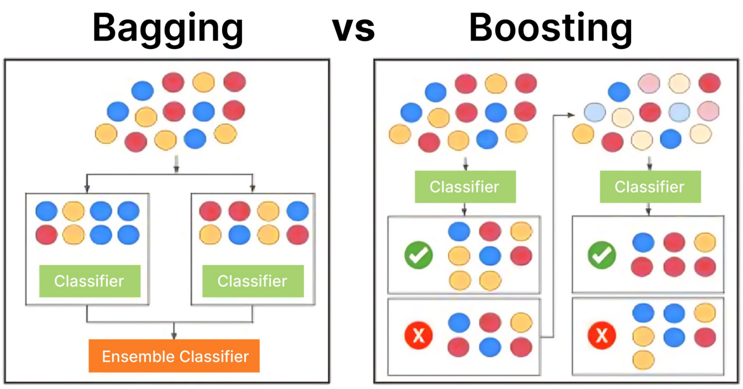

Figure 2.2: Boosting vs Bagging

The most widely used ensemble strategies are Bagging and Boosting:

- **Bagging (Bootstrap Aggregating):** This technique generates several versions of a training set using bootstrap sampling (i.e., sampling with replacement) and trains a separate model on each. The final output is obtained by aggregating predictions using majority voting for classification or averaging for regression. Bagging is particularly effective at reducing variance. A well-known application is the Random Forest algorithm, which enhances bagging by introducing randomness in feature selection at each split, thus decreasing correlation among trees and improving performance (Breiman, 2001).

- **Boosting:** Unlike bagging, boosting is a sequential technique in which each model is trained to correct the errors of its predecessor. The algorithm focuses more on difficult cases by adjusting the weights of misclassified instances. This process typically leads to a reduction in bias and an increase in predictive power. Popular boosting techniques include AdaBoost, which adjusts weights adaptively, and Gradient Boosting, which optimizes a loss function using gradient descent. Advanced variants such as XGBoost, LightGBM, and CatBoost improve efficiency, scalability, and performance through features like regularization, native handling of missing values, and fast execution (Broby, 2024; Sexton, 2022).

Relevance in Credit Risk Modeling: Ensemble methods have demonstrated high effectiveness in credit scoring and risk prediction. Algorithms such as random forests and gradient boosting machines are frequently cited in recent literature for their ability to deliver state of the art results(Shi et al., 2022; Broby, 2024; Martin et al., 2024). Their main advantages include high predictive accuracy, resistance to noise, ability to model complex interactions, and the provision of feature importance metrics. Moreover, they can be adapted to address class imbalance, which is a common issue in default prediction tasks (He et al., 2018).

2.6 Sequential Models in Finance

Hidden Markov Models (HMMs) are used to work with data that changes over time. They are based on the idea that what we observe (like market behavior) is influenced by "hidden" states that we cannot see directly. At each time step, the model assumes that the system is in one of these hidden states, which produces the observed outcome. The change from one state to another follows simple rules, where the next state depends only on the current one not on everything that happened before (Rabiner, 1989).

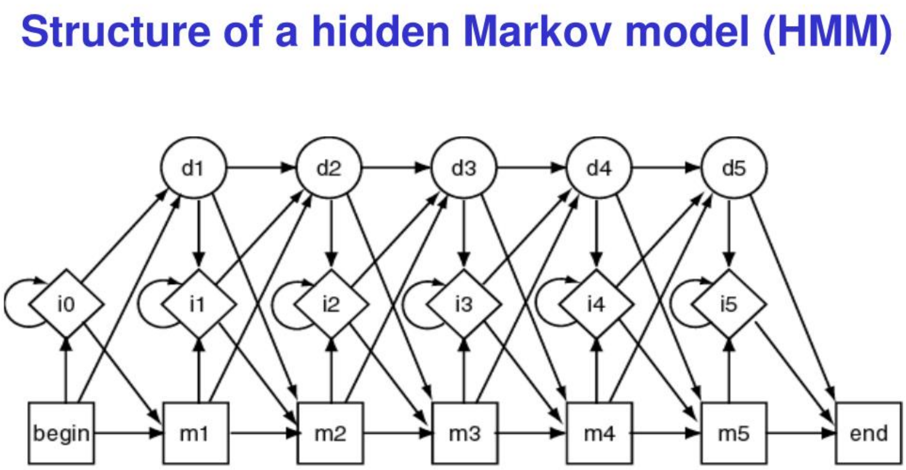

Figure 2.3: Hidden Markov Model

In finance, HMMs are useful for modeling shifts in market behavior, such as moving between bull and bear markets, or changes in credit risk levels (Oelschläger and Adam, 2024). By training an HMM on past financial data, it is possible to estimate which hidden states likely caused the observed patterns, and how likely the system is to switch from one state to another. This helps analysts better understand market regimes and adjust strategies accordingly. Although HMMs are relatively easy to interpret and use, they have some limitations. For example, they assume that the number of hidden states is fixed in advance, and they only look at the most recent state to predict the next one. This can be a problem when the data shows long-term trends or more complex behavior (Khalifa et al., 2021). Also, in very noisy or highly non-linear datasets, HMMs might not perform as well as more advanced deep learning models. Still, HMMs remain a solid choice for detecting regime shifts, and researchers have started combining them with deep learning methods to get better results (Avinash et al., 2024)..

2.7 Identified Gaps and Project Motivation

This Background highlights an area of research where different modeling techniques have been used to predict credit risk, including machine learning techniques and probabilistic models.

- **Comparative Evaluation on Temporal Financial Data:** Although many studies assess machine learning models for credit risk, there is limited work comparing probabilistic graphical models such as Static and Dynamic Bayesian Networks with sequence based models like Hidden Markov Models and ensemble methods like Random Forest and XG-Boost. Existing research often focuses on default prediction using binary outcomes or relies on static datasets that may not reflect the complexity of real financial data.
- **Handling Clients with Different History Lengths:** In real datasets, some clients have records over several years, while others appear only once. Using Static Bayesian Networks for one-time clients and Dynamic Bayesian Networks for those with longer histories is an interesting idea, but not much research has tested how well this works in practice.
- **Matching Models to Real Use Cases:** A model's performance can depend on how the data is structured, how the target variable is defined, and what the bank needs from the model. Many studies use standard datasets, but fewer take into account how the model will actually be used in a specific banking context.
- **From Prediction to Real-World Use:** It's not enough for a model to be accurate—it also needs to be easy to use and apply in a real system. Research often separates model development from practical deployment, which makes it harder to judge the real impact of the model.

This project aims to fill these gaps through a structured comparison of four models Bayesian Networks, Hidden Markov Models, Random Forest, and XGBoost using a dataset covering years of client information from a financial institution. The goal is to predict the client's credit risk category 1 year in advance using standard metrics such as classification accuracy and more nuanced measures such as the Quadratic Weighted Kappa (QWK). By conducting this comparison within a consistent framework and developing a working prediction tool, the study aims to provide practical information on which models are best suited for data-driven credit risk assessment.

2.8 Conclusion

This chapter gave a clear overview of credit risk and the main modeling techniques used to assess it. We explained what credit risk is, why it matters, and how machine learning is being increasingly used in the financial sector. We also compared traditional methods with newer, data-driven approaches. Special focus was placed on models like Bayesian networks, ensemble methods , and sequential models. Each was discussed in terms of how it can be used to evaluate credit risk. This review helped identify gaps in the existing research, which supports the purpose and direction of our project.

Chapter 3

Methodology

This chapter outlines the methodological framework adopted for the credit risk prediction project. To ensure a structured and systematic approach, the work is based on the Cross-Industry Standard Process for Data Mining (CRISP-DM), a widely recognized methodology in data science. CRISP-DM divides the data mining process into six key phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Each step guides the project by defining objectives, understanding the available data, preparing it appropriately, selecting and designing suitable models, and finally planning the evaluation and deployment strategy. The aim of this chapter is to present the rationale behind each phase, the decisions made, and the general strategy prior to implementation. Specific technical details and coding aspects are addressed in the next chapter.

1. **Business Understanding:** is the initial phase that focuses on understanding the project objectives and requirements from a business perspective and then converting this knowledge into a definition of data mining problems and a preliminary plan designed to achieve these objectives.

2. **Data Understanding:**This phase involves collecting and exploring the data to become familiar with it, identify data quality problems, and first discover insights into the data.

3. **Data Preparation:**This phase covers all activities to construct the final dataset from the initial raw data.Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include data selection,cleaning, construction of new attributes,integration of data, and formatting.

4. **Modeling:**In this phase, various modeling techniques are selected and applied. Before applying a modeling technique, there are often specific requirements on the form of data. Therefore, it is often necessary to step back to the data preparation phase.

5. **Evaluation:** At this stage, the model is thoroughly evaluated, and the steps executed to construct the model are reviewed to ensure that it achieves the business objectives.

6. **Deployement:**In this phase , the knowledge gained from training a model will need to be organized and presented in a way that the customer can use it.

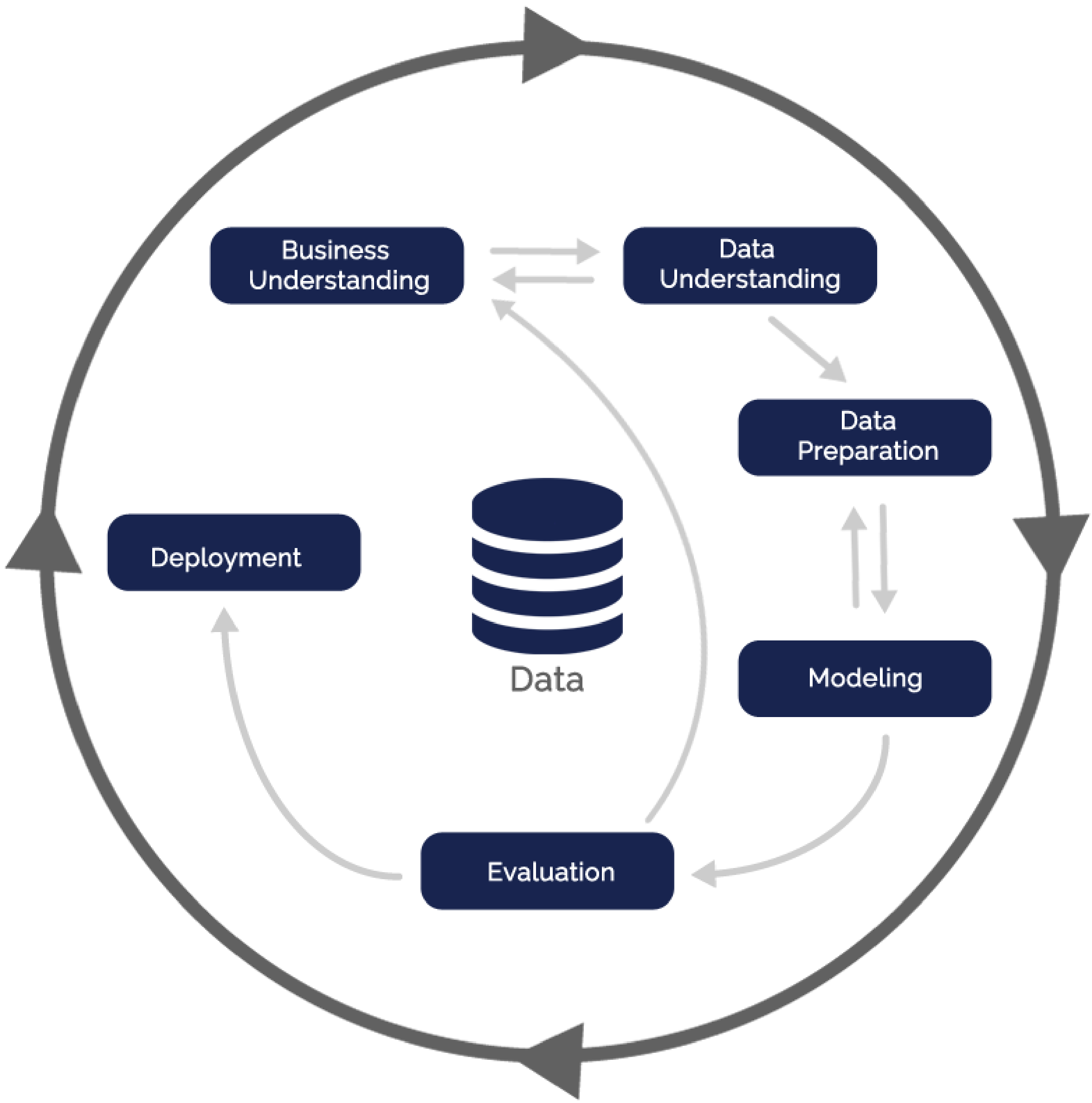

Figure 3.1: CRISP-DM Flow Chart

The phases of the CRISP-DM methodology and their application within this project are presented in the following sections.

3.1 Business Understanding

In recent years, Tunisia's economy has experienced significant fluctuations, directly impacting how banks assess the creditworthiness of their clients. In this uncertain environment, financial institutions must proactively identify potential risks and take early actions to decrease them. The primary objective of this project is to assist in this effort by developing a system that can predict a client's future credit risk level based on their historical financial and behavioral data. By anticipating changes in client risk profiles, banks can improve their credit management strategies, focus on risky accounts, and make informed decisions.This approach draws inspiration from stress testing, a commonly used technique in the banking sector to evaluate institutional performance under adverse economic conditions. However, rather than simulating macroeconomic scenarios, this study adopts a data-driven methodology, using machine learning models and probabilistic frameworks to analyze client behavior over time. By incorporating temporal information, the project aims to deliver personalized and dynamic risk predictions that reflect not only a client's current situation but also their likely future evolution. This allows for a more detailed understanding of credit risk, aligning with modern risk management practices that emphasize agility.

3.2 Data Understanding and Preparation

This section presents a comprehensive overview of the dataset used in this study and the steps taken to prepare it for analysis. Understanding the origin, structure, and characteristics of the data is essential before applying any machine learning techniques. Likewise, proper preparation ensures the dataset is clean, consistent, and suitable for modeling. The combination of these two phases lays the groundwork for building reliable predictive models.

3.2.1 Source and Time Coverage

The dataset used for this project was provided by Al-Baraka Bank, a financial institution specializing in credit risk management. It contains detailed financial and behavioral information about clients, collected over a four year period from 2020 to 2024. Each client is uniquely identified by a client code, which enables tracking their profile and behavior over time.
Thanks to this temporal organization, the dataset supports longitudinal analysis observing how a client's financial situation and credit risk evolve across quarters. This feature is particularly valuable in credit risk modeling, where historical behavior is often a key indicator of future default risk.

3.2.2 Overview of the Main Features

The dataset includes a range of features relevant to credit risk assessment. These features can be grouped into two broad categories:

- **Categorical features:** Describing qualitative attributes such as business sector and activity type.
- **Numerical features:** Representing financial indicators like unpaid amounts, provisions, and guarantees.

Here is a summary of the main features included in the dataset:

- **Client's ID:** A unique identifier for each client.
- **Time Slice:** Indicates the specific quarter, enabling tracking over time.
- **Credit Risk Class T:** The risk class assigned to the client at a given time.
- **Credit Risk Class T+1:** The future risk class used as the target variable for prediction.
- **Business Sector:** Categorized into Professionals, Retail, and Banking/Financial Institutions.
- **Activity:** Specifies the client's line of business, covering 13 distinct categories such as Commerce, Transport, Construction, and Real Estate.
- **Unpaid Amount:** Total overdue or unpaid financial obligations.
- **Total Engagement Amount:** Sum of balance sheet and off-balance sheet commitments.
- **Guarantee:** Value of guarantees provided by the client.
- **Provision:** Funds set aside to cover expected credit losses.
- **Reserved Profits:** Retained earnings not yet distributed.

Together, these variables offer a comprehensive picture of each client's financial condition and behavior, which is critical for building predictive models of credit risk.

3.2.3 Temporal Data Structure

One of the dataset's key features is its temporal structure. Client information is recorded over multiple quarters, allowing us to distinguish between two types of clients:

- **Single-period clients:** Clients who appear in only one year.
- **Multi-period clients:** Clients with data spanning several years , enabling the study of behavioral trends and changes in risk class over time.

This temporal aspect is crucial for generating meaningful features that reflect how a client's financial status evolves an important factor in credit risk prediction.

3.2.4 Preprocessing Steps

To prepare the dataset for analysis, several key preprocessing operations were carried out:

- **Standardization of Column Names:** All column names were cleaned and unified to remove spaces or special characters that could interfere with analysis or model training.
- **Handling Missing Values:** Specific imputation strategies were applied to deal with incomplete data:
  - Categorical features were imputed using the most frequent category.
  - Numerical features were imputed using the median to minimize the effect of outliers.
- **Encoding Categorical Variables:** Categorical data like Business Sector and Activity were transformed into numeric format through label encoding, ensuring compatibility with machine learning algorithms.
- **Discretization of Numerical Variables:** To apply probabilistic models like Bayesian Networks, continuous features were converted into discrete bins. For example, variables like 'Unpaid Amount' and 'Provision' were categorized into qualitative levels (e.g., Low, Medium, High), improving interpretability and model compatibility.
- **Correlation Matrix Analysis:** A correlation matrix was computed to understand the relationships and interdependencies between the various features. This analysis was crucial for identifying highly correlated variables, which could indicate multicollinearity and inform subsequent feature selection or engineering decisions. The insights gained from the correlation matrix helped in ensuring that the features used for model training were well-suited and not redundant
- **Generation of Temporal Features:** Leveraging the dataset's longitudinal structure, additional features were engineered to capture client behavior over time:
  - **Data History:** Past financial values were included to reflect prior behavior.
  - **Trends:** Percentage change between time periods was computed to capture movement.
  - **Risk Class Stability:** Duration of consecutive periods spent in the same class.
  - **Risk Class Change:** Direction and frequency of class transitions.
  - **Volatility:** Standard deviation across time to measure behavioral variability.
  - **Historical Min/Max:** Extreme values reached by each variable to capture risk peaks.

By integrating both static and temporal information,the preprocessing phase refined the dataset into a more analysis ready format by enhancing consistency, performing discretization, and engineering informative features. This not only improved model readiness but also enhanced the ability to capture subtle changes in client behavior, which are crucial for effective credit risk prediction.

3.3 Modeling

This section presents the machine learning models trained to classify and predict clients' future credit risk levels. Each model was selected based on its ability to handle specific data characteristics, particularly the temporal dimension and the need for interpretability in a financial context.

3.3.1 Static Bayesian Network:

A Static Bayesian Network is a graphical model used to represent the probabilistic relationships among a set of variables at a single point in time. In this model, each variable is represented as a node, and the edges indicate conditional dependencies. These dependencies are quantified using Conditional Probability Tables (CPTs). In the context of this project, the SBN is suited for clients who only appear once in the dataset. Since there is no historical information for these cases, their credit risk must be assessed using only current financial and behavioral data. The SBN infers the probability of a client falling into each risk class by modeling how different variables interact at that moment. While this model provides useful insights, it is limited in that it does not account for changes over time. It treats each client as a static observation, which can reduce its predictive power when compared to models that incorporate historical trends. Therefore to overcome this limitation and to capture the temporal structure present in the data, we extended the model to a Dynamic Bayesian Network. This approach allows us to represent how clients' financial attributes evolve over time, offering a more accurate and realistic framework for credit risk prediction.

3.3.2 Dynamic Bayesian Network:

The Dynamic Bayesian Network extends the SBN by incorporating the time dimension. It models how a clients variables and risk status change across sequential time points. This is particularly effective for clients with data spanning multiple quarters. The DBN captures the transitions between risk states by modeling dependencies not only within a time step but also across time steps. For example, it can learn how a change in unpaid debt or provisions today may affect the client's risk level in future periods. This makes it a valuable tool for understanding and forecasting credit risk based on temporal patterns. By leveraging the sequence of historical observations, the DBN offers more precise predictions for clients with evolving financial behaviors.

3.3.3 Hidden Markov Model:

The Hidden Markov Model is another model used to capture time-based dynamics. Unlike the DBN, which models relationships among all variables, the HMM assumes that the underlying risk status of a client is a hidden state that cannot be directly observed. Instead, the model infers this hidden state from the observed variables, such as unpaid amounts or provisions. The application of HMM to credit risk evolution allows for:

- Understanding the probability that a client moves from one risk class to another based on their observations.
- inferring the most probable sequence of hidden states, the model can predict a clientś future risk class.

- Analyzing emission probabilities can reveal which observations are most indicative of a particular risk state

3.3.4 XGBOOST Model:

XGBoost was applied to predict the future credit risk class of clients. By learning from historical financial and behavioral data, the model aimed to classify clients into risk categories based on how their situation is expected to evolve. The strength of XGBoost lies in its ability to handle non-linear relationships, interactions among features, and temporal trends through engineered features.
In this study, XGBoost was not just used with static data. Instead, it was adapted to take into consideration the temporal structure of the dataset by incorporating features that describe the clients past behavior, such as historical trends, volatility, and changes in financial indicators. These features enriched the representation of temporal patterns and contributed to improving the model's ability to capture complex dynamics in client behavior.

Hyperparameter Tuning and Class Imbalance Handling: XGBoost offers a wide range of hyperparameters that can be tuned to optimize model performance. This tuning process involves adjusting parameters such as `learning_rate`, `n_estimators`, `max_depth`, `subsample`, and `colsample_bytree` to find the optimal combination that minimizes prediction error.Thus, Cross-validation technique is employed for this purpose. XGBoost was trained on a dataset balanced to address the significant class imbalance typical in credit risk data. By generating synthetic samples for minority classes, SMOTE improved the model's ability to detect rare events such as defaults, without relying on internal class weighting parameters.

3.3.5 Random-Forest:

In this project, the Random Forest algorithm was used to model the evolution of credit risk over time by predicting the client's future risk class . The dataset was prepared as a sequence of client records over multiple time periods, allowing the model to capture temporal patterns in client behavior. The Random Forest model was trained on features extracted from each time slice, excluding identifiers like client code and time step. When clients had multiple records across different quarters, the model learned to predict the future risk level using both current financial indicators and past patterns. For each prediction instance, the input features came from time T, and the target variable was the risk class at T+1. To improve the model performance, hyperparameters such as the number of trees, maximum depth, and minimum samples per leaf were tuned using the Optuna framework systematically explores combinations of hyperparameters such as the number of trees, maximum depth, and minimum sample size per leaf node.

3.4 Evaluation Metrics

To evaluate the performance of the implemented machine learning models, several metrics were used, particularly suited for imbalanced and multi-class classification tasks in credit risk assessment. These metrics are derived from the confusion matrix, which serves as the basis for most classification performance evaluations.

Confusion Matrix: The confusion matrix is a tabular representation that compares predicted class labels with actual labels. It provides a breakdown of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), allowing for a detailed error analysis:

$$\begin{bmatrix} \mathrm{TP} & \mathrm{FP} \\ \mathrm{FN} & \mathrm{TN} \end{bmatrix}$$

In multi-class classification problems, the confusion matrix is extended to an $n \times n$ matrix, where n is the number of classes.In our case 6×6 since we have 6 classes from 0:low risk to 5: high risk. Diagonal elements represent correct predictions, while off-diagonal elements indicate misclassifications. To assess the performance of the implemented machine learning models, a comprehensive set of evaluation metrics was utilized.These metrics help evaluate the different models , especially when the data is unbalanced and there are several risk categories, which is common in credit risk analysis.

- **Quadratic Weighted Kappa (QWK):**The Quadratic Weighted Kappa is a useful metric that measures how well the model's predictions match the actual labels. It is often used in multi classification problems where the classes follow an order. Unlike regular accuracy, QWK gives more importance to the size of the error — the further the prediction is from the true class, the bigger the penalty. A QWK score of 1 means perfect agreement, 0 means the model is guessing randomly, and a negative score means it's doing worse than random guessing.

- **Accuracy:** Accuracy is one of the most straightforward evaluation metrics, representing the proportion of correctly classified instances (both true positives and true negatives) out of the total number of instances. While intuitive, accuracy can be misleading in imbalanced datasets, as a model might achieve high accuracy by simply predicting the majority class.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

- **Precision:**Precision, also known as Positive Predictive Value, measures the proportion of true positive predictions among all positive predictions made by the model. It focuses on the quality of positive predictions, answering the question: "Of all instances predicted as positive, how many were actually positive?" High precision indicates a low rate of false positives

- **Recall:** Recall, also called Sensitivity or True Positive Rate, shows how many actual positive cases the model was able to find. It answers the question: "Out of all the real positive cases, how many did the model correctly detect?" A high recall means the model missed very few positive cases, which means it has a low number of false negatives.

- **F1-Score:**The F1-Score combines precision and recall into one number that shows how well a model is doing overall. It's especially helpful when the groups you're trying to predict aren't equal in size. The F1-Score makes sure the model does well at both finding the right cases (recall) and being accurate when it makes a positive prediction (precision). A higher F1-Score means the model is doing a good job balancing both.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.5 Deployment

Once the models were developed and evaluated, the final phase focused on defining how these models could be used in a practical, operational context. The objective of this stage was to outline a deployment strategy that would allow non-technical users, such as risk analysts, to interact with the predictive system and make informed decisions based on the model outputs. The deployment strategy was structured around three key components: user interaction, model integration, and the functional role of the system within the bank's existing credit risk management workflow.

3.5.1 User Interaction

To ensure accessibility and ease of use, the system was designed to include a web-based interface. This interface aims to guide users through the process of entering relevant financial and behavioral client information. The emphasis was placed on usability, with a simple and intuitive form that requires no technical knowledge. Upon submitting the input, the system provides the user with a prediction of the client's credit risk level.

3.5.2 Model Integration

The interface is conceptually linked to a backend layer responsible for interacting with the trained machine learning models. This component handles the receipt of user inputs, executes necessary preprocessing, selects the appropriate model (e.g., XGBoost or Bayesian Network), and returns the prediction. This integration ensures that the system remains responsive, reliable, and aligned with the decision-making needs of the risk analysis process.

3.5.3 Functional Overview

The deployed solution is designed to support practical risk evaluation tasks, enabling seamless input of client data, automated generation of predictions, and integration with broader credit risk management procedures. It also aims to facilitate interpretation of results and decision-making through a user-friendly interface and data visualizations.

Figure 3.2 below provides a high-level overview of the deployed application's architecture, illustrating the interaction between the user, the frontend interface, the backend logic, and the machine learning models used for prediction.

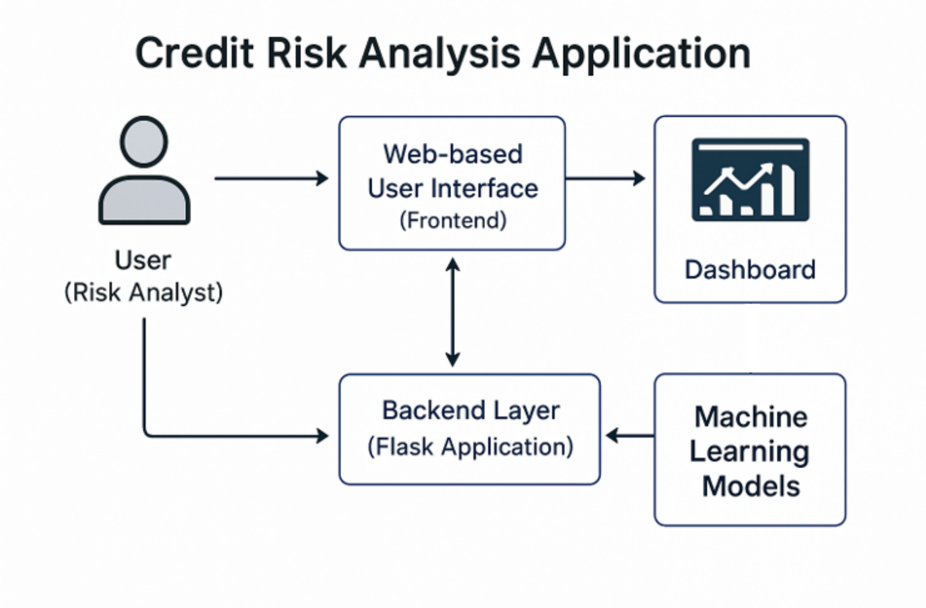

Figure 3.2: System architecture of the credit risk analysis application

The technical implementation details of this deployment strategy, including the specific architecture, and user interface components will be presented in the implementation chapter.

3.6 Conclusion

This chapter explained the steps followed to build a credit risk prediction system using the CRISP-DM approach. Each phase was carefully planned and carried out. We started by analyzing the dataset, focusing on its time based structure. Then we prepared the data by cleaning it, filling in missing values, and creating new features to capture how client behavior changed over time. Several models were tested. Probabilistic models were used to model risk patterns, while tree-based models helped capture complex relationships. To improve results we tuned their parameters and used methods to handle class imbalance. Finally, a web based system was created so users can enter client data and get instant risk predictions.

Chapter 4

Implementation

This chapter describes the practical implementation of the credit risk classification models based on the methodology outlined in the previous chapter. It focuses on the concrete steps taken during the development process, including data loading, preprocessing, feature engineering, model training, and prediction. Each section details how the conceptual framework from Chapter 3 was translated into working code using appropriate tools and libraries. The objective is to provide a clear and transparent view of how the models were developed and applied to real world data to predict credit risk. Through this implementation, we aim to demonstrate the construction of a functional system that leverages historical data to support risk based decision making.

4.1 Development Environment and Libraries

4.1.1 Technical Stack and Development Setup

The implementation was carried out primarily using the Python programming language, chosen for its flexibility and extensive ecosystem of libraries dedicated to data science and machine learning. The development environment was structured around Visual Studio Code (VS Code), a versatile source code editor that supports a wide range of programming languages and tools. For the construction and visualization of the Static Bayesian Network, GeNIe Modeler software was utilized. GeNIe provides a user-friendly graphical interface for building probabilistic graphical models and facilitates reasoning under uncertainty. When combined with programming languages such as Python, it allows for both visual model design and automated analysis (BayesFusion, 2024).
To develop the user interface and the backend Application Programming Interface (API) for model deployment, the following web technologies were employed:

- **HTML:** Used to define the structure and content of the web application interface.
- **CSS:** Applied to style and visually enhance the frontend components, ensuring a responsive and consistent design.
- **JavaScript:** Enabled dynamic behavior and client-side interactivity, serving as the foundation for React-based development.
- **React:**A JavaScript library used for building interactive and dynamic user interfaces for the frontend application.

- **Node.js:**A JavaScript runtime environment that allows for the execution of JavaScript code on the server-side, often used in conjunction with React for building full-stack applications.
- **Flask:**A lightweight Python web framework used to build RESTful APIs for deploying the trained machine learning models, enabling communication between the frontend and the backend models.

4.1.2 Key Python Libraries for Data Science and Machine Learning

Table 4.1: Libraries Used in the Study

Library	**Purpose**
`Pandas`, `NumPy`	Used for data manipulation, cleaning, and numerical computations.
`Scikit-learn`	Provides preprocessing tools such as `StandardScaler` and `LabelEncoder`, enables data splitting, and supports model evaluation using metrics like `cohen_kappa_score`, `confusion_matrix`, and `classification_report`. Also used to implement the `Random Forest` model.
`XGBoost`	Utilized to implement the eXtreme Gradient Boosting algorithm for classification tasks.
`hmmlearn`	Applied for modeling sequential data using Hidden Markov Models (HMMs).
`pgmpy`	Employed to build and perform inference on Static and Dynamic Bayesian Networks.
`imbalanced-learn`	Used to address class imbalance across the six credit risk classes by generating synthetic samples for underrepresented classes.
`Matplotlib`, `Seaborn`	Used for creating various plots and visual representations of data and model results.
`Optuna`	Facilitates the optimization of hyperparameters through an efficient search strategy.
`Flask-CORS`	Enables Cross-Origin Resource Sharing (CORS) to allow frontend applications to communicate with the Flask backend.
`flask_jwt_extended`	Provides JSON Web Token (JWT) support for Flask APIs, enabling secure authentication and authorization mechanisms across user sessions.
`pickle`	Used for serializing and saving trained machine learning models to disk, allowing for model persistence and reuse.

4.2 Data Engineering & Preprocessing in Practice

As outlined in Chapter 3, data understanding and preparation are key components of the CRISP-DM methodology. This section presents how these preparatory steps were implemented in practice using Python.
Data preparation is a critical phase in any machine learning project, as it transforms raw data into a format suitable for training and evaluating models. This section describes the specific steps carried out in this project, including data loading, cleaning, handling of missing values, encoding, discretization, and the generation of temporal features.

4.2.1 Handling Missing Values

Missing data was treated based on the type of feature:

- **Categorical features:** Missing values were filled with the mode, which is the most frequently occurring category.

- **Numerical features:** Missing values were replaced with the median, as it is less sensitive to outliers than the mean.

4.2.2 Encoding and Discretization

- **Encoding categorical variables:** Textual attributes such as `Rubrique` and `Activité` were converted into numerical form using Scikit-learn's `LabelEncoder`, which assigns a unique integer to each category. An example is shown in (A.1).

- **Discretization of numerical variables:** For Bayesian models (both Static and Dynamic), continuous values were converted into discrete intervals such as *Very Low*, *Low*, *Medium*, *High*, and *Very High*, as illustrated in (A.2). This transformation is necessary because Bayesian networks work with discrete probability tables and cannot directly process continuous variables.

4.2.3 Correlation Matrix Analysis

As illustrated in Figure (4.1), the correlation matrix revealed several key points. For instance, it highlighted the moderate correlation between 'Provision' and 'Total Bilan&Hbilan'. This analysis was instrumental in identifying potential multicollinearity and guiding our feature selection process, ensuring that the features used were well suited and not redundant, thereby enhancing the robustness of our models.

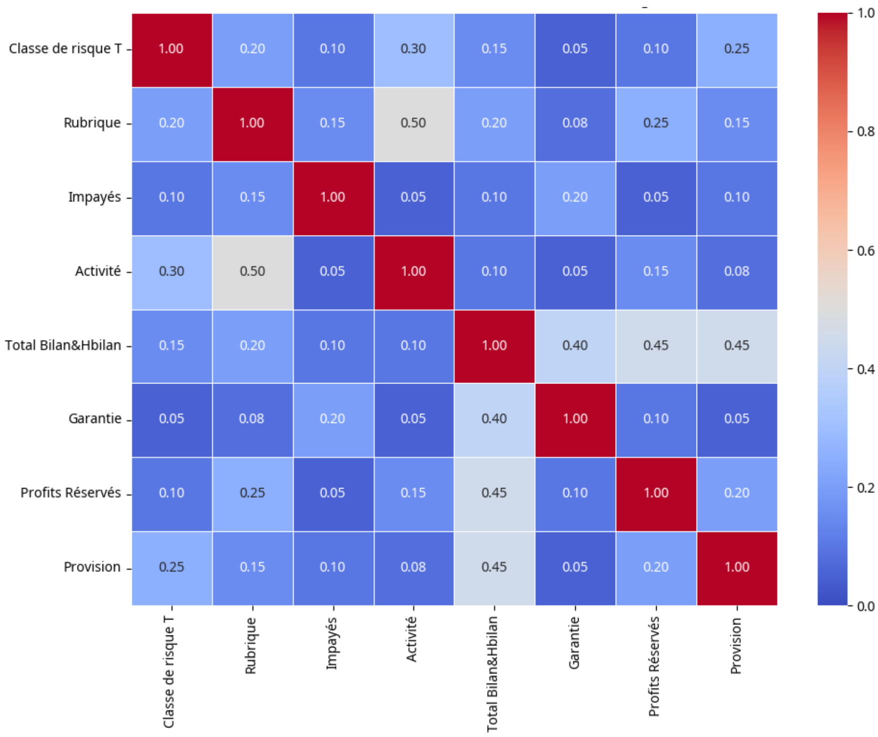

Figure 4.1: Correlation Matrix

4.2.4 Generation of Temporal Features for XGBoost

Following the methodology described in Section 3.2.4, this part details the technical implementation of temporal feature engineering for XGBoost.
For time-aware models such as XGBoost, additional features were created to capture how each client's financial behavior evolves over time. These include:

- **Data history:** Integration of past values of key features.
- **Trends:** Computation of changes in financial variables between time periods.
- **Risk class stability and change:** Indicators for how long a client remained in the same risk class and whether transitions occurred.
- **Volatility:** Standard deviation of financial values to represent behavioral variability.
- **Historical Min/Max:** Recording the lowest and highest values reached by financial indicators during the observation window.

A detailed illustration of this temporal feature engineering process is provided in (B.2).

4.3 Model Implementation

Each model follows clear steps from setup to training and evaluation.

4.3.1 XGBoost Model

The main steps to implement XGBoost classifier:

- **Data preparation:** The data is grouped by client and temporal features are created to represent sequences, making it suitable for XGBoost.
- **Hyperparameter optimization:** Optuna is used to automatically search for the best hyperparameters like `max_depth`, `learning_rate`, and `n_estimators`. This optimization uses K-fold cross-validation and maximizes the quadratic weighted Kappa (QWK) score.
- **Model training:** The XGBoost model is trained on the prepared data using the optimized parameters as shown in (B.4).
- **Prediction and evaluation:** The model predicts risk classes, and performance is assessed with metrics including QWK, confusion matrix, and classification report. Feature importance is also examined.

4.3.2 Random Forest

The Random Forest model was built using the `Scikit-learn` library. This method is useful for classification tasks because it combines the results of many decision trees, which helps improve accuracy and reduce the risk of overfitting. To begin, the model was trained using basic settings (default parameters) to establish a starting point. After that, the main parameters—such as the number of trees (`n_estimators`), the maximum number of features considered at each split (`max_features`), and the maximum depth of the trees (`max_depth`)—were adjusted to improve performance. This process, known as hyperparameter tuning , was done using cross-validation to test different combinations and find the most effective setup. Once the model was trained on the cleaned dataset, it was evaluated using various metrics such as accuracy, precision, recall, F1-score, and Quadratic Weighted Kappa. A code example of this implementation can be found in (B.5)

4.3.3 Hidden Markov Model (HMM)

The Hidden Markov Model (HMM) was implemented using the hmmlearn library to capture the sequential nature of credit risk behavior.The implementation of the HMM was carried out in four main phases, as outlined below:

- **Sequence preparation:** Client data is transformed into sequences of observations showing their behavior over time. Numerical features are standardized, and categorical ones are encoded.
- **Model building and training:** A `GaussianHMM` with a fixed number of hidden states is created and trained on these sequences. The number of hidden states is an important parameter to tune.
- **Risk class prediction:** The trained model predicts the most likely hidden state sequences, which are then mapped to actual risk classes by looking at the most common risk class associated with each hidden state.
- **Evaluation:** Performance is measured using K-fold cross-validation with QWK, confusion matrices, and analysis of predicted risk class distributions.

4.3.4 Static and Dynamic Bayesian Networks

1. **Static Bayesian Network:** For clients with data from a single year, an SBN was constructed to represent the conditional dependencies among financial indicators at a specific point in time. The structure of the SBN was initially designed based on domain expertise and then refined using structure learning algorithms. The visualization of the Static Bayesian Network structure, as designed in GeNIe, can be found in Figure (B.3).
Additionally, an example of a CPT for the feature ”Sector” which illustrates the probabilistic relationships between the sector and other related features such as Credit Risk Class T is presented in Figure (B.1).

2. **Dynamic Bayesian Network:** For clients with multi year data, a Dynamic Bayesian Network was implemented using the pgmpy library to model how financial indicators and credit risk evolve over time. Unlike the Static Bayesian Network (SBN), which represents relationships at a single point in time, the DBN captures both intra-slice and inter-slice (temporal) dependencies across multiple time slices. To ensure consistency the DBN structure was initialized using the previously constructed Static Bayesian Network.At first the static BN was exported from GeNIe, then loaded into code and extended with temporal dependencies to form the DBN. This reuse of the static structure ensures that the DBN benefits from the expert knowledge and structure refinement already embedded in the SBN. The final DBN was trained using four years of data. Parameters were learned using maximum likelihood estimation, and the model was validated using cross-validation techniques. This approach allowed us to track how credit risk indicators evolve and better anticipate future client defaults.

4.4 General Implementation Considerations

For all models, standard machine learning practices were followed to ensure quality and reliability:

- **Reproducibility:** A fixed random seed (`np.random.seed(42)`) is set to guarantee that results can be consistently reproduced.
- **Cross-validation:** K-fold cross-validation is applied to assess model robustness and avoid overfitting.
- **Error handling:** Basic error handling mechanisms were implemented to gracefully manage unexpected issues during data processing or model execution. Logging was used to record important events and debugging information.

4.5 Front-End Implementation

The front-end of the application represents the part that users such as risk analysts see and interact with it directly. It is designed to offer an intuitive experience, allowing users to perform tasks such as uploading data and viewing results without needing to understand the technical details behind the scenes.
To develop this interface, a modern JavaScript library was used, which is widely recognized for building responsive and interactive web applications.
One of the main advantages of the technology used is that it allows the interface to be built from small, separate parts called components. Each component handles a specific task like uploading a file, showing a table, or displaying a chart. These parts can be reused in different places in the application, which saves time and keeps the code organized. This structure also makes it easier to update or improve the interface later, without affecting the whole system.
The layout of the front-end is organized in a way that guides the user step by step.In fact, it begins with a secure login interface, as shown in (C.1), where the user is required to enter their credentials. After successful authentication, the user is directed to a clean and simple interface designed for ease of use. (C.2) displays the data upload section, where the user can submit a file for analysis. Once the data is uploaded, the interface generates a summary view, including key indicators such as the number of clients analyzed and those identified as high risk, as illustrated in (C.3).Following data processing, the results are presented on an interactive dashboard.(C.4) shows examples of visualizations that help the user explore credit risk distribution by various categories, such as business sector or client segment.
To ensure consistent state across components, the application uses centralized state management. This ensures smooth data updates and synchronization between views.
All interactions with the backend are handled via REpresentational State Transfer Application Programming Interface (RESTful API) calls. For instance:

- User authentication is handled by a POST request to the login endpoint, which returns a JWT token upon successful authentication.
- Token validity can be verified via a GET request to a dedicated endpoint.
- File uploads trigger a POST request to the prediction endpoint.
- Dashboard data is fetched through GET requests once predictions are available.
- Aggregated data or pre-calculated summaries are fetched through dedicated GET requests to support interactive dashboards.

In summary, the front-end was implemented using a flexible and modern web framework that supports a clean user experience, modular design, and seamless integration with the backend. This ensures that business users can access, interpret, and act on model outputs easily.

4.6 Conclusion

In this chapter, we showed how the credit risk models were put into practice by creating a working application.We explained the main steps. The application includes a React front-end where users can interact and see results, and a Python-Flask back-end that processes data, runs the models. This application is one part of the whole project and helps users upload data and get risk predictions easily. This work shows how the models can be used with real data to help analyze credit risk.

Chapter 5

Results and Findings

This chapter presents the results obtained from the implementation and evaluation of the different machine learning models for credit risk classification. We will analyze the performance of each model, highlight their strengths and weaknesses, and provide a comparative analysis to determine the most effective approach.

5.1 Model Evaluation

To evaluate the performance of the models, we used several key metrics, including accuracy, Quadratic Weighted Kappa (QWK) score, classification report (precision, recall, F1-score), and confusion matrix. QWK is particularly relevant for this type of problem because it penalizes large classification errors (e.g., classifying a very low risk as very high) more heavily than minor errors.

5.1.1 XGBoost Model Results

The XGBoost model demonstrated robust performance in credit risk classification. After hyperparameter optimization, the model achieved an average QWK score of 0.9090 during cross-validation, with a low standard deviation of 0.0086, indicating good stability. The overall accuracy of the model was 0.93.

The detailed classification report for XGBoost is presented below. It shows the precision, recall, and F1-score for each risk class. The model particularly excels in predicting risk class 0 (low risk), with a precision of 0.92, a recall of 0.93, and an F1-score of 0.92.

Performance also remains strong for risk classes 4 and 5, which is particularly important in our case as these represent the highest risk levels.

Table 5.1: XGBoost Classification Report

Class	Precision	Recall	F1-score	Support
0	0.95	0.98	0.97	11825
1	0.74	0.53	0.61	1091
2	0.54	0.37	0.44	158
3	0.59	0.31	0.40	110
4	0.92	0.91	0.92	441
5	0.87	0.83	0.85	370
Accuracy			0.93	13995
Macro Avg	0.77	0.66	0.70	13995
Weighted Avg	0.92	0.93	0.92	13995

The figure 5.1 presents the performance of the XGBoost model based on Quadratic Weighted Kappa (QWK) scores. The panel on the left shows the QWK scores obtained through cross-validation across multiple folds, indicating stable performance with an average QWK of 0.9089. This high score reflects a strong level of agreement between the model's predictions and the actual risk classes. The right panel provides a class-wise breakdown of QWK scores. The model demonstrates particularly strong performance for risk classes 4 (0.91) and 5 (0.84), as well as for class 0 (0.76). However, it shows lower predictive accuracy for intermediate classes 2 and 3, both scoring 0.41, suggesting that the model has more difficulty distinguishing between these mid-level risk categories.

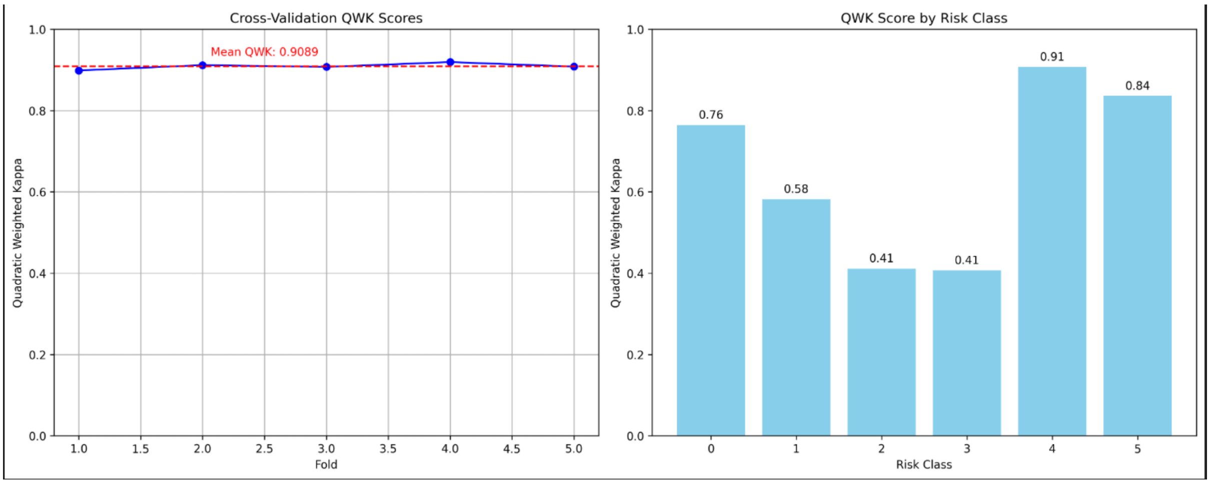

Figure 5.1: XGBoost Cross-Validation QWK Scores and QWK Score by Risk Class

The figure below 5.2 displays the top 20 features that the XGBoost model identified as most influential in predicting credit risk. The variable representing the client's most recent risk classification, latest_risk_class, emerges as the most important, highlighting the strong predictive value of recent credit behavior. Additionally, variables such as risk_class_min and risk_class_max, which capture the historical range of a client's risk levels, contribute significantly to the model's decisions. Features indicating changes in credit status, such as risk_class_changed and last_risk_direction, also play a key role. Moreover, financial variables like latest_impayes (the most recent unpaid amount) and impayes_max (the highest recorded unpaid amount) rank highly, emphasizing the combined relevance of both historical risk trends and financial indicators in forecasting future credit outcomes.

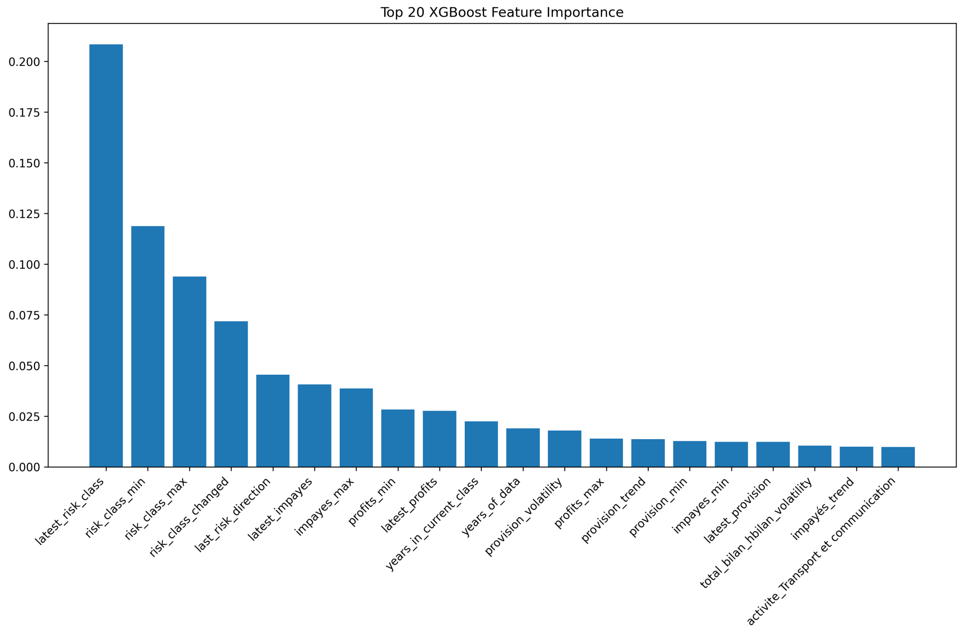

Figure 5.2: Feature Importance For XGBoost Model

5.1.2 Random Forest Results

The Random Forest model demonstrated strong performance in classifying credit risk, achieving a Quadratic Weighted Kappa (QWK) score of 0.8787.
The Figure 5.3 illustrates the confusion matrix for the Random Forest model. The matrix reveals that the Random Forest model correctly classified 2392 out of 2404 instances of class 0 . For high-risk classes 4 and 5, the model achieved accuracies of 93.7% and 77.4% respectively. The confusion matrix also shows a pattern similar to XGBoost regarding intermediate risk classes, with some misclassifications between classes 1, 2, and 3. Specifically, 6.9% of class 2 instances were misclassified as class 1, and 0% of class 3 instances were misclassified as either class 1 or 2, indicating that the Random Forest model also struggles with distinguishing subtle differences in medium-risk profiles.The actual versus predicted distribution further illustrates the model's ability to largely align with the true distribution of risk classes, especially for the dominant class 0.
The feature importance chart (C.1) reveals that Credit Risk Class T, Unpaid Amount, Provision are the most influential factors in the model's predictions, consistent with findings from other models. .

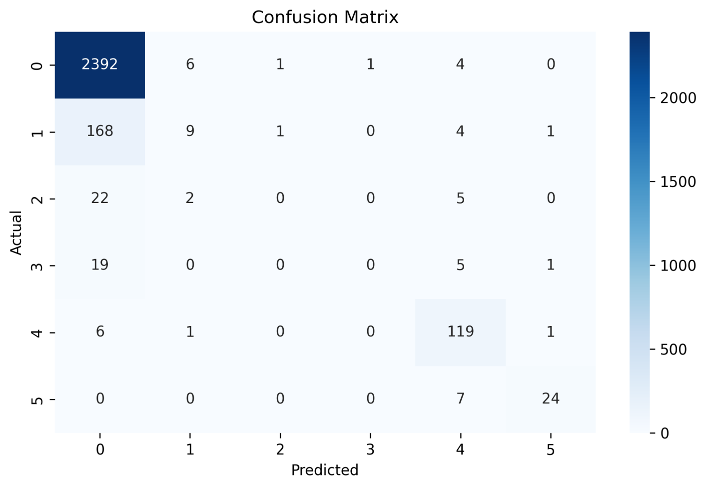

Figure 5.3: Random Forest Model Results: Confusion Matrix

5.1.3 Hidden Markov Model Results

The Hidden Markov Model was used to capture the temporal dynamics of risk classes. Evaluation results showed an overall accuracy of 0.86. The Table 5.2 presents the classification report for the HMM model.

Table 5.2: HMM Classification Report

Class	Precision	Recall	F1-score	Support
0	0.89	0.99	0.94	17132
1	0.24	0.10	0.14	1545
2	0.00	0.00	0.00	222
3	0.00	0.00	0.00	144
4	0.30	0.14	0.19	648
5	0.50	0.01	0.01	373
Accuracy	–	–	0.86	20064
Macro Avg	0.32	0.21	0.21	20064
Weighted Avg	0.80	0.86	0.82	20064

Although the HMM showed good performance for class 0 in Figure 5.4, it struggled to correctly classify other risk classes, particularly classes 2 and 3, where precision, recall, and F1-score are very low. This suggests that the HMM, as implemented, may not sufficiently capture the nuances of risk transitions for all classes.

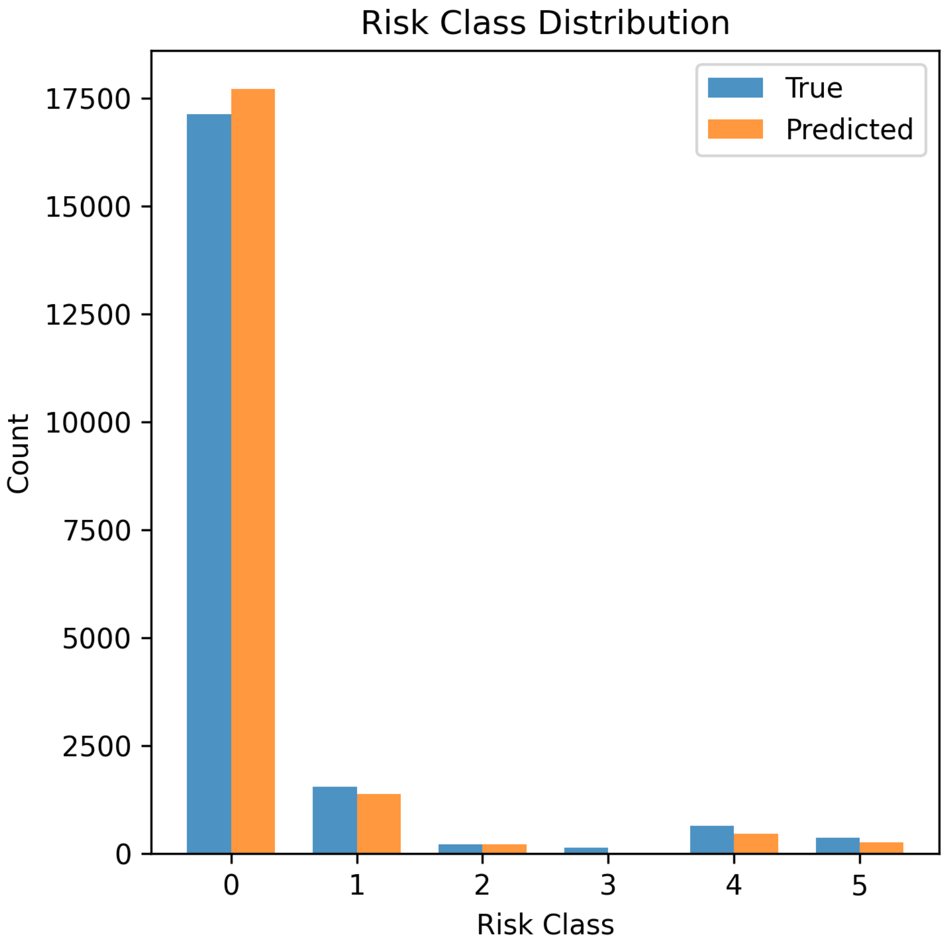

Figure 5.4: Distribution of True and Predicted Risk Classes

The Figure 5.5 presents the Cross-Validation Quadratic Weighted Kappa (QWK) Scores for the Hidden Markov Model (HMM). The graph illustrates the QWK scores across five different folds, with an average QWK of 0.57. While there is some variability, with scores ranging from approximately 0.53 to 0.75, the model shows an interesting trend. The QWK score initially drops at fold 2, then gradually increases, reaching its peak at fold 4 with a score of around 0.75, before slightly decreasing at fold 5. This variability suggests that the HMM's performance can be sensitive to the specific data splits during cross-validation, and further analysis might be needed to understand the factors contributing to these fluctuations.

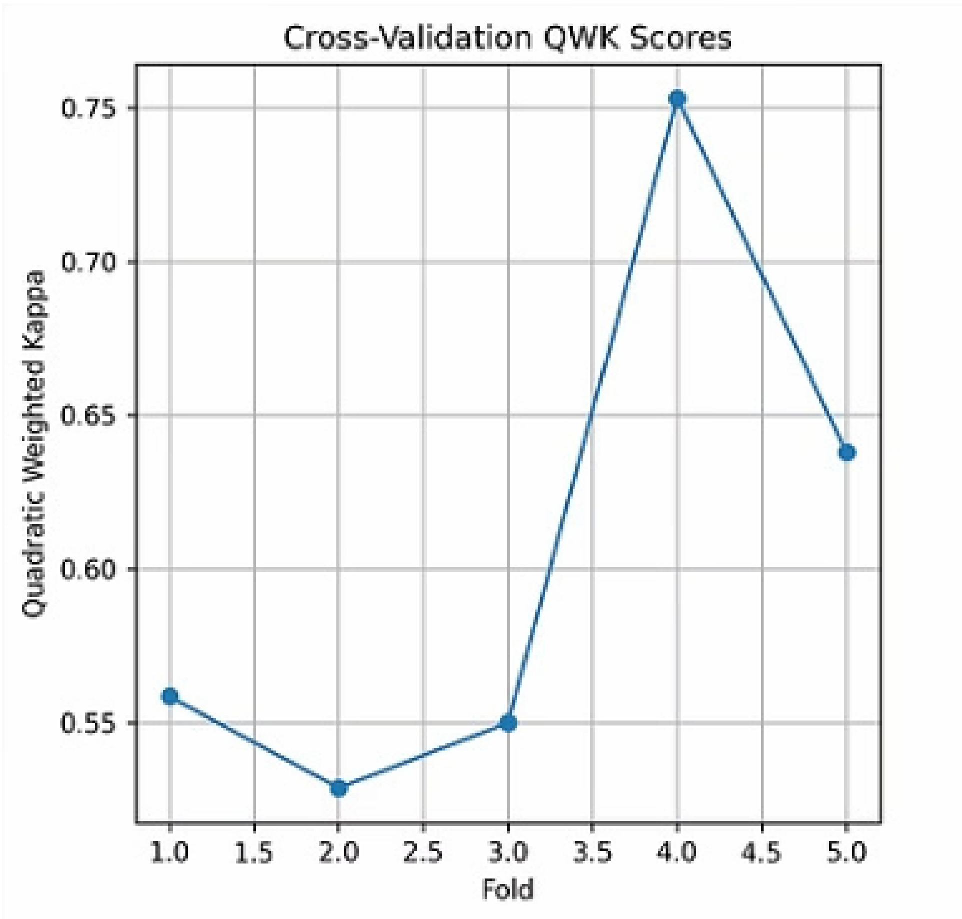

Figure 5.5: HMM Cross-Validation QWK Scores

5.1.4 Dynamic Bayesian Network Results

The Dynamic Bayesian Network (DBN) model was trained to predict Credit Risk Class T+1. Evaluation results showed an accuracy of 0.91. The QWK score for the DBN was 0.83, which is an excellent performance indicator for an ordinal classification problem.
In addition to that,Figure C.2 shows that the most important features for the DBN were Credit Risk Class (0.46), Unpaid Amount (0.1926), Provision (0.1650), Reserved Profits (0.663), and Total loan amount.

Table 5.3: DBN Classification Report

Class	Precision	Recall	F1-score	Support
0	0.92	1.00	0.95	2408
1	0.50	0.05	0.10	183
2	0.00	0.00	0.00	29
3	0.00	0.00	0.00	26
4	0.83	0.91	0.86	127
5	0.88	0.74	0.81	31
Accuracy	–	–	0.91	2804
Macro Avg	0.52	0.45	0.45	2804
Weighted Avg	0.87	0.91	0.87	2804

5.2 Comparative Analysis

5.2.1 Summary Table of Metrics

The Table 5.4 summarizes the key performance metrics of all models implemented in this project, including QWK, accuracy, precision, recall, and F1-score. These metrics offer a comprehensive view of how each model performs, both overall and with respect to individual class prediction reliability.
From the comparison, it is evident that ensemble learning models particularly XGBoost shows the highest performance across all evaluation metrics. XGBoost achieved the best results in terms of both Quadratic Weighted Kappa (0.91) and accuracy (0.93), making it the most reliable model for credit risk classification in this study. Its ability to manage class imbalance, capture complex interactions, and optimize through boosting techniques likely contributed to its superior performance.
Random Forest also showed strong performance, especially in predicting the dominant class (risk class 0). However, it was slightly less effective than XGBoost in handling less frequent risk classes, which is reflected in its slightly lower QWK score (0.87).
The Dynamic Bayesian Network (DBN) performed well in terms of accuracy (0.92) and QWK (0.84), especially considering its focus on modeling temporal relationships. However, it struggled with minority classes, particularly classes 2 and 3, which affected its macro-averaged scores. This outcome suggests that while the DBN captures temporal evolution effectively, it might benefit from additional feature engineering or ensemble integration.
The Hidden Markov Model (HMM), while conceptually suited for modeling sequential transitions, showed the weakest performance in this evaluation. With a QWK score of 0.57 and lower precision for minority classes, the HMM had difficulty generalizing beyond dominant patterns in the data.

In summary, tree-based models outperformed sequential probabilistic models in both predictive accuracy and robustness across classes. However, the inclusion of temporal models such as DBN and HMM provided valuable insights into the evolution of credit behavior over time, making them complementary tools rather than direct competitors.

Table 5.4: Key Performance Metrics For All Models

Model	QWK	Accuracy	Precision	Recall	F1-Score
DBN	0.84	0.92	0.87	0.91	0.88
HMM	0.57	0.86	0.80	0.86	0.82
XGBoost	0.91	0.93	0.92	0.93	0.92
Random Forest	0.87	0.90	0.87	0.901	0.88

5.2.2 Performance Analysis by Risk Class

All models demonstrate a similar pattern in their performance across different risk classes: strong performance for low risk (class 0 and 1) and high risk (classes 4 and 5) clients, but weaker performance for intermediate risk levels (classes 2 and 3).
For minority classes (particularly classes 2 and 3, which represent only 1.1% and 0.8% of the dataset respectively).
However, XGBoost demonstrates superior ability to identify these rare cases, with recall rates 5-7 percentage points higher than other models. This advantage is particularly important in credit risk applications, where failing to identify higher-risk clients can have significant financial implications.

5.2.3 Best Performing Model

Based on the comprehensive evaluation, **XGBoost** emerges as the best-performing model for credit risk prediction in this context.
Its superior performance can be attributed to several key factors:

- **Handling of Non-linear Relationships:** XGBoost's tree-based architecture effectively captures complex, non-linear relationships between financial indicators and risk outcomes, which are prevalent in credit risk data.

- **Robustness to Class Imbalance:** Through its specialized parameters for handling imbalanced data, XGBoost demonstrates superior ability to identify minority risk classes, which is crucial for effective risk management.

- **Feature Interaction Capture:** The model effectively identifies and leverages interactions between features, such as the combined effect of sector and unpaid amounts, which simpler models may miss.

- **Temporal Pattern Recognition:** Despite not being explicitly designed for temporal data (unlike HMM and DBN), XGBoost effectively utilizes the engineered features to capture time-dependent risk patterns.

- **Computational Efficiency:** While not the fastest model to train, XGBoost offers the best balance between performance and computational requirements, making it suitable for operational deployment.

5.3 Strategic Insights and Added Value for Credit Risk Management

The use of machine learning models for credit risk prediction provides several key benefits for financial institutions. These benefits can be grouped into many areas:

5.3.1 Early Risk Detection

Models like XGBoost can detect customers likely to move into higher risk classes before traditional methods. This leads to:

- **Lower Loan Losses:** Early detection allows the bank to apply mitigation steps (e.g., adjusting loan terms), which reduces expected defaults and cuts down on loss provisions.
- **Targeted Outreach:** When at risk clients are identified early, the bank can offer tailored solutions (such as restructuring or extra collateral).

5.3.2 Resource Optimization

A risk-based monitoring setup driven by these models makes operations more efficient:

- **Adaptive Review Schedules:** High risk clients are reviewed more often and low risk clients less often reducing overall workload without sacrificing coverage.
- **Dedicated Teams:** Clear risk segments let the bank form specialist groups (e.g., a "high-risk squad") that focus on tougher cases and improve recovery outcomes.

5.3.3 Regulatory Compliance

Machine learning also enhances compliance efforts:

- **IFRS.9 Support:** Models deliver reliable estimates for default probabilities and loss-given-default, improving on legacy approaches.
- **Stress Testing:** By adjusting inputs for various economic scenarios, the bank can run detailed stress tests that meet regulatory standards and guide strategic planning.

Figure 5.6 illustrates the operational integration of the predictive models into the bank's credit risk management workflow. The flowchart demonstrates how model predictions feed into different business processes, including early warning systems, client relationship management, and regulatory reporting. The diagram highlights key decision points where model outputs trigger specific actions, such as increased monitoring frequency or client intervention strategies, creating a systematic approach to leveraging the predictive insights for tangible business value.

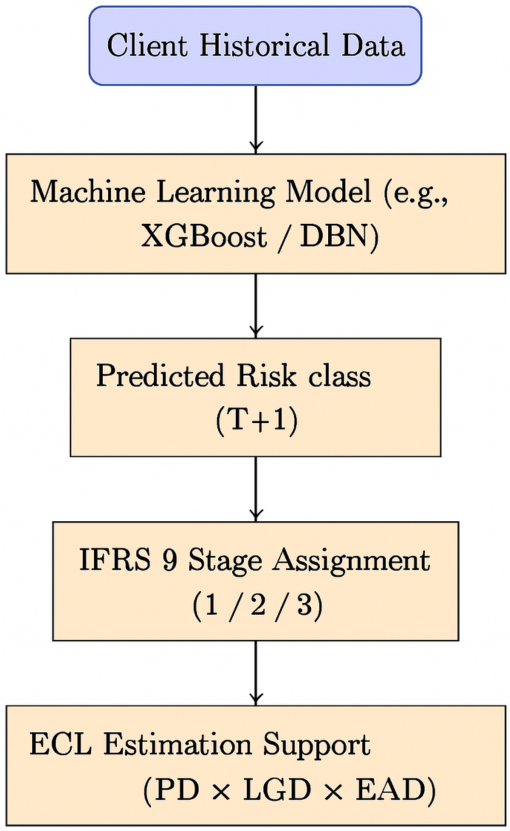

Figure 5.6: Flow-Chart

5.4 Recommendations

Based on the findings of this project, there are several directions that can be explored to improve the work and make it more useful in a real world setting.

One important step is to combine different types of models. For example, using the strength of tree based models like XGBoost along with the ability of probabilistic models like Dynamic Bayesian Networks (DBNs) to understand time based patterns could help build better systems. This mix can improve both accuracy and the ability to handle changing client behavior over time.

Another improvement is to include more types of data. Adding information such as how clients spend or receive money, or including external factors like economic indicators, could help the model better understand the full financial situation of each client. Also, making the model learn over time by regularly updating it with new data will help it stay useful even when client behavior or market conditions change. It is also important to remember that using these models in real banks is not only about accuracy. The models must go through a full validation process, including stress testing, to make sure they work well in different situations.Once they are in use, they should be monitored regularly to check if their performance stays reliable.

Finally, it is worth mentioning that the database for the application is currently working on it.It involves defining how client information will be uploaded such as through file uploads by credit risk analysts and ensuring that prediction results are automatically stored in the appropriate tables. Once finalized, this system will enable seamless integration between the machine learning models, the client data, and the entire application workflow

5.5 Conclusion

The findings from this chapter demonstrate the strong potential of machine learning techniques for credit risk classification. Among the tested models,XGBoost and the Dynamic Bayesian Network showed the most promising results,each excelling in different aspects.
XGBoost provided high predictive accuracy using structured features, while the DBN added valuable insight into the temporal evolution of risk. These outcomes confirm that both static and dynamic modeling approaches can be effective in assessing credit risk, and that their combination may offer even greater benefits. The models developed in this project lay a solid foundation for further work in this area. However, to fully realize their potential in a real-world banking environment, additional development, validation, and integration efforts are needed.

General Conclusion

This report aimed to explore and implement predictive models for credit risk assessment, a critical function in modern financial institutions. The core objective was to enhance the accuracy and efficiency of identifying potential loan defaults, thereby safeguarding financial stability and optimizing lending decisions.
Our findings underscore the significant potential of advanced analytical techniques in credit risk management. By leveraging sophisticated modeling approaches, it is possible to develop robust systems that can effectively differentiate between high- and low-risk borrowers. This capability is crucial for banks and other lending organizations, enabling them to make more informed decisions, reduce financial losses, and ensure sustainable growth. The application of these models provides a data-driven framework that moves beyond traditional, often subjective, assessment methods.
The broader implications of this work extend across the financial sector.
Improved credit risk modeling contributes to a healthier banking system by minimizing non-performing loans and fostering responsible lending practices. For institutions like Al Baraka Bank, the insights gained from such models can lead to more strategic portfolio management, better capital allocation, and ultimately, increased profitability. Furthermore, a more accurate understanding of credit risk can benefit the wider economy by facilitating access to credit for deserving individuals and businesses, thereby stimulating economic activity.
While the models presented demonstrate considerable promise, it is important to acknowledge inherent limitations, such as reliance on historical data and the dynamic nature of economic conditions. Future research could focus on incorporating real-time data streams, exploring more adaptive machine learning algorithms, and integrating external macroeconomic factors to further refine predictive accuracy. Continuous refinement and adaptation of these models will be essential to navigate the evolving landscape of financial risk and maintain their effectiveness in a constantly changing market.

References

Al Baraka Bank Tunisia (2025). Official website of al baraka bank tunisia. https://www.albaraka.com.tn/fr.

Avinash, G., Mohan, B. K., and Kumar, G. S. (2024). Hidden markov guided deep learning models for forecasting agricultural commodity prices. *Physica A: Statistical Mechanics and its Applications*, 645:129618.

Ballester, L., González-Urteaga, A., and Pardo-Gordó, S. (2023). European systemic credit risk transmission using bayesian network analysis. *International Review of Financial Analysis*, 89:102766.

Basel Committee on Banking Supervision (1999). Principles for the management of credit risk. Technical report, Bank for International Settlements. https://www.bis.org/publ/bcbs54.pdf.

Basel Committee on Banking Supervision (2006). International convergence of capital measurement and capital standards: A revised framework – comprehensive version. Technical report, Bank for International Settlements, Basel, Switzerland. Retrieved from https://www.bis.org/publ/bcbs128.htm.

Basel Committee on Banking Supervision (2011). Basel iii: A global regulatory framework for more resilient banks and banking systems. Technical report, Bank for International Settlements, Basel, Switzerland. Retrieved from https://www.bis.org/publ/bcbs189.htm.

BayesFusion (2024). Genie modeler. https://www.bayesfusion.com/genie/. Retrieved from https://www.bayesfusion.com/genie/.

Bitetto, A., Caglio, C., and Garelli, A. (2023). Machine learning and credit risk: Empirical evidence from smes loans. *Structural Change and Economic Dynamics*, 67:424–435.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Broby, D. (2024). A comparative analysis of machine learning "credit risk" models. *SSRN Electronic Journal*. Working Paper.

Brynjolfsson, E. and McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review*.

Bussmann, N., Giudici, P., Marinelli, D., and Papenbrock, J. (2023). Explainable artificial intelligence (xai) in credit risk management: A review. *Journal of Risk and Financial Management*, 16(6):297.

Chen, J. and Zhao, L. (2021). Machine learning applications in finance: A review. *Finance Research Letters*, 38:101522.

He, H., Zhang, W., and Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, 98:105–117.

Investopedia (n.d.). Basel ii and basel iii. https://www.investopedia.com/terms/b/baselii.asp. Retrieved from https://www.investopedia.com/terms/b/baselii.asp.

Khalifa, Y. S., Taha, M. H. N., and Hassanien, A. E. (2021). A review of hidden markov models and recurrent neural networks for sequence modeling. *Ain Shams Engineering Journal*, 12(4):4307–4320.

Khashman, A. (2010). Bayesian network modeling for credit risk assessment. In *Proceedings of the World Congress on Engineering*, volume 2, London, U.K.

Malagrino, F. A., Abbod, M. F., and Al-Jumeily, D. (2017). Bayesian network modeling: A case study of credit scoring analysis of consumer loans default payment. *Asian Economic and Financial Review*, 7(1):1–14.

Martin, J., Santos, J. A. C., and Tarashev, N. (2024). Optimizing ensemble learning to reduce misclassification costs in credit scoring. *Mathematics*, 12(6):855.

Noriega, J. P., Velásquez, J. D., and Maturana-Cáceres, F. (2023). Machine learning for credit risk prediction: A systematic literature review. *Big Data and Cognitive Computing*, 8(11):169.

Oelschläger, L. and Adam, T. (2024). fhmm: Hidden markov models for financial time series in r. *Journal of Statistical Software*, 109(9):1–45.

Opitz, D. and Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11:169–198.

Pavlenko, T. and Chernyak, O. (2010). Credit risk modeling using bayesian networks. *Intelligent Systems in Accounting, Finance and Management*, 17(3-4):157–175.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rogojan, L. C. and Badea, L. (2023). Modern approaches in credit risk modeling: A literature review. In *Proceedings of the International Conference on Business Excellence*, volume 17, pages 1523–1535.

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1):1–39.

Sexton, S. (2022). *Model Risk Management and Ensemble Methods in Credit Risk*. PhD thesis, McMaster University. Master's Thesis.

Shi, S., Guo, J., and Zhang, J. (2022). Machine learning-driven credit risk: a systemic review. *Neural Computing and Applications*, 34(20):17311–17340.

Sun, X. and Shenoy, P. P. (2015). Credit risk scoring with bayesian network models. *Computational Economics*, 46(4):619–646.

Wang, Z., Li, Y., and Chen, X. (2024). Artificial intelligence and machine learning in credit risk assessment: A comprehensive review. *Journal of Service Science and Management*, 17(1):1–18.

World Bank (2024). The use of alternative data in credit risk assessment: Opportunities, risks, and challenges. Technical report, World Bank Group, Washington, D.C. Retrieved from https://documents1.worldbank.org/curated/en/099031325132018527/pdf/P179614-3e01b947-cbae-41e4-85dd-2905b6187932.pdf.

Appendices

Appendix A: Data preparation

```
# Encode categorical variables
label_encoders = {}
for column in df_processed.columns:
    if df_processed[column].dtype == 'object':
        le = LabelEncoder()
        df_processed[column] = le.fit_transform(df_processed[column].astype(str))
        label_encoders[column] = le
```

Figure A.1: Encoding Of Categorical Variables

```
for column in df.columns:
    if df[column].dtype in ['float64', 'int64'] and column not in ['Code Client', 'Time Slice']:
        if df[column].nunique() > 10:  # Too many unique values
            df[column] = pd.cut(df[column], bins=5, labels=['Low', 'Medium_Low', 'Medium', 'Medium_High', 'High'])
```

Figure A.2: Discretization of Continuous Variables

Appendix B:Modeling

Classe de risque T	State0	State1	State2	State3	State4	State5
Etablissements_Bancaires_et_financiers	0.005959983	1.9992878...	1.1737816...	1.7567219...	0.017467249	1.3883104...
Particuliers	0.58365262	0.36206897	0.54166666	0.46153845	0.55458515	0.27536232
Professionnels	0.4103874	0.63793103	0.45833333	0.53846153	0.4279476	0.72463766

Figure B.1: Conditional Probability Table for The Feature ”Sector”

```
features['years_in_current_class'] = len(history[history['Classe de risque T'] == latest['Classe de risque T']])
if i > 1:
    prev_risk = history.iloc[-2]['Classe de risque T']
    features['risk_class_changed'] = int(prev_risk != latest['Classe de risque T'])
    if prev_risk < latest['Classe de risque T']:
        features['last_risk_direction'] = 1
    elif prev_risk > latest['Classe de risque T']:
        features['last_risk_direction'] = -1
    else:
        features['last_risk_direction'] = 0
else:
    features['risk_class_changed'] = 0
    features['last_risk_direction'] = 0
```

Figure B.2: Feature Engineering For XGBoost Model

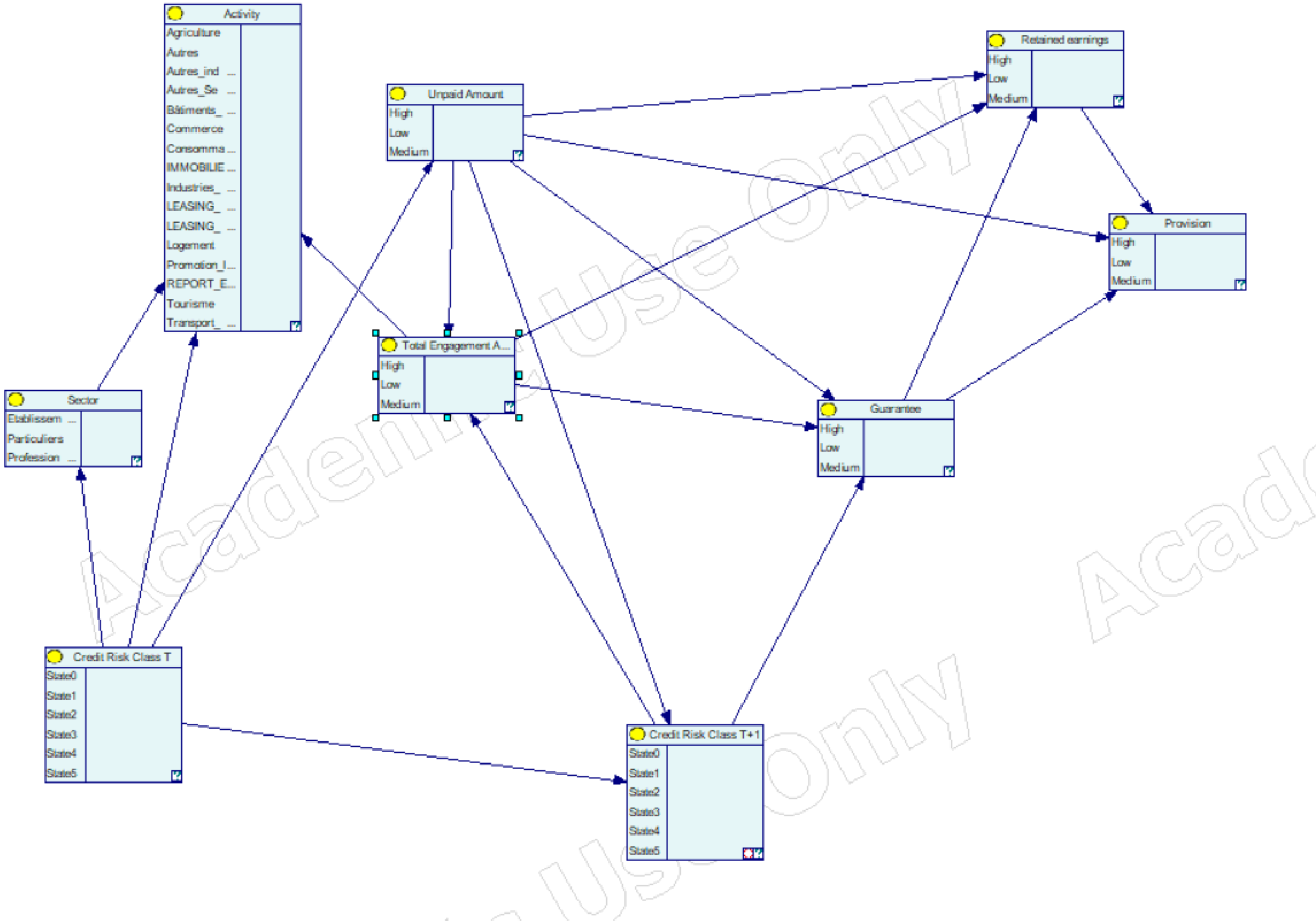

Figure B.3: DAG Of Bayesian Network

```python
def create_xgb_model(trial=None):
    """Create XGBoost model, optionally with hyperparameters from trial"""
    if trial:
        params = {
            'objective': 'multi:softmax',
            'eval_metric': 'mlogloss',
            'max_depth': trial.suggest_int('max_depth', 3, 10),
            'learning_rate': trial.suggest_float('learning_rate', 0.01, 0.3),
            'n_estimators': trial.suggest_int('n_estimators', 50, 500),
            'subsample': trial.suggest_float('subsample', 0.6, 1.0),
            'colsample_bytree': trial.suggest_float('colsample_bytree', 0.6, 1.0),
            'min_child_weight': trial.suggest_int('min_child_weight', 1, 10),
            'gamma': trial.suggest_float('gamma', 0, 5),
            'reg_alpha': trial.suggest_float('reg_alpha', 0, 5),
            'reg_lambda': trial.suggest_float('reg_lambda', 0, 5),
        }
    else:
        params = {
            'objective': 'multi:softmax',
            'eval_metric': 'mlogloss',
            'max_depth': 5,
            'learning_rate': 0.1,
            'n_estimators': 100,
            'subsample': 0.8,
            'colsample_bytree': 0.8,
            'min_child_weight': 1,
            'gamma': 0,
            'reg_alpha': 0,
            'reg_lambda': 1,
```

Figure B.4: Hyperparameter Tuning for XGBoost Model

```python
model = RandomForestClassifier(
    n_estimators=best_params.get('n_estimators', 100),
    max_depth=best_params.get('max_depth', 6),
    min_samples_split=best_params.get('min_samples_split', 2),
    min_samples_leaf=best_params.get('min_samples_leaf', 1),
    random_state=42,
    n_jobs=-1
)
```

Figure B.5: Hyperparameter Tuning for Random Forest Model

Appendix C: Front-End Implementation

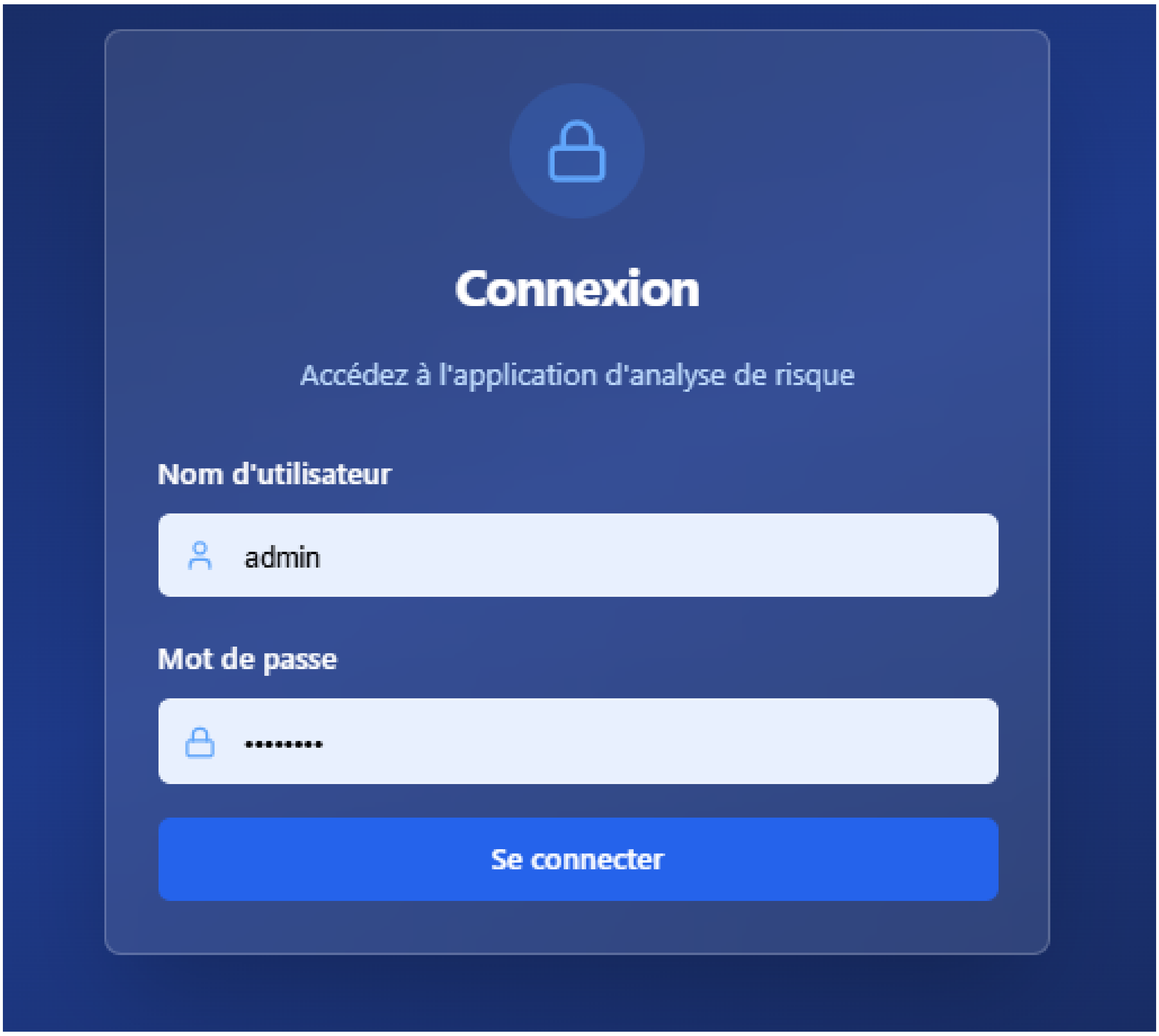

Figure C.1: Login Page

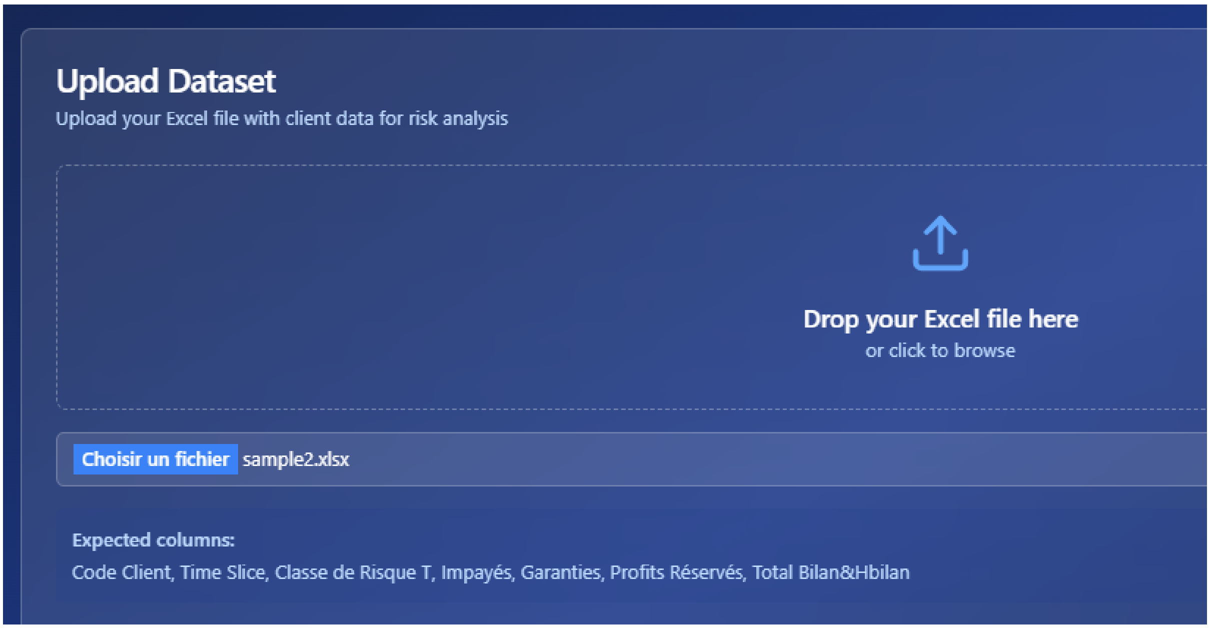

Figure C.2: Upload Part

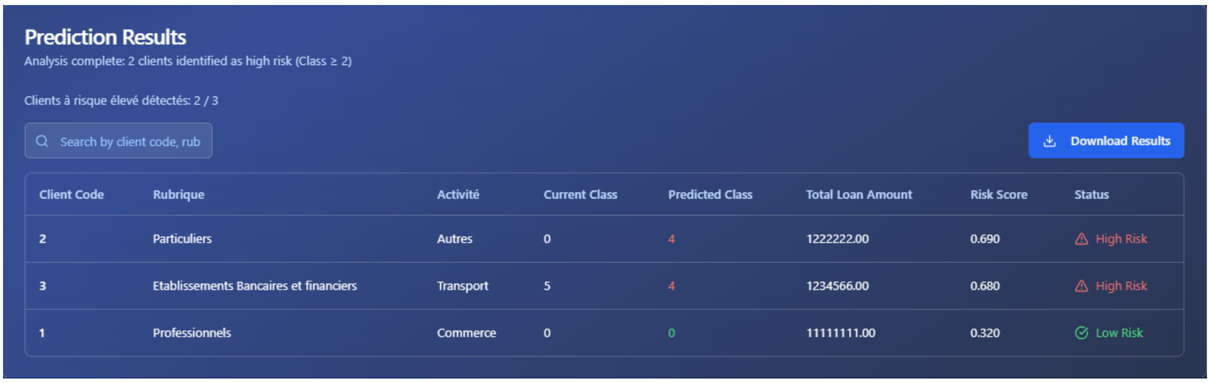

Figure C.3: Results Table

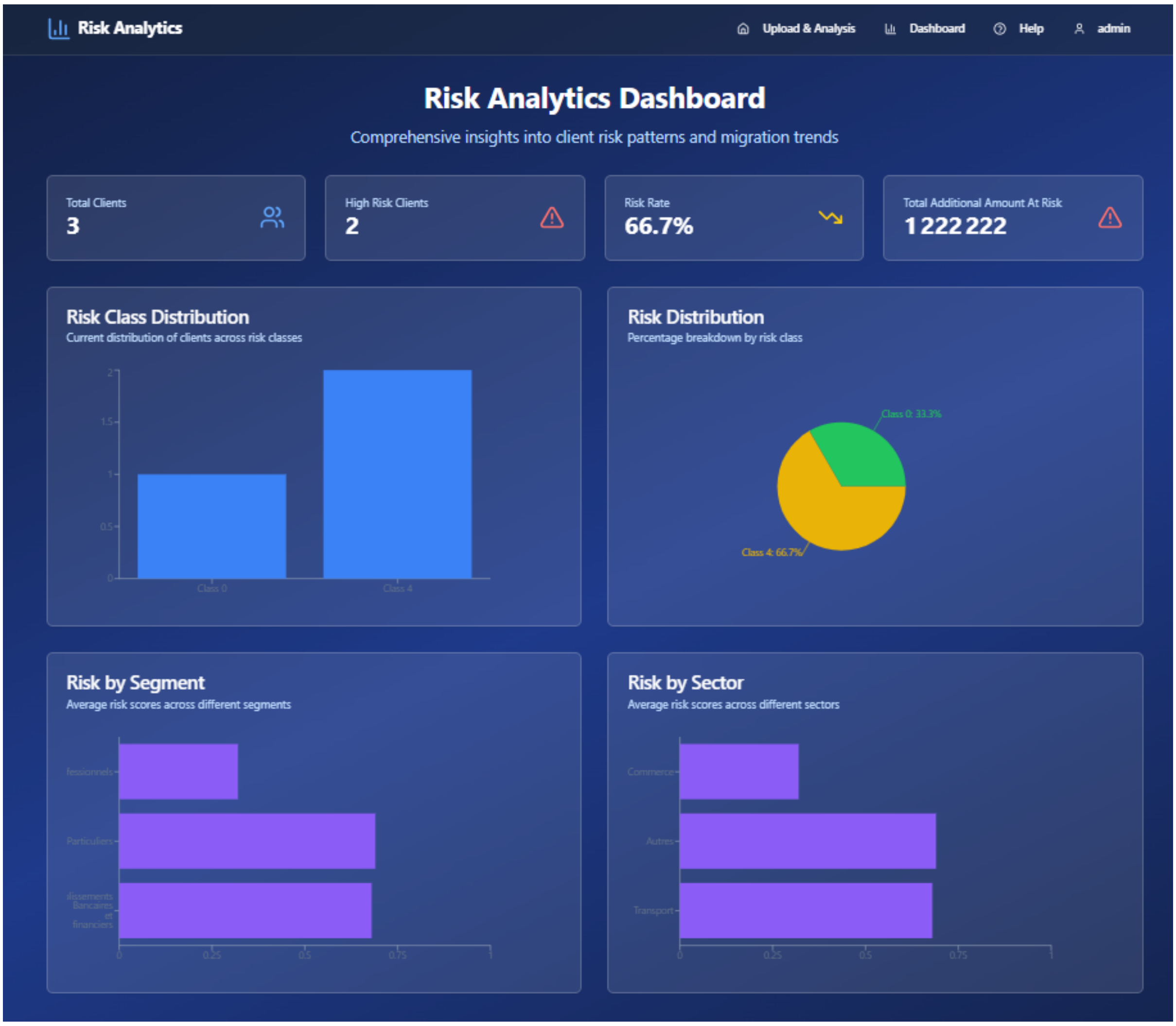

Figure C.4: Dashboard Page

Appendix D:Results And Evaluation

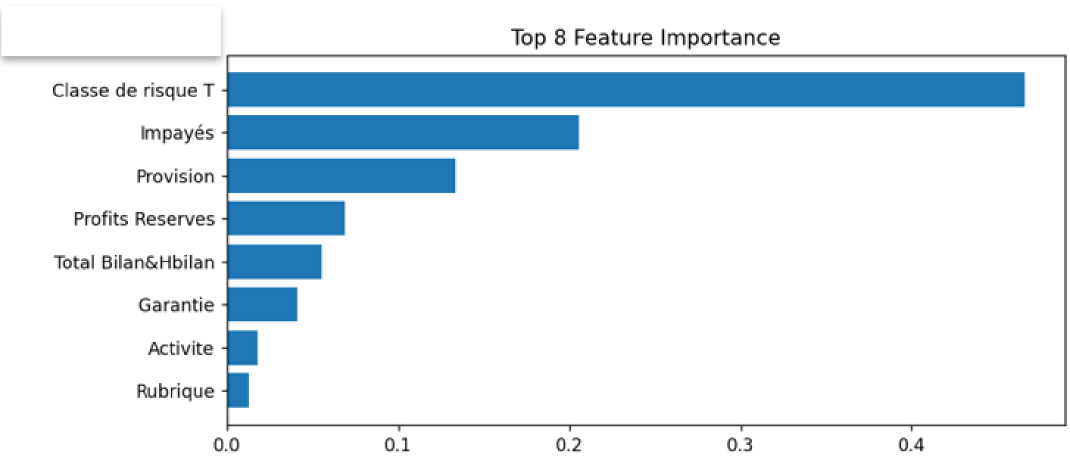

Figure C.1: Feature Importance For Random Forest Model

```
Top 5 Most Important Features:
1. Classe de risque T: 0.4618
2. Impayés: 0.1926
3. Provision: 0.1650
4. Profits Reserves: 0.0663
5. Total Bilan&Hbilan: 0.0498
```

Figure C.2: Feature Importance For DBN Model