

Article

Not peer-reviewed version

Innovative Credit Scoring and Sales Accounting Solutions for SMEs in Kazakhstan

[Gulnaz Zakariya](#), [Olzhas Akylbekov](#), [Aiman Moldagulova](#)*, [Ryskhan Satybaldiyeva](#)

Posted Date: 21 November 2025

doi: 10.20944/preprints202511.1621.v1

Keywords: credit scoring; data integration; digital transformation; financial technology; hybrid models; real-time analytics; sales accounting





Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Innovative Credit Scoring and Sales Accounting Solutions for SMEs in Kazakhstan

Gulnaz Zakariya ¹, Olzhas Akylbekov ¹, Aiman Moldagulova ^{2,*}
and Ryskhan Satybaldiyeva ²

¹ Department of Software Engineering, Satbayev University, Almaty, Kazakhstan

² Department of Cybersecurity, information processing and storage, Satbayev University, Almaty, Kazakhstan

* Correspondence: a.moldagulova@satbayev.university; Tel.: +7-701-727-9025

Abstract

The paper examines the combination of traditional banking credit assessment techniques with contemporary internal sales accounting systems in Kazakhstan, aiming to augment the precision and resilience of financial assessments pertaining to SMEs. The proposed model consists of two discrete components: a traditional credit scoring module that employs logistic regression and a supplementary sales analytics module that leverages ensemble machine learning methodologies - random forests and gradient boosting algorithms. The outputs generated by these components are amalgamated through an ensemble strategy, where optimal weighting coefficients are ascertained via cross validation. An empirical analysis was conducted on a dataset encompassing 41,000 SME records from a prominent Kazakhstan bank alongside daily transactional sales data from 150 SMEs gathered between the years 2021 and 2024. The integrated hybrid model demonstrated a statistically meaningful enhancement in predictive efficacy, as evidenced by an increase in the area under the ROC curve from 0.76 to 0.87 and a decrease in mean squared error from 0.12 to 0.08 relative to the traditional methodology. The investigation delves into the transformative influence of digitalization on innovation within SMEs, elucidating that improved real-time data integration not only sharpens risk assessment processes but also promotes adaptive lending strategies and operational efficiencies.

Keywords: credit scoring; data integration; digital transformation; financial technology; hybrid models; real-time analytics; sales accounting

1. Introduction

The effectiveness of credit assessment is of paramount importance for financial stability and economic growth. For producers and importers of goods, credit assessment plays a crucial role in securing the capital needed for expansion, investment, and risk reduction. However, traditional credit assessment methods often fail to provide an accurate estimation of risks, leading to financial losses for lenders, restrictive credit conditions for borrowers, and missed opportunities for business development [1,2]. Classic scoring models, despite their simplicity and interpretability, inadequately capture the complex, often nonlinear, nature of borrower behavior. They are poorly adapted to dynamic financial conditions and overlook numerous factors influencing a company's solvency, such as supply chain characteristics, seasonal fluctuations in demand, and foreign economic conditions. This inefficiency is especially problematic for small and medium-sized enterprises (SMEs), which often struggle to demonstrate a stable credit history or sufficient collateral [3].

In the domain of credit scoring, extensive research has been conducted to enhance the accuracy and efficiency of credit evaluation models. A significant body of literature emphasizes behavioral scoring, which has gained prominence following the global financial crisis and the subsequent Basel Committee recommendations for banks to adopt more effective credit evaluation systems. Traditional statistical approaches such as Linear Discriminant Analysis and Logistic Regression remain widely used; however,

the integration of machine learning (ML) and artificial intelligence (AI) techniques—such as neural networks, support vector machines, and random forests—has demonstrated substantial improvements in predictive performance [4, 5]. Moreover, hybrid models that combine behavioral and credit valuation techniques have emerged as a promising direction. For example, models utilizing neural networks and data mining methods can effectively analyze borrower behavior based on transaction frequency, monetary value, and recency, thereby improving risk differentiation [6].

At the same time, the landscape of financial data management has undergone a dramatic transformation, particularly affecting SMEs. The introduction and widespread adoption of advanced sales accounting and management systems—such as Umag and Alsep—have revolutionized the way businesses capture, record, and monitor transactional data in real time. These systems provide detailed, granular datasets that extend far beyond traditional financial indicators, encompassing operational metrics such as daily sales volumes, inventory turnover rates, and cash flow dynamics. This evolution not only streamlines financial processes but also enhances the ability of firms to identify seasonal patterns and short-term fluctuations in business activity. Such capabilities enable more accurate forecasting, better risk management, and improved financial resilience, offering a rich source of data for modern credit assessment models. Against this backdrop, modern research highlights that credit risk assessment remains a multifaceted challenge. While traditional models are interpretable and regulator-friendly, they struggle to reflect nonlinear borrower behavior or adapt to rapidly changing market conditions [7]. This limitation is particularly severe for manufacturers and importers, whose financial reliability often depends on external factors—trade flows, supply chain disruptions, and global market volatility—that conventional scoring systems fail to capture. Consequently, many enterprises encounter significant barriers to financing, limiting their ability to scale operations and contribute to broader economic development [8].

To overcome these shortcomings, scholars increasingly explore the integration of machine learning, explainable artificial intelligence (XAI), and reinforcement learning (RL) within predictive modeling frameworks [9]. ML and AI algorithms can uncover complex dependencies in borrower data, achieving superior predictive accuracy, while XAI techniques enhance transparency and accountability. Meanwhile, reinforcement learning introduces dynamic optimization, enabling lending policies to evolve based on real-world repayment outcomes. By rewarding desired borrower behaviors, RL systems can refine credit allocation strategies over time, which is particularly relevant in volatile sectors such as trade finance. Studies demonstrate that combining RL with predictive modeling forms a foundation for intelligent credit scoring platforms tailored to the needs of producers and importers within the context of Industry 4.0 [10].

Despite significant progress, several critical questions remain open: how to balance the accuracy of ML models with the transparency required for regulatory and practical adoption; how to adapt RL mechanisms to the requirements of credit underwriting; and what architectural solutions are most suitable for building flexible, scalable platforms designed for SMEs operating in high-risk industries such as import and manufacturing. Therefore, the development of an intelligent small business lending system that integrates reinforcement learning and predictive modeling represents a relevant and timely research direction. The focus on small businesses is motivated by their high level of informational uncertainty and limited access to banking resources. Unlike larger firms, SMEs often lack comprehensive financial statements or a stable credit history, reducing the effectiveness of traditional scoring models. In this context, intelligent data analysis and self-learning algorithms can significantly improve the accuracy and objectivity of creditworthiness assessment, thereby promoting inclusive and sustainable financial development.

2. Literature Review

The importance of credit scoring in modern financial transactions is widely recognized, particularly for small and medium-sized enterprises (SMEs). In recent years, credit scoring systems based on machine learning (ML) and artificial intelligence (AI) methods have emerged as cutting-edge tools

for increasing credit availability and improving risk assessment. This is particularly relevant for countries like Kazakhstan, where SMEs play a key role in economic development [11]. Recent research demonstrates the high effectiveness of ML algorithms, including XGBoost and LightGBM, in building robust and accurate credit scoring models. These models use alternative data sources to predict default probabilities, improving the accuracy of lending decisions and reducing default rates [12].

Empirical studies confirm the successful implementation of complex credit scoring models in various contexts. For example, the use of neural networks and support vector machines has improved the accuracy of credit risk assessment, indicating significant potential for integrating innovative methods into banking practices [13]. However, despite technological advances, small and medium-sized businesses in Kazakhstan continue to face barriers to obtaining loans due to information asymmetry, insufficient data transparency, and limited digitalization [14,15].

The multi-criteria credit scoring model proposed by Roy and Shaw [15] uses a hybrid best-worst method (BWM) and the TOPSIS method, which allows for the consideration of multiple factors, from credit history to cash flow liquidity. The results showed that traditional commercial models are not always adapted to the unstructured data typical of SMEs. In a similar vein, Zhang et al. [16] developed a neural network-based credit risk assessment model for micro, small, and medium-sized enterprises (MSMEs), which addresses the problem of information asymmetry and adapts to market conditions.

Current research in Kazakhstan, particularly the work of Tyumambayeva and Abdeshova [14], reveals persistent structural problems in bank lending to SMEs, despite government support measures (tax incentives, interest rate subsidies, etc.). The authors emphasize that access to credit remains limited, and the existing lending system requires modernization to improve the effectiveness of enterprise creditworthiness assessment.

In parallel, the implementation of sales accounting systems and accounting information systems (AIS) as a means of increasing financial transparency is being explored. Kahyani et al. [17] showed that the integration of automated accounting systems contributes to improved financial management, but the application of AI and credit scoring in the context of SMEs in Kazakhstan remains understudied.

Recent research highlights that the use of AI methods in sales accounting and credit scoring systems can significantly improve the operational efficiency and sustainability of businesses. Predictive analytics and personalized AI-based solutions have already demonstrated a positive impact on customer experience and sales effectiveness [18].

Despite the advantages of machine learning and reinforcement learning for forecasting, their practical implementation in the financial sector is limited by transparency and regulatory compliance requirements. The paper [19] examines the issues of testability and explainability of ML-based credit scoring models. While modern algorithms provide high forecasting accuracy, they often function as "black boxes," which hinders their application in highly regulated environments. Researchers note the need to develop explainable and testable models, particularly in the banking sector, where decisions directly impact clients' access to financing.

The challenges of balancing accuracy and interpretability, as well as the high computational cost of explainable algorithms, remain unresolved. This is due to both the technical limitations of existing approaches and the heterogeneity of the data used in credit scoring systems. A promising direction is the development of hybrid systems that combine the predictive power of complex models with ex post facto explanation tools. Table 1 provides a systematic overview of key studies reflecting the evolution of credit scoring approaches—from traditional statistical models to modern machine learning algorithms and explainable AI. Early work [2], [1] focused on interpretable statistical methods such as logistic regression and discriminant analysis, which provided transparency but suffered from limited adaptability and low predictive power. Subsequent studies [4], [5] demonstrated the advantages of ensemble and nonlinear models, which can significantly improve forecasting accuracy. However, high accuracy was accompanied by decreased explainability and increased computational costs.

Modern approaches [9] focus on developing explainable machine learning (XAI) concepts and integrating ex post facto explanation tools, which provides a basis for balancing interpretability and

accuracy. Thus, the analysis of the table confirms the need to develop hybrid, explainable, and adaptive models capable of meeting both regulatory requirements and the practical needs of credit institutions. However, as shown in Table 1, most existing studies focus on improving forecast accuracy, ignoring the aspects of data diversity, adaptability and real-time learning.

Table 1. Comparative analysis of existing studies of creditworthiness assessment methods.

Authors (year)	Methods/Models	Object of study	Key contributions/Results	Restrictions/Gaps in research
DJ Hand and W. E. Henley (1997) [2]	Methods of statistical classification (discriminant analysis, logistic regression)	Consumer credit scoring	A classic review summarizing early statistical approaches	No machine learning is used; static models are data-dependent
J. Bastos (2009) [3]	Loss Given Default (LGD) Prediction Based on Regression	Bank loan portfolios	Quantitative Loss Given Default (LGD) Model	Narrow focus (default loss only); no behavioral or adaptive modeling
A. Khandani, A. Kim, and A. Lo (2010) [8]	Machine learning (SVM, RF, boosting)	Consumer lending risk	Innovative application of machine learning to predict credit risk	Does not address transparency or regulatory compliance

Table 1. Comparative analysis of existing studies of creditworthiness assessment methods.

Authors (year)	Methods/Models	Object of study	Key contributions/Results	Restrictions/Gaps in research
O.M. Soloshenko (2015) [10]	K-Plus Nearest Neighbor (K+NN)	Consumer loan	Improved adaptation of KNN to evaluation tasks	High computational complexity; lack of interpretability
S. Lessmann et al. (2015) [4]	Comparative analysis of machine learning algorithms (support vector machines, random forests, neural networks)	Credit scoring datasets	Demonstrated superiority of ML methods over classical statistical models	High accuracy but limited interpretability and adaptability
R. Florez and J. Ramon (2015) [5]	Ensemble learning (correlated-corrected decision forest)	Credit risk assessment	Improved balance between accuracy and interpretability	Still static; no self-learning or real-time adjustments
N. Siddiqui (2017) [1]	Development of traditional statistical indicator systems; logistic regression; expert systems	Consumer and retail credit	A comprehensive methodology for constructing interpretable scorecards used in banking practice	Limited adaptability and automation; lack of machine learning and dynamic learning
M. Bucker et al. (2020) [9]	Explainable Machine Learning (XAI, SHAP/LIME)	Credit scoring models	Focus on transparency, verifiability and explainability	Trade-off between accuracy and interpretability; lack of adaptability
W. Byun et al. (2023) [19]	Practical Application of Deep Reinforcement Learning to Optimal Trade Execution	Optimally execute large stock orders over varying time horizons in realistic market conditions	PPO-LSTM-based deep reinforcement learning model	Lack of stress-scenario evaluation

Thus, the identified gaps confirm the need to develop a **scalable, explainable, and integrated credit scoring system** adapted to the conditions of the Kazakhstani financial market. This will not only increase confidence in credit analysis models but also expand SME access to financial resources, contributing to the country's sustainable economic development.

3. Materials and Methods

3.1. Traditional Credit Scoring Module

The Traditional Credit Scoring Module is meticulously crafted to evaluate and interpret conventional financial metrics that have long been integral to the assessment of credit risk. This module employs logistic regression, a robust and widely respected statistical technique for binary classification, to estimate the probability of default or financial distress based on a comprehensive analysis of historical credit data. At the core of this model are several critical predictor variables that act as indicators of an enterprise's financial stability and creditworthiness. These variables include Credit History Score, Annual Revenue and Collateral Value.

Credit History Score metric provides a quantitative evaluation of an enterprise's historical borrowing behavior. It encompasses a range of factors, such as instances of previous defaults, the timeliness of repayments, and an overall assessment of creditworthiness. A higher credit history score typically reflects a responsible borrowing pattern and a lower likelihood of future defaults.

Serving as a proxy for the size and financial strength of the enterprise, annual revenue is a fundamental determinant of its ability to meet debt obligations. Higher revenue figures suggest greater financial capacity to service debt, thereby enhancing the enterprise's credit profile.

Collateral Value variable represents the estimated worth of assets pledged as security for loans. By providing collateral, enterprises offer lenders an additional layer of protection against potential defaults. The greater the collateral value, the more secure lenders feel, which can positively influence credit decisions.

The statistical significance of each predictor variable is rigorously evaluated through hypothesis testing, alongside comprehensive measures of goodness-of-fit, which collectively ensure the model's reliability and predictive accuracy. This thorough analytical approach underscores the importance of data-driven decision-making in the realm of credit risk assessment.

3.2. Internal Sales Analytics Module

The Internal Sales Analytics Module is a sophisticated tool meticulously designed to enhance the operational effectiveness of small and medium-sized enterprises (SMEs) by providing real-time insights derived from comprehensive sales and transactional datasets. This module harnesses the power of advanced machine learning algorithms, particularly random forests and gradient boosting, to uncover complex, non-linear relationships within the data that traditional analytical methods may overlook. At the core of the module's functionality are key variables extracted from the internal sales dataset, which include Daily Sales Volumes, Seasonal Indices and Real-Time Cash Flow Trends.

Daily Sales Volumes. This high-frequency metric serves as a critical indicator of daily revenue generation and the overall operational activities of the SME. By analyzing daily sales, businesses can identify trends, assess performance fluctuations, and make informed decisions to optimize their sales strategies.

Seasonal Indices. These quantitative constructs are essential for capturing periodic variations in sales, allowing businesses to understand and anticipate seasonal patterns and cyclicity in their revenue streams. By recognizing these trends, SMEs can strategically align their inventory management, marketing efforts, and resource allocation to maximize profitability during peak periods.

Real-Time Cash Flow Trends. Continuous monitoring of cash inflows and outflows provides invaluable insights into the liquidity position and operational efficiency of the enterprise. This real-time analysis enables SMEs to make proactive financial decisions, ensuring they maintain sufficient cash reserves to meet operational needs and capitalize on growth opportunities.

To achieve optimal performance, the machine learning algorithms integrated within this module undergo rigorous refinement through grid search and cross validation techniques. These methodologies ensure precise parameter estimation and significantly reduce the risk of overfitting, thereby enhancing the reliability of the predictive models. The culmination of this analytical process is a dynamically computed risk score that encapsulates the current financial performance and operational trends of the SME. This innovative output serves as a timely enhancement to the static characteristics of traditional credit scoring systems, providing businesses with a more nuanced understanding of their financial health and enabling them to make data-driven decisions that foster growth and sustainability.

3.3. Ensemble Integration

The concluding phase of the hybrid model represents a sophisticated integration of outputs generated from both the Traditional Credit Scoring Module and the Internal Sales Analytics Module, achieved through the application of advanced ensemble methodologies. This ensemble integration is carried out using a weighted aggregation technique, which allows for the determination of a comprehensive risk score as a convex combination of two distinct components: the credit-based risk score (R_c) and the sales-based risk score (R_s).

The formula for the total risk score is:

$$R_{\text{total}} = w_c \cdot R_c + w_s \cdot R_s, \quad (1)$$

where

R_{total} is the final comprehensive risk score;

R_c is the credit-based risk score;

R_s is the sales-based risk score;

w_c is the weight assigned to the credit-based risk score;

w_s is the weight assigned to the sales-based risk score.

For this aggregation to be a convex combination, the weights must satisfy:

$$w_c \geq 0, \quad w_s \geq 0, \quad w_c + w_s = 1. \quad (2)$$

The primary objective of this procedure is to identify the optimal weighting that minimizes prediction error, which is quantitatively assessed using metrics such as the Mean Squared Error (MSE) and the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC-ROC). The ensemble strategy harnesses the strengths of both modules: the traditional credit scoring technique, which offers stability and valuable historical insights, and the sales analytics module, which provides agility and real-time relevance. Such a dual-faceted approach ensures that the final risk score is not only robust but also reflective of contemporary market dynamics. The integration process is visually elucidated in Figure 1, which illustrates the sequential data flow from initial collection through the distinct processing phases of each module, culminating in the ensemble integration that yields the final comprehensive risk score.

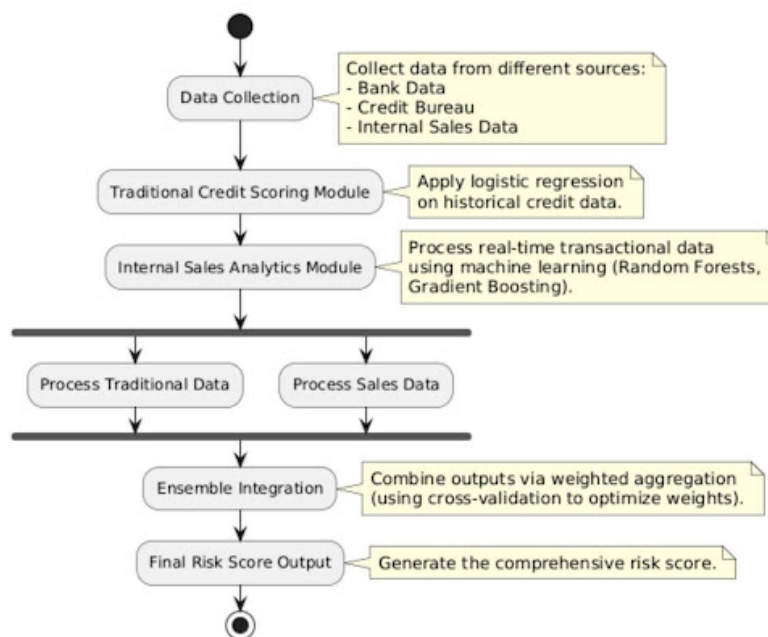


Figure 1. Diagram of the Hybrid Model Integration Process.

By synthesizing these complementary methodologies, the hybrid model establishes a more nuanced and precise risk assessment framework, specifically designed for Small and Medium-sized Enterprises (SMEs) operating within volatile and seasonal markets. This integrative approach signifies a substantial advancement over conventional credit scoring techniques, as it not only enhances predictive performance but also provides critical operational insights. These insights are essential for developing dynamic lending strategies, particularly in emerging economies where market conditions can fluctuate significantly. The hybrid model's innovative framework represents a paradigm shift in risk assessment methodologies, combining traditional and contemporary analytical techniques to deliver superior predictive capabilities and actionable insights that are crucial for informed decision-making in the lending landscape. The following sections will delve deeper into the experimental setup and results, articulated in a fully academic and scientific style, to further elucidate the efficacy and implications of this advanced risk assessment model.

4. Experiments and Results

4.1. Data Sample and Preparation

To methodically evaluate the effectiveness of the proposed hybrid model for credit risk assessment, we meticulously curated a comprehensive dataset that integrates traditional banking credit information with internal sales transaction records. This dataset comprises 41,000 entries from small and medium-sized enterprises (SMEs) spanning the years 2021 to 2024. The banking data captures a wealth of information, including detailed credit histories, repayment behaviors, collateral evaluations, and annual revenue figures. These variables are essential for constructing a robust historical profile of financial performance and creditworthiness, forming the foundation of established credit scoring methodologies. In parallel to the banking data, we collected daily transaction records from 150 SMEs that employ advanced digital sales accounting systems, such as Umag and Asep, throughout the same period. This sales data is particularly rich and granular, encompassing operational metrics such as daily sales volumes, inventory turnover rates, seasonal variations, and real-time cash flow dynamics. Each SME contributes multiple records, and the mapping was established via unique SME's identifiers in the dataset. These high-frequency data points are invaluable for understanding the dynamic operational behavior of SMEs, especially in environments characterized by significant seasonal fluctuations. To ensure the integrity and compatibility of our integrated dataset, we undertook rigorous data cleaning

procedures. In our model, we aggregated raw sales transactions into meaningful time windows to reduce noise and ensure alignment with credit events.

4.2. Experimental Setup

The experimental configuration was designed with meticulous attention to detail, aiming to conduct a thorough and exhaustive evaluation of the efficacy of both the discrete model components and the integrated hybrid model. To achieve this, the complete dataset was systematically partitioned into two distinct subsets: a training set comprising 70% of the total observations and a testing set that accounted for the remaining 30%. This stratified division was strategically implemented to ensure that the distribution of critical variables was preserved across both segments, thereby enhancing the reliability and validity of the subsequent model assessments. During the training phase, a conventional credit scoring model was developed using logistic regression, a statistical method well-suited for binary classification tasks. This model was specifically tailored to the financial institution's credit data, enabling it to estimate the probability of default. Key predictor variables were meticulously selected and included factors such as credit history scores, annual income, and collateral valuation, each of which plays a significant role in determining an individual's creditworthiness. In parallel, an internal sales analytics module was constructed utilizing advanced ensemble machine learning techniques, notably Random Forests and Gradient Boosting. These methodologies were carefully chosen for their ability to capture complex patterns in the data. The models were trained on a high-frequency sales dataset, and hyperparameter optimization was rigorously performed through grid search combined with extensive cross validation. Once the individual modules were independently developed, an ensemble integration methodology was employed to combine their outputs into a cohesive framework. This integration was executed through a weighted aggregation technique, where the final risk score was computed as a convex combination of the traditional credit-based risk score and the sales-based risk score. The weighting parameter, which signifies the relative contribution of each module to the overall score, was meticulously determined through rigorous cross validation on the training set. This ensured that the ensemble model achieved maximal predictive accuracy.

The experimental protocol was structured around a series of sequential procedures, each designed to build upon the previous step:

1. Data Collection and Data Pre-processing. These steps took 80% overall methodology.
2. Training of the Traditional Model. Logistic regression was applied to the bank's credit data to establish baseline risk scores, serving as a foundation for comparison.
3. Development of the Sales Analytics Module. Advanced machine learning techniques, specifically random forests and gradient boosting, were utilized to extract meaningful predictive features from the sales data, thereby enhancing the model's robustness.
4. Ensemble Integration. The outputs from the traditional credit scoring model and the sales analytics module were integrated using a weighted ensemble approach, allowing for a more comprehensive risk assessment.
5. Evaluation. The performance of both the individual models and the integrated hybrid model was rigorously evaluated on the testing set using a comprehensive suite of performance metrics, ensuring a thorough understanding of each model's effectiveness.

This methodological framework was meticulously crafted to ensure that both historical and real-time data are synergistically leveraged, ultimately enhancing the precision and reliability of credit risk evaluations.

4.3. Performance Metrics

To thoroughly evaluate the predictive efficacy of our models, we employed critical metrics that provide a comprehensive assessment of classification performance. These metrics include Area Under the Curve (AUC), Accuracy, Sensitivity, Specificity, and Mean Squared Error (MSE). The AUC of the Receiver Operating Characteristic (ROC) curve is a vital metric that reflects the model's ability to

effectively differentiate between positive and negative classes. A higher AUC value indicates superior model performance, suggesting that the model is adept at correctly identifying true positives while minimizing false positives. Accuracy measures the proportion of correct predictions made by the model across all classes. Sensitivity, or the true positive rate, assesses the model's ability to correctly identify positive instances, while specificity, or the true negative rate, evaluates its competency in recognizing negative instances. For continuous risk score predictions, the Mean Squared Error serves as a key indicator of model accuracy. It quantifies the average of the squares of the errors—that is, the average squared difference between the predicted risk scores and the actual outcomes. A lower MSE signifies that the model's predictions are closely aligned with the true values, indicating a higher level of precision in its forecasting capabilities.

4.4. Results

In this segment, we offer an exhaustive synthesis of the descriptive statistics pertaining to the two principal data sources employed in our research. Table 2 delivers an all-encompassing depiction of essential variables derived from both the banking credit dataset and the internal sales dataset.

The empirical analysis of bank credit data, encompassing 41,000 records from small and medium-sized enterprises (SMEs), reveals a mean credit score of 650 accompanied by a standard deviation of 45, signifying a notably stable credit performance across the selected enterprises. The mean annual revenue for these entities approximates 210 (± 45) million tenge, whereas the average collateral values are estimated to be around 126 million tenge. Understanding these patterns not only enhances the accuracy of risk assessments but also empowers financial institutions to tailor their lending strategies, ultimately fostering a more resilient ecosystem for small and medium enterprises. The seasonality index, presenting a mean of 1.5 and a standard deviation of 0.3, encapsulates the intrinsic seasonal variances in sales - an essential element that conventional credit scoring models frequently neglect.

Table 2. Descriptive Statistics for Credit and Sales Data (in million tenge).

Variable	Credit Data	Sales Data
Mean Credit Score	650 \pm 45	-
Mean Annual Revenue	210 \pm 45	218 \pm 49
Average Daily Sales Volume	-	150 \pm 20
Seasonality Index	-	1.5 \pm 0.3
Collateral Value	126 \pm 21	-

The evaluative analysis of the predictive efficacy of the discrete modules alongside the amalgamated hybrid model was conducted utilizing the designated testing set. Figure 2 illustrates the Receiver Operating Characteristic (ROC) curves corresponding to both the conventional credit scoring framework and the amalgamated hybrid model. The ROC curve associated with the hybrid model signifies a considerable enhancement in discriminative capability, attaining an Area Under the Curve (AUC) of 0.87 in contrast to an AUC of 0.76 for the traditional framework. The hybrid model not only improves predictive accuracy but also enhances decision-making processes for lenders and financial institutions when assessing creditworthiness in SMEs.

Table 3 offers a detailed comparative analysis of the key performance metrics for both models.

Table 3. Model Performance Metrics.

Metric	Traditional Model	Hybrid Model
AUC	0.76	0.87
Accuracy%	0.78	0.85
Mean Squared Error (MSE)	0.12	0.08
Sensitivity	0.75	0.83
Specificity	0.80	0.87

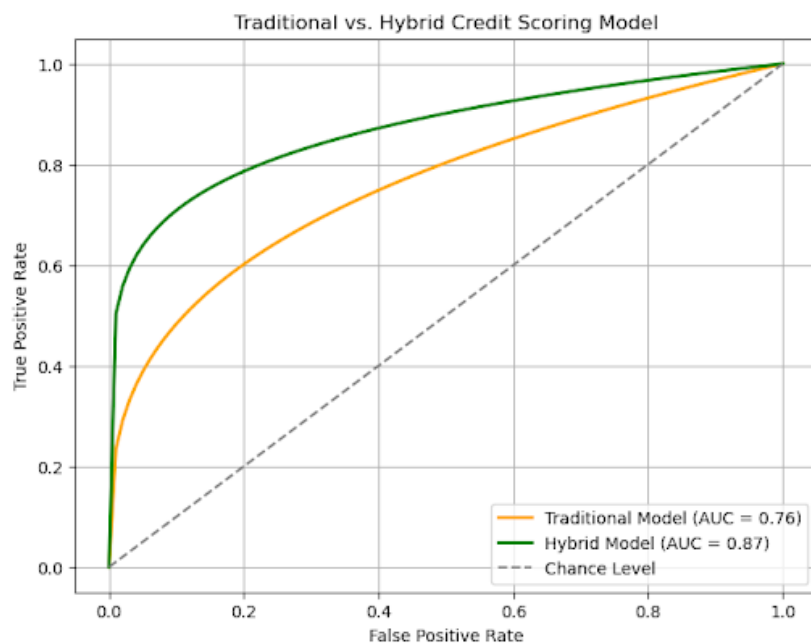


Figure 2. ROC Curves for Traditional vs. Hybrid Model.

The hybrid model represents a significant advancement in the realm of credit risk assessment, achieving an impressive overall accuracy rate of 85%. This figure stands in stark contrast to the 78% accuracy rate recorded by traditional models, underscoring the hybrid approach's superior performance. Furthermore, the model exhibits notable enhancements in both sensitivity and specificity, with sensitivity increasing to 83% from the previous 75% and specificity rising to 87% compared to the conventional model's 80%. These improvements are not merely incremental; they are indicative of a more robust analytical framework. The reduction in the Mean Squared Error (MSE) from 0.12 to 0.08 serves as a quantitative testament to the hybrid model's precision in predicting risk scores. A lower MSE signifies that the model's predictions are closer to the actual outcomes, thereby enhancing the reliability of the risk assessments it generates. The compelling findings of this study provide substantial evidence that the integration of real-time internal sales data with traditional credit information can significantly improve the effectiveness of credit risk evaluations. By leveraging both historical financial indicators and current operational trends, the hybrid model offers a nuanced and comprehensive view of the creditworthiness of small and medium-sized enterprises (SMEs). This dual approach allows for a more sophisticated analysis that takes into account the dynamic nature of business operations, which is particularly crucial in the rapidly evolving economic landscape of emerging markets.

5. Discussion

5.1. Integration Benefits

The findings of this investigation provide a thorough and nuanced understanding of how the integration of real-time internal sales metrics with established credit scoring frameworks can significantly enhance the accuracy and comprehensiveness of credit risk evaluations. The hybrid model developed in this study leverages the combined strengths of both historical financial data and real-time operational insights, resulting in a more sophisticated and flexible paradigm for assessing credit risk. By incorporating real-time sales information, this model enables financial institutions to achieve a finer level of risk differentiation. This is particularly important as it equips lenders with the tools necessary to discern between temporary cash flow disruptions and more persistent financial challenges that could adversely affect long-term viability. Such a capability is crucial in environments like those faced by small and medium-sized enterprises (SMEs) in Kazakhstan, where seasonal fluctuations and informal financial practices can obscure the true risk profile of a business. In this context, the ability to

analyze real-time sales data allows financial institutions to make more informed lending decisions, ultimately leading to better risk management and reduced default rates. The study underscores the importance of adapting credit evaluation methods to reflect the dynamic nature of business operations, particularly in emerging markets where traditional metrics may fall short. This innovative approach not only enhances the precision of credit assessments but also fosters greater financial inclusion by enabling SMEs to access the funding they need to thrive, even amidst economic uncertainties. The hybrid model's dynamic adjustment mechanism plays a pivotal role in the continuous enhancement of risk evaluations. By incorporating a wide array of operational data - ranging from daily sales figures and seasonal fluctuations to real-time cash flow trends - the model adapts and evolves in alignment with the shifting landscapes of business environments. This inherent adaptability not only significantly improves the accuracy of predictions but also empowers the model to identify early warning signs that could indicate potential declines in a borrower's creditworthiness. The outputs generated by this advanced model are characterized by a notably higher area under the Receiver Operating Characteristic (ROC) curve (AUC) and a reduced mean squared error (MSE). These metrics provide lenders with a sophisticated tool for effectively distinguishing between high-risk and low-risk borrowers, thereby enabling more informed credit decisions and flexible pricing strategies. The integration of diverse data sources within the hybrid framework fosters increased confidence among lenders, as the accuracy of the risk scores produced enhances the robustness of risk management protocols. This, in turn, leads to more optimal capital allocation and a reduction in default rates. Furthermore, the model affords lenders a comprehensive view of a small and medium-sized enterprise's operational efficiency and historical credit behavior. This holistic approach is particularly beneficial in emerging markets, where traditional credit indicators often fall short in providing a complete picture of a borrower's financial health. The empirical evidence presented in this study underscores the model's exceptional discriminative power and predictive reliability, which can significantly bolster the resilience of lending portfolios. Ultimately, this contributes to the establishment of a more stable and sustainable financial ecosystem, fostering greater economic growth and stability.

5.2. Broader Implications

The findings of this investigation extend far beyond the confines of the Kazakhstan market, offering valuable insights applicable to a variety of emerging economies. The demonstrated effectiveness of the integration approach underscores the potential for similar hybrid models to be successfully adopted in other developing markets that are characterized by economic volatility and a significant reliance on informal financial practices. In these contexts, traditional credit data often fall short of capturing the full spectrum of an enterprise's operational intricacies. By leveraging real-time data, financial institutions can significantly enhance the accuracy of their risk assessments, leading to more informed decision-making. In light of these insights, financial institutions operating in such dynamic environments are encouraged to consider the implementation of hybrid models to improve their credit risk evaluation processes. By integrating comprehensive operational data into their risk assessment frameworks, banks can not only strengthen their ability to predict defaults but also design lending solutions that are more tailored and responsive to the unique needs of borrowers. This strategic shift has the potential to reduce default rates and enhance the overall quality of credit offerings. Moreover, the adoption of integrative models can facilitate the development of dynamic pricing strategies that align more closely with the real-time performance metrics of small and medium-sized enterprises (SMEs). This alignment allows for the optimization of interest rates and loan terms based on current market conditions, thereby fostering a more responsive and equitable lending environment. Ultimately, the implications of this research suggest a transformative approach to credit risk management that could significantly benefit both financial institutions and their clients in emerging markets. The implications of adopting integrative lending methodologies extend far beyond their immediate application, significantly influencing macroeconomic stability and promoting financial inclusiveness across diverse sectors. By leveraging hybrid paradigms that enhance the accuracy of risk assessments, financial institutions can optimize the allocation of credit throughout the economic landscape. This strategic

distribution of resources is particularly vital for small and medium-sized enterprises (SMEs), which play a crucial role in driving economic growth and fostering innovation. The refined risk assessment framework not only improves the precision with which lenders evaluate potential borrowers but also empowers SMEs to access the capital necessary for expansion and development. As these enterprises flourish, they contribute to job creation, technological advancement, and overall economic dynamism, underscoring their importance as a cornerstone of sustainable economic progress. Moreover, this sophisticated approach to risk evaluation can serve as a catalyst for the development of robust regulatory frameworks and best practices that prioritize transparency and resilience within credit markets. Such frameworks are especially critical in regions marked by economic volatility, where traditional lending practices may fall short. By establishing guidelines that encourage responsible lending and informed decision-making, stakeholders can foster a more stable financial environment, ultimately enhancing trust and participation in the credit market. This comprehensive strategy not only mitigates risks but also cultivates a more inclusive financial ecosystem, ensuring that diverse economic actors can thrive in an increasingly interconnected world.

5.3. Digitalization and Innovation Impacts

The integration of internal sales accounting frameworks with traditional credit scoring methodologies represents not just a slight improvement in operational efficiency; it signifies a profound and transformative shift in how financial risk is conceptualized and quantified in today's rapidly evolving digital environment. This transformation is emblematic of a broader trend of digitalization that is fundamentally reshaping business operations and financial governance practices across the globe. As highlighted by Opoku and Aribigbola [20], the extent to which small and medium-sized enterprises (SMEs) embrace digitalization has a significant impact on their capacity for both product and process innovations. Their research reveals that SMEs operating within sectors that are digitally advanced experience markedly greater benefits from the process of digital transformation compared to those entrenched in more traditional industries. This disparity underscores the critical importance of digital adoption in fostering competitive advantages, enhancing operational capabilities, and driving innovation in an increasingly interconnected marketplace. The implications of this shift are vast, suggesting that organizations that fail to adapt to the digital landscape risk being left behind, while those that leverage these advancements can unlock new opportunities for growth and efficiency. Thus, the amalgamation of these frameworks is not merely a technical enhancement; it is a strategic imperative that redefines the very essence of financial management and risk assessment in the modern age. Engagement in digital value chains has been shown to significantly foster process innovations by creating enhanced opportunities for operational improvements and efficiency optimizations. This is particularly relevant in today's fast-paced business environment, where agility and adaptability are paramount. For instance, the integration of sophisticated sales accounting systems generates real-time data that equips small and medium-sized enterprises (SMEs) with the tools necessary to continuously monitor key performance indicators. This capability allows these organizations to make rapid adjustments in response to dynamic market conditions, thus maintaining a competitive edge. Furthermore, research indicates that organizations with a higher proportion of employees possessing graduate degrees tend to reap more substantial benefits from digitalization. This correlation can be attributed to the fact that these individuals are typically more skilled in leveraging advanced digital tools and interpreting complex data analytics. Their proficiency enables them to harness the full potential of digital technologies, driving innovation and improving decision-making processes within the organization. In addition, SMEs that are engaged in international markets and operate in a globalized economic landscape demonstrate a stronger relationship between digitalization and innovation. These enterprises benefit from their exposure to diverse market environments and best practices from around the world. This cross-border engagement not only broadens their understanding of various consumer behaviors but also inspires the adoption of innovative approaches and solutions that can be tailored to meet the needs of different markets. Consequently, the interplay between digital transformation and innovation becomes a critical driver of success for these globally active SMEs,

positioning them to capitalize on emerging opportunities and navigate the complexities of the modern economy with greater efficacy. Moreover, government-sponsored initiatives designed to promote digitalization have proven to be instrumental in strengthening the connection between digital transformation and innovation outcomes. These programs provide critical infrastructural support, which is essential for small and medium-sized enterprises (SMEs) seeking to modernize their operations. By incentivizing SMEs to adopt advanced digital tools and technologies, these initiatives significantly enhance their operational efficiency and bolster their competitive positioning in the marketplace. The implications of these findings are substantial and far-reaching. The integration of digital tools into traditional financial frameworks not only improves the accuracy and effectiveness of risk assessment but also serves as a catalyst for continuous innovation within SMEs. This dual benefit underscores the urgency of embracing digital transformation as a strategic priority. By doing so, SMEs can not only thrive in a rapidly evolving economic landscape but also contribute to sustained economic growth and resilience in emerging markets. In essence, the role of digital transformation transcends mere technological adoption; it is a vital component for fostering a culture of innovation that can drive long-term economic sustainability. As SMEs leverage digitalization, they position themselves not just to survive but to flourish, ultimately shaping the future economic landscape of their respective regions.

6. Conclusion

In conclusion, this investigation offers compelling evidence that the integration of traditional banking credit scoring methodologies with internal sales accounting frameworks significantly enhances the accuracy and responsiveness of credit risk assessments for small and medium-sized enterprises (SMEs) in Kazakhstan. The proposed hybrid model effectively captures both the historical financial stability and the dynamic operational performance of SMEs, thereby facilitating a more comprehensive and timely evaluation of credit risk. The empirical findings reveal notable improvements in key predictive performance metrics. Specifically, there is a marked increase in the area under the receiver operating characteristic curve (AUC), which signifies enhanced model discrimination between creditworthy and non-creditworthy borrowers. Additionally, there is a substantial reduction in the mean squared error (MSE) related to risk score predictions, indicating that the model generates more precise forecasts of credit risk. This dual approach not only strengthens the reliability of credit evaluations but also equips financial institutions with the necessary tools to better support the growth and sustainability of SMEs in the region.

The hybrid model serves as a pivotal advancement in the methodological framework for credit risk assessment, offering profound implications for financial institutions navigating the complexities of modern markets. As digital transformation continues to reshape the landscape of emerging economies, the adoption of innovative methodologies—such as the one articulated in this discussion—becomes increasingly vital. This approach not only enhances access to financing for small and medium-sized enterprises (SMEs) but also plays a crucial role in reducing default rates and fostering a more resilient financial ecosystem. The integration of digital tools into credit scoring methodologies acts as a powerful catalyst for widespread digital transformation within SMEs.

Acknowledgments: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19675226).

References

1. Siddiqi, N. *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards*; Wiley: Hoboken, NJ, USA, 2017. <https://doi.org/10.1002/9781119282396>.
2. Hand, D. J.; Henley, W. E. Statistical classification methods in consumer credit scoring: A review. *J. R. Stat. Soc. A* **1997**, *160* (3), 523–541. <https://doi.org/10.1111/j.1467-985X.1997.00078.x>.
3. Bastos, J. Forecasting bank loans loss-given-default. *J. Bank. Finance* **2009**, *34*, 2510–2517. <https://doi.org/10.1016/j.jbankfin.2010.04.011>.

4. Lessmann, S.; Baesens, B.; Seow, H.-V.; Thomas, L. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. Oper. Res.* **2015**. <https://doi.org/10.1016/j.ejor.2015.05.030>.
5. Florez, R.; Ramon, J. Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment: A correlated-adjusted decision forest proposal. *Expert Syst. Appl.* **2015**, *42*. <https://doi.org/10.1016/j.eswa.2015.02.042>.
6. Rogoan, L.; Croicu, A.; Iancu, L. Modern approaches in credit risk modeling: A literature review. *Proc. Int. Conf. Bus. Excell.* **2023**, *17*, 1617–1627. <https://doi.org/10.2478/picbe-2023-0145>.
7. Ali, M.; Razaque, A.; Yoo, J.; Kabievna, U.; Moldagulova, A.; Satybaldiyeva, R.; Zhuldyz, K.; Kassymova, A. Designing an intelligent scoring system for crediting manufacturers and importers of goods in Industry 4.0. *Logistics* **2024**, *8* (1), 33. <https://doi.org/10.3390/logistics8010033>.
8. Khandani, A.; Kim, A.; Lo, A. Consumer credit-risk models via machine-learning algorithms. *J. Bank. Finance* **2010**, *34*, 2767–2787. <https://doi.org/10.1016/j.jbankfin.2010.06.001>.
9. Bücken, M.; Szepannek, G.; Gosiewska, A.; Biecek, P. Transparency, auditability and explainability of machine learning models in credit scoring. *J. Oper. Res. Soc.* **2020**, *73* (1), 70–90. arXiv:2009.13384. <https://doi.org/10.48550/arXiv.2009.13384>.
10. Soloshenko, O. M. K-plus-nearest neighbor method development for credit scoring machine learning tasks. *East.-Eur. J. Enterp. Technol.* **2015**, *3* (9), 29–38. <https://doi.org/10.15587/1729-4061.2015.43730>.
11. Nwaimo, C. S.; Adegbola, A. E.; Adegbola, M. D. Predictive analytics for financial inclusion: Using machine learning to improve credit access for underbanked populations. *Comput. Sci. IT Res. J.* **2024**, *5* (6), 1358–1373. <https://doi.org/10.51594/csitrj.v5i6.1201>.
12. Chang, A.; Yang, L.-K.; Tsaih, R.-H.; Lin, S.-K. Machine learning and artificial neural networks to construct P2P lending credit-scoring model: A case using Lending Club data. *Quant. Finance Econ.* **2022**, *6* (2), 303–325. <https://doi.org/10.3934/qfe.2022013>.
13. Zhao, J.; Li, B. Credit risk assessment of SMEs in supply chain finance based on SVM and BP neural network. *Neural Comput. Appl.* **2022**, *34* (15), 12467–12478. <https://doi.org/10.1007/s00521-021-06682-4>.
14. Tyumambayeva, A.; Abdeshova, A. B. Current state of bank lending to small and medium-sized businesses in Kazakhstan. *Bull. Kazakh Univ. Econ. Finance Int. Trade* **2022**, *4* (49), 13. [https://doi.org/10.52260/2304-7216.2022.4\(49\).13](https://doi.org/10.52260/2304-7216.2022.4(49).13).
15. Roy, P. K.; Shaw, K. A multicriteria credit scoring model for SMEs using hybrid BWM and TOPSIS. *Financ. Innov.* **2021**, *7* (1), 12. <https://doi.org/10.1186/s40854-021-00295-5>.
16. Zhang, L.; Wei, M.; Yi, Z. Credit decision system for MSMEs based on neural network and nonlinear programming. In *Proceedings of the International Conference on Signal Processing (ICSP); 2022*; pp 1500–1506. <https://doi.org/10.1109/ICSP54964.2022.9778636>.
17. Cahyani, D.; Hazmi, Y.; Khairun Zuhra, N.; Wildani, R. R. Evaluation of the implementation of the credit sales accounting information system. *West Sci. Soc. Humanit. Stud.* **2024**, *2* (07), 109–1098. <https://doi.org/10.58812/wsshs.v2i07.1067>.
18. Kedi, W. E.; Ejimuda, C.; Idemudia, C.; Ijomah, T. I. AI software for personalized marketing automation in SMEs: Enhancing customer experience and sales. *World J. Adv. Res. Rev.* **2024**. <https://doi.org/10.30574/wjarr.2024.23.1.2159>.
19. Byun, W. J.; Choi, B.; Kim, S.; Jo, J. Practical application of deep reinforcement learning to optimal trade execution. *FinTech* **2023**, *2*, 414–429. <https://doi.org/10.3390/fintech2030023>.
20. Opoku, E., & Aribigbola, M. Enhancing small and medium-sized businesses through digitalization. *World Journal of Advanced Research and Reviews* **2024**, *23*(2), 239–249. <https://doi.org/10.30574/wjarr.2024.23.2.2313>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.