Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 24, No. 1, 2026

Research on Dynamic Credit Risk Prediction in Commercial Banks Based on a Hybrid Scorecard-LSTM-Attention Model

Ziyan Cheng

College of Economics and Management, Nanjing Forestry University, Nanjing, 210037, China

chengzy@njfu.edu.com

Abstract: With the increasing fluctuations in the global economy and the development of digital finance in China, commercial banks are facing two major challenges: the declining quality of credit assets and the growth of bad loans. Traditional credit scoring models based on Logistic Regression are easy to explain, but they struggle to handle complex data that is high-dimensional or changes over time. To improve the accuracy of credit risk prediction, this study develops a hybrid model called Scorecard-LSTM. This model combines traditional credit scoring methods with LSTM networks, also added an Attention Mechanism to help the model focus more on key customer behaviors. The study uses data from a commercial bank in Taiwan, including basic customer information, financial status, and repayment history. By using methods like WOE encoding, feature binning, and joint training, the model connects the clear logic of traditional scorecards with the powerful learning ability of deep learning. The experimental results show that this hybrid model is much better than traditional models in both accuracy and stability. With an AUC value of 0.7741, the model performs very well in identifying bad loans and providing early warnings fo r potential defaults.

Keywords: Credit Risk, Credit Scorecard, Attention Mechanism, LSTM, Neural Network.

1. Research Background and Significance

With the deepening reform of China's financial industry and the rapid development of financial technology, commercial banks are facing a double challenge. While credit services support the economy, banks are also dealing with a decline in asset quality and an increase in bad loans. According to data from the National Financial Regulatory Administration, the scale of non-performing and "special mention" loans in China's commercial banks has continued to grow recently. This shows that credit risks are becoming more complex and diverse. This trend suggests that traditional risk control systems are starting to show their limitations when facing the dynamic changes of the Big Data era and a complex economic environment.

Traditional methods for assessing credit risk mainly rely on statistical modeling, such as credit scoring cards built with Logistic Regression [1]. These models have a clear structure and are easy to explain, which is why they have played a central role in bank approvals and regulatory compliance for a long time. However, today's financial market involves high-dimensional data, non-linear changes in customer behavior, and uncertainty in the external environment. Traditional linear models struggle to fully capture these hidden risk relationships. As a result, they often react too slowly to risks and suffer from a decrease in prediction accuracy.

Driven by both global economic changes and digital transformation, commercial banks urgently need to use intelligent tools to improve their risk management systems [2]. Machine learning and deep learning technologies offer new solutions for financial risk control because they are very good at learning data features and recognizing patterns. Compared to traditional models, machine learning can automatically find hidden information from different types of data and build models for complex credit behaviors. In particular, LSTM (Long Short-Term Memory) in deep learning can effectively track changes in customer behavior over time. This provides a more flexible and forward-looking framework for predicting credit risks. At the same time, the Attention mechanism augments the network's information selection capacity, enabling the model to prioritize critical features with greater precision. Consequently, this optimization leads to superior outcomes when addressing high-complexity objectives [3].

Therefore, exploring a hybrid model which combines traditional scorecards with neural networks is very important. This approach can balance the need for clear explanations with high prediction accuracy [4]. It also meets financial regulations and follows the trend of digital development in the banking industry. The goal of the study is to build a dynamic credit risk prediction system based on the Scorecard-LSTM-Attention hybrid model. This will provide commercial banks with an intelligent and scalable path for managing risks [5].

2. Research Status

Early credit risk assessment mainly relied on statistical methods to build scorecard models. These models are widely used in commercial bank credit approvals because they can explain and use simply.

Although research on machine learning and credit risk prediction in China started relatively late, it has developed quickly in recent years and formed a complete research system. There are now many directions for using machine learning in this field. Feng (2025) analyzed cases in banking, insurance, and securities to prove the practical value of Big Data technology in risk management, which can effectively reduce potential losses through real-time monitoring and warning systems [6]. Cui (2025) argued that algorithms such as Support Vector Machines (SVM), decision trees, and neural networks have clear advantages in improving prediction accuracy, optimizing decision processes, and lowering costs [7]. These methods are especially useful for risk control in small and medium-sized banks. Xu (2024) improved the use of Random Forest in personal credit assessment and emphasized the importance of including non-traditional features like social behavior [8]. Zhang (2024) improved the Gradient Boosting Decision Tree (GBDT) by introducing a special learning mechanism, which effectively solved the problem of "overfitting" caused by unbalanced credit data [9].

Deep learning can also be combined with hybrid models to focus on dynamic risk assessment, relationship-based risk control, and automated credit decision systems. For example, some studies use Graph Neural Networks (GNN) to model the guarantee relationships between companies, which helps provide early warnings for systemic risks. Zhang et al. (2023) built a GNN model for corporate guarantee networks. By analyzing the network of relationships between companies, this model successfully predicted 83% of chain default events. It issued warnings 2 to 3 quarters earlier than traditional methods [10]. Similarly, the DynamicGraphRisk framework developed by Kumar et al. (2022) updates the social network graphs of borrowers in real-time to track how risk spreads. When this model was used on a digital lending platform in India, it reduced the bad debt rate by 28% [11].

3. Theoretical Overview

3.1. Basic Theory of Credit Scorecards

In credit risk management, scoring technology uses statistical methods to analyze large amounts of historical data, identify clues related to credit risk, and build a reasonable model [12]. The model then gives a numerical score to a user's credit status. Since credit scoring is based on the analysis of big data, it is highly accurate and reliable. The development of credit scorecards has gone through three main stages: First, focusing on user grading and behavior; Second, focusing on selecting and building mathematical models .Finally, focusing on choosing the factors and indicators that affect credit [13]. Basically, there are three kinds of methods that have been studied in credit scoring; classification techniques, Markov Chain and Survival analysis [14].

3.2. Basic Theory of LSTM

Recurrent neural networks (RNNs) have been widely adopted in research areas concerned with sequential data, such as text, audio, and video. However, RNNs consisting of sigma cells or tanh cells are unable to learn the relevant information of input data when the input gap is large. By introducing gate functions into the cell structure, the long short-term memory (LSTM) could handle the problem of long-term dependencies well [15].

Long Short-Term Memory (LSTM) is a special type of Recurrent Neural Network (RNN). It consists of an input layer, one or more hidden layers, and an output layer. The LSTM model uses a forget gate, an input gate, and an output gate to selectively control information. It can decide which information to remove or add to the memory unit, allowing the model to effectively save important information from a long time ago [16].

In credit risk prediction, customer transactions, repayment records, and credit limit usage all show clear "time-series" features. LSTM can learn the time-dependent relationships in borrower behavior through its memory mechanism, which helps it capture risk trends more accurately. Compared to static models like Logistic Regression or Random Forest, LSTM has a strong ability to model dynamic features and non-linear patterns. It is also very robust; even in financial scenarios with missing data, it still performs well. Additionally, it can be combined with scorecards or Attention mechanisms to achieve both clear explanations and high accuracy. LSTM networks are to a certain extend biologically plausible and capable to learn more than 1,000 timesteps, depending on the complexity of the built network.

3.3. Basic Theory of Attention Mechanism

The attention mechanism can be used as a tool to explain incomprehensible neural architecture behavior [17]. In neural networks, the Attention mechanism calculates attention weights to measure how important each part of the input is to the current task. A common method is to perform weighted calculations on Query, Key, and Value vectors. The weights are usually calculated based on the similarity between the Query and the Key, and then normalized using the softmax function [18].

The classic "Self-Attention" mechanism allows each position in a sequence to combine information from all other positions. This helps the model understand global relationships. This mechanism is the core of the Transformer architecture, which has led to breakthroughs in tasks like machine translation and speech recognition. The Attention mechanism improves the network's ability to select information, allowing the model to focus on key features more efficiently. This results in better performance in complex tasks.

4. Data Preparation and Indicator System Construction

The data for this study comes from the Default of Credit Card Clients dataset provided by the UCI Machine Learning Repository. It contains information on credit card defaults from a bank in Taiwan. The dataset includes 23 features from 30,000 customers, such as personal attributes, credit limits, bill statements, and repayment records. The goal is to predict whether a customer will default on their payment next month. Key features include gender, education, marriage status, age, and the bill amounts and actual payments for the last six months. As summarized in Table 1, the dataset consists of several categories of variables.

Table 1. Variable Meanings

Category	Variable Name	Data Type	Description & Value Definitions
Identity	ID	Integer	Unique identifier for each customer (usually excluded from modeling).
Demographics	SEX	Categorical	1 = Male, 2 = Female.
	EDUCATION	Categorical	1 = Graduate School, 2 = University, 3 = High School, 4 = Others.
	MARRIAGE	Categorical	1 = Married, 2 = Single, 3 = Others.
	AGE	Numerical	Age of the customer in years.
Credit Limit	LIMIT_BAL	Numerical	Amount of given credit (individual and family/supplementary).
History of Past Payment	PAY_n (n=0, 2-6)	Categorical	Repayment status for the last 6 months (-1=Pay duly, 1=Delay 1 month, 2=Delay 2 months, etc.).
Amount of Bill Statement	BILL_AMT_n (n=1-6)	Numerical	Amount of bill statement for the last 6 months (n=1 is most recent).
Amount of Previous Payment	PAY_AMT_n (n=1-6)	Numerical	Amount of actual payment made in the last 6 months.
Target Variable	Y	Binary	Default payment next month (1 = Yes, 0 = No).

During the data preprocessing stage, the study used the pandas library in Python to read and check the raw data. After checking, it found no missing values, but some categorical variables had unusual codes. For example, in the "EDUCATION" category, the codes 0, 5, and 6 were combined into "Others". For continuous variables, the study used StandardScaler to normalize the data. This ensures that all features have a mean of 0 and a variance of 1, which helps the model learn faster and eliminates the impact of different scales.

Since LSTM models are suitable for time-series data, the study organized the customer's repayment status, bill amounts, and payment amounts from the last six months into a time-based sequence. Each customer's input is structured as a (6, 3) matrix, representing six time steps and three types of dynamic features. This method reflects how a customer's credit behavior changes over time.

Finally, the data was divided into a train set (80%) and a test set (20%). The study used stratified sampling to make sure the proportion of default cases remained the same in both sets. After cleaning and organizing the data, a complete and standardized dataset was ready, providing a solid foundation for the LSTM-Attention credit risk model. Table 2 presents the descriptive statistics for key variables in the dataset.

Table 2. Descriptive Statistics

Variable	Obs	Mean	Std. dev.	Min	Max
id	30,000	15000.5	8660.398	1	30000
limit_bal	30,000	167484.3	129747.7	10000	1000000
sex	30,000	1.603733	.4891292	1	2
education	30,000	1.842267	.7444945	1	4
marriage	30,000	1.557267	.5214048	1	3
age	30,000	35.4855	9.217904	21	79
pay_0	30,000	-.0167	1.123802	-2	8
pay_2	30,000	-.1337667	1.197186	-2	8
pay_3	30,000	-.1662	1.196868	-2	8
pay_4	30,000	-.2206667	1.169139	-2	8
pay_5	30,000	-.2662	1.133187	-2	8
pay_6	30,000	-.2911	1.149988	-2	8
bill_amt1	30,000	51223.33	73635.86	-165580	964511
bill_amt2	30,000	49179.08	71173.77	-69777	983931
bill_amt3	30,000	47013.15	69349.39	-157264	1664089
bill_amt4	30,000	43262.95	64332.86	-170000	891586
bill_amt5	30,000	40311.4	60797.16	-81334	927171
bill_amt6	30,000	38871.76	59554.11	-339603	961664
pay_amt1	30,000	5663.581	16563.28	0	873552
pay_amt2	30,000	5921.163	23040.87	0	1684259
pay_amt3	30,000	5225.681	17606.96	0	896040
pay_amt4	30,000	4826.077	15666.16	0	621000
pay_amt5	30,000	4799.388	15278.31	0	426529
pay_amt6	30,000	5215.503	17777.47	0	528666
y	30,000	.2212	.4150618	0	1

5. Empirical Analysis

This study uses the "Default of Credit Card Clients" dataset from the UCI Machine Learning Repository, which contains credit card default records from a bank in Taiwan. The dataset includes a total of 30000 completed loan repayment records. Among these records, 6636 show that the customer defaulted the following month, while 23,364 show that the customer paid on time. All of this data is used for training and testing the models. The goal is to predict the probability of default based on each customer's personal features, credit limit, and historical repayment performance.

5.1. Credit Scorecard Modeling

In this study, Y (Default payment next month) is used as the dependent variable. The study analyzed 24 independent variables, including loan amount, loan terms, and interest rates.

First, an automated feature selection process was performed based on the following standards:

IV < 0.02: Information Value (IV) measures the predictive power of a variable. Variables with an IV of less than 0.02 were automatically removed because they provide almost no useful information to distinguish between good and bad customers.

Missing Rate > 50%: Since the dataset is very complete, this rule mainly acts as a safety filter.

Correlation > 0.7: If two variables are highly correlated, the function keeps the one with the higher IV value. This helps avoid the problem of multicollinearity.

After the selection process, the variables SEX and MARRIAGE were removed. Finally, 22 effective features were kept for the model.

Next, the study performed WOE (Weight of Evidence) binning. After binning, the following calculation was performed for the i-th group:

$$WOE_i = \ln\left(\frac{Py_i}{Pn_i}\right) = \ln\left(\frac{y_i/y_T}{n_i/n_T}\right) \quad (1)$$

Table 3 presents a sample of the dataset after variable binning, showing the WOE-transformed values for selected features.

Table 3. Sample Data After Variable Binning

	bad_loan	PAY_5_woe	PAY_AMT6_woe	AGE_woe	PAY_6_woe	LIMIT_BAL_woe	BILL_AMT4_woe	PAY_2_woe	PAY_AMT1_woe
3	0	-0.244304	0.019319	-0.016469	-0.233450	0.197458	0.166510	-0.404764	-0.017578
7	0	-0.244304	0.019319	0.246734	-0.233450	0.197458	-0.010213	-0.404764	0.552881
13	1	-0.244304	0.273671	-0.160764	1.351641	0.197458	-0.104337	1.501138	-0.017578
16	1	1.481982	0.273671	0.246734	1.351641	0.676765	0.166510	-0.404764	-0.017578
17	0	-0.244304	-0.693093	0.172317	-0.233450	-0.338483	-0.104337	-0.404764	-0.485817

Finally, a personal credit scorecard model was built based on the Logistic Regression model. The dataset was divided into a training set (24,000 records) and a test set (6,000 records) using an 8:2 ratio. They study used Logistic Regression to build the model and then combined the trained regression coefficients with the WOE results to generate the final credit scorecard. Figure 1 illustrates the distribution of credit scores for both the training and test sets.

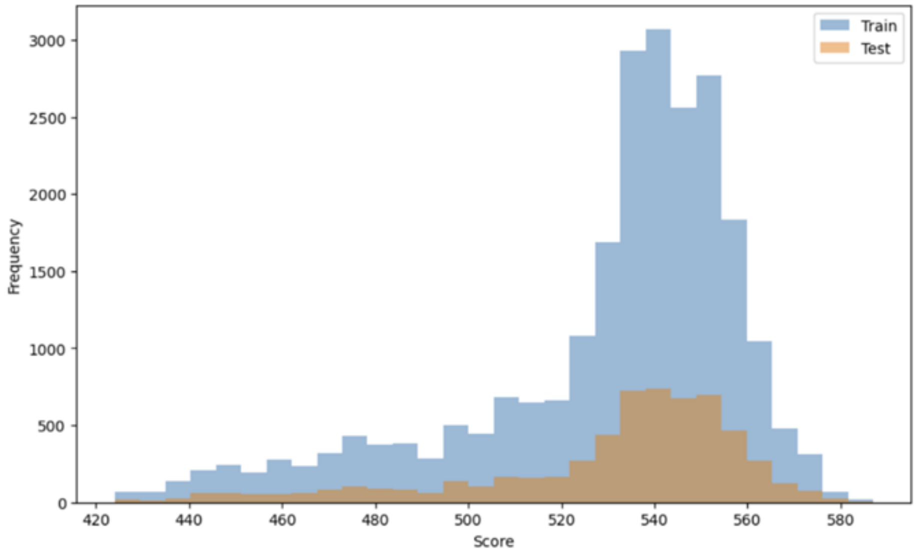

Figure 1. Credit Score Distribution

After building the credit scorecard model based on WOE encoding and Logistic Regression, the study standardized the output scores.

To solve this and keep the features within a unified range, the study used the Min-Max normalization method. The parameters for normalization were calculated using the training set, and then the same rules were applied to the test set to avoid information leakage.

The standardized credit scores were then used as static risk features in the next model. They provide overall information about a customer's credit background and offer useful context for modeling time-series behavior features.

5.2. Hybrid Modeling of Static Credit Scores and Time-Series Features

To fully explore the dynamic changes in customer credit risk, the study built behavior feature representations based on time series. The study organized the dynamic features into a specific shape of (30,000, 6, 3). This represents 30,000 customers, each with a 6-month time window and 3 types of behavioral features (PAY, BILL_AMT, and PAY_AMT).

For example, for the first customer in the dataset, the input sequence was structured as a nested list to represent the changes from 6 months ago to the most recent month:

[-2, 0, 0] (6 months ago)
[-2, 0, 0] (5 months ago)
[-1, 0, 0] (4 months ago)
[-1, 689, 0] (3 months ago)
[2, 3102, 689] (2 months ago)
[2, 3913, 0] (Most recent month)

In credit risk modeling, traditional scorecard models provide stable and easy-to-understand static risk assessments. Meanwhile, deep learning models based on time series are good at capturing dynamic changes in customer behavior. To combine the advantages of both, the study proposes a fusion method for static credit scores and time-series features. This approach keeps the interpretability of traditional scorecards while improving the ability to describe the risk evolution process. Table 4 illustrates the fused feature sequence for the first customer, where the static credit score remains constant across each time step, combined with dynamic features such as payment status (PAY), bill amounts (BILL), and payment amounts (AMT, PAY_AMT).

Table 4. Fused Feature Sequence for the First Customer

Month	PAY	BILL_AMT	PAY_AMT	Static_Score
T-6	-2.0	0.0	0.0	0.220859
T-5	-2.0	0.0	0.0	0.220859
T-4	-1.0	0.0	0.0	0.220859
T-3	-1.0	689.0	0.0	0.220859
T-2	2.0	3102.0	689.0	0.220859
T-1	2.0	3913.0	0.0	0.220859

Specifically, the study treats the calibrated default probability from the scorecard model as the customer's static risk profile. The static credit score was expanded across the time dimension to remain constant at each time step. In this way, the static score acts as stable background information in the time series, forming the model input together with the dynamic behavior features.

5.3. Construction of the Credit Risk Prediction Model Based on LSTM and Attention Mechanism

After completing the construction of time-series behavior features and the extraction of static credit scores, the study designed a deep learning model that integrates both types of information to improve the accuracy of default risk prediction. The model uses a dual-input structure to process the behavioral time-series features and the static credit scores from the scorecard separately, performing fusion modeling in a high-dimensional space.

First, for the six-month behavior sequence, the study used two LSTM layers for feature extraction. The first LSTM layer captures basic time dependencies, such as trends in bill amounts and fluctuations in repayment behavior. The second LSTM layer learns higher-level temporal dynamics to describe the evolution of the customer's credit risk. To prevent overfitting and improve the model's generalization ability, a Dropout mechanism was introduced between the two LSTM layers.

Building on the LSTM outputs, the study introduced an Attention Mechanism to aggregate the hidden states from different time steps through weighted calculations. This mechanism automatically identifies key periods that contribute significantly to default prediction, allowing the model to focus more on months with strong risk signals while reducing dependence on time steps with less information. Through attention weighting, a feature vector that comprehensively reflects the temporal risk of customer behavior is obtained. Meanwhile, the static credit score from the scorecard model, which represents the customer's long-term credit level, is fed into a fully connected layer to form a high-dimensional credit representation vector.

Finally, the behavioral time-series feature vector and the static credit feature vector are concatenated. Through this fusion method, the model can make a comprehensive judgment by combining the customer's behavioral patterns with their overall credit background.

5.4. Empirical Results

To verify the effectiveness of the proposed fusion model combining static credit scores and time-series behavior features, the study conducted comparative experiments. The performance of the traditional scorecard model, the LSTM-Attention model using only time-series features, and the proposed hybrid fusion model were evaluated separately. AUC (Area Under the Curve) was used as the primary metric to measure the overall performance of these models. Table 5 presents the comparative results of the three models across multiple evaluation metrics, including Accuracy, Precision, Recall, F1 Score, and AUC.

Table 5. Comparison of Prediction Performance Results

Prediction Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC
Credit Scorecard	53.48	27.07	65.11	38.24	0.7642
LSTM-Attention	77.83	49.90	54.33	52.02	0.7729
Scorecard- LSTM-Attention	81.17	54.31	49.43	56.86	0.7741

From the experimental results, it can be observed that the Scorecard- LSTM-Attention model outperforms individual models in terms of Accuracy, Precision, and F1-score. Specifically, the Accuracy reached 81.17% and the F1-score reached 56.86%. These results indicate that integrating static credit scores with dynamic behavior features can effectively enhance the ability to identify defaults.

Although the improvement in AUC is limited, the Scorecard- LSTM-Attention model demonstrates superior classification stability under the default threshold. This suggests that static and dynamic are complementary, providing a more balanced and reliable risk assessment than using a single data source alone.

Figure 2 illustrates the training dynamics of the Scorecard-LSTM-Attention model, showing the trends in loss and AUC for both the training and validation sets over 2,000 epochs.

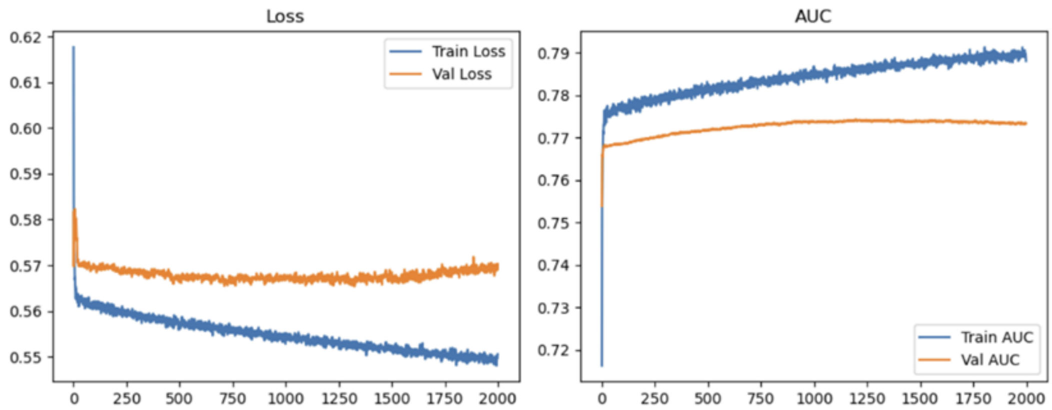

Figure 2. Training Results of the Scorecard-LSTM-Attention Model

6. Conclusion

The study proposes a hybrid model framework (Scorecard-LSTM-Attention) that integrates traditional scorecards with deep learning. In modeling dynamic customer behavior data, such as repayment records and credit utilization rates, the study introduces Long Short-Term Memory (LSTM) to capture time-dependent features. Simultaneously, the study utilizes an attention mechanism to enhance the model's ability to focus on key behavioral changes.

Furthermore, the study incorporates the scoring results generated by the scorecard as part of the neural network input. This approach effectively merges the interpretability of traditional models with the predictive power of deep learning models, providing a more forward-looking and robust solution for credit risk prediction in commercial banks. Through experimental validation, the study demonstrates that the model is expected to achieve a favorable balance among risk prediction accuracy, model stability, and practical business implementation.

References

[1] Peussa, A. (2016). Credit risk scorecard estimation by logistic regression [Unpublished manuscript]. University of Helsinki, Faculty of Science, Department of Mathematics and Statistics.

[2] Ma, H. (2023). Risk management of commercial banks under the background of big data. Market Modernization, (5), 95–97.

[3] Zhu, X., Cheng, D., Zhang, Z., Lin, S., & Dai, J. (2019). An empirical study of spatial attention mechanisms in deep networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6688–6697). IEEE. https://doi.org/10.1109/ICCV.2019.00678

[4] Dang, Y., Miao, Z. Q., Zhang, T., & Feng, D. F. (2022). Application progress of big data methods in systemic financial risk monitoring and early warning. Journal of Financial Development Research, 2(3).

[5] Liu, P. (2025). Application of LSTM-Attention hybrid model in financial risk prediction. Information & Computer, 37(13), 43–45.

[6] Feng, Y. C. (2025). Research on the application and optimization of big data-based risk management framework in financial institutions. In Proceedings of the Academic Seminar on Smart Building and Intelligent Economic Construction (I) (pp. 248–251).

[7] Cui, X. N. (2025, February 6). Application of machine learning algorithms in financial market risk analysis and prediction. Henan Economic News, p. 9.

[8] Xu, E. X. (2024). Research on personal credit risk based on random forest model. Audit and Wealth Management, (9), 55–58.

[9] Zhang, Z. T. (2024). Application research of gradient boosting method in credit risk assessment. Automation Application, 65(13), 26–31.

[10] Zhang, L., et al. (2023). Graph neural networks for systemic risk assessment in credit guarantee networks. Nature Computational Science, 3(1), 78–92. https://doi.org/10.1038/s43588-022-00386-7

[11] Kumar, A., et al. (2022). Dynamic graph risk: Real-time credit risk prediction using evolving social networks. In 2022 IEEE International Conference on Big Data (pp. 345–358). IEEE. https://doi.org/10.1109/BigData55660.2022.9975643

[12] Xu, H. Y. (2024). Research on credit risk prediction based on machine learning [Master's thesis]. Jilin University of Finance and Economics.

[13] Zhang, J. L., Guo, S. Y., Ren, C. P., & Ma, Q. (2024). Research on personal credit scoring card model based on logistic regression. Modern Information Technology, 8, 12–16.

[14] Dong, G., Lai, K. K., & Yen, J. (2010). Credit scorecard based on logistic regression with random coefficients. Procedia Computer Science, 1(1), 2463–2468. https://doi.org/10.1016/j.procs.2010.04.273

[15] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural Computation, 31(7), 1235–1270. https://doi.org/10.1162/neco_a_01199

[16] Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM—A tutorial into long short-term memory recurrent neural networks [Preprint]. arXiv. https://arxiv.org/abs/1909.09586

[17] Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. Neurocomputing, 452, 48–62. https://doi.org/10.1016/j.neucom.2021.04.069

[18] Brauwers, G., & Frasincar, F. (2023). A general survey on attention mechanisms in deep learning. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3279–3298. https://doi.org/10.1109/TKDE.2021.3067332