International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 OPEN ACCESS
Available Online at : www.ijsrcseit.com
doi : https://doi.org/10.32628/CSEIT25112761

Synergizing Generative AI and Machine Learning for Financial Credit Risk Forecasting and Code Auditing

Bhushan Chaudhari, Santhosh Chitraju Gopal Verma
Senior Tech Lead
Software Developer

ARTICLEINFO

Article History:

Accepted : 29 March 2025
Published: 01 April 2025

Publication Issue
Volume 11, Issue 2
March-April-2025

Page Number
2882-2893

ABSTRACT

Financial stability and efficiency of costs and time require the credit risk assessment process to evaluate models and comparisons while assessing future business impacts on the commercial banking sector. Accurate credit risk evaluation remains fundamental because financial institutions need it to prevent defaults while developing superior lending methods. A new AI framework based on Generative AI coupled with BERT technology presents itself for financial credit risk forecasting tasks. The model advances data representation by producing synthetic information and improves generalization through expert feature choice mechanisms while delivering fairness through automatic code evaluation systems. The Bank Credit Card dataset evaluation shows BERT surpasses conventional models to deliver 99.31% accuracy together with 99.61% precision 99.76% recall and 99.87% F1-score. BERT produces superior classification results than SVM and Decision Tree in addition to Logistic Regression as verified through comparative analysis. In order to better adapt to changing financial market conditions, future research will focus on creating hybrid models and real-time credit risk monitoring. The study's findings support the application of deep learning in financial risk management.

Keywords—Banking, Credit Risk Forecasting, Financial Risk Assessment, Generative AI, BERT, Bank Credit Card data.

Introduction

The banking system is a vital component that helps evaluate the correctness of datasets for loan applicant classification between good and bad credit risk categories. Loans possessed by applicants with good credit tend to guarantee repayment with high probability but bad credit applicants demonstrate minimal repayment potential as potential defaulters. Credit risk evaluation techniques are used to reduce the defaulter rate [1]. Small increases in credit evaluation accuracy lead to substantial financial profit reductions[2]. Strong credit risk datasets help organizations achieve better outcomes by lowering costs, boosting decisions speed and reducing loan collection perils.

The rapidly changing financial environment makes exact forecasting and thorough risk assessment increasingly vital because of its benefits in modern business operations. Standard financial analysis methods face difficulties when it comes to following the rising quantity of data and field complexity[3][4]. Predictive modeling finds its essential practical application in credit risk analysis to predict whether financial institutions will earn or lose money through their applicant loan approvals. Credit becomes available to different entities like people, businesses, and nonprofits for equipment investments home buying and consumer loans[5]. Financial institutions offer three types of credit services: credit cards and loans and payment delays. Financial institutions give credit for anticipated timely payments and interest payments that establish their rates through risk assessment of default.

The importance of credit risk prediction as a financial analytic topic has significantly increased during the previous decades. Credit card default prediction is one of the critical problems that creditors must solve effectively[6]. Before the advent of technology, credit risk modeling used statistical analysis and expert assessment methods. Modern risk assessment methods have been made possible by the significant technological advancements of AI and ML[7][8] since their appearance[9]. The large mismatch in credit risk data sets' default and non-default transaction counts leads machine learning models to show systematic biases when making predictive classifications.

A. Motivation and Contribution

Financial institutions require credit risk assessment to successfully set up long-term lending operations and decrease defaults. Despite relying on rule-based systems and human-made characteristics, traditional risk assessment approaches fail to spot the intricate patterns in large credit datasets. Credit risk forecasts become problematic when unbalanced datasets are present because they lead to biased classification methods. By efficiently comprehending the contextual links in financial data, transformers adopting BERT topologies show great promise for this type of analysis. A research project has been developed to boost credit risk prediction by implementing a BERT-based classification system while improving dataset balancing and feature selection to advance financial institution decision processes. The main contributions are:

- Using the Bank credit card dataset for credit risk prediction.
- Implement SMOTE for class imbalance handling, improving model performance in imbalanced datasets.
- Application of the Chi-square test to improve model interpretability and performance.
- Propose a BERT-based classification model for credit risk prediction, leveraging contextual token representations.
- To ensure a thorough evaluation, Metrics including accuracy, precision, recall, F1-score, and ROC-AUC are used to assess the model's performance.

B. Justification and Novelty

The justification for this research lies in the increasing need for accurate, efficient, and interpretable credit risk forecasting models that can address the limitations of traditional statistical methods. This study proposes a novel AI-driven credit risk assessment framework by integrating Generative AI with Machine Learning, leveraging BERT for superior structured data representation. Unlike traditional models, our approach addresses class imbalances, dynamic risk factors, and complex data structures through synthetic data generation, advanced feature selection, and automated code auditing, ensuring improved generalization, fairness, and transparency.

C. Structure of the paper

The Study of the Structured is as follows: Section II discusses pertinent research on financial credit risk forecasting, while Section III lays out the methods and resources utilized. Here in Section IV, we offer the experimental results of the model. Section V presents a summary of the investigation's findings and serves as its conclusion.

Literature Review

This section discusses some review articles on Machine Learning for Financial Credit Risk Forecasting. Table I highlights the paper, methods, dataset, key findings, and limitations/future work.

Bhandari et al. (2024) identify such risks and prevent them beforehand. The ANN model outperforms other models in terms of accuracy and the ROC curve, achieving an astounding accuracy of 85.88%. The XG-Boost confusion matrix is second to the Random Forest model regarding balance. Synthetic data points were generated to balance the imbalanced dataset, and feature engineering techniques is used to eliminate less relevant features[10].

Su et al. (2024) random forest algorithm is used as the main modeling and prediction tool. A significant quantity of credit-related data is gathered, and the data is preprocessed and feature-engineered, which includes data standardization, feature selection, and missing value processing. Analyse and contrast the model's performance. The accuracy percentage for passing the test ranges from 90% to 98%[11].

Yemmanuru, Yeboah and Nti. (2024) evaluated using various performance metrics. The best accuracy (80.67%) and strong recall (93.55%) were attained using Support Vector Machines (SVM). Additionally, SVM showed the greatest accuracy (83.57%), F1 score (84.70%), recall (88.84%), and ROC AUC score (83.44%) when dimensions were decreased to 10 (from 20) using Principal Component Analysis. The open-source Extreme Gradient Boosting technique produced synthetic records using SMOTE and obtained the best accuracy score of 83.3% with a ROC AUC of 83.29%[12].

Tang et al. (2023) The model gathers the pertinent financial market data, creates the deep learning network structure based on the training samples, uses the ANN-based forecast model to train the sample set, produces the best answer, and finishes the financial risk forecast. Simulation results show that after several rounds, this method's accuracy is unquestionably better than that of the traditional SVM algorithm. Compared to the SVM approach, the accuracy was 20.55% higher at 96.84%, and the error was reduced by 32.84%. Thus, an appropriate and practical evaluation model is the ANN-based financial market forecasting model[13].

Wanjale et al. (2023) give a summary of financial risk prediction while highlighting the advantages and disadvantages of various approaches. The research covers a range of machine learning methods, such as logistic regression, naive Bayes, decision trees, random forest, XG Boost, and KNN. Many types of consumer data, including income, credit amount, and payment history, are used in financial risk prediction[14].

TABLE I. SUMMARY OF BACKGROUND STUDY FOR FINANCIAL CREDIT RISK FORECASTING USING MACHINE LEARNING TECHNIQUES

Reference	Methods Used	Dataset	Key Findings	Limitations & Future Work
Bhandari et al., (2024)	ANN, Random	Credit-related data	ANN achieved the highest accuracy (85.88%).	Future research might investigate additional deep

Reference	Methods Used	Dataset	Key Findings	Limitations & Future Work
	Forest, XGBoost	(imbalanced, balanced using SMOTE)	Random Forest had the most balanced confusion matrix. XGBoost also performed well.	learning models and feature engineering strategies to improve accuracy.
Su et al., (2024)	Random Forest	Large-scale credit data	Random Forest accuracy ranged from 90% to 98%. Feature engineering and data preprocessing improved performance.	Future work may include testing deep learning models and optimizing hyperparameters further.
Yemmanuru, Yeboah and Nti, (2024)	SVM, XGBoost, PCA, SMOTE	Credit-related dataset	SVM had the highest accuracy (80.67%) and recall (93.55%). PCA improved accuracy to 83.57%. XGBoost with SMOTE achieved 83.3% accuracy.	Further improvements could focus on hybrid models combining SVM with deep learning.
Tang et al., (2023)	ANN, SVM	Financial market data	ANN outperformed SVM with 96.84% accuracy, reducing error by 32.84%.	Future work may refine ANN architectures and introduce reinforcement learning for better predictions.
Wanjale et al., (2023)	XGBoost, KNN, Random Forest, Decision Trees, Naïve Bayes, and Logistic Regression	Consumer financial data (income, credit amount, payment history)	Evaluation of machine learning models for predicting financial risk in comparison.	Deep learning models and ensemble learning strategies can be investigated in future studies.

Methodology

The research methodology for financial credit risk forecasting involves several steps shown in Figure 1. The study utilizes a Kaggle bank credit card dataset comprising applications linked via a unique ID. Using min-max normalization to scale numerical features, removing outliers, and addressing missing values via mode imputation are all examples of data preparation. Feature selection is performed using the Chi-square test to retain relevant predictors. Class imbalance is addressed by using the SMOTE. Thirty percent is used for testing, while seventy percent is for training. A BERT-based classification model is implemented, leveraging token representation, self-attention, multi-head attention, and feedforward networks. The final categorization layer predicts credit risk. The model is evaluated using ROC-AUC, F1-score, accuracy, precision, and recall. The findings show that the suggested strategy successfully improves credit risk prediction by reducing class imbalance, improving feature selection, and aiding in better financial decision-making.

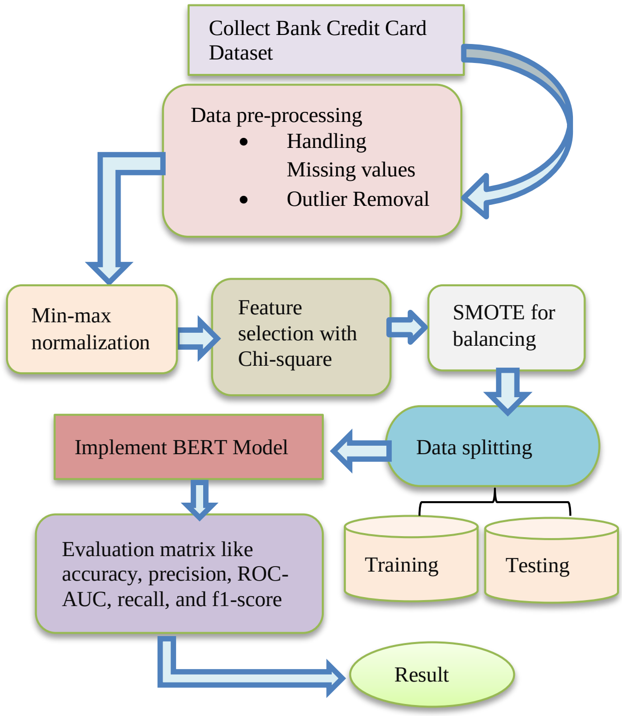

Figure 1. Flowchart for Financial Credit Risk Forecasting

The steps in the suggested technique are described in short below:

A. Data collection

The study is predicated on a Kaggle dataset. This dataset consists of real bank credit card information shown online after removing sensitive client data. There are two different data files: application_record.csv and credit_record.csv. The applicants' data, which might be utilized as predictive features, is included in the initial application record dataset. The second one, the credit record dataset, records customers' credit history (or credit card usage patterns). The ID is the main key, or linking column, between the application and credit record databases. The visualization of data is provided in below:

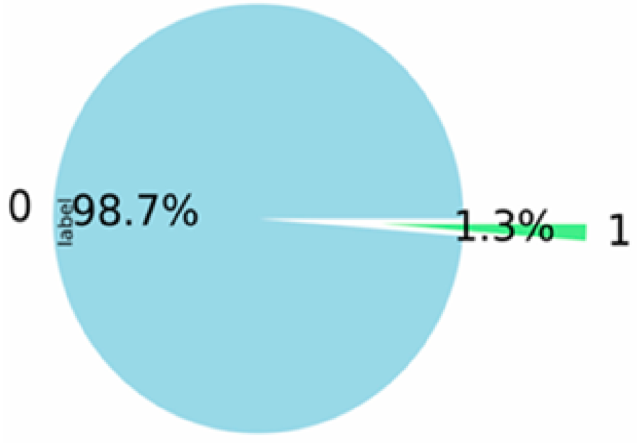

Figure 2. Pie chart for Target distribution

The results are remarkably unbalanced, as seen in Figure 2, with a rate of 98.7 for excellent customers and 1.3% for negative consumers. 0 and 1 for each label.

Figure 3. Distribution of important variables

The distribution of age and job experience is shown in Figure 3, and income concerning credit status. Lower-risk individuals (label 0) tend to be older and have longer employment durations, while higher-risk individuals (label 1) exhibit shorter work experience with more outliers. Income distribution remains similar across both groups, with lower-risk applicants showing a slightly higher median income. These patterns highlight key demographic and financial factors influencing credit risk assessment.

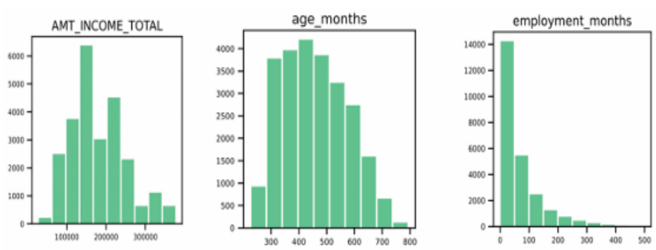

Figure 4. Histogram of features

Figure 4 illustrates the histograms of key features in the bank credit card dataset, which includes age months, employment months, and AMT_INCOME_TOTAL. AMT_INCOME_TOTAL's distribution shows a right-skewed trend, suggesting that there are more applicants with lower income levels. The age months histogram follows a near-normal distribution, with most individuals aged between 300 and 600 months. Conversely, employment months demonstrates a highly right-skewed distribution, suggesting that most applicants have relatively short employment durations. These distributions provide insights into credit card applicants' demographic and financial characteristics, essential for credit risk assessment.

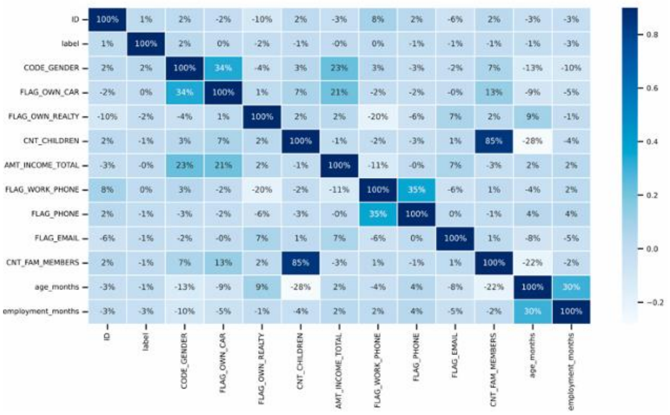

Figure 5. Correlation matrix

The correlations between different variables in the bank credit card dataset are visually shown by the correlation heatmap in Figure 5. Darker hues indicate greater positive or negative associations. The color's intensity indicates the correlation's strength and direction. It is noteworthy that although "FLAG_PHONE" and "FLAG_WORK_PHONE" show a stronger link, characteristics like "FLAG_OWN_CAR" and "FLAG_OWN_REALTY" show a modest correlation. The "employment months" variable also shows a notable correlation with "age," suggesting a dependency between work experience and age. This analysis aids in understanding feature interactions and their potential impact on credit risk prediction.

B. Data preprocessing

Pre-processing the data is necessary to get high-quality results from the knowledge discovery method in question. Most of the next procedures are completed in several iterations to get the desired outcomes. The first phases in the preprocessing process are as follows:

- **Handling Missing values:** Value mode replaces null values in numerical columns.There were no duplicates due to the controlling strategies used by both central and commercial banks.
- **Outlier Removal:** Outliers can have an impact on the research's quality since they can have an impact on all statistical data. In this case, eliminate the outliers from the dataset.

C. Min-max normalization

The Bank credit dataset's numerical column values were converted to a typical 0–1 scale using the min-max normalization method, which prevented the value ranges from being distorted. Using Equation (1).

$$Y = a - \frac{\min(a)}{\max(a)} - \min(a) \tag{1}$$

where stands for the original value and Y for the normalized value.

D. Feature selection with Chi-square

In feature selection, extraneous predictor variables are eliminated or changed to make the final dataset more attributed to the issue description and modeling method[15]. The Chi-square statistics forms part of the filter methods and includes calculating the degree to which one category variable is independent of another. The Chi-square statistic is given as equation (2).

$$x^2 \text{parent} \tag{2}$$

where m is the number of levels or groups, k is the number of classes, A_{ij} is the number of observations in the time period between class I and class J, and E_{ij} is the expected frequency of A_{ij}.

E. SMOTE for Balancing

A dataset with an unequal distribution of observations for the predicted variable is considered unbalanced in classification. Getting beyond the unbalanced dataset problem, which may have detrimental impacts on performance, the SMOTE was employed to see if it produced the superior outcomes displayed below:

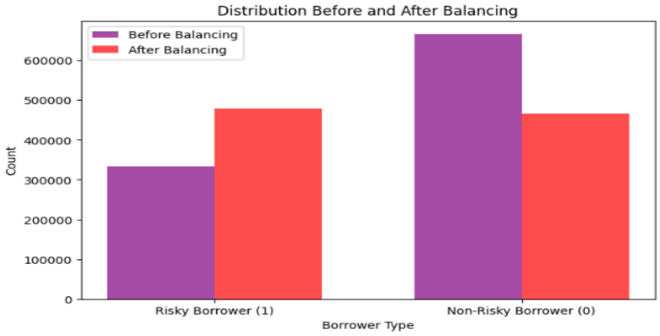

Figure 6. Plot for Data balancing

Figure 6 presents a comparative bar chart illustrating the distribution of risky and non-risky borrowers before and after data balancing. The "Before Balancing" chart on the left shows an imbalanced dataset with 334,330 risky borrowers and 665,670 non-risky borrowers. On the right, the "After Balancing" chart depicts a more balanced dataset, where the number of risky borrowers has increased to 478,050, while non-risky borrowers have been adjusted to 466,050. This balancing ensures a more equitable distribution, reducing model bias and improving classification performance.

F. Data splitting

The dataset utilized in this study was split into two sections: 70% training and 30% testing (70:30)

G. Proposed BERT model

The BERT model is among the most cutting-edge tools available for natural language processing[16][17]. The Transformer architecture is the foundation for the BERT paradigm, which processes text in both directions via self-attention techniques[18][19]. The fundamental components of BERT can be represented mathematically as follows:

Token Representation in BERT: BERT processes text inputs by first converting words into numerical representations[20]. Each input token x_i is transformed into a token embedding using Equation (3):

$$T_i = E_{(x_i)} + P_i \tag{3}$$

where: T_i is the token representation, $E_{(x_i)}$ is the word embedding, P_i is the positional encoding to retain word order information.

Self-Attention Mechanism: Self-attention is one of BERT's main characteristics, which enables words to attend to other words in a sentence, regardless of their position. The self-attention mechanism is defined as Equation (4):

$$\text{Attention}\,(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

Where Soft max makes sure that attention scores add up to 1, which indicates the weight of each word in the context; d sub k is the dimensionality of the key vectors; and Q, K, and V are query, key, and value matrices obtained from input tokens.

Multi-Head Attention Mechanism Instead of using a single attention mechanism, Because BERT uses multi-head self-attention, the model may concurrently collect numerous context representations Equation (5):

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{Head}_1, \text{Head}_2, \text{Head}_h)W^o \tag{5}$$

where: h is the number of attention heads, W^o is an output transformation matrix.

Transformer Feedforward Network: Each Transformer block in BERT includes a feedforward network (FFN) that refines the token representations Equation (6):

$$\text{FNN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{6}$$

Where: $W_1 W_2$ are weight matrices, $b_1 b_2$ are biases, ReLU activation ($\max(0, x)$) introduces non-linearity.

Output Layer for Classification: BERT uses the hidden state of the [CLS] token for classification tasks, processed through a SoftMax layer to generate class probabilities Equation (7):

$$\text{CapP}(y|(X) = \text{softmax}(W_o h_{CLS} + b_o \tag{7}$$

Where: h_{CLS} is the [CLS] token's hidden state, WW_o b_o are trainable parameters. The final prediction determines whether a customer is at low or high risk for credit default.

H. Evaluation metrics

Various metrics are available to gauge a model's efficacy and the caliber of its predictions. According to true positive, true negative, false positive, and false negative classifications, measures such as accuracy, precision, recall, and F1 score are offered. ROC-AUC curves are also provided as an alternative way to evaluate the model's efficacy. The equations of metrics, as shown in (8) to (12), are based on the fundamental measuring parameters of the confusion matrix are listed in below:

- **True positive (TP):** The likelihood of being a good customer is precisely anticipated.
- **False positive (FP):** When negative clients are mistakenly assumed to be positive ones.
- **True negative (TN):** When it's possible to forecast which consumers will be nasty.
- **False negative (FN):** The misprediction of excellent customers as negative customers.
- **Accuracy:** The easiest metric to use is accuracy. It is computed as the number of correct forecasts divided by the total number of events. The formula Equation (8):

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \tag{8}$$

Precision: It is computed by taking the number of classifier-predicted positive results and dividing it by the number of accurate positive outcomes. It is stated as. It is given by Equation (9).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{9}$$

Recall: The TPR for the error estimation is determined by calculation. It shows how well the test can determine the rate. We computed sensitivity using the confusion matrix. It is sometimes referred to as recall or TPR. It is given by Equation (10):

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{10}$$

F1-score: It is considered as the number of frauds that get predicted into the actual total number of cases as described in equation (11):

$$\text{F1} - \text{Score} = \frac{2(\text{Precision}*\text{Recall})}{\text{Precision+Recall}} \tag{11}$$

ROC Curve and AUC Score: The AUC score and ROC curve are two of the most important evaluation tools for assessing how well classification models function. ROC is a probability curve, whereas AUC is a measure of separability. It illustrates how well the model differentiates across various classifications. The better the model predicts that bad credit card users will be bad and good credit card users will be good, the higher the AUC., it is given as Equation (12)

$$\text{AUC} = \frac{1+\frac{\text{TP}}{\text{TP+FN}}-\frac{\text{FP}}{\text{FP+TN}}}{2} \tag{12}$$

These matrices are used to identify the deep learning and machine learning models.

Result Analysis and Discussion

To evaluate the ML and DL algorithms' prediction ability for credit risk classification, this part starts with the experiment design. The simulations were run on a machine with an Intel® Core TM i5-1035G1 CPU running at 1.19 GHz and 8 GB of RAM using the Jupiter Lab® software, version 3.2.1, and the Python programming language, version 3.9.13 The performance is evaluated using the following metrics: AUC, f1-score, recall, accuracy, and precision. The proposed Generative AI Model performance for Financial Credit Risk Forecasting is shown below.

TABLE II. BERT MODEL ON BANK CREDIT CARD DATASET FOR FINANCIAL CREDIT RISK FORECASTING

Performance Measures	Bidirectional Encoder Representations from Transformers (BERT)
Accuracy	99.31
Precision	99.61
Recall	99.76
F1-score	99.87
ROC-AUC	0.99

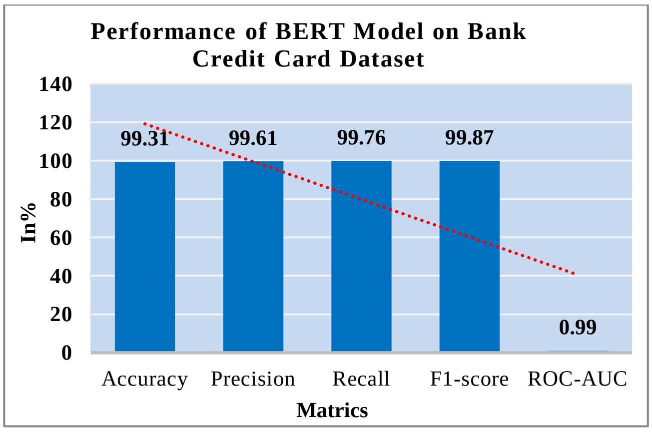

Figure 7. Bar Graph for BERT Model Performance

Table II and Figure 7 above show the BERT model's performance for financial credit risk forecasting. The model demonstrates exceptional predictive capability, achieving an accuracy of 99.31%, indicating a high proportion of correctly classified instances. The model's 99.61% accuracy indicates that it can reduce false positives, and its 99.76% recall indicates that it successfully detects true positive instances. The 99.87% F1 score shows a good mix between recall and accuracy. The ROC-AUC score of 0.99 further validates the model's exceptional capacity to differentiate between creditworthy clients and those who are not.

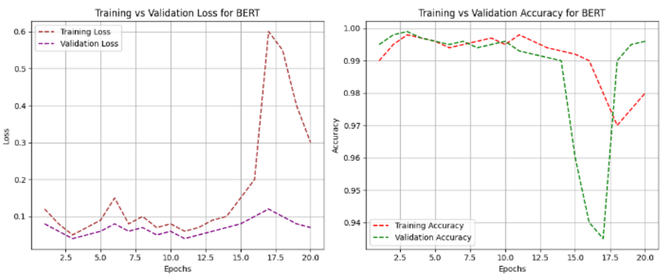

Figure 8. Accuracy and Loss plot for BERT Model

A BERT model's training and validation results across 20 epochs are shown in Figure 8. The left plot shows loss, where training loss (red dashed) remains stable around 0.1 until epoch 15, spikes to 0.6 at epoch 17, and then declines. Validation loss (purple dashed) follows a steadier trend with slight increases. The right plot tracks accuracy, with training (red dashed) and validation (green dashed) accuracy staying above 0.99, except for a sharp validation drop to 0.935 at epoch 17 before recovery. This suggests temporary instability or overfitting, but the model stabilizes by epoch 20.

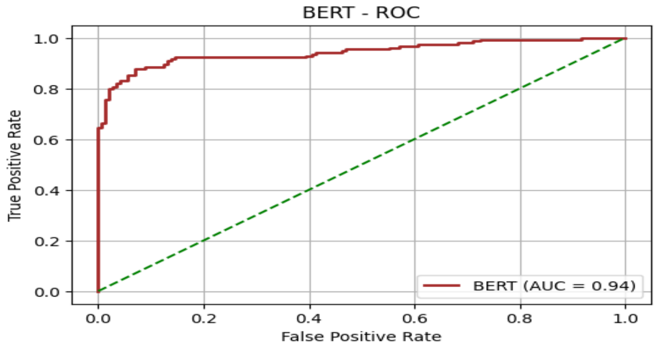

Figure 9. ROC Curve for BERT model

The BERT model's ROC curve, which shows how well it distinguishes between classes, is shown in Figure 9. The x-axis displays the FPR, while the y-axis displays the TPR. The dashed green diagonal line is the random classifier baseline, while the solid red curve is the BERT model. The model's great classification performance is demonstrated by its AUC value of 0.94, demonstrating its ability to correctly differentiate between favorable and unfavorable circumstances.

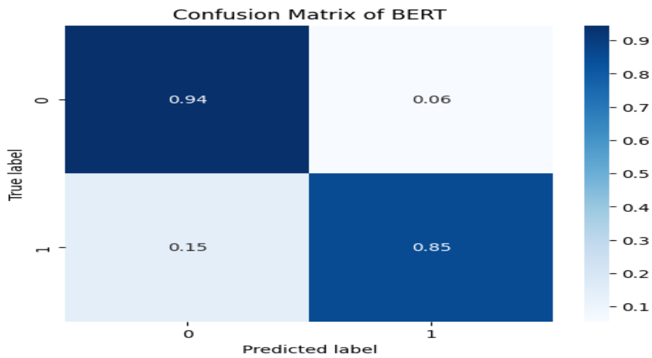

Figure 10. Confusion Matrix of BERT model

The BERT classifier's confusion matrix is shown in Figure 10. The true negative rate (1.00) is at the top-left of the matrix, followed by the false positive rate (0.00) at the top-right, the false negative rate (0.01) at the bottom-left, and the true positive rate (0.99) at the bottom-right. These numbers show that the model has a high classification accuracy and a low misclassification rate.

	precision	recall	f1-score	support
0	0.990	0.997	0.993	7422
1	0.997	0.990	0.993	7406
accuracy			0.993	14828
macro avg	0.993	0.993	0.993	14828
weighted avg	0.993	0.993	0.993	14828

Figure 11. Classification Report of BERT model

Figure 11 shows the BERT categorization report using bank credit card data. According to the report, the model achieves 99% precision, 99% recall, 99% f1-score, and Support 7422 for class 0, whereas class 1 achieves 99% precision, 99% recall, 99% f1-score, and Support 7406.

A. Comparative Analysis and Discussion

The comparative analysis for Financial Credit Risk Forecasting in this section contains the dataset for bank credit cards. Table III compares the proposed BERT model to SVM, DT, and LR concerning several performance metrics. These metrics include recall, accuracy, precision, f1-score, and AUC-ROC.

TABLE III. ML AND DL MODELS PERFORMANCE COMPARISON ON THE BANK CREDIT CARD DATASET FOR FINANCIAL CREDIT RISK FORECASTING

Measures	SVM[21]	LR[22]	DT[23]	BERT
Accuracy	90.60	78	81	99.31
Precision	92.85	85	80	99.61
Recall	96.35	82	81	99.76
F1-score	94.57	84	80	99.87

Table III compares the performance of various models, including SVM, LR, DT, and BERT, on the Bank Credit Card dataset for financial credit risk forecasting. With a recall of 99.76%, an F1 score of 99.87%, and an accuracy of 99.31% (matching precision of 99.61%), BERT outperforms all conventional ML models. A strong performance emerges from SVM since it achieved 90.60% accuracy and a 94.57% F1-score, demonstrating its ability to excel in classification applications. The prediction abilities of LR and DT remain restricted because they achieve accuracies of 78% and 81% and F1-scores of 84% and 80%. The experimental findings prove that BERT achieves better financial credit risk prediction results because it understands financial data structures with complex patterns better than common machine learning models.

The financial credit risk forecasting approach built upon BERT yields better accuracy levels, improved feature representations, and enhanced generalization capability. Within its system the BERT-based framework controls imbalanced classes effectively while reducing false detection and provides automatic code evaluation for maintaining fairness standards. Financial institutions can rely on this framework as a superior ML model that delivers high-performance functionality along with scalability capabilities.

Conclusion and Future Work

The banking sector experiences a transformation through GenAI because it generates operational efficiency, enriches customer interactions, and drives industry advancements. The financial sector is significantly improved by using AI and ML for credit risk modeling. Credit risk management is essential to banking operations because poor risk management can harm all other banking-related industries. This paper shows how the BERT model outperforms conventional machine learning techniques regarding predictive power by presenting an AI-driven financial credit risk forecasting framework. The proposed model surpasses SVM, LR, and DT with impressive results: 99.31% accuracy, 99.61% precision, 99.76% recall, and 99.87% F1 score. The ROC-AUC value of 0.99 further validates its robust classification ability, effectively distinguishing between creditworthy and non-creditworthy customers. BERT's deep contextual understanding enhances risk assessment accuracy while reducing misclassification. Despite minor training fluctuations, it ensures stability and generalization, making it reliable for financial institutions. Future work may explore hybrid models and real-time credit risk monitoring for improved adaptability.

References

[1]. T. N. Pandey, A. K. Jagadev, S. K. Mohapatra, and S. Dehuri, "Credit risk analysis using machine learning classifiers," in 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017, 2018. doi: 10.1109/ICECDS.2017.8389769.

[2]. S. Tyagi, "Analyzing Machine Learning Models for Credit Scoring with Explainable AI and Optimizing Investment Decisions," vol. 5, no. 01, pp. 5–19, 2022.

[3]. K. Hussain, "Harnessing AI for Predictive Accuracy in Financial Forecasting and Risk Assessment," no. November 2023, 2024, doi: 10.13140/RG.2.2.17070.73281.

[4]. A. Immadisetty, "Real-Time Fraud Detection Using Streaming Data in Financial Transactions," J. Recent TRENDS Comput. Sci. Eng., vol. 13, no. 1, pp. 66–76, Feb. 2025, doi: 10.70589/JRTCSE.2025.13.1.9.

[5]. I. Emmanuel, Y. Sun, and Z. Wang, "A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method," J. Big Data, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00882-0.

[6]. S. Tyagi, T. Jindal, S. H. Krishna, S. M. Hassen, S. K. Shukla, and C. Kaur, "Comparative Analysis of Artificial Intelligence and its Powered Technologies Applications in the Finance Sector," in Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022, 2022. doi: 10.1109/IC3I56241.2022.10073077.

[7]. S. Moolchandani, "Advancements in Credit Risk Modelling: Exploring the role of AI and Machine Learning," Int. J. Manag. IT Eng., vol. 13, p. 7, 2023.

[8]. Laith Haleem Al-Hchemi, "Evaluating Generative AI in enhancing banking services efficiency," Econ. FORUM, vol. 14, no. 4, pp. 47–54, 2024.

[9]. R. Dattangire, R. Vaidya, D. Biradar, and A. Joon, "Exploring the Tangible Impact of Artificial Intelligence and Machine Learning: Bridging the Gap between Hype and Reality," 2024 1st Int. Conf. Adv. Comput. Emerg. Technol. ACET 2024, pp. 2024–2025, 2024, doi: 10.1109/ACET61898.2024.10730334.

[10]. T. Bhandari, P. Payne, M. K. Gourisaria, S. Shekhar Patra, and V. Petli, "Guardians of Credit: 'A Comprehensive Analysis of Machine Learning Models for Credit Risk Detection,'" in 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS), 2024, pp. 1–6. doi: 10.1109/ICICACS60521.2024.10498659.

[11]. Y. Su, "Financial Technology Credit Risk Modeling and Prediction based on Random Forest Algorithm," in 2024 Second International Conference on Data Science and Information System (ICDSIS), 2024, pp. 1–4. doi: 10.1109/ICDSIS61070.2024.10594215.

[12]. P. K. Yemmanuru, J. Yeboah, and I. K. Nti, "Customer Credit Risk: Application and Evaluation of Machine Learning and Deep Learning Models," in 2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI), IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/ICMI60790.2024.10585896.

[13]. D. Tang, "Optimization of Financial Market Forecasting Model Based on Machine Learning Algorithm," in Proceedings - 2023 International Conference on Networking, Informatics and Computing, ICNETIC 2023, 2023. doi: 10.1109/ICNETIC59568.2023.00104.

[14]. K. Wanjale, S. Saraf, K. Thakre, A. Chitre, and R. Mahajan, "Financial Risk Prediction Using Consumer Information and Different Machine Learning Algorithms: A Comparative Analysis," in 2023 3rd Asian Conference on Innovation in Technology, ASIANCON 2023, 2023. doi: 10.1109/ASIANCON58793.2023.10270196.

[15]. S. Murri, "From Raw to Refined: The Art and Science of Data Engineering," p. 286, 2025.

[16]. Y. H. Rajarshi Tarafdar, "Finding majority for integer elements," J. Comput. Sci. Coll., vol. 33, no. 5, pp. 187–191, 2018.

[17]. S. Pahune and N. Rewatkar, "Healthcare: A Growing Role for Large Language Models and Generative AI," Int. J. Res. Appl. Sci. Eng. Technol., vol. 11, no. 8, pp. 2288–2301, Aug. 2023, doi: 10.22214/ijraset.2023.55573.

[18]. J. A. Benítez-Andrades, J. M. Alija-Perez, M. E. Vidal, R. Pastor-Vargas, and M. T. García-Ordas, "Traditional Machine Learning Models and Bidirectional Encoder Representations From Transformer (BERT)-Based Automatic Classification of Tweets About Eating Disorders: Algorithm Development and Validation Study," JMIR Med. Informatics, 2022, doi: 10.2196/34492.

[19]. J. Q. Gandhi Krishna, "Implementation Problems Facing Network Function Virtualization and Solutions," IARIA, pp. 70–76, 2018.

[20]. S. Nokhwal, S. Nokhwal, S. Pahune, and A. Chaudhary, "Quantum Generative Adversarial Networks: Bridging Classical and Quantum Realms," in 2024 8th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI), New York, NY, USA: ACM, Apr. 2024, pp. 105–109. doi: 10.1145/3665065.3665082.

[21]. F. Doko, S. Kalajdziski, and I. Mishkovski, "Credit Risk Model Based on Central Bank Credit Registry Data," J. Risk Financ. Manag., 2021, doi: 10.3390/jrfm14030138.

[22]. V. Chang, Q. A. Xu, S. H. Akinloye, V. Benson, and K. Hall, "Prediction of bank credit worthiness through credit risk analysis: an explainable machine learning study," Ann. Oper. Res., 2024, doi: 10.1007/s10479-024-06134-x.

[23]. M. B. Seitshiro and S. Govender, "Credit risk prediction with and without weights of evidence using quantitative learning models," Cogent Econ. Financ., vol. 12, no. 1, p., 2024, doi: 10.1080/23322039.2024.2338971.