

# Deep Learning-Based Credit Scoring and Risk Assessment System

K. Krupa sagari<sup>1</sup>, K. Pavani<sup>2</sup>, P. Gandhi<sup>3</sup>

#1 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

**Abstract:** Credit scoring prediction is a critical task in the financial sector, enabling institutions to evaluate borrower creditworthiness and manage financial risk effectively. Traditional machine learning techniques often struggle to capture temporal dependencies and complex nonlinear relationships present in modern financial and behavioral datasets. To address these limitations, this project proposes an enhanced credit scoring prediction framework based on a Hybrid Long Short-Term Memory (LSTM) network integrated with a self-attention mechanism.

The proposed system utilizes advanced preprocessing techniques, including Min-Max normalization, Synthetic Minority Oversampling Technique (SMOTE), Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA), to improve data quality, handle class imbalance, and reduce dimensionality. The Hybrid LSTM model captures temporal patterns in sequential financial data, while the self-attention layer selectively focuses on the most influential features, improving contextual understanding and predictive performance.

For comparative analysis, traditional machine learning algorithms such as Random Forest and XGBoost are also implemented as baseline models. The enhanced Hybrid LSTM with Attention model is deployed through a Flask-based web application that enables real-time credit score prediction by processing uploaded financial data. Experimental evaluation on the Credit Risk dataset demonstrates that the proposed model achieves an accuracy of 89.87%, outperforming conventional machine learning and standard deep learning approaches. The developed framework provides an accurate, scalable, and interpretable solution for modern credit risk assessment, supporting financial institutions in making reliable lending decisions and improving overall risk management...

**Index terms** - Credit Scoring Prediction, Deep Learning, Hybrid LSTM, Self-Attention Mechanism, Credit Risk Assessment, Financial Risk Management, Machine Learning, Random Forest, XGBoost, Feature Selection, Principal Component Analysis (PCA), Synthetic Minority Oversampling Technique (SMOTE), Creditworthiness Evaluation, Predictive Analytics, Financial Data Mining, Explainable AI (XAI), Flask Web Application, Classification Models, Loan Approval Prediction, Artificial Intelligence in Finance.

## 1. INTRODUCTION

Credit scoring prediction is one of the most important processes in the financial sector, as it assists banks, lending institutions, and financial organizations in evaluating the creditworthiness of borrowers. Accurate credit scoring helps financial institutions make informed decisions regarding loan approvals, credit limits, interest rates, and risk management strategies. Traditional credit scoring systems primarily rely on predefined rules, statistical methods, and historical financial records to assess borrower risk. Although these approaches have been widely adopted, they often struggle to capture the complex and dynamic nature of modern financial behavior.

The rapid growth of digital banking, online transactions, and financial technology (FinTech) platforms has resulted in the generation of large volumes of structured and unstructured financial data. These datasets contain valuable information such as transaction histories, spending patterns, repayment behavior, and customer interactions. Traditional machine learning techniques such as Random Forest and XGBoost can process structured financial information effectively; however, they have limitations in capturing temporal dependencies and evolving behavioral patterns that influence credit risk over time.

Recent advances in Artificial Intelligence (AI) and Deep Learning have provided powerful tools for analyzing complex financial datasets. Long Short-Term Memory (LSTM) networks, a specialized type of Recurrent Neural Network (RNN), are particularly effective in modeling sequential and time-dependent data. LSTM models can learn long-term

dependencies and hidden patterns within financial records, making them suitable for credit scoring prediction. However, standard LSTM models may not always focus on the most relevant information when processing lengthy financial sequences.

To overcome these challenges, this project proposes an enhanced Hybrid LSTM model integrated with a Self-Attention mechanism for credit scoring prediction. The attention layer enables the model to identify and prioritize the most important financial features and temporal patterns, improving contextual understanding and prediction accuracy. The framework also incorporates advanced preprocessing techniques such as Min-Max Normalization, Synthetic Minority Oversampling Technique (SMOTE), Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA) to improve data quality, handle class imbalance, and reduce dimensionality.

The proposed system is implemented using Python and deployed through a Flask-based web application that allows users to upload financial data and obtain real-time credit score predictions. The model classifies borrowers into Good Credit or Bad Credit categories, assisting financial institutions in making reliable lending decisions. Experimental evaluation demonstrates that the Hybrid LSTM with Self-Attention outperforms traditional machine learning and deep learning models, achieving an accuracy of 89.87%. The proposed framework therefore provides a scalable, accurate, and intelligent solution for modern credit risk assessment and financial decision-making.

## 2. LITERATURE SURVEY

### *3.1 Integration of Convolutional Neural Network Models and Machine Learning Methods in Credit Score Classification*

Accurate credit score classification is essential for financial institutions to evaluate customer creditworthiness and manage financial risks effectively. Traditional classification techniques often experience limitations when handling large-scale datasets, resulting in lower prediction accuracy and increased processing time. To address these challenges, an advanced framework combining Convolutional Neural Network models with Machine Learning algorithms was introduced for credit score classification.

In this approach, one-dimensional sequential financial data is transformed into two-dimensional grayscale images to enable efficient feature extraction using Convolutional Neural Network architectures. Multiple deep learning models were implemented to learn complex patterns from the transformed data and improve classification performance. The extracted features from these models were further classified using Machine Learning algorithms such as K-Nearest Neighbor, Linear Discriminant Analysis, Naive Bayes, and Support Vector Machine.

Performance evaluation was carried out using metrics including accuracy, precision, sensitivity, specificity, and F-score. Feature selection techniques were also applied to reduce computational complexity and improve overall classification efficiency. Experimental results demonstrated that integrating deep learning models with Machine Learning classifiers significantly improved credit score prediction accuracy and reduced processing overhead.

### *3.2 Innovative Credit Risk Assessment Using Social Media Data*

The rapid growth of financial technology has encouraged the development of more inclusive credit scoring systems capable of evaluating individuals without traditional credit histories. Conventional credit assessment methods mainly depend on banking records, collateral assets, and historical loan information, making it difficult for younger individuals and underserved populations to access financial services.

To overcome this limitation, a modern credit scoring framework utilizing social media analytics and Machine Learning techniques was proposed. The system analyzes demographic details, behavioral characteristics, psycholinguistic information, and social network data collected from online platforms to evaluate financial reliability.

The proposed framework generates predictive credit scores and classifies borrowers into different credit risk categories based on behavioral and social indicators. By incorporating digital footprints and alternative data sources, the system improves financial inclusion and expands access to credit for individuals lacking formal financial records. This approach demonstrates the potential of Machine Learning and social media analysis in transforming modern credit scoring systems.

### *3.3 Addressing Bias and Data Privacy Issues in Artificial Intelligence-Based Credit Scoring*

The increasing use of Artificial Intelligence in credit scoring systems has raised concerns regarding algorithmic bias, fairness, and data privacy.

Automated decision-making systems may unintentionally produce biased predictions, leading to unfair loan approval outcomes among different demographic groups. In addition, cybersecurity and data privacy risks have become major concerns in financial applications.

To address these issues, a cybersecurity risk assessment framework was developed to evaluate bias, privacy vulnerabilities, and security threats in Artificial Intelligence-driven credit scoring systems. Statistical fairness analysis, probabilistic modeling, and simulation-based approaches were applied to measure discrimination patterns and identify cybersecurity risks within financial systems.

The findings highlighted the importance of integrating fairness-aware Machine Learning techniques and advanced cybersecurity protection mechanisms into automated financial decision-making systems. The proposed approach improves transparency, security, regulatory compliance, and trustworthiness in Artificial Intelligence-based credit assessment applications.

### *3.4 Automated Credit Assessment Framework Using Extraction, Transformation, and Loading Process and Machine Learning*

In modern business environments, real-time analysis of financial data is essential for effective decision-making and risk management. Automated data integration and Business Intelligence systems play a significant role in processing enterprise data efficiently and generating actionable insights.

An automated credit assessment framework was developed using Extraction, Transformation, and

Loading processes combined with Machine Learning classification techniques. The automated data integration process continuously collects and organizes financial data into a centralized storage system for analysis.

Machine Learning algorithms such as Logistic Regression and Neural Network models were applied to classify customer credit risk and support automated lending decisions. The framework improves operational efficiency, minimizes manual intervention, and assists financial institutions in complying with regulatory standards while managing financial risk effectively.

### *3.5 Machine Learning-Based Credit Scoring for Informal Business Merchants*

Access to financial support and working capital remains a major challenge for informal business sectors due to the lack of formal credit histories and appropriate financial products. Improved credit scoring methods can help financial institutions provide better lending support to small businesses and informal merchants.

To address this issue, Machine Learning-based credit scoring models were developed for informal merchants using supervised learning techniques. Financial transaction data from merchants was analyzed using algorithms such as Logistic Regression and Support Vector Machine to predict loan repayment behavior and default risks.

Multiple predictive models were evaluated to determine the most effective approach for default prediction and credit risk assessment. The proposed system demonstrated improved accuracy in

identifying creditworthy merchants and reducing lending risks for financial institutions. This approach highlights the potential of Machine Learning in supporting financial inclusion and improving access to credit for underserved business communities.

### 3. METHODOLOGY

#### i) Proposed Work:

The proposed system introduces an advanced credit scoring prediction framework based on a Hybrid Long Short-Term Memory (LSTM) network integrated with a Self-Attention mechanism. The primary objective of the system is to improve the accuracy and reliability of credit risk assessment by effectively capturing temporal dependencies, nonlinear relationships, and complex behavioral patterns present in modern financial datasets. Unlike traditional machine learning algorithms such as Random Forest and XGBoost, which primarily focus on structured financial data, the proposed model can analyze both sequential and heterogeneous information to generate more accurate creditworthiness predictions.

The system begins with a comprehensive data preprocessing stage that includes handling missing values, removing inconsistencies, and preparing the dataset for model training. Min-Max Normalization is applied to scale numerical attributes into a uniform range, ensuring stable model convergence. To address class imbalance issues commonly found in credit risk datasets, the Synthetic Minority Oversampling Technique (SMOTE) is employed to generate balanced training samples. Recursive Feature Elimination (RFE) is then used to identify the most relevant features, while Principal Component

Analysis (PCA) reduces dimensionality and improves computational efficiency.

After preprocessing, the optimized dataset is provided to multiple machine learning and deep learning models, including Random Forest, XGBoost, Hybrid LSTM, and the proposed Extension Hybrid LSTM with Self-Attention. The Hybrid LSTM model is designed to capture long-term temporal dependencies in borrower behavior, transaction histories, repayment trends, and financial activities. The integrated Self-Attention mechanism enhances the model by assigning greater importance to critical features and significant time steps, enabling more effective learning of contextual relationships within the data.

The trained model is deployed using a Flask-based web application that provides a user-friendly interface for real-time credit score prediction. Authorized users can securely upload customer financial data files, and the system automatically processes the input data through the trained model. Based on the learned patterns, the system predicts whether the applicant belongs to the Good Credit or Bad Credit category and displays the results instantly.

The proposed framework offers several advantages over traditional credit scoring systems. It improves prediction accuracy by capturing both temporal and nonlinear patterns, enhances interpretability through the attention mechanism, reduces the impact of irrelevant features, and provides scalable deployment through a web-based platform. Experimental results demonstrate that the Extension Hybrid LSTM with Self-Attention achieves an accuracy of 89.87%, outperforming conventional machine learning and standard deep learning models. Therefore, the

proposed system provides an intelligent, reliable, and efficient solution for modern credit risk assessment and financial decision-making.

## ii) System Architecture:

The above architecture illustrates the workflow of the proposed Hybrid LSTM with Self-Attention model developed for credit scoring prediction in the financial sector. The framework combines deep learning components, feature transformation mechanisms, expert learning modules, and optimization strategies to improve the accuracy of credit risk assessment.

The process begins with the Input Layer, where financial and behavioral data are provided to the system after preprocessing. The input data undergoes feature embedding to transform raw attributes into meaningful numerical representations that can be efficiently processed by deep learning models. To improve model stability and training efficiency, Root Mean Square Normalization (RMS Norm) is applied, ensuring balanced feature distributions throughout the network.

The normalized features are then passed through a Hybrid Loss Function and Deep Learning Architecture, which help the model learn complex nonlinear relationships and optimize classification performance. The architecture integrates Long Short-Term Memory (LSTM) units to capture temporal dependencies and sequential patterns present in customer financial activities, repayment histories, and transaction behaviors. This enables the model to analyze long-term financial trends that influence creditworthiness.

To further enhance learning capability, a Mixture of Experts (MoE) module is incorporated into the framework. The MoE layer consists of multiple expert networks, where each expert specializes in learning different aspects of financial behavior. A gating mechanism dynamically selects the most relevant experts for each input sample, improving prediction accuracy and reducing computational complexity.

The architecture also employs a Feature Transformation Layer, which converts extracted features into higher-level representations. This transformation process enables the model to identify hidden relationships among variables and improve contextual understanding of financial data. The transformed features are then combined using an Add + RMS Norm operation to maintain stable gradient flow and prevent information loss during training.

A key component of the proposed framework is the Multi-Head Self-Attention Mechanism. The attention layer generates Query (Q), Key (K), and Value (V) vectors and computes attention scores to identify the most important features influencing credit score predictions. By focusing on significant financial attributes and temporal patterns, the self-attention mechanism enhances the model's ability to capture contextual dependencies and improve interpretability.

The output embeddings generated by the attention module are further processed through Feed Forward Neural Network (FFN) Heads and Feature Fusion Mechanisms, which integrate information from multiple representation spaces. This fusion process ensures that both local and global feature interactions are effectively captured before final classification.

Finally, the optimized representations are passed through the Output Layer, where the model predicts the applicant's creditworthiness by classifying the credit score into categories such as Good Credit or Bad Credit. The architecture is trained using advanced optimization techniques and multi-task learning strategies, enabling the system to achieve high predictive accuracy while maintaining robustness and scalability.

Overall, the proposed Hybrid LSTM with Self-Attention architecture provides an intelligent and efficient framework for credit scoring prediction by combining temporal learning, attention mechanisms, expert networks, and feature fusion techniques to enhance financial risk assessment and decision-making.



Fig1 proposed architecture

### iii) Modules:

#### 1. Importing the Packages

This module imports all necessary Python libraries and classes required for data processing, visualization, machine learning, and deep learning operations. Important libraries such as NumPy, Pandas, Scikit-learn, TensorFlow, and Matplotlib are

loaded to support data analysis, preprocessing, model training, evaluation, and graphical representation. These packages ensure smooth execution of all system modules and maintain compatibility throughout the implementation process.

#### 2. Exploring the Dataset

This module loads the Credit Risk dataset and examines its structure, attribute values, and statistical information. It helps understand the dataset by identifying feature names, data types, class distributions, and possible inconsistencies. Initial dataset exploration assists in planning suitable preprocessing, feature engineering, and model development strategies.

#### 3. Data Processing

The data processing module handles missing values, duplicate records, inconsistent entries, and noisy information present in the dataset. It ensures that all financial records are clean, valid, and suitable for machine learning and deep learning model training. Proper preprocessing improves data quality, reduces prediction errors, and enhances overall model performance.

#### 4. Visualization

This module generates graphical visualizations of data distributions, class labels, and relationships between categorical and numerical features. Visualization techniques help identify important patterns such as class imbalance, customer financial behavior, loan distribution, and repayment trends. These insights support better feature engineering and model preparation.

#### 5. Label Encoding

The label encoding module converts categorical attributes into numerical format using encoding techniques. Since machine learning and deep learning algorithms require numerical input, this step ensures that categorical data can be processed and interpreted effectively by predictive models.

#### 6. Feature Selection

This module applies Recursive Feature Elimination and Principal Component Analysis to select the most important financial features and reduce dataset dimensionality. Feature selection improves model accuracy, decreases computational complexity, and focuses the learning process on the most relevant attributes contributing to credit score prediction.

#### 7. Split the Data into Train and Test

The processed dataset is divided into training and testing subsets, commonly using an 80:20 ratio. The training dataset is used for model learning, while the testing dataset evaluates prediction performance on unseen data. This division ensures fair and unbiased model evaluation.

#### 8. Model Training

This module trains multiple prediction models including Random Forest, Extreme Gradient Boosting, Proposed Hybrid Long Short-Term Memory, and Extension Hybrid Long Short-Term Memory with Attention models. These algorithms learn financial and behavioral patterns from the dataset to predict customer creditworthiness accurately while handling both structured and sequential data effectively.

#### 9. Evaluation

The evaluation module measures model performance using metrics such as accuracy, precision, recall, and F1-score. It compares the effectiveness of different algorithms and identifies the best-performing model for credit score prediction tasks.

#### 10. SHAP Analysis

This module performs SHAP analysis to interpret and explain prediction results generated by machine learning and deep learning models. SHAP identifies the importance of individual features and shows how each attribute influences the predicted credit score, improving transparency and interpretability of the system.

#### 11. Flask Server

The Flask Server module initializes a web-based environment for deploying the credit scoring application online. It enables users to access the trained prediction models through a web interface and generate real-time credit score predictions efficiently.

#### 12. User Login

This module authenticates users by verifying login credentials before granting system access. It ensures secure usage of the application and restricts operations to authorized users, protecting sensitive financial and customer information.

#### 13. Credit Score Prediction

The credit score prediction module processes uploaded or entered customer financial data and predicts whether the credit score belongs to the Good or Bad category. Prediction results are displayed

clearly, helping financial institutions make informed loan approval and risk management decisions.

#### 14. Logout

This module securely terminates active user sessions and closes system access after usage. It helps maintain application security and prevents unauthorized access to confidential financial information.

#### iv) Algorithms:

##### 1. Random Forest

Random Forest is a supervised ensemble learning algorithm that creates multiple decision trees and combines their predictions through a majority voting mechanism. The algorithm is widely used for classification tasks because of its robustness, ability to handle high-dimensional data, and resistance to overfitting. In the proposed credit scoring system, Random Forest serves as a baseline model for analyzing structured financial attributes such as income, loan amount, employment status, and repayment history. The model helps classify borrowers into Good Credit and Bad Credit categories by identifying complex relationships among financial variables. Experimental evaluation achieved an accuracy of 87.22%, demonstrating reliable performance for credit risk assessment.

##### 2. XGBoost (Extreme Gradient Boosting)

XGBoost is an advanced gradient boosting algorithm that builds decision trees sequentially, where each new tree attempts to minimize the errors generated by previous trees. The algorithm incorporates regularization techniques, making it highly efficient

and capable of preventing overfitting. In this project, XGBoost is used to analyze financial records and identify nonlinear relationships affecting creditworthiness. The model effectively handles missing values and large datasets while providing strong predictive performance. Experimental results show that XGBoost achieved an accuracy of 86.65%, making it a competitive machine learning model for credit score prediction.

##### 3. Proposed Hybrid LSTM

The Proposed Hybrid Long Short-Term Memory (LSTM) model is designed to overcome the limitations of traditional machine learning algorithms in handling sequential financial data. LSTM is a specialized Recurrent Neural Network (RNN) architecture capable of learning long-term dependencies through memory cells and gating mechanisms. The model processes temporal financial behaviors such as spending habits, repayment trends, and transaction histories. To improve performance, the Hybrid LSTM is integrated with preprocessing techniques including Min-Max Normalization, SMOTE, Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA). This enables the model to learn meaningful temporal patterns and improve credit score classification accuracy. The Hybrid LSTM achieved an accuracy of 88.54%, outperforming traditional machine learning methods.

##### 4. Extension Hybrid LSTM with Self-Attention

The Extension Hybrid LSTM with Self-Attention is the proposed novel contribution of the project. This model enhances the standard Hybrid LSTM architecture by incorporating a Self-Attention mechanism that enables the network to focus on the

most relevant features and critical time-dependent patterns within financial data. The attention layer assigns importance weights to different inputs, allowing the model to emphasize influential financial behaviors while reducing the impact of irrelevant information.

The combination of temporal memory from LSTM and contextual feature selection from Self-Attention improves predictive accuracy and interpretability. The model effectively captures complex nonlinear interactions and evolving financial patterns, resulting in more reliable credit risk assessment. Experimental evaluation demonstrated that the Extension Hybrid LSTM with Self-Attention achieved the highest performance among all models with an accuracy of 89.87%, precision of 85.75%, recall of 83.51%, and F1-score of 84.55%. Therefore, it was selected as the final deployment model for real-time credit scoring prediction through the Flask-based web application.

#### 4.EXPERIMENTAL RESULTS

The experimental evaluation of the proposed system was carried out using IoT network traffic datasets containing both normal and malicious traffic records. The system was implemented using Python, Flask framework, and machine learning libraries such as Scikit-Learn, NumPy, and Pandas. Machine learning algorithms including Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were trained using preprocessed IoT traffic data to perform anomaly detection and attack classification. The experimental results demonstrate that the proposed system successfully identifies abnormal traffic behavior and classifies various

cyber-attacks such as DoS, DDoS, malware attacks, reconnaissance attacks, botnet activities, and data theft attacks with improved accuracy and reduced false positive rates.

The developed web-based application provides real-time anomaly detection, live input analysis, traffic monitoring, attack confidence evaluation, and risk reference analysis through an interactive user interface. Experimental outputs show that the system effectively analyzes network traffic samples, detects suspicious activities, and generates attack alerts along with confidence scores and traffic feature analysis. The system also supports manual live input analysis and reference-based risk prediction for identifying high-risk and low-risk network traffic conditions. The obtained results confirm that the proposed machine learning framework provides efficient, scalable, and intelligent security protection for IoT environments against modern cyber threats.

**Accuracy:** A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

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

$$\text{Accuracy} = \frac{(TN + TP)}{T}$$

**Precision:** Classification accuracy or positive cases constitute precision. The formula for accuracy is:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

**Recall:** A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

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

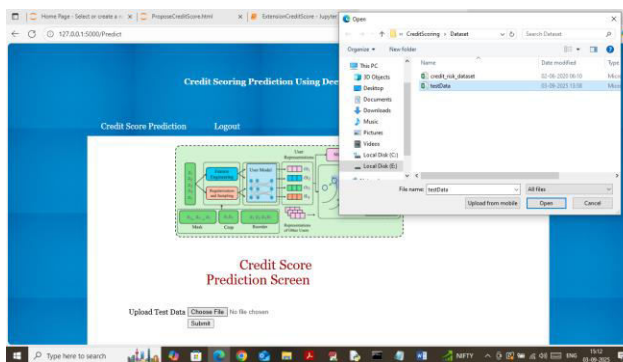
**mAP:** Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

**$AP_k$  = the AP of class  $k$**   
 **$n$  = the number of classes**

**F1-Score:** A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot Precision)}{(Recall + Precision)}$$



**Fig 2:** In above screen selecting and uploading test data file and then click on buttons to get below page

Test Data	Predicted Credit Score
[25 9600 'MORTGAGE' 1.0 'MEDICAL' 'C' 5500 [12.87 0.57 'N' 3]	Bad
[25 78000 'RENT' 5.0 'HOMEIMPROVEMENT' 'C' 3200 [12.99 0.04 'N' 4]	Good
[24 83000 'RENT' 3.0 'HOMEIMPROVEMENT' 'A' 3200 [6.99 0.04 'N' 3]	Good
[26 77100 'RENT' 8.0 'EDUCATION' 'B' 35000 [12.42 0.45 'N' 3]	Good
[23 78000 'RENT' 3.0 'EDUCATION' 'B' 3200 [10.99 0.04 'N' 3]	Good
[24 85000 'RENT' 5.0 'PERSONAL' 'B' 3200 [10.25 0.04 'N' 4]	Good
[24 78956 'RENT' 5.0 'MEDICAL' 'B' 35000 [11.11 0.44 'N' 4]	Bad
[23 92111 'RENT' 7.0 'MEDICAL' 'T' 35000 [20.25 0.32 'N' 4]	Bad
[24 12000 'RENT' 2.0 'HOMEIMPROVEMENT' 'B' [12.50 9.99 0.27 'N' 2]	Bad
[21 35000 'MORTGAGE' 2.0 'VENTURE' 'A' 8000	

**Fig 3:** In above screen in first column can see test data values and in second column can see predicted Credit score as 'Good or Bad'. This score can be utilized by banks to approved or reject loans

#### 4. CONCLUSION

This project successfully developed an advanced credit scoring prediction system using deep learning techniques for financial risk assessment. The proposed framework combines traditional machine learning models, including Random Forest and XGBoost, with deep learning architectures such as Hybrid Long Short-Term Memory (LSTM) networks to improve the accuracy and reliability of credit score prediction. The system effectively addresses the limitations of conventional credit scoring methods by capturing temporal dependencies, nonlinear relationships, and evolving financial behavior patterns present in modern financial datasets.

To enhance data quality and model performance, several preprocessing techniques including Min-Max Normalization, Synthetic Minority Oversampling Technique (SMOTE), Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA) were incorporated into the framework. These techniques improved feature representation, handled class imbalance, reduced dimensionality, and increased the overall effectiveness of the predictive models.

The major contribution of this work is the implementation of an Extension Hybrid LSTM model integrated with a Self-Attention mechanism. The attention layer enables the model to focus on the most significant financial features and temporal patterns, resulting in improved contextual understanding and more accurate creditworthiness evaluation. Experimental results demonstrated that the Extension Hybrid LSTM with Self-Attention achieved the highest performance among all evaluated models, obtaining an accuracy of 89.87%, outperforming Random Forest, XGBoost, and the standard Hybrid LSTM model.

Furthermore, the trained model was successfully deployed through a Flask-based web application, allowing users to upload financial data and obtain real-time credit score predictions. The system classifies applicants as Good Credit or Bad Credit, assisting financial institutions in making informed lending decisions and improving risk management strategies.

Overall, the proposed framework provides an intelligent, scalable, and efficient solution for modern credit scoring prediction. By combining temporal feature learning, attention-based contextual analysis, and advanced preprocessing techniques, the system

enhances prediction accuracy, supports fair credit evaluation, and contributes to more reliable financial decision-making in the banking and lending sectors..

## 5. FUTURE SCOPE

To better capture long-range temporal correlations and complicated feature interactions, new deep learning architectures like transformer-based models can improve the credit scoring system. Multiple data sources including transaction histories, social media behavior, and other credit indicators can improve forecast accuracy and generalization. Expanding explainable AI strategies can improve financial decision-making and regulatory compliance insights. Real-time adaptive learning frameworks can update models with new data to prevent financial behavior concept drift. Finally, adding dynamic risk assessment, fraud detection, and cross-institutional data integration would create a comprehensive, scalable, and robust financial credit evaluation system.

## REFERENCES

- [1] Gür, Y. E., Toğaçar, M., & Solak, B. (2025). Integration of CNN models and machine learning methods in credit score classification: 2D image transformation and feature extraction. *Computational Economics*, 65(5), 2991-3035.
- [2] Alamsyah, A., Hafidh, A. A., & Mulya, A. D. (2025). Innovative Credit Risk Assessment: Leveraging Social Media Data for Inclusive Credit Scoring in Indonesia's Fintech Sector. *Journal of Risk and Financial Management*, 18(2), 74.
- [3] Salami, I. A., Adesokan-Imran, T. O., Tiwo, O. J., Metibemu, O. C., Olutimehin, A. T., &

Olaniyi, O. O. (2025). Addressing bias and data privacy concerns in AI-driven credit scoring systems through cybersecurity risk assessment. *Asian Journal of Research in Computer Science*, 18(4), 59-82.

[4] Biswas, N., Mondal, A. S., Kusumastuti, A., Saha, S., & Mondal, K. C. (2025). Automated credit assessment framework using ETL process and machine learning. *Innovations in Systems and Software Engineering*, 21(1), 257-270.

[5] Kazimoto, D., & Baadel, S. (2025). Empowering Africa's Disfranchised SMEs: Machine Learning-Based Credit Scoring for Informal African Merchants. *Human-Centric Intelligent Systems*, 1-9.

[6] T. Schick and H. Schütze, "Exploiting cloze-questions for few-shot text classification and natural language inference," in *Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguistics*, Main Volume, 2021. [Online]. Available: <https://arxiv.org/abs/2001.07676>

[7] S. Hu, N. Ding, H. Wang, Z. Liu, J. Wang, J. Li, W. Wu, and M. Sun, "Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification," in *Proc. Annu. Meeting Assoc. Comput. Linguistics*, 2022. [Online]. Available: <https://arxiv.org/abs/2108.02035>

[8] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang, "Text classification via large language models," in *Proc. Findings Assoc. Comput. Linguistics (EMNLP)*, 2023. [Online]. Available: <https://arxiv.org/abs/2305.08377>

[9] X.Han, W.Zhao, N.Ding, Z. Liu, and M. Sun, "PTR: Prompt tuning with rules for text

classification," *AI Open*, vol. 3, pp. 182–192, Jan. 2022.

[10] Z. Wen and Y. Fang, "Augmenting low-resource text classification with graph-grounded pre-training and prompting," in *Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2023, pp. 506–516,

### Author Profiles



**Ms k.krupa sagari** is working as an Assistant Professor in the Department of MCA at SRK Institute of Technology, Vijayawada. She has completed her MCA and M.Sc. (Mathematics). She has 6 years of teaching experience at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include Artificial Intelligence, Machine Learning, and Cybersecurity.



**Ms. K. Pavani** is working as an Assistant and Head of Department of MCA, in SRK Institute of technology in Vijayawada. She completed her MCA and M.Tech in Computer Science. She has 10 years of teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR

District. Her areas of interest include AI and ML,



**Mr.P.Gandhi** is MCA Student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in B.Sc. (Computers) from VKR Degree College Gannavaram His area of interest are DBMS and Machine Learning with Python.