AI-Driven Financial Advisory and Credit Risk Prediction: A Machine Learning Approach to Personal Finance Management

Jyothsna Karuparthi (jk773)
Jasleen Kaur (jl2448)
Afsheen Khan (ak2788)
Bhanu Sai Nikhil Vinnakota (bv258)

Submitted to: Dr. Arashdeep Kaur

Contents

1 Performance Evaluation Parameters

1.1 Introduction

In the context of Phase 3 of the project, we need to design an evaluation metric for our AI-powered financial advisory system that is easy to implement and can be achieved using the computer resources available to the team. This paper refines key executability metrics, enabling them to be efficiently measurable, adequately optimized, and feasibly executable. We contrast theoretical best practices with real-world feasibility, concentrating on pragmatic approaches to measuring model accuracy, recommendation effectiveness, user engagement, and system performance.

1.2 Model Performance Metrics (Credit Risk Prediction System)

1.2.1 Accuracy (Overall Correctness)

Accuracy provides a brief overview of model performance in predicting credit defaults, albeit potentially misleading for imbalanced datasets. It is calculated using the equation 1 as:

$$\text{Acuuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{1}$$

In our dataset, where non-defaulters significantly outnumber defaulters, accuracy alone may exaggerate the performance score due to class imbalance. Therefore, it should be used in combination with other metrics such as F1-score and Precision-Recall AUC. A feasible target for accuracy is set at **75%**, benchmarking against relative financial risk models.

1.2.2 Precision (Avoiding False Alarms)

Precision calculates the correctly identified predictions among those predicted as positives. High precision ensures that financially secure users are not mistakenly flagged as high risk, which can erode trust in the system. It is defined by equation 2 as:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

To reduce false positives and maintain system credibility, we aim for a precision score of at least **80%**. Balancing precision and recall using the **F1-score** is crucial for achieving reliable predictions.

1.2.3 Recall (Cost of Missing Defaulters)

Recall measures the system's ability to correctly identify defaulters. A false negative (incorrectly classifying a defaulter as low risk) can result in severe financial consequences. Recall is given by equation 3 as:

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

To ensure sufficient detection of actual defaulters, a recall rate of at least **65-70%** is targeted.

1.2.4 F1-Score (Balanced Precision & Recall)

F1-score serves as a harmonic mean of precision and recall, balancing both aspects. It is calculated by equation 3 as:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \tag{4}$$

A feasible target for F1-score is **72%** to ensure both high precision and recall.

1.2.5 ROC-AUC (Overall Model Ranking Performance)

The ROC-AUC score measures how well the model ranks high-risk users against low-risk ones. A target of at least **0.80** is set to ensure effective classific

1.2.6 Logarithmic Loss (Probability Calibration)

Logarithmic loss evaluates how well the predicted probabilities align with actual outcomes. A lower log loss indicates better-calibrated probability estimates. The target is to achieve a log loss of **0.45**.

1.3 Metrics for Measuring Financial Advisory Effectiveness

1.3.1 Recommendation Adoption Rate

This metric assesses whether users act on the AI-generated financial recommendations. A target adoption rate of at least **60%** ensures that the advisory system provides actionable and trusted advice.

1.3.2 Financial Improvement Score

This measures whether users experience financial improvements after following recommendations. A feasible target is a **risk score reduction of at least 15%**.

1.3.3 User Satisfaction Score

User satisfaction is critical for system credibility. A target score of at least **4.0 out of 5** ensures that users find the recommendations valuable.

1.4 User Experience & Engagement Metrics

1.4.1 Session Duration

A longer session duration indicates higher engagement with the recommendations. A **minimum average of 3 minutes per session** is set as a feasibility target.

1.4.2 Click-Through Rate (CTR)

CTR measures how often users engage with AI recommendations. A target of at least **20%** ensures that recommendations are relevant and compelling.

1.4.3 Retention Rate

Retention rate indicates user loyalty and repeated engagement with the platform. A feasible **monthly retention rate (MoM) of 50%** is targeted.

1.5 Measures of Computational Efficiency

1.5.1 Inference Time

Inference time is critical to ensure real-time performance. The goal is to maintain inference times of **500ms per prediction**.

1.5.2 Memory Consumption

Efficient memory usage is essential for system scalability. The target is to keep memory consumption **1GB RAM per user session**.

1.5.3 API Response Time

Fast API response times improve user experience. The feasibility target is **800ms per API request**.

1.6 Conclusion

This performance assessment framework provides practical, achievable metrics that align with real-world constraints and available computing resources. By prioritizing realistic targets over theoretically optimal numbers, we account for hardware limitations, dataset characteristics, and user behavior. This approach ensures accurate risk predictions, helpful financial advice, and a positive user experience.

2 Literature Review

2.1 The Evolution of AI in Credit Risk Assessment

The landscape of credit risk assessment has undergone a remarkable transformation, shifting from traditional statistical models to sophisticated AI-driven solutions. With increasing financial complexity and the availability of vast datasets, machine learning has become a pivotal tool in credit evaluation. This evolution has allowed financial institutions to move beyond rigid, rule-based systems to more dynamic, adaptive frameworks that provide real-time insights into borrower risk.

2.2 Foundations of Credit Risk Prediction

Historically, credit risk assessment relied on conventional statistical methods such as logistic regression and decision trees. Early research focused on structured financial datasets to predict default probabilities, identifying patterns in borrower behavior. However, these models were often limited in their predictive capabilities, struggling to account for non-linear relationships and the nuanced financial behaviors of individuals and businesses. Additionally, imbalanced datasets—where defaulters were significantly outnumbered by non-defaulters—posed a major challenge, often leading to biased risk evaluations (Alam et al., 2020) [1].

2.3 The Rise of Machine Learning in Credit Scoring

As the need for more sophisticated risk assessment grew, machine learning techniques gained traction. Researchers explored a variety of classification algorithms, including neural networks and support vector machines (SVMs), which showed promise in improving predictive accuracy. However, the challenge of class imbalance persisted, necessitating the development of new data-balancing techniques to enhance model reliability and fairness (Li, 2019) [2].

2.4 Tackling Data Imbalance: A Turning Point

To mitigate the issues associated with skewed datasets, techniques such as the Synthetic Minority Oversampling Technique (SMOTE) and its advanced counterpart, Adaptive Synthetic Sampling (ADASYN), were introduced. These methods generated synthetic examples of minority class instances, allowing models to learn more effectively from limited default cases. The study *"Credit Risk Prediction Based on Improved ADASYN Sampling and Optimized LightGBM"*proposed a novel approach—KM-ADASYN-TL-FLLightGBM (KADT-FLightGBM)—which incorporated K-Means clustering to refine synthetic sample generation, significantly improving predictive accuracy while minimizing noisy data points (Song et al., 2024) [7].

2.5 Gradient Boosting: A Game-Changer in Credit Risk Analysis

With financial transactions becoming increasingly complex, researchers turned to more powerful algorithms like gradient boosting to refine credit risk assessment. Models such as LightGBM, XGBoost, and CatBoost demonstrated superior performance, particularly in handling large-scale datasets (Tareaf et al., 2024) [10].

A groundbreaking study using a 50GB dataset from American Express provided fresh insights into data chunking strategies and hyperparameter tuning, revealing how CatBoost excelled in structured financial datasets, while LightGBM and XGBoost required strategic handling of missing values to maintain high accuracy. This research emphasized the importance of model customization based on dataset characteristics, a crucial factor in enhancing real-world applicability.

2.6 Feature Selection and Explainability: Making AI Transparent

While machine learning models significantly improved predictive power, the financial industry demanded greater transparency in AI-driven decisions. The study *"Credit Risk Analysis Using Machine Learning Techniques"* introduced a hybrid feature selection methodology combining correlation-based filtering with Recursive Feature Elimination (RFE). By isolating the most influential financial indicators, this approach reduced data noise while enhancing interpretability (Satheeshkumar et al., 2024) [11].

Meanwhile, *"Performance of Machine Learning Algorithms in Digital Finance Risk Prediction"* explored Explainable AI (XAI) techniques, ensuring that financial institutions and borrowers could understand the reasoning behind credit decisions. This push toward explainability became a cornerstone of regulatory compliance, fostering trust between AI-driven financial systems and their users (Shao, 2024) [8].

2.7 The Evolution of AI-Driven Financial Advisory Systems

Beyond credit risk prediction, AI has expanded into financial advisory, transforming passive data analysis tools into intelligent financial assistants. Research such as *"Credit Risk Prediction Based on Machine Learning Methods"* introduced real-time analytics and deep learning-based dashboards, enabling AI-driven systems to provide personalized financial guidance (Li, 2019) [2]. These innovations bridged the gap between raw financial data and actionable insights, empowering users with smarter financial decision-making capabilities.

A notable advancement in this domain was the integration of Natural Language Processing (NLP), enabling AI-driven chatbots to deliver financial advice in a user-friendly conversational format. The study *"Machine Learning and Financial Planning"* showcased an interactive advisory system capable of not only analyzing credit risk but also offering tailored recommendations to users based on their spending patterns and financial goals (Yemmanuru et al., 2024) [12].

2.8 The Power of Stacked Generalization

Recent research has explored stacked generalization, a meta-learning technique that combines multiple machine learning models to enhance predictive performance. In *"Credit Risk Analysis Using Machine Learning Al-*

gorithms", stacked ensemble models significantly outperformed individual classifiers, particularly in assessing credit risk for Small and Medium Enterprises (SMEs). This approach reinforced the growing trend of hybrid AI frameworks that integrate diverse algorithms to maximize predictive accuracy and reliability (Naik, 2021) [9].

2.9 Future Prospects: The Road Ahead for AI in Finance

As AI-driven financial systems continue to evolve, several promising directions are emerging:

- **Reinforcement Learning for Dynamic Credit Strategies** – AI models that learn and adapt financial strategies based on long-term user behavior and market fluctuations.
- **Blockchain Integration for Secure Transactions** – Enhancing the security and transparency of AI-powered financial analytics.
- **Federated Learning for Privacy-Preserving AI** – Enabling AI models to learn from decentralized data sources while ensuring user privacy and compliance with regulatory standards.

Additionally, *"Regulatory Compliance in AI-Based Financial Systems"* underscored the importance of ethical AI, emphasizing the necessity for transparent, unbiased models that adhere to evolving financial regulations (Zhuang & Wei, 2024) [4].

2.10 Conclusion: AI-Powered Financial Intelligence in the Modern Era

The transition from traditional credit scoring to AI-driven financial intelligence has been marked by continuous innovation, with each breakthrough refining predictive accuracy, transparency, and usability. Today, AI-driven financial assistants not only assess risk but also serve as proactive financial guides, helping individuals and businesses navigate complex financial landscapes with confidence.

By integrating real-time risk assessment, personalized financial coaching, and advanced analytics, AI has revolutionized credit evaluation, bridging the gap between raw data and meaningful financial insights. As machine learning models continue to evolve, the future of AI-driven finance promises even greater adaptability, personalization, and ethical governance, ensuring that technology remains a force for financial stability and empowerment.

3 Comparison Table

Table 1: Comparison of Different Credit Risk Prediction Models

S.No.	Ref.No.	Year	Methodology	Accuracy	AUC	KS	F1-score	Recall	Specificity	Precision	G-mean	Convergence Rate	Confidence Rate
1	3	2007	Evolutional Neural Network + Dempster-Shafer Algorithm + Grey Incidence Analysis	84.40%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	2	2019	Logistic Regression	N/A	85.18%	55.69%	N/A	N/A	N/A	N/A	N/A	N/A	N/A
			XGBoost	N/A	90.06%	64.95%	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3	1	2020	Gradient Boosted Decision Tree	88.7% (Taiwan credit card dataset)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
4	6	2020	Logistic Regression (highest)	81.43%	N/A	N/A	N/A	70.27%	N/A	N/A	63.40%	N/A	N/A
			Random Forest (highest)	82.86%	N/A	N/A	N/A	67.57%	N/A	N/A	67.57%	N/A	N/A
			Neural Networks (highest)	79.29%	N/A	N/A	N/A	81.08%	N/A	N/A	57.69%	N/A	N/A
5	9	2021	LGBM	95.35%	99%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	5	2023	Neural Networks	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7	14	2023	SVM	94.07%	N/A	N/A	84.58%, 82.11%	76.33%	N/A	N/A	N/A	N/A	N/A
8	15	2023	XGBoost (best performing)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
			Random Forest	94.07%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
9	4	2024	WassersteinGAN-GP oversampling with LightGBM (best performing model)	N/A	79%	87%	63%	48%	N/A	94.84%, 95.91%	N/A	N/A	N/A
10	7	2024	KM-ADASYN-TL-FLLightGBM (KADT-FLightGBM)	N/A	99.08%	N/A	99.07%	N/A	N/A	N/A	99.08%	99%	98.30%
11	8	2024	Random Forest	98.80%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
12	10	2024	CatBoost (best performing)	86.51% (avg)	92.57% (avg)	N/A	72.80% (avg)	N/A	N/A	N/A	N/A	N/A	N/A
13	11	2024	XGBoost (best performing)	95.85%	N/A	N/A	95.84%	95.85%	N/A	N/A	N/A	N/A	N/A
14	12	2024	SVM (best performing)	90.67%	N/A	N/A	N/A	93.55%	N/A	N/A	N/A	N/A	N/A
15	13	2024	SVM + KNN	82.50%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Note: Certain values presented are averages.

4 References

[1]T. M. Alam et al., “An Investigation of Credit Card Default Prediction in the Imbalanced Datasets,” IEEE Access, vol. 8, pp. 201173–201198, 2020, doi: https://doi.org/10.1109/access.2020.3033784.

[2]Y. Li, “Credit Risk Prediction Based on Machine Learning Methods,” 2019 14th International Conference on Computer Science Education (ICCSE), Aug. 2019, doi: https://doi.org/10.1109/iccse.2019.8845444.

[3]T.-N. Chou, “A Novel Prediction Model for Credit Card Risk Management,” International Conference on Innovative Computing, Information and Control, Sep. 2007, doi: https://doi.org/10.1109/icicic.2007.68.

[4]Y. Zhuang and H. Wei, “Design of a Personal Credit Risk Prediction Model and Legal Prevention of Financial Risks,” IEEE Access, pp. 1–1, Jan. 2024, doi: https://doi.org/10.1109/access.2024.3466192.

[5]Parth Pangavhane, Shivam Kolse, Parimal Avhad, Tushar Gadekar, N. K. Darwante, and S. V. Chaudhari, “Transforming Finance Through Automation Using AI-Driven Personal Finance Advisors,” Dec. 2023, doi: https://doi.org/10.1109/iccakm58659.2023.10449538.

[6]“Credit Risk Analysis Using Machine Learning Techniques — IEEE Conference Publication — IEEE Xplore,” ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/9096854

[7]M. Song, H. Ma, Y. Zhu, and M. Zhang, “Credit Risk Prediction Based on Improved ADASYN Sampling and Optimized LightGBM,” Journal of Social Computing, vol. 5, no. 3, pp. 232–241, Sep. 2024, doi: https://doi.org/10.23919/jsc.2024.0019.

[8]Y. Shao, “Performance of Machine Learning Algorithms in Digital Finance Risk Prediction,” pp. 195–199, Oct. 2024,
doi: https://doi.org/10.1109/icdacai65086.2024.00043.

[9]K. S. Naik, “Predicting Credit Risk for Unsecured Lending: A Machine Learning Approach,” arXiv (Cornell University), Jan. 2021, doi: https://doi.org/10.48550/arxiv.2110.02206.

[10]Raad Bin Tareaf, M. AbuJarour, and F. Zinn, “Revolutionizing Credit Risk: A Deep Dive into Gradient-Boosting Techniques in AI-Driven Finance,” Jan. 2024, doi: https://doi.org/10.1109/icoin59985.2024.10572140.

[11]S Satheeshkumar, M Dakshana, K Gunalan, P Anandan, R Saveetha, and M Nithya, “Leveraging Machine Learning and Forecasting Techniques to Enhance Credit Risk Analysis and Prediction,” pp. 781–786, Oct. 2024, doi: https://doi.org/10.1109/icssas64001.2024.10760746.

[12]Prathibha Kiran Yemmanuru, J. Yeboah, and I. K. Nti, “Customer Credit Risk: Application and Evaluation of Machine Learning and Deep Learning Models,” pp. 1–5, Apr. 2024,
doi: https://doi.org/10.1109/icmi60790.2024.10585896.

[13]H. Wang, L. Chen, L. Chen, and Y. Wu, “Construction of Family Financial Risk Management System Based on Artificial Intelligence and Big Data Analysis,” Jan. 2024,
doi: https://doi.org/10.1109/ssaic61213.2024.00029.

[14]C. Fang, T. Bu, and F. Fang, “Research on credit-risk models via machine-learning algorithms and logistic regression for predicting CBA consumer behaviour,” 2022 International Conference on Computers, Information Processing and Advanced Education (CIPAE), pp. 344–350, Aug. 2023, doi: https://doi.org/10.1109/cipae60493.2023.00073.

[15]Kirti Wanjale, S. Saraf, Kshitij Thakre, Abhijit Chitre, and R. Mahajan, “Financial Risk Prediction Using Consumer Information and Different Machine Learning Algorithms: A Comparative Analysis,” pp. 1–5, Aug. 2023, doi: https://doi.org/10.1109/asiancon58793.2023.10270196.