**Revolutionizing Financial Analytics: The Role of Artificial Intelligence in Market Prediction, Customer Retention, and Churn Modelling**

Bhavya Vermaa, Aryan Khandelwalb, Pranay Nepaliac Aayushi Rajputd, Aaditya Varshneye

\* Kiran Salunkef, Sugam Shivhareg, Vivekanand Bagalh

Department of Computer Science, SVKM’s NMIMS, MPSTME, Shirpur, Dist. Dhule, Maharashtra a, b, c, d

Department of Computer Engineering, SVKM’s NMIMS, MPSTME, Shirpur, Dist. Dhule, Maharashtra e

\*Assistant Professor, SVKM’s NMIMS, MPSTME, Shirpur, Dist. Dhule, Maharashtra. E-mail: [kiran.salunke@nmims.edu](mailto:kiran.salunke@nmims.edu) f

Assistant Professor, SVKM’s NMIMS, MPSTME, Shirpur, Dist. Dhule, Maharashtra. E-mail: [sugam.shivhare@nmims.edu](mailto:sugam.shivhare@nmims.edu) g

Assistant Professor, SVKM’s NMIMS, MPSTME, Shirpur, Dist. Dhule, Maharashtra. E-mail: [vivekanand.bagal@nmims.edu](mailto:vivekanand.bagal@nmims.edu) h

**Abstract**

In today’s world, customer relationship management (CRM) analytics, real-time insights and predictions are empowered by Artificial Intelligence (AI) in the financial industry. This research investigates the implications of AI for financial datasets, market prediction, and customer retention whilst concentrating on churn modelling. Machine learning prototypes, enabling AI-driven algorithms, are capable of confirming difficult, non-linear patterns in both structured and unstructured financial data to improve credit scoring, risk assessment, and algorithmic trading. The use of alternate data sources combined with real-time updates in AI-based tools provides dynamic insights that optimize investment strategies and mitigate risks. The paper further discusses customer churn modelling in banking by examining a data set with 10,000 customers to find the most important predictors of churn. Machine learning models are used to analyze demographic, financial and behavioural factors to create an accurate classification framework. The result intends to bring illegitimate insights for financial institutions, refining customer safeguarding policies and implementing long-term development in an AI-driven monetary ecosystem.

***Keywords:*** *Credit Scoring Models, Customer Churn Prediction, Machine Learning in Finance, AI-Driven Risk Assessment, Financial Behaviour Analysis*

# 1. Introduction

### The development of AI has been rapid and it has brought positive changes in financial markets and institutions to improve decision-making, decreasing costs and risk management.) Generative AI (GenAI) is this change that can produce complex outputs such as fraud detection insights and automated financial reports, which are very hard for traditional systems to replicate. This paper aims to establish that AI is still critical in finance and its application presents ongoing opportunities for innovation [15].

New innovations in AI such as ChatGPT and generative models have changed the way we create content, analyze data and converse with customers. The rapid adoption of ChatGPT indicates the increasing importance of AI in various sectors, especially in finance where it improves customer interaction and operational processing. However, there are risks such as bias, data privacy, cyber threats and semi-transparency of decision making which requires strong governance frameworks [1]. In addition to automation, AI is also used in predictive modelling of historical and real time data to forecast market trends, measure risks and optimize customer behaviour strategies. It facilitates algorithmic trading, portfolio management and financial risks assessments, and it facilitates the generation of artificial data and process automation [2]. With more than 95% of major banks adopting AI-based risk management, data quality, compliance, and ethics policy developments are emerging [16] the AI-enabled churn modelling, which helps in customer retention by identifying potential churners and offering proactive interventions, is also becomes a reality [3]. It is therefore crucial to balance the innovation and governance of AI in finance to make it sustainable

Credit scoring and risk assessment form the backbone of any financial system. Such models are crucial in helping banks and other financial institutions determine the wealth status of individuals and businesses before they grant loans, issue credit cards, or provide other financial products. These systems have shifted from a purely traditional, rule-based approach to a data-driven one utilizing advanced artificial intelligence AI and machine learning ML algorithms. These tools have revolutionized how financial institutions evaluate credit risk, thereby dramatically affecting the lending environment [4][5]. Table 1 illustrates the comparison between traditional and modern credit scoring models.

***2.1 Customary Credit Scoring Models***

The FICO score is a traditional credit scoring model that uses statistical techniques to predict the likelihood of default based on historical financial behaviour. The FICO model reviews five key factors: payment history (35%), amounts owed (30%), credit age (15%), new credit (10%), and credit mix (10) to offer a standardized risk assessment too. But it has its restrictions; it does not incorporate real-time financial data like utility payments or rental history, thereby limiting its suitability for changing borrower profiles [6]. Logistic regression is also widely applied in credit risk modelling to quantify probability of default (PD) based on historical data. Simple to use but fails to handle intricate financial activities and non-linear data relationships. However, conventional models have several major limitations, including limited data coverage, prejudice, fairness issues, and inflexibility in risk analysis [7]. In the future, advanced AI-based credit models may offer a more holistic and real-time risk assessment model.

***2.2 Contemporary Credit Scoring Models:***

Through accelerated insight, organizations and governments are able to improve services and fight fraud with the help of data analytics. They achieve efficiency, and healthcare organizations are better equipped to deliver patient care and cut costs. Future business success will be driven by innovations like IoT, blockchain, and data analytics. Borrowers’ credit, income and employment history are analyzed by machine learning models for decision trees [12]. It refines decision trees to reduce variance and improve predictive accuracy on high-risk/low-risk loan classifying. GBMs and XGBoost improve accuracy by boosting and iteratively learning weak learners to model nonlinear credit risk dependencies [13, 14]. High-risk and low-risk borrowers are effectively identified by SVMs which are useful in determining the optimal separations in the multidimensional space for credit risk assessment [8].

***2.2 AI-Driven Credit Scoring Models:***

It uses real time data to arrive at credit scoring from traditional models and therefore provides more accurate and dynamic credit analyses. Static models rely on historical behaviour but AI looks at real time financial transactions and therefore models that use fresh insights [9]. Such as social media activity, mobile or online usage, and transactions, enhance AI models which can be beneficial for people with limited credit history. However, the lack of transparency of deep learning, known as the black box effect, is an issue. SHAP and LIME techniques are proposed to enhance the interpretability to gain the AI decision-making trust of the stakeholders [10]. AI-based systems provide real-time credit ratings by constant tracking of the borrower’s behaviour and this provides the latest evaluations to the lenders. This is opposed to conventional models which are based on older financial information, hence making AI a game changer in credit risk analysis [11].

**Table 1 Comparison of Traditional and Modern Credit Scoring Models**

| **Model** |  | **Proposed Model** | **Advantages** | **Drawbacks** |
| --- | --- | --- | --- | --- |
| **FICO Score** |  | Founded on credit history, payment behaviour, and credit mix. | Standardized, broadly acknowledged. | Restricted scope, eliminates real-time data. |
| **Decision Trees** |  | Breaches data into decision nodes to assess risk. | Easy to construe. | Susceptible to overfitting, subtle to noise. |
| **Random Forests** |  | Sums several decision trees for accuracy. | Diminishes overfitting, and high accuracy. | Computationally expensive, less interpretable. |
| **XGBoost** |  | Optimized gradient boosting model. | Fast, and handles sparse data well. | Involves careful tuning, computationally severe. |
| **AI-based Models** |  | Uses real-time, alternative data for assessment. | Dynamic, inclusive, and highly accurate. | Lacks transparency, and raises privacy concerns. |
| **Alternative Data Models** |  | Influences non-traditional data sources. | Profits those with limited credit history. | Regulatory and privacy challenges. |

Despite advancements in AI and machine learning for credit scoring, several research gaps remain. A key issue is the interpretability of AI models, which can be addressed by developing better explainability techniques like SHAP and LIME to enhance transparency and trust. Additionally, the integration of alternative data raises privacy and ethical concerns, requiring research into frameworks that balance innovation with privacy protections and ethical considerations.

**3. Churn Modelling and Application in Credit Scoring**

Customer churn, the tendency of a customer to leave a financial institution, is one of the biggest challenges in the industry. Using machine learning and AI, predictive churn modelling is beneficial for banks to keep customers, decrease turnover, and increase profits. These models look at demographics, financial behaviour, account activity, and service satisfaction to predict attrition. For example, customers with low balances, high levels of debt, or irregular payments are more likely to churn. Churn models traditional leverage neural networks, decision trees, and logistic regression to determine behavioural trends and to quantify churn risk. Enhancement of the prediction accuracy is achieved through the application of NLP and deep learning in the analysis of unstructured data from social media and customer reviews. A case study of 10,000 bank customers is provided to show how AI-driven analytics can identify the key churn indicators that can be used for proactive retention tactics. It is important for financial institutions to maintain long-term customer relationships and business growth by predicting the insights with personalized interventions.

***3.1 Dataset Overview***

This analysis is performed on the Bank Customer Churn Dataset made available by Kaggle, thus adhering to the ethical and privacy guidelines. The dataset does not reveal customer’s personal identifiers but incorporates essential information in its structured format to support the churn prediction. The Customer ID identifies each record while the Row Number is used to order the data. Although Surname is provided, it is used as a reference rather than as a feature in the prediction model. Model training requires feature selection and the demographic characteristics like Gender and Age can help in the understanding of the churn behaviour. Tenure is a measure of how long a customer has been with the bank, while financial engagement variables such as Balance, NumOfProducts and HasCrCard give an overview of account activity. The IsActiveMember feature clearly differentiates between the active and inactive users and is therefore directly related to the churn probability. Also, EstimatedSalary can be used as an economic stability indicator. The Exited variable is the target variable which classifies the customers as churned (1) and retained (0) to enable prediction for churn analysis and decision making processes [17].

***3.2 Conceptual Framework***

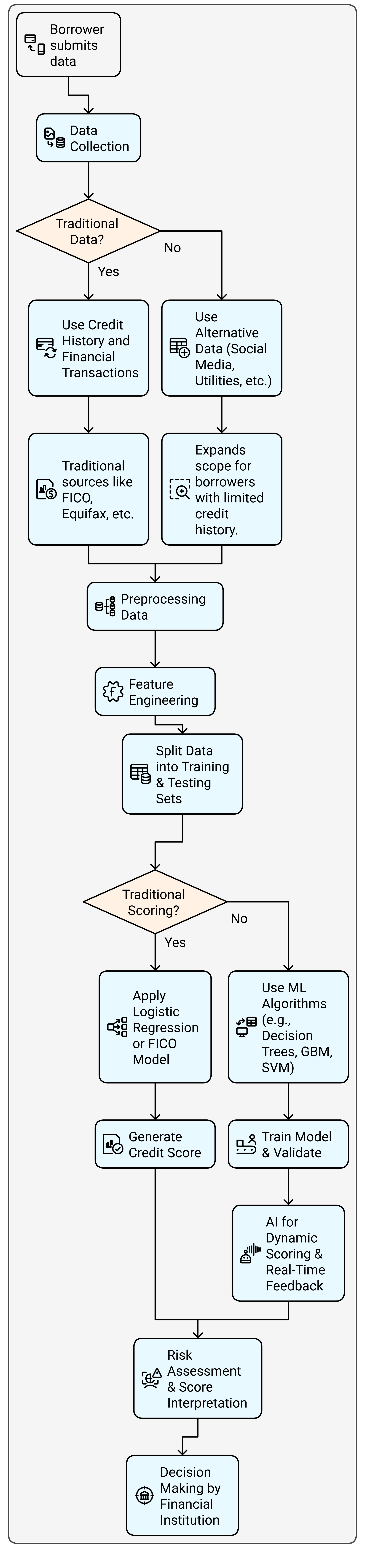
This study’s conceptual framework as shown in figure 1 classifies the key factors that cause customer churn into three categories: Demographic, Financial, and Behavioural factors. The model development process is divided into a clear pipeline, which starts from data collection and pre-processing, which involves imputing missing data, one-hot encoding of categorical variables, and standardization of numerical features. The feature selection keeps only the essential predictors and the data is divided into 80-20 for training and testing purposes. For traditional credit scoring, logistic regression is used because of its simplicity. At the same time, other machine learning models, including decision trees, gradient boosting (XGBoost) and SVM, improve the predictive accuracy. An ensemble model based on VotingClassifier combines Logistic Regression, Random Forest and XGBoost with soft voting for more accurate churn prediction. AI based dynamic scoring offers real time feedback to risk assessment. This framework improves credit assessment by incorporating non-traditional data sets (e.g. social media sites), thus enhancing financial inclusion while at the same time offering stable and data driven churn predictions that can be used to support decision making processes in financial institutions.

**4. Results**

Figure 2 demonstrates the relationship between Customer Churn Percentage and Tenure by credit card ownership (HasCrCard: 0 or 1). Rate of churn is slightly higher for customers without a credit card (18.21%) than for those with one (16.80%) at tenure 7, indicating that financial features do affect retention. The overall Churn Percentage by Tenure is shown in Figure 3, with a decline through tenure 5, then a peak at tenure 8. This indicates that the customers in the middle are most stable than those who are new or long-time customers. These understandings are useful for enhancing AI-based churn prediction models that can enhance risk evaluation and credit scoring processes. Churn prediction is based on demographic and tenure based insights not on the financial behaviour. Figure 3 shows churn trends by tenure: The highest initially at 24% and then lowest at tenure 5 and then highest at tenure 8, suggesting volatility at early and late tenure. This behaviour based model helps to improve the churn predictions.

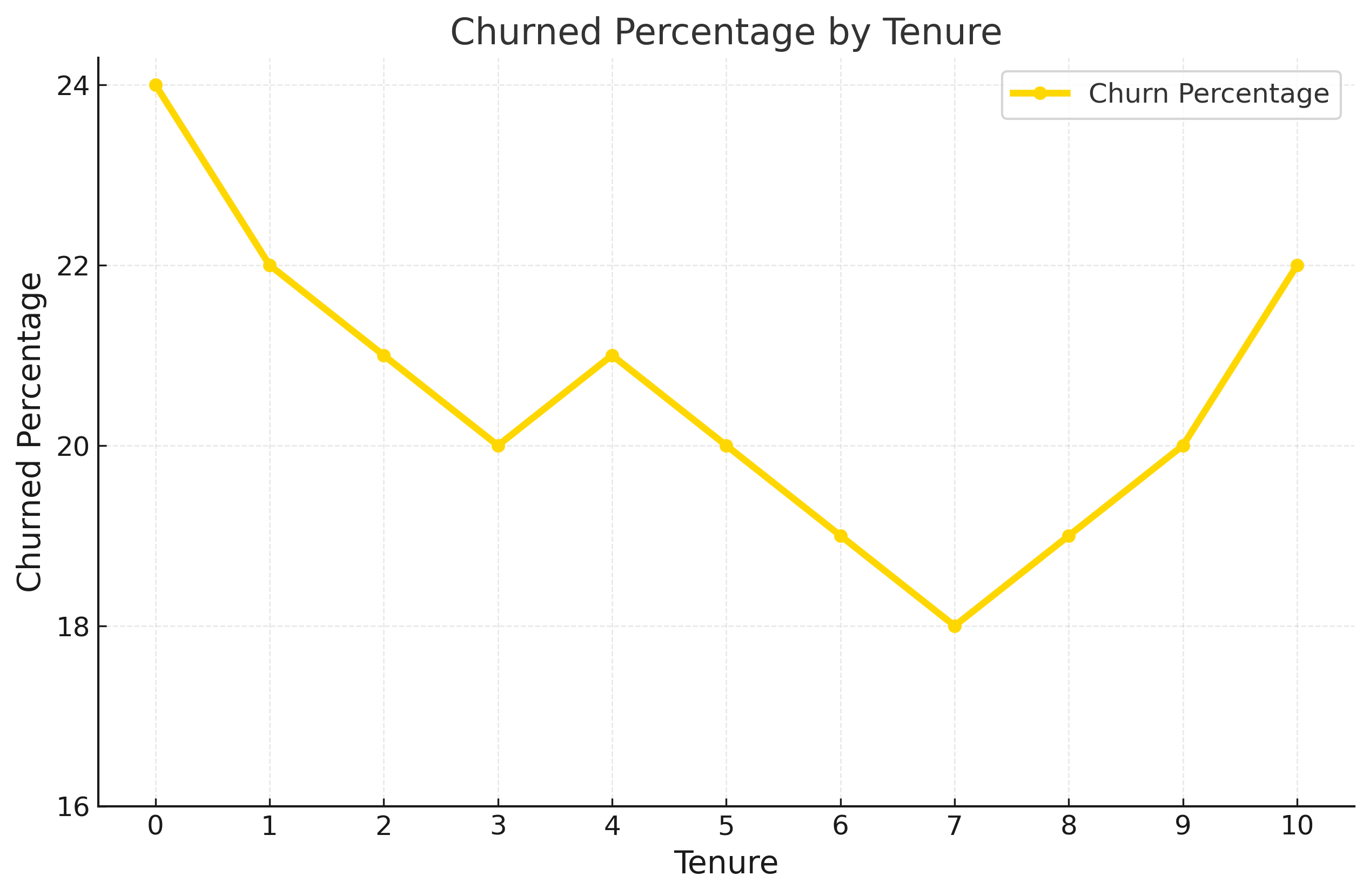
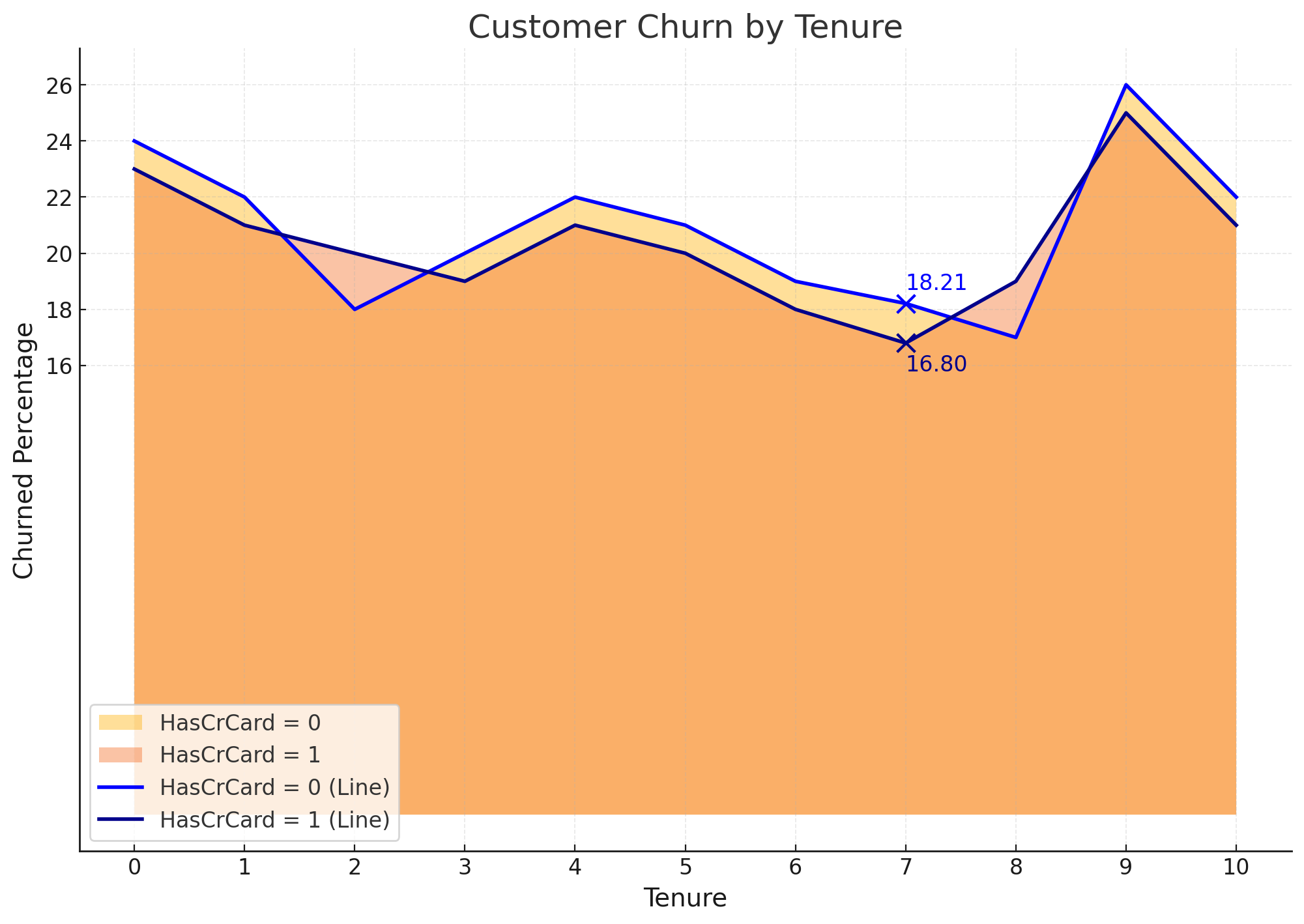
Figure 4 presents a credit score histogram with a concentration in the range of 600 – 800 and a sharp peak for highly creditworthy customers. The AI driven analytics improve the credit risk models, using the non-traditional data, the dynamic scoring and the personalized products, which accelerate the loan approvals and decrease the risk of default.

Figure 5 looks at customer activity, geography and balances. Germany has the highest mean balances while France and Spain have lower mean balances but significant variation in active vs. inactive customers. Regional strategies driven by AI can enhance customer retention, loyalty, and financial health.



**Fig. 1 Proposed Model**

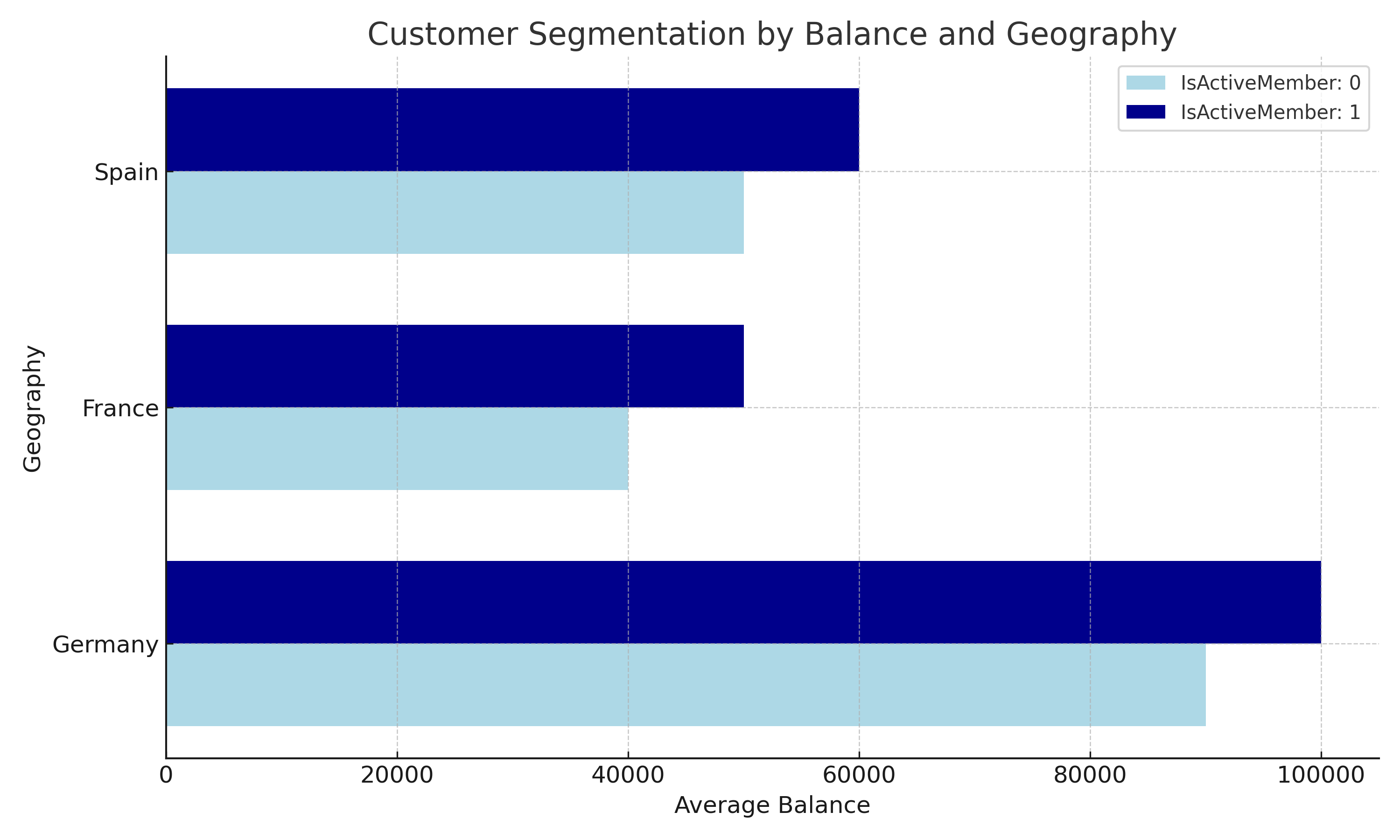
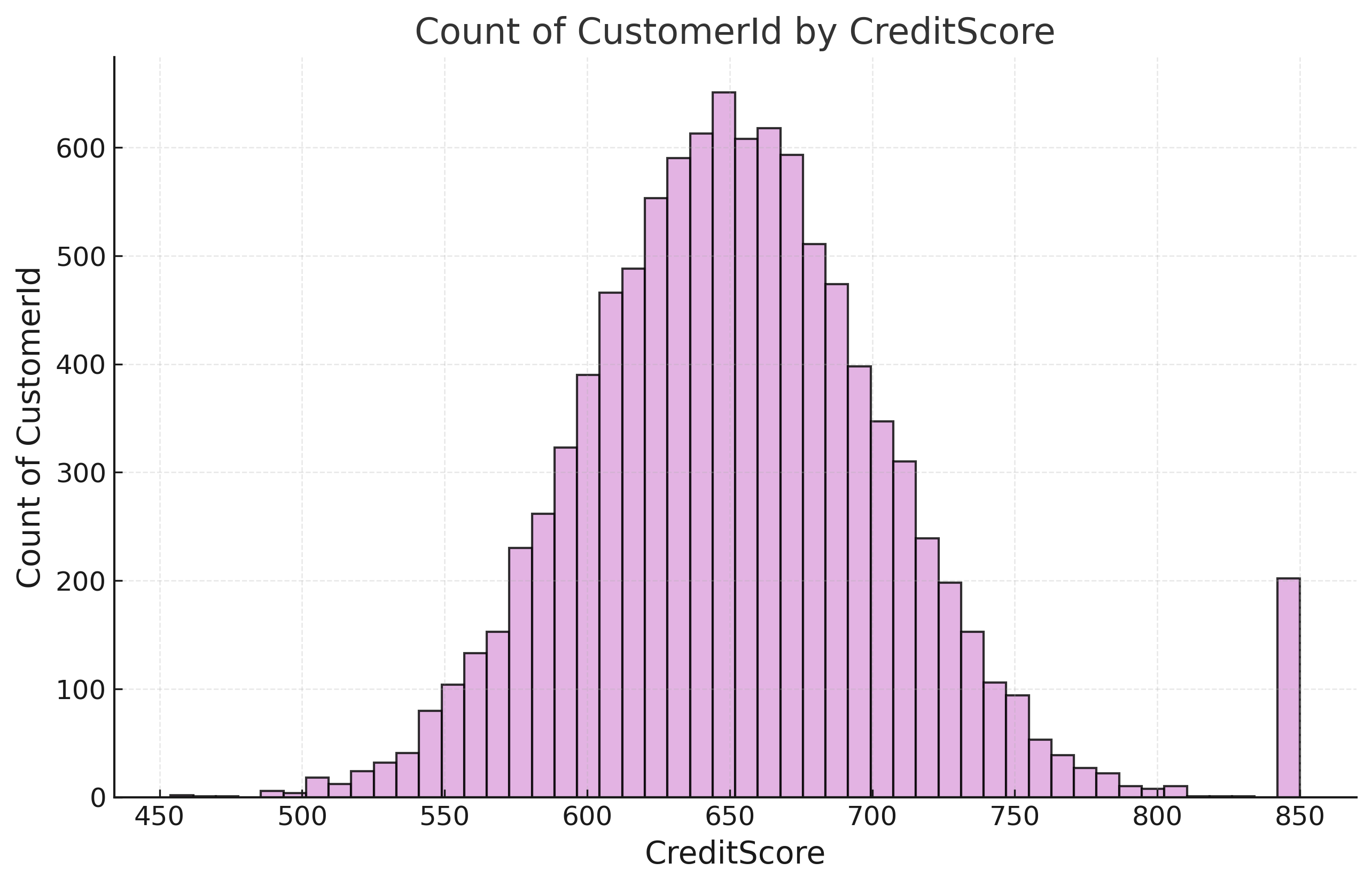
The AI driven credit scoring improves loan evaluations, tenure based retention and equitable decision making. It means risk assessment is better, customer loyalty is better encouraged, and the financially excluded such as those with sparse credit records are considered.

****

**Fig. 2 Customer Churn Percentage and Tenure Fig. 3 Churn Percentage by Tenure**

The spike at a credit score of 850 indicates that it acted as a false roof rather than a real step at the end of the dataset. There is a sharp discontinuity at 850 there are no scores above it and the industry practices that indicate this is deliberate. They regard 850 as a high level of credibility and all top scorers are clustered in one category. It is also a structured limitation which is rather specific to credit rating systems than to some statistical anomaly.

In the analysis of traditional and modern scoring models, the proposed optimized random forest model comes first with precision of 87%, which is 2.75% better than the random forest model. After that, XGBoost and Ensemble models are close behind at 86%, exhibiting good predictive capability. Logistic Regression lags at 72%, having difficulty with the non-linearity. In general, tree-based models are more precise and are therefore suitable for classification tasks.

****

**Fig. 4 Distribution of customer credit scores**  **Fig. 5 Relationship between Customer Activities**

**5. Future Scope**

Customer retention with AI powered dynamic chatbots will continue to develop with NLP for easier conversation and XAI for the clarity of the process. In subsequent iterations, the technology can be applied to the telecommunications, healthcare, and e-commerce sectors to improve the predictive analytics of consumer behaviour. the ability to integrate real-time learning mechanisms will allow the chatbots to grow dynamically, optimizing the customer interaction management to make the retention methods more accurate and efficient in order to develop stronger relationships with the customers while upholding the principles of ethical AI and predictive modelling.

**6. Conclusion**

The development of credit scoring and risk analysis has integrated the conventional statistical approaches and the new generation of AI and ML techniques. FICO scores and logistic regression are mostly used because they are easy to understand and implement, but they have their limitations and restrictions regarding the availability of data. Decision Trees, Random Forest, and Gradient Boosting based AI-models are very effective at identifying patterns in data sets and making real-time risk assessments. new types of data such as data from social media or utility companies increases the availability of credit, allowing people without a traditional credit history to get credit. future work needs to be done on the problems of the black box AI, which can be unfair and non-transparent and which raise regulatory concerns. The use of the traditional approach combined with the new AI developments helps to maintain the effectiveness, fairness, and risk management of the credit ecosystem. These improvements are a contribution in the improvement of the assessment of creditworthiness with a view of enhancing access to financial services, minimizing credit risks and defining the future of financial decisions.

**References**

1. Luo, M., Y. Liu, and H. Hu. 2023. "Artificial Intelligence in Finance: Risks and Opportunities." *Fintech Notes* 2023 (006): 1–12. Accessed August 2023. <https://www.elibrary.imf.org>.
2. Shabsigh, G., and E. B. Boukheroua. 2023. "Generative Artificial Intelligence in Finance." *IMF Library*, August. Accessed August 2023. <https://www.elibrary.imf.org>.
3. UK Finance. 2023. "AI Risk Management in Financial Services." *UK Finance Report*. Accessed 2023. <https://www.ukfinance.org>.
4. Liu, Z., X. Zhang, and J. Li. 2023. "A Comprehensive Review of AI Applications in Credit Scoring." *Journal of Financial Technology* 12 (4): 234–245.
5. Zhang, Y., S. Wang, and J. Xu. 2023. "AI in Credit Risk Assessment: A Review of Current Approaches." *Journal of Finance and AI* 8 (1): 99–115.
6. Boukherouaa, E., and G. Shabsigh. 2023. "Artificial Intelligence and Credit Scoring: The Road Ahead." *Financial Services Review* 27 (2): 45–62.
7. Choi, H., K. Lee, and M. Song. 2023. "Machine Learning Techniques for Credit Risk Prediction." *Financial Risk and AI* 5: 131–145.
8. Pławiak, P., M. Abdar, and U. R. Acharya. 2019. "Application of New Deep Genetic Cascade Ensemble of SVM Classifiers to Predict the Australian Credit Scoring." *Applied Soft Computing* 84: 105740.
9. Wang, H., Y. Chen, and X. Lu. 2021. "Leveraging Alternative Data Sources for Credit Scoring: A Machine Learning Approach." *IEEE Access* 9: 51204–51214.
10. Lundberg, S., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 4765–4774. Montreal, QC, Canada.
11. Gupta, M., A. Verma, and R. Sharma. 2022. "Real-Time Credit Scoring Using AI-Driven Transaction Monitoring." *IEEE Transactions on Computational Social Systems* 7 (3): 423–432.
12. Quinlan, J. R. 1986. "Induction of Decision Trees." *Machine Learning* 1 (1): 81–106.
13. Friedman, J. H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29 (5): 1189–1232.
14. Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. San Francisco, CA, USA.
15. Chlouverakis, Kostis. "How Artificial Intelligence Is Reshaping the Financial Services Industry." Accessed February 7, 2025. <https://www.ey.com/en_gr/insights/financial-services/how-artificial-intelligence-is-reshaping-the-financial-services-industry>.
16. Mackintosh, J., and L. Quest. 2023. "The Impact of AI in Financial Services." *UK Finance and Oliver Wyman*, November 22. Accessed November 22, 2023. <https://www.ukfinance.org.uk/system/files/2023-11/The%20impact%20of%20AI%20in%20financial%20services.pdf>.
17. Sharma, K. D. "Banking Churn Analysis & Modelling." Accessed 2025. <https://www.kaggle.com/code/kdsharma/banking-churn-analysis-modeling>.