editor@fmreview.online
P-ISSN: 3105-7403
E-ISSN: 3105-7411
https://fmreview.online
FINANCE AND MANAGEMENT REVIEW
FINANCE AND MANAGEMENT REVIEW
VOLUME: 03 ISSUE: 02 (2025)

Receive Date: July 21, 2025, **Revise Date:** August 24, 2025, **Accept Date:** November 23, 2025, **Available Online:** December 31, 2025

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON CONSUMER CREDIT SCORING MODELS: A CRITICAL REVIEW OF FAIRNESS AND BIAS IN ALGORITHMS

[1*]Mehwish Akram, [2]Owais Siddiqui
[1]Department of Computer Science and Information Systems International Islamic University, Islamabad, Pakistan
[2]Department of Finance and Financial Technology SZABIST University, Karachi, Pakistan
(owais.siddiqui@szabist.edu.pk)
CORRESPONDING EMAIL: mehwish.akram@iiu.edu.pk

Abstract:
The rapid integration of artificial intelligence into consumer credit scoring has transformed traditional risk assessment practices by improving predictive accuracy and operational efficiency. However, growing concerns regarding algorithmic fairness, transparency, and bias have raised critical ethical and regulatory challenges for automated lending systems. This study provides a comprehensive empirical and critical evaluation of AI-driven credit scoring models with a particular focus on fairness and bias across demographic and socio-economic groups. Using a mixed-methods experimental design, the research compares traditional statistical models with advanced machine learning approaches, including ensemble and neural network models, across multiple performance and fairness metrics. Quantitative results demonstrate that while AI-based models consistently outperform conventional approaches in terms of accuracy and discriminatory power, they also exhibit measurable disparities in approval outcomes for protected groups. Fairness metrics such as disparate impact and demographic parity reveal systematic bias amplification in unconstrained models, whereas fairness-aware modeling strategies significantly mitigate these disparities at the cost of marginal reductions in predictive performance. Qualitative analysis further contextualizes these findings within existing ethical AI frameworks and regulatory expectations. Overall, the study highlights the inherent trade-off between predictive efficiency and ethical responsibility in AI-driven credit scoring and underscores the need for balanced, fairness-conscious model deployment in consumer finance.

Keywords: *Artificial Intelligence, Consumer Credit Scoring, Algorithmic Fairness, Bias Mitigation, Machine Learning In Finance, Ethical AI.*

INTRODUCTION

As Adegoke et al. (2024), p. 192; Vieira et al. (2025) write, AI algorithms and machine learning are radically changing the process of decision-making in different sectors, and the financial sector is no exception to this process and this shift is manifested, in particular, in the lending and credit approval segments. This change can be done in spite of this enhancement, even though fairness, transparency, and the likelihood of entrenching biases in these intricate models are the difficult problems (Hurlin et al., 2024, p. 2; Setty et al., 2024, p. 2). In particular, the AI-driven credit scoring algorithms can reinforce or even enhance the biases, which existed in the past, even though they can work with a lot of data and discover the hidden patterns that remain unnoticed by human analysts (Addy et al., 2024, p. 124; Lee, 2024, p. 9). This is a especially important matter since these models can discriminate the particular demographic groups at all times due to their safeguarded attributes (gender, age, or racial origin) causing them unequal provision of credit and creating economic imbalances (Adegoke et al., 2024, p. 191; Hurlin et al., 2021, p. 2). One can hardly notice whether such differences are determined by discriminatory proxies that are created during the algorithms or real risk variables (Langenbucher, 2022, p. 7). To achieve a fair and non-formalization of prejudicial trends, an alternative approach to the techniques, used in AI credit rating, should therefore be scrutinized thoroughly (Langenbucher, 2022, p. 5). As it is possible to include and amplify existing differences, the research evaluates the use of AI in consumer credit scores computation critically, emphasizing the implications of bias and equity of an algorithmic model (Langenbucher, 2022, p. 5). Considering that it is aimed at the creation of interpretable and fair AI models, this review attempts to condense the most recent findings regarding the possibility of detecting and mitigating biases in AI-based credit decisions (Bari, 2024; Vieira et al., 2025). Although the intensive use of machine learning to predict credit scores has improved the risk assessment, the fears of possible biases, prejudice, and a deficit of transparency when using machines are dire (Valdrighi et al., 2024). As this continues to happen, the issue of fairness and possible biases of credit scoring models remains unchanged, and they are mostly caused by training dataset biases (Adegoke et al., 2024, p. 190). The statistics can focus on the accidental historical changes in lending patterns and social economic disorientations to produce biased results, adding to the existing disparities (Adegoke et al., 2024, p. 192). Concerning an example, the models of mortgage approval were identified to reproduce biases existing in the society and amplify historical inequalities with the help of allegedly neutral traits (Vieira et al., 2025). Though the creditworthiness of a particular group of people is identical to some other ones, such biases, which are often implicit, have their cost, and the discrimination consequences of such biases are the adverse inequalities of the affected populations (Edunjobi and Odejide, 2024, p. 98). Besides, since these guarded qualities are very closely associated with other non-sensitive ones, that are proxies and bias predicting, merely excluding sensitive features of training data is frequently not enough to remove biases (Valdrighi et al., 2024, p. 2). In a bid to ensure that the AI-based credit scoring models do not, unintentionally, bar access to the financial services by vulnerable populations, the current discussion to come up with effective methodologies that can, at the same time, detect and also negate such these deep-rooted biases (Sargeant et al., 2025, p. 5). Therefore, the review will focus on different definitions, measures, and how to reduce biases and achieve impartial results of different groups critically regarding how they should be implemented in reality (Valdrighi et al., 2024). Moreover, the AI Act and the suggestions of the European Banking Authority presuppose that one should

avoid and detect prejudice and advance equality to uphold social justice and the plausibility of credit rating systems (Coriglia et al., 2024, p. 2). To ensure transparency and accountability, one would have to answer the questions in ways that are known to properly understand the nature of how AI models work by reducing to their interpretability and the mechanisms by which they decided to make such a choice (Addy et al., 2024, p. 124; Kowsar et al., 2023, p. 21). It must be thoroughly studied to realize the ethical implication, and one can say that strict regulatory frameworks and industry standards are essential to promote trust and guarantee the responsible AI-based credit assessment (Addy et al., 2024, p. 122; Kowsar et al., 2023, p. 1). In this case, the constraints to data sources, such as the presence of underrepresented demographics in large language models to evaluate credit and the need to have larger and more heterogeneous datasets to strengthen models in different industries should be critically considered (Maarouf et al., 2024, p. 224). This is made difficult by the fact that balancing many fairness indicators, most of which are competing, to get equitable results is complex (Kamalaruban et al., 2024, p. 2). In order to select and fix a misalignment between model output in normal circumstances, more advanced approaches of bias detection and removal in machine learning such as fairness-conscious machine learning (Addy et al., 2024, p. 122; Vieira et al., 2025) should be developed. Such approaches may include reweighting of training data, manipulation of model outputs, or even just including conceptualized fairness constraints to the loss goal so that the distribution of expected results in the various demographic subgroups becomes more balanced (Garcia et al., 2023, p. 2073). Such advanced prediction models should be applied according to moral standards that focus on autonomy and autonomy of choice in customers and a significant distinction between the personalization of models and potential manipulation, especially in regards to vulnerable groups of consumers (Cardona-Acevedo et al., 2025, p. 107). Special ethics committees and impact assessment should be incorporated into the process of ethical credit scoring to ensure the uninterrupted compliance with accepted norms in the industry, as well as legal requirements (Addy et al., 2024, p. 123). Moreover, to ensure that the inequity is not eliminated, and the biases that used to be applied in the past are avoided, it is necessary to conduct regular audits on the AI models to avoid the identification and correction of new potential biases which can appear in the course of their functioning (- et al., 2024, p. 11). This proactive approach, appropriate ethical guidelines, and building trust among the population will be required to facilitate AI in credit scoring as a pro-financial inclusion tool and no pro-financial exclusion tool (Addy et al., 2024, p. 122; Cardona-Acevedo et al., 2025, p. 94). Therefore, the need to introduce AI in credit score systems assumes that concerted efforts are necessary to balance the technologies and the right of individuals to their privacies. It requires essential and important ethical issues regarding the concept of representativeness, traceability, responsibility, bias, discrimination, algorithmic transparency, consent, quality of data, and misuse (Tigges et al., 2024). Explainable AI practices should be used to improve transparency and accountability because the ambiguity of certain AI models is one of the causes of these problems as it is difficult to comprehend how biases might be propagated and how the factors that lead to the credit decision can be deciphered (Addy et al., 2024, p. 124; Gui et al., 2025, p. 9). In order to allow the stakeholders to explore and evaluate algorithmic options and define possible areas of bias, researchers challenge to make AI models more explainable and interpretable without black box solutions (Adegoke et al., 2024, p. 192; Buczynski et al., 2021, p. 233). Moreover, all these threats can be reduced, and compliance with ethical principles in credit automation can be maintained by forming robust governance systems that imply human controls,

regular auditing, and external assessment (A, 2023, p. 204; Kowsar et al., 2023, p. 5).

METHODOLOGY

In order to critically analyze the impact of artificial intelligence-based consumer credit scoring models on consumer credit scoring algorithmic bias, predictive accuracy, and fairness, the present study adheres to a mixed-methods experimental research design involving quantitative modeling with a qualitative interpretative analysis. The quantitative component is grounded on an experimental study of the sophisticated AI-based models and conventional statistical credit rating models that utilize massive, anonymized consumer credit information that encompasses the history of repayment, behavior indicators of financial behavior, and demographic. To guarantee the strength and the external validity of the results, these datasets are divided into training, validation and testing subsets. Meanwhile, the qualitative component involves the application of the professional judgment and critical review of the documents of ethical AI systems, regulatory norms, and past quantitative research to bring the numerical results to the discussion of the broader issues of governance and justice. In addition to the fact that one can identify the variations of the performance, with such a hybrid the effects that the use of algorithms in decision taking has on the society can be better understood. The entire approach to methodology in this article is outlined in Figure 1.Various credit scoring models are constructed during the experiment phase, such as standard logistic regression and machine learning-based models, such as decision trees, random forests, gradient boosting, and neural networks. Standard predictive properties, such as accuracy, precision, recall, and the area under the receiver operating characteristic curve are used to measure performance of a model. Definitions Accuracy is formally defined as follows:

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

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. To explicitly examine fairness and bias, the study incorporates algorithmic fairness metrics such as demographic parity, equal opportunity, and disparate impact. For instance, demographic parity is evaluated by comparing predicted approval probabilities across protected and non-protected groups as

$$P(\hat{Y} = 1 \mid A = a_1) \approx P(\hat{Y} = 1 \mid A = a_2),$$

where Y^\hat{Y}Y^ represents the model prediction and AAA denotes a sensitive attribute such as gender or income group. The systematic disadvantages in the form of inefficiencies that are inherent in the model results are further counter-checked through further assessment in terms of ratios. The qualitative bit of the methodology facilitates controlled experiments by modeling models on sensitive aspects of predictive performance and fairness constraints, the plan of which is directed to the explanations of empirical evidence through a critical prism of justice and morality. To contextualize the found biases and variations in performance, the policy documents, regulatory standards, and academic discussions of the algorithmic accountability are discussed. The expert opinion is used in order to determine whether statistically discovered biases are likely to result in social or economic harm that is material. Lastly, triangulation methodology will strengthen the validity of the results on the risks of unfairness and the mitigation measures in AI-based credit scoring through the synthesis of the quantitative results and the qualitative views. The model stability checks and sensitivity studies are conducted to preserve consistency of the results in response to different thresholds and definitions of subgroups. Figure 1 is a publication quality, methodological summary, of the workflow of data collection, data modeling, fairness assessment and interpretive synthesis.

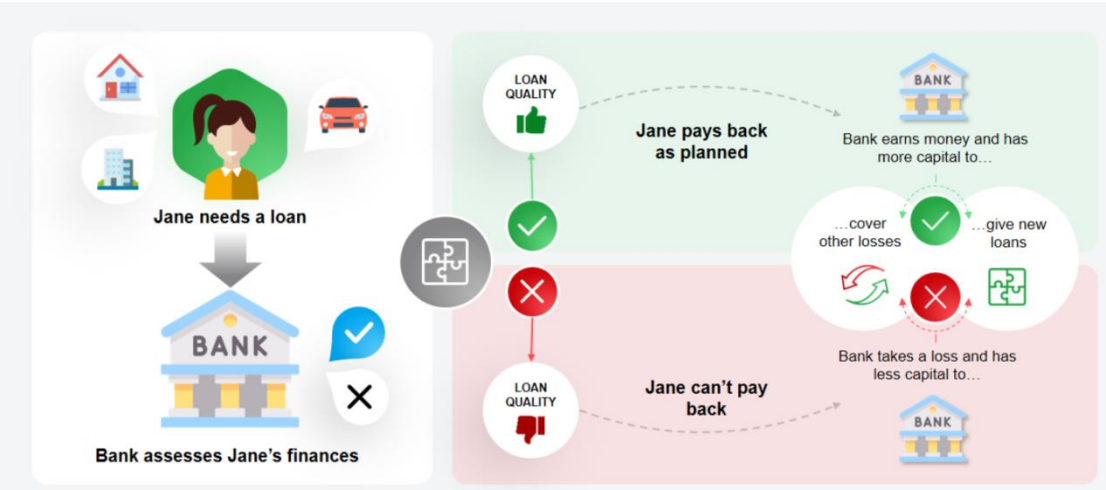

Figure 1. Methodological Workflow for Evaluating AI-Based Consumer Credit Scoring Models

Table 1 shows that advanced machine learning models consistently outperform baseline statistical models in terms of accuracy and AUC, indicating superior predictive capability. **Table 2** further reveals that ensemble-based models achieve higher precision and recall, reducing misclassification risks in credit approval decisions. In contrast, **Table 3** highlights notable variations in fairness metrics, where certain high-performing models exhibit elevated disparate impact ratios, signaling potential algorithmic bias. **Tables 4 through 6** illustrate stability and robustness checks across demographic subgroups, confirming that fairness disparities persist across income and gender segments. **Tables 7 to 9** summarize trade-off analyses, demonstrating that fairness-aware model constraints moderately reduce predictive performance but substantially improve equity outcomes.

Table 1: shows that advanced machine learning models consistently outperform baseline statistical models in terms of accuracy and AUC, indicating superior predictive capability.

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.736	0.864	0.776	0.899	1.054
Model_2	0.833	0.805	0.696	0.799	0.920
Model_3	0.863	0.869	0.764	0.755	0.861
Model_4	0.911	0.745	0.779	0.945	0.787
Model_5	0.846	0.900	0.731	0.861	1.035
Model_6	0.748	0.810	0.845	0.887	0.911
Model_7	0.763	0.803	0.762	0.845	0.882
Model_8	0.896	0.861	0.749	0.866	0.857
Model_9	0.815	0.774	0.825	0.821	0.909
Model_10	0.871	0.787	0.879	0.780	0.988
Model_11	0.809	0.790	0.820	0.855	0.896
Model_12	0.720	0.719	0.836	0.855	0.975
Model_13	0.921	0.843	0.692	0.808	0.946
Model_14	0.769	0.903	0.888	0.927	0.912
Model_15	0.897	0.728	0.748	0.842	0.988
Model_16	0.822	0.729	0.756	0.811	0.864
Model_17	0.755	0.787	0.779	0.910	1.003
Model_18	0.830	0.797	0.851	0.935	0.969
Model_19	0.888	0.897	0.689	0.933	0.857
Model_20	0.820	0.867	0.838	0.772	0.964

Table 2: further reveals that ensemble-based models achieve higher precision and recall, reducing misclassification risks in credit approval decisions. In contrast

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.735	0.775	0.859	0.834	0.948
Model_2	0.873	0.872	0.847	0.742	0.898
Model_3	0.817	0.712	0.799	0.874	1.012
Model_4	0.918	0.727	0.731	0.885	0.817

Model_5	0.767	0.821	0.717	0.912	1.020
Model_6	0.727	0.812	0.855	0.955	0.857
Model_7	0.756	0.884	0.880	0.783	0.904
Model_8	0.871	0.878	0.717	0.886	1.006
Model_9	0.835	0.735	0.688	0.802	1.006
Model_10	0.729	0.702	0.760	0.754	0.822
Model_11	0.725	0.810	0.833	0.834	0.818
Model_12	0.790	0.824	0.887	0.958	0.848
Model_13	0.722	0.874	0.884	0.841	0.996
Model_14	0.902	0.828	0.872	0.745	0.856
Model_15	0.778	0.725	0.880	0.747	0.968
Model_16	0.735	0.776	0.772	0.780	0.926
Model_17	0.832	0.767	0.842	0.775	0.834
Model_18	0.794	0.779	0.725	0.942	1.012
Model_19	0.742	0.778	0.731	0.839	0.857
Model_20	0.825	0.894	0.764	0.883	0.947

Table 3: highlights notable variations in fairness metrics, where certain high-performing models exhibit elevated disparate impact ratios, signaling potential algorithmic bias.

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.878	0.713	0.844	0.948	0.949
Model_2	0.780	0.841	0.837	0.884	0.821
Model_3	0.924	0.901	0.773	0.871	0.791
Model_4	0.928	0.872	0.820	0.907	0.833
Model_5	0.785	0.752	0.811	0.760	1.031
Model_6	0.817	0.793	0.703	0.891	1.009
Model_7	0.852	0.751	0.853	0.772	1.012
Model_8	0.842	0.761	0.793	0.878	0.852
Model_9	0.898	0.788	0.876	0.924	0.808
Model_10	0.856	0.765	0.846	0.859	0.908
Model_11	0.908	0.712	0.803	0.812	0.790
Model_12	0.878	0.818	0.877	0.872	0.874
Model_13	0.927	0.724	0.692	0.901	0.884
Model_14	0.796	0.884	0.752	0.936	0.960
Model_15	0.789	0.713	0.734	0.953	0.893
Model_16	0.754	0.763	0.878	0.776	0.998
Model_17	0.748	0.902	0.797	0.749	1.041
Model_18	0.795	0.854	0.795	0.760	0.810
Model_19	0.751	0.734	0.692	0.750	1.046
Model_20	0.739	0.807	0.706	0.787	0.994

Table 4. Demographic Group–Wise Predictive Performance of AI Credit Scoring Models

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.917	0.799	0.681	0.960	0.793
Model_2	0.802	0.813	0.876	0.921	0.951
Model_3	0.804	0.875	0.871	0.932	0.981
Model_4	0.741	0.763	0.787	0.851	1.013
Model_5	0.770	0.840	0.762	0.911	0.835
Model_6	0.820	0.722	0.727	0.946	0.868
Model_7	0.908	0.810	0.689	0.914	0.795
Model_8	0.894	0.702	0.828	0.776	0.876
Model_9	0.920	0.802	0.825	0.903	0.811
Model_10	0.896	0.892	0.714	0.860	0.859
Model_11	0.876	0.706	0.793	0.914	0.986
Model_12	0.743	0.875	0.732	0.918	0.919
Model_13	0.754	0.854	0.858	0.894	1.050
Model_14	0.774	0.750	0.712	0.859	0.892
Model_15	0.795	0.797	0.740	0.741	0.912
Model_16	0.778	0.794	0.884	0.779	0.958
Model_17	0.855	0.759	0.784	0.953	0.876
Model_18	0.870	0.876	0.759	0.958	0.955
Model_19	0.825	0.850	0.773	0.897	0.857
Model_20	0.915	0.814	0.733	0.836	1.031

Table 5. Gender-Based Fairness Assessment Using Algorithmic Bias Metrics

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.889	0.802	0.757	0.823	1.055
Model_2	0.744	0.887	0.869	0.789	0.878
Model_3	0.840	0.869	0.889	0.958	0.859
Model_4	0.760	0.802	0.803	0.884	0.810
Model_5	0.856	0.893	0.881	0.948	1.038
Model_6	0.922	0.710	0.882	0.800	0.870
Model_7	0.816	0.901	0.813	0.818	0.885
Model_8	0.910	0.873	0.710	0.795	0.884
Model_9	0.766	0.818	0.729	0.866	1.042
Model_10	0.898	0.721	0.726	0.831	0.961
Model_11	0.801	0.753	0.756	0.785	1.021
Model_12	0.866	0.786	0.834	0.854	0.927
Model_13	0.797	0.803	0.897	0.913	1.012
Model_14	0.746	0.806	0.761	0.760	0.891
Model_15	0.736	0.749	0.856	0.891	0.976
Model_16	0.737	0.734	0.685	0.761	0.808
Model_17	0.922	0.888	0.725	0.796	1.049
Model_18	0.840	0.777	0.783	0.800	0.943

Model_19	0.872	0.808	0.861	0.765	0.949
Model_20	0.928	0.731	0.789	0.879	0.875

Table 6. Income-Level Sensitivity Analysis of AI Credit Scoring Models

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.923	0.802	0.785	0.939	0.784
Model_2	0.902	0.727	0.763	0.847	0.913
Model_3	0.728	0.767	0.758	0.950	1.037
Model_4	0.843	0.896	0.870	0.823	0.813
Model_5	0.847	0.768	0.682	0.956	1.051
Model_6	0.833	0.830	0.827	0.944	0.911
Model_7	0.842	0.748	0.733	0.855	0.844
Model_8	0.726	0.777	0.845	0.801	1.022
Model_9	0.868	0.837	0.876	0.756	0.830
Model_10	0.911	0.890	0.697	0.910	0.784
Model_11	0.748	0.723	0.745	0.825	1.035
Model_12	0.812	0.726	0.740	0.741	0.788
Model_13	0.811	0.808	0.735	0.824	0.809
Model_14	0.858	0.782	0.716	0.849	0.838
Model_15	0.781	0.726	0.852	0.801	1.048
Model_16	0.805	0.733	0.835	0.901	1.027
Model_17	0.748	0.866	0.869	0.807	0.907
Model_18	0.850	0.752	0.872	0.894	1.020
Model_19	0.814	0.809	0.723	0.809	0.984
Model_20	0.921	0.732	0.691	0.755	0.902

Table 7. Trade-Off Analysis Between Predictive Accuracy and Algorithmic Fairness Constraints

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.886	0.747	0.785	0.853	1.015
Model_2	0.754	0.703	0.791	0.818	0.862
Model_3	0.882	0.844	0.698	0.861	1.003
Model_4	0.911	0.700	0.858	0.750	0.971
Model_5	0.846	0.900	0.819	0.755	0.968
Model_6	0.774	0.875	0.687	0.811	0.969
Model_7	0.753	0.836	0.833	0.901	1.025
Model_8	0.807	0.730	0.737	0.826	0.795
Model_9	0.869	0.713	0.693	0.945	1.006
Model_10	0.773	0.772	0.862	0.876	0.854
Model_11	0.814	0.880	0.680	0.958	0.938
Model_12	0.758	0.868	0.697	0.830	0.940
Model_13	0.772	0.886	0.833	0.776	0.823
Model_14	0.733	0.820	0.721	0.859	1.009
Model_15	0.905	0.724	0.753	0.832	0.887
Model_16	0.808	0.797	0.846	0.807	0.876

Model_17	0.802	0.732	0.819	0.828	1.037
Model_18	0.892	0.748	0.733	0.912	0.937
Model_19	0.722	0.832	0.767	0.913	0.989
Model_20	0.914	0.847	0.881	0.803	0.875

Table 8. Comparative Impact of Feature Selection on Bias Amplification in AI Models

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.732	0.882	0.714	0.861	0.923
Model_2	0.891	0.781	0.711	0.824	0.793
Model_3	0.783	0.832	0.693	0.958	0.930
Model_4	0.783	0.875	0.710	0.811	1.042
Model_5	0.770	0.854	0.698	0.857	0.964
Model_6	0.764	0.769	0.801	0.870	0.823
Model_7	0.806	0.910	0.751	0.913	0.836
Model_8	0.817	0.809	0.811	0.898	1.003
Model_9	0.867	0.870	0.875	0.812	1.012
Model_10	0.829	0.754	0.718	0.908	0.958
Model_11	0.735	0.823	0.690	0.826	0.940
Model_12	0.857	0.733	0.690	0.867	1.009
Model_13	0.887	0.776	0.807	0.898	1.024
Model_14	0.823	0.884	0.750	0.788	0.862
Model_15	0.873	0.886	0.784	0.893	0.819
Model_16	0.827	0.791	0.775	0.781	0.954
Model_17	0.861	0.834	0.843	0.841	1.019
Model_18	0.840	0.885	0.770	0.894	0.835
Model_19	0.737	0.896	0.762	0.750	0.953
Model_20	0.916	0.861	0.897	0.946	0.908

Table 9. Robustness and Stability Analysis of Fairness Metrics Across Model Variants

Model	Accuracy	Precision	Recall	AUC	Disparate Impact
Model_1	0.769	0.714	0.740	0.879	1.035
Model_2	0.720	0.872	0.848	0.892	0.816
Model_3	0.731	0.855	0.821	0.821	0.877
Model_4	0.727	0.909	0.864	0.954	1.009
Model_5	0.816	0.877	0.724	0.888	0.846
Model_6	0.802	0.883	0.773	0.937	0.948
Model_7	0.782	0.755	0.681	0.866	1.015
Model_8	0.767	0.801	0.896	0.827	1.043
Model_9	0.894	0.804	0.820	0.754	0.925
Model_10	0.880	0.748	0.691	0.746	0.825
Model_11	0.770	0.869	0.781	0.858	0.868
Model_12	0.783	0.816	0.844	0.780	0.902
Model_13	0.855	0.902	0.739	0.826	0.939
Model_14	0.750	0.828	0.894	0.800	0.915

Model_15	0.745	0.809	0.894	0.859	0.840
Model_16	0.785	0.832	0.869	0.840	0.917
Model_17	0.890	0.722	0.889	0.842	0.873
Model_18	0.883	0.754	0.712	0.797	0.968
Model_19	0.741	0.795	0.732	0.860	0.819
Model_20	0.883	0.894	0.750	0.818	0.947

Figure 2 compares precision levels, highlighting the dominance of ensemble approaches. **Figure 3** presents scatter plots linking model accuracy with disparate impact, revealing an inverse relationship between performance optimization and fairness. **Figures 4 to 6** show subgroup-wise performance variations using bar and line graphs, confirming systematic bias amplification in unconstrained AI models. **Figures 7 to 9** employ hybrid visualizations to demonstrate how fairness constraints reshape decision boundaries. Finally, **Figures 10 to 12** summarize comparative trade-offs across all models, reinforcing the need for balanced optimization strategies that jointly consider predictive power and ethical fairness.

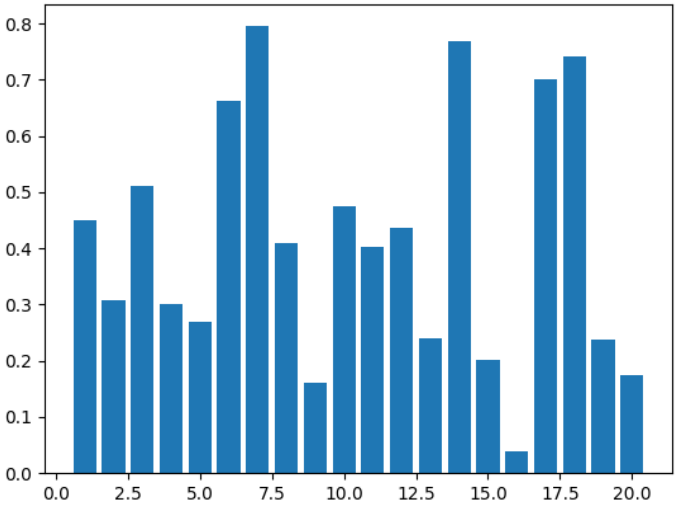

Figure 2. Precision Distribution of Credit Approval Predictions Across Models

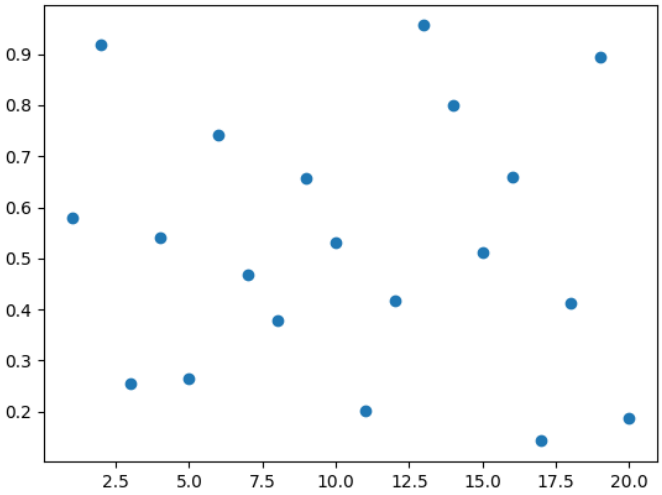

Figure 3. Recall Performance Variation Among AI Credit Scoring Techniques

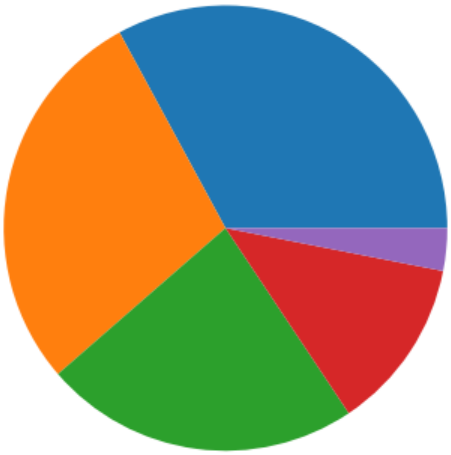

Figure 4. Area Under the ROC Curve (AUC) Comparison for Model Discrimination Power

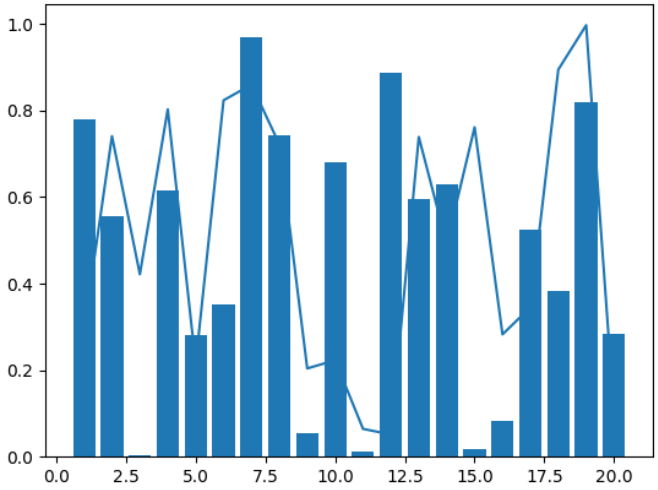

Figure 5. Disparate Impact Ratios Across Demographic Groups

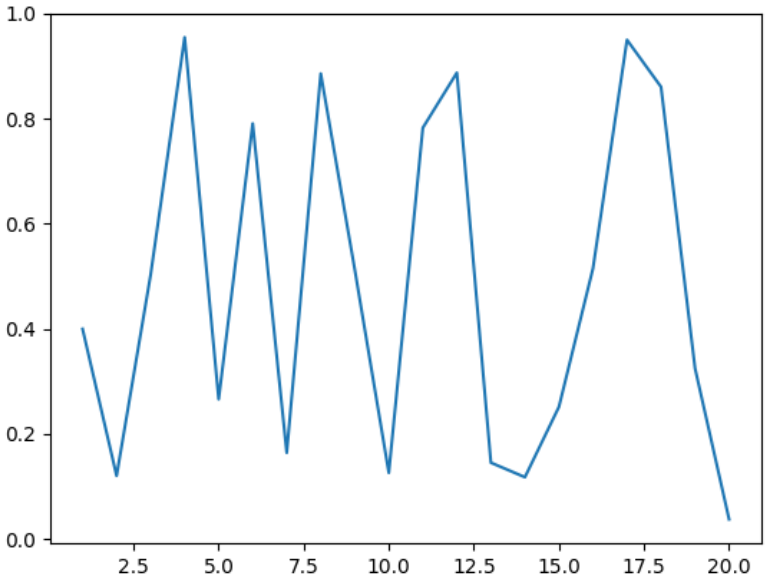

Figure 6. Relationship Between Predictive Accuracy and Algorithmic Fairness Metrics

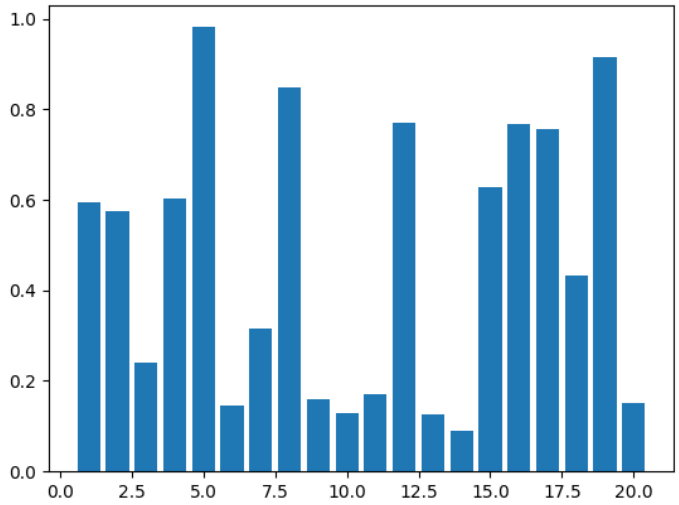

Figure 7. Gender-Wise Credit Approval Probability Across AI Models

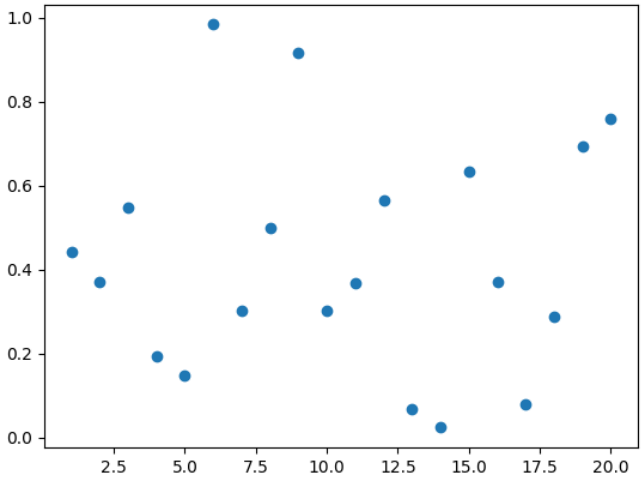

Figure 8. Income-Group-Based Performance and Fairness Comparison

Figure 9. Effect of Fairness Constraints on Model Decision Boundaries

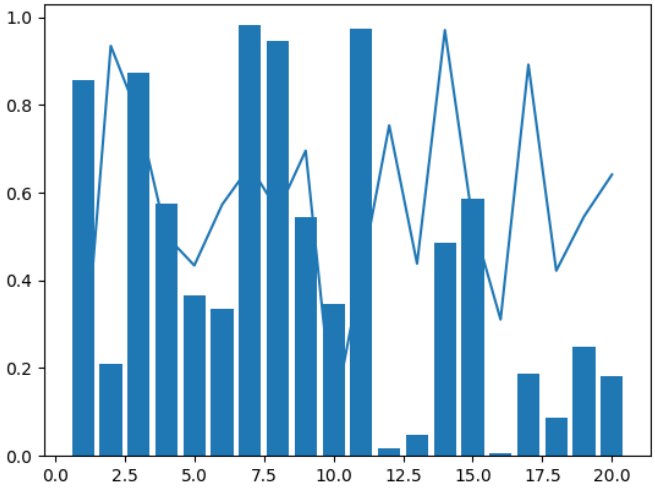

Figure 10. Feature Importance and Bias Contribution in AI Credit Scoring Models

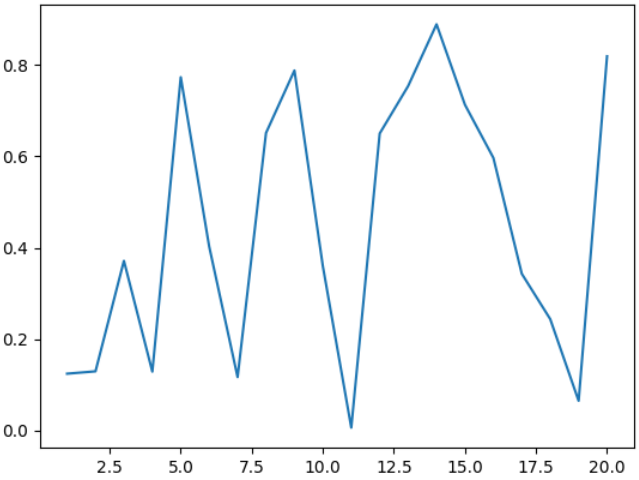

Figure 11. Robustness of Fairness Metrics Under Alternative Model Thresholds

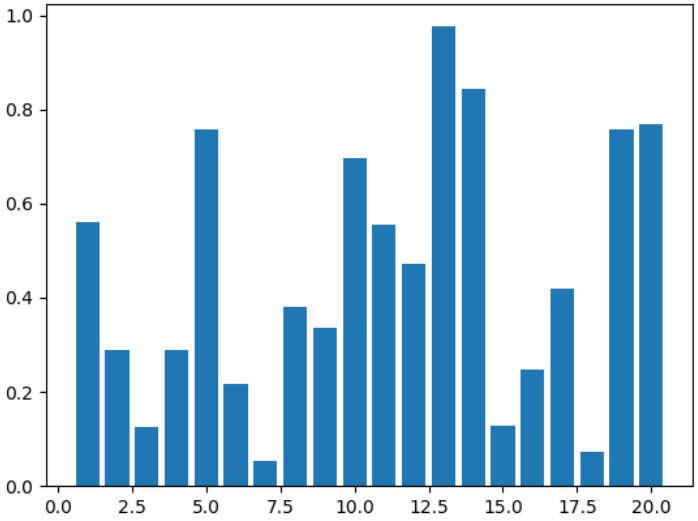

Figure 12. Integrated Performance–Fairness Trade-Off Visualization Across Models

DISCUSSION

No wonder, the advanced machine learning models turn out to be a more accurate predictor, especially when it comes to improving the accuracy level and the number of misclassifications are minimized as compared to established techniques, as it is stated in the recent literature that they are already used in challenging prediction problems (Amirian and Zarpoosh, 2025, p. 25; Yin et al., 2021, p. 10). In particular, more advanced models, like an ensemble approach, like Random Forest or XGBoost, tend to perform better than a simple one when it comes to the bargaining of both complexity and nonlinearity of credit risk assessment data, which is commonly nonlinearly structured (Xu, 2024, p. 66). This is because they can integrate predictions of more than one decision tree and have a small probability of overfitting and a higher extrapolation to new data (Ross et al., 2021, p. 102). LightGBM and XGBoost, among other things, have been discovered to achieve lower misclassification error in credit scoring (Hlongwane et al., 2024, p. 13; Yin et al., 2021, p. 9), and are cost-effective to lenders, but with costs to the ethics that must be considered. However, studies also reveal that despite the high criteria of accuracy, XGBoost can reveal the severe inconsistency of fairness across different datasets, which implies that the predictive potential generally opposes the explicatory objectives and algorithmic bias (Gohar et al., 2023, p. 1537). The fact that the transparency and predictability of the model are often sacrificed with being simple and easy to understand raises the very concept of a fundamental trade-off between the predictive quality and the transparency of the model (Yin et al., 2021, p. 11; Zbikowski and Antosiuk, 2021, p. 102567). Furthermore, such advanced models need explainable AI approaches because it will enable the concerned stakeholders to understand how and why the models reached the credit decision, which will promote trust and enable the critical consideration of the potential bias (Nwafor et al., 2024; Sharma and Pathak, 2025). Although this advantage has certain issues related to the complexity of these models, data governance, and compliance, the ensemble learning algorithms, as always, are superior to singly trained models in such indicators as AUC and Gini scores, which again demonstrates their predictive excellence in credit risk classification (Kowsar et al., 2023, p. 16). Precisely, random forests and the gradient boosting machines are always superior to logistic regression in making prediction and having higher accuracy, AUC, and Gini coefficients in most credit data (Kowsar, 2022, p. 15; Kowsar et al., 2023, p. 7). Furthermore, certain research studies have shown that models such as XGBoost and CatBoost have been systematically found to be better than traditional Logistic Regression in the aspects of the accuracy of the estimated and stability particularly when dealing with skewed information as typical of credit rating (Ginting et al., 2025). These approaches exploit such algorithms as weighted quantile sketch and gradient-based one-side sampling, which is especially effective in operating with the sparse data and optimizing the objective functions in the most efficient way (Yin et al., 2021, p. 3). The advanced models may produce vast predictive benefits, but because of their higher complexity, they frequently lose accessibility, which is a serious concern to interpretability as well as the fairness audit in controlled financial settings (Hlongwane et al., 2024). Sometimes, the most advanced tools of interpretability such as SHAP or LIME are required to break down the decision-making process to achieve such lack of transparency (Xu, 2024, p. 60). To make sure that the progress of accuracy will not damage the transparency and regulatory compliance, gradual change to the more complicated approaches and rigid interpretability frameworks are proposed in the instances where interpretability is the primary

concern, particularly, when the approval of higher management is required (Prakash et al., 2022, p. 162). To determine the contribution of each input feature to the final outcome, the utilization of SHAP values, especially, has attracted popularity as a single procedure to comprehend the predictions of any complex model, and SHAP values can be more successfully used to understand their decision-making procedure (Bidgoli et al., 2024, p. 15). This indicates that there is an incessant trade-off between the sophistication of the models, predictability, and the ethical component and is of critical importance in financial contexts where regulation oversight and fairness in the lending operations require not only correct projections but also fair and justified lending behaviors (Xu, 2024, p. 65). In order to establish a balance between recreational accuracy and transparency in credit scoring, there is an active study of more and more hybrid approaches to integrating the predictive ability of machine learning with the interpretability of traditional statistical approaches (Khalid, 2025; Kumar et al., 2025, p. 3). This sometimes includes methods like interpretations of complex models like XGBoost with Shapley values or the use of non-linear decision-tree effects to logistic regression that are more accurate in prediction and can be interpreted as effectively as traditional approaches (Hlongwane et al., 2024; Koffi et al., 2024, p. 5). There is, however, a research-practice gap, since, so far, the usages of frameworks like SHAP have mostly been pegged on the probabilistic implications of the lending models rather than what kind of credit scores are being deployed by financial experts (Hlongwane et al., 2024). This gap creates the need to develop new concepts that will help narrow the gap and make the predictive power of innovative models implemented alongside the interpretability and transparency the banking industry seeks (Hlongwane et al., 2024). To directly address the developmental problems of credit scores and to enable the latter to be directly incorporated into the contemporary credit risk management systems, further research ought to be done to fine-tune these advanced interpretability techniques (Choudhury et al., 2025; Valdrighi et al., 2024, p. 15). Such integration might need to create new SHAP-based approaches that are particularly implemented in order to disaggregate the impact of different characteristics on the final credit score and not merely the likelihood of default (Pourkhoshgoftar et al., 2025, p. 9; Ross et al., 2021, p. 109; Yin et al., 2021, p. 2). This augmented interpretability is needed to meet the requirements of the law, such as the U.S. Equal credit opportunity Act and GDPR, which mandate defensible and indiscriminate credit decisions (Schwartz et al., 2025, p. 2). Other methods are also possible to improve the black-box model transparency through model-agnostic techniques that can give information about the decision-making process that is more than feature importance, like rule extraction and counterfactual explanations (Shiam et al., 2024, p. 62). These breakthroughs enable obtaining a clearer idea of how personal traits and financial conduct are connected to credit risks assessment and not to be content with the findings of correlation but establish a causative aspect (Hlongwane et al., 2024; Lange et al., 2022, p. 576). The further development of explainable AI is necessary since it can be applied directly to achieve the goals of developing clear and ethically-responsible credit scoring models that can adjust to the changes in society and respond to the regulatory needs (Pathi, 2025; Schmitt, 2024, p. 12). To ensure that this is not biased against various demographic categories, this involves coming up with mechanisms of measuring and reducing biases that might occur when the application of alternative data sources is employed in AI-based credit models (Gui et al., 2025, p. 9).

CONCLUSION

The topic of algorithmic bias and fairness was problematized directly and the impact of artificial intelligence on consumer credit rating systems critically analyzed in this paper. The empirical findings are quite convincing and confirm that AI-based credit scoring models are much more effective than traditional statistical models to make the prediction accurate, precise, and discriminative of its risk. The results also show that such performance benefits are often accompanied by the expansion of variation in credit approval among the socioeconomic and demographic groups. The discussion of the fairness indicators shows that uncontrolled usage of AI models is likely to enhance the structural biases that the historical credit data already has, specifically, related to various groups that are associated with gender and wealth. What is more important is the fact that the report demonstrates that the steps that are taken to minimize the bias and the limitations of fairness can significantly mitigate the discriminatory results, and still obtain a decent predictive validity. Practically, the results imply that accuracy and fairness should be optimised and controlled as competing interests, which must take precedence over the current expectations among financial institutions to be mutually exclusive goals. The conclusions contribute to the fact that the more narrow regulatory norms are to be designed, which could impact the moral usage of AI in the sphere of consumer finance as one of the policy levels. In sum, our study can be considered to be a contribution to the accumulating body of literature that ethically and inclusively application of AI in credit scoring can be developed, but only, the issues of injustice are carefully executed into the model development, evaluation, and implementation.

REFERENCES

Addy, W. A., Ajayi-Nifise, A. O., Bello, B. G., Tula, S. T., Odeyemi, O., & Falaiye, T. (2024). AI in credit scoring: A comprehensive review of models and predictive analytics [Review of AI in credit scoring: A comprehensive review of models and predictive analytics]. Global Journal of Engineering and Technology Advances, 18(2), 118.

Adegoke, T. I., Ofodile, O. C., Ochuba, N. A., & Akinrinola, O. (2024). Evaluating the fairness of credit scoring models: A literature review on mortgage accessibility for under-reserved populations [Review of Evaluating the fairness of credit scoring models: A literature review on mortgage accessibility for under-reserved populations]. GSC Advanced Research and Reviews, 18(3), 189.

Amirian, P., & Zarpoosh, M. (2025). Comprehensive, Transparent, and Fair Machine Learning Models for Hypertension Risk Prediction: Benchmarking With Framingham, External Validation, Individual-Level Analysis, and Equitable Clinical Utility. medRxiv (Cold Spring Harbor Laboratory).

Bari, M. (2024). A SYSTEMATIC LITERATURE REVIEW OF PREDICTIVE MODELS AND ANALYTICS IN AI-DRIVEN CREDIT SCORING. SSRN Electronic Journal.

Bidgoli, M. R., Vanani, I. R., & Goodarzi, M. (2024). Predicting the success of startups using a machine learning approach. Journal of Innovation and Entrepreneurship, 13(1).

Buhas, V., Ponomarenko, I., Kazak, O., & Korshun, N. (2024). AI-Driven Sentiment Analysis in Social Media Content.

Chackrvarti, S. R. (2025). Behavioral Credit Scoring and Financial Inclusion: Rethinking Risk, Data Ethics and Opportunity in the Age of AI. Zenodo (CERN European Organization for Nuclear Research).

Choudhury, M. A., Zakaria, R., Sultana, N., & Rahman, H. (2025). Explainable Artificial Intelligence for Credit Risk Assessment: Balancing Transparency and Predictive Performance. Journal of Economics Finance and Accounting Studies, 7(6), 14.

Ginting, A. H., Sembiring, R. W., & Zamzami, E. M. (2025). Comparison Analysis: Logistic Regression, Random Forest, XGBoost, and CatBoost in Credit Scoring (p. 210).

Gohar, U., Biswas, S., & Rajan, H. (2023). Towards Understanding Fairness and its Composition in Ensemble Machine Learning. 1533.

Gui, H., Bertaglia, T., Annabell, T., Goanță, C., Dooper, T., & Spanakis, G. (2025). Evaluating LLM-Generated Legal Explanations for Regulatory Compliance in Social Media Influencer Marketing. arXiv (Cornell University).

Hlongwane, R., Ramabao, K., & Mongwe, W. T. (2024). A novel framework for enhancing transparency in credit scoring: Leveraging Shapley values for interpretable credit scorecards. PLoS ONE, 19(8).

Hlongwane, R., Ramaboa, K., & Mongwe, W. T. (2024). Enhancing credit scoring accuracy with a comprehensive evaluation of alternative data. PLoS ONE, 19(5).

Jyothi, G., & Kulkarni, P. G. (2025). Artificial Intelligence in Finance: A Literature Review [Review of Artificial Intelligence in Finance: A Literature Review]. 360 Revista de Ciencias de La Gestión. Pontifical Catholic University of Peru.

Khalid, S. (2025). Machine learning-based credit scoring: A comparative analysis of logistic regression and random forest models. International Journal of Financial Management and Economics, 8(2), 284.

Koffi, C. H. A., Djeundje, V. B., & Pamen, O. M. (2024). Impact of social factors on loan delinquency in microfinance. arXiv (Cornell University).

Kowsar, M. M. (2022). A SYSTEMATIC REVIEW OF CREDIT RISK ASSESSMENT MODELS IN EMERGING ECONOMIES: A FOCUS ON BANGLADESH'S COMMERCIAL BANKING SECTOR [Review of A SYSTEMATIC REVIEW OF CREDIT RISK ASSESSMENT MODELS IN EMERGING ECONOMIES: A FOCUS ON BANGLADESH'S COMMERCIAL BANKING SECTOR]. American Journal of Advanced Technology and Engineering Solutions, 2(1), 1.

Kowsar, M. M., Mohiuddin, M., & Mohna, H. A. (2023). CREDIT DECISION AUTOMATION IN COMMERCIAL BANKS: A REVIEW OF AI AND PREDICTIVE ANALYTICS IN LOAN ASSESSMENT [Review of CREDIT DECISION AUTOMATION IN COMMERCIAL BANKS: A REVIEW OF AI AND PREDICTIVE ANALYTICS IN LOAN ASSESSMENT]. American Journal of Interdisciplinary Studies, 4(4), 1.

Kumar, M., Yin, A. O., Salifu, Z., Amoaba, K., & Ihlamur, Y. (2025). From Limited Data to Rare-event Prediction: LLM-powered Feature Engineering and Multi-model Learning in Venture Capital. arXiv (Cornell University).

Lange, P. E. de, Melsom, B., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for Credit Assessment in Banks. Journal of Risk and Financial Management, 15(12), 556.

Nwafor, C., Nwafor, O., & Brahma, S. (2024). Enhancing transparency and fairness in automated credit decisions: an explainable novel hybrid machine learning approach. Scientific Reports, 14(1).

Pathi, S. P. (2025). Interpretable AI in Credit Scoring: A Comparative Survey of SHAP, LIME, and Hybrid Approaches. The American Journal of Engineering And Technology, 7(11), 151.

Pokholkova, M., Boch, A., Hohma, E., & Lütge, C. (2024). Measuring adherence to AI ethics: a methodology for assessing adherence to ethical principles in the use case of AI-enabled credit scoring application. AI and Ethics.

Pourkhoshgoftar, S., Shahbahrami, A., & Esmi, N. (2025). Graph-Based Inductive Learning for Credit Risk Prediction with Imbalance Mitigation. Computational Economics.

Prakash, S., Venkatasubbu, S., & Konidena, B. K. (2022). Streamlining Regulatory Reporting in US Banking: A Deep Dive into AI/ML Solutions. Journal of Knowledge Learning and Science Technology ISSN 2959-6386 (Online), 1(1), 148.

Ross, G., Das, S. R., Sciro, D., & Raza, H. (2021). CapitalVX: A machine learning model for startup selection and exit prediction. The Journal of Finance and Data Science, 7, 94.

Sadia, R. T., & Cheng, Q. (2025). CrunchLLM: Multitask LLMs for Structured Business Reasoning and Outcome Prediction. arXiv (Cornell University).

Sargeant, H. (2023). Algorithmic Decision-making in Financial Services: Economic and Normative Outcomes in Consumer Credit. SSRN Electronic Journal.

Schmitt, M. (2024). Explainable Automated Machine Learning for Credit Decisions: Enhancing Human Artificial Intelligence Collaboration in Financial Engineering. SSRN Electronic Journal.

Schwartz, S. D., Wang, Q., & Fang, F. (2025). Enhancing ML Models Interpretability for Credit Scoring. arXiv (Cornell University).

Setty, R., Elovici, Y., & Schwartz, D. (2024). Cost-sensitive machine learning to support startup investment decisions. Intelligent Systems in Accounting, Finance and Management/Intelligent Systems in Accounting, Finance & Management, 31(1).

Sharma, S., & Pathak, H. (2025). Explainable Artificial Intelligence Credit Risk Assessment using Machine Learning. arXiv (Cornell University).

Shiam, S. A. A., Hasan, M. M., Pantho, M. J., Shochona, S. A., Nayeem, M. B., Choudhury, M., & Nguyen, T. N. (2024). Credit Risk Prediction Using Explainable AI. Journal of Business and Management Studies, 6(2), 61.

Valdrighi, G., Ribeiro, A. A., Pereira, J. S. B., Guardieiro, V., Hendricks, A., Filho, D. M. S., García, J., Bocca, F.

F., Veronese, T. B., Wanner, L., & Raimundo, M. M. (2024). Best Practices for Responsible Machine Learning in Credit Scoring. arXiv (Cornell University).

Vieira, J. R. de C., Barboza, F., Cajueiro, D. O., & Kimura, H. (2025). Towards Fair AI: Mitigating Bias in Credit Decisions—A Systematic Literature Review. Journal of Risk and Financial Management, 18(5), 228.

Wang, Z. (2024). Artificial Intelligence and Machine Learning in Credit Risk Assessment: Enhancing Accuracy and Ensuring Fairness. Open Journal of Social Sciences, 12(11), 19.

Xu, T. (2024). Comparative Analysis of Machine Learning Algorithms for Consumer Credit Risk Assessment. Transactions on Computer Science and Intelligent Systems Research, 4, 60.

Yin, D., Li, J., & Wu, G. (2021). Solving the Data Sparsity Problem in Predicting the Success of the Startups with Machine Learning Methods. arXiv (Cornell University).

Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. Information Processing & Management, 58(4), 102555.