Scientific Research Publishing

Open Journal of Social Sciences, 2024, 12, 19-34
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Artificial Intelligence and Machine Learning in Credit Risk Assessment: Enhancing Accuracy and Ensuring Fairness

Zhiqin Wang

Department of Banking and Finance, Faculty of Business and Economics, Monash University, Melbourne, Australia
Email: wang.zhiqin@foxmail.com

How to cite this paper: Wang, Z. Q. (2024). Artificial Intelligence and Machine Learning in Credit Risk Assessment: Enhancing Accuracy and Ensuring Fairness. *Open Journal of Social Sciences, 12,* 19-34.
https://doi.org/10.4236/jss.2024.1211002

Received: October 10, 2024
Accepted: November 2, 2024
Published: November 5, 2024

Abstract

Credit risk assessment has become one of the major concerns in modern finance regarding informed lending decisions. Although several studies have used traditional logistic regression and linear discriminant analysis techniques, these have increasingly become inadequate tools in today's complex and data-rich environment. Such models often struggle with large datasets and nonlinear relationships, thus reducing their predictive power and adaptability. Artificial Intelligence (AI) and Machine Learning (ML) provide two of the most innovative approaches to credit risk modeling. This paper reviews a few ML models applied to improve the accuracy and efficiency of credit risk assessment, from Random Forests and Support Vector Machines to Neural Networks. Compared to the more traditional models, AI models can enhance predictive accuracy by using a wealth of structured and unstructured information, including alternative information sources such as social media activities and transaction history. However, despite noticeable advantages, there are some challenges concerning the use of AI in credit risk assessment, including model opaqueness, bias, and regulatory compliance. The nature of such a "black box", especially for deep learning algorithms, can limit their interpretability and complicate regulatory compliance and decision rationalization. To solve problems induced by this "black box" nature, explainable AI techniques, namely Shapley values and LIME, have been implemented to enhance the transparency of models and raise stakeholder trust in support systems for decision-making. This review aims to evaluate the current applications of AI and ML in credit risk assessment, weigh the strengths and limitations of various models, and discuss the ethical considerations and regulatory challenges linked to their adoption by credit institutions.

Keywords

Artificial Intelligence, Machine Learning, Credit Risk Assessment, Data Bias

1. Introduction

Accurate and fair assessment of credit risks is rudimentary in modern finance for informed lending decisions. Whereas the traditional models included logistic regression and linear discriminant analysis, these have recently become inadequate in today's complex and data-rich environments. Though transparent and regulator-friendly, these methods usually cannot capture borrowers' nuanced nonlinear behaviors, compromising predictive accuracy.

The change in credit risk modeling heralds AI and ML. Predictive power has radically improved by accessing large datasets and the ability to discern patterns not accessible to traditional analytics. However, interpretability, ethical issues, and regulatory compliance become issues with the translucent nature of "black box" models such as deep neural networks.

This review critically analyzes various state-of-the-art AI and ML models in credit risk assessment by comparing them with traditional methods and discussing the potential of such models to reshape financial decision-making. We analyze different models, such as Random Forests, Support Vector Machines, and Neural Networks, to discuss how these technologies can improve accuracy by considering ethical and regulatory considerations proper for widespread adoption. The paper will help achieve a balanced understanding of the responsible implementation of AI, hoping that technological advancements will translate into improvements in the consistency and transparency of credit risk assessment.

1.1. Traditional Credit Risk Modeling

For years, traditional models such as logistic regression and linear discriminant analysis have formed the foundation of credit risk assessment, a core function of the financial industry. These models conventionally rely on structured data, including borrowers' income, credit scores, and debt levels, to predict the likelihood of default (Breeden, 2020). Although these methods have served well for many decades, they are based on assumptions of linearity, which often fail to capture the complexity of borrowers' behavior in today's data-rich environment.

For instance, logistic regression assumes a linear relationship between the input variables and the outcome default or non-default, which can oversimplify real-world credit risk (Bussmann et al., 2020). Furthermore, traditional models often require significant human intervention to extract key features, potentially overlooking meaningful nonlinear interactions between variables (Addo et al., 2018).

Despite these deficiencies, traditional models remain popular, because they are transparent and acceptable to many regulators. Their interpretability advantages over more complex model types, neural networks are a good example of them

appealing to regulators explicitly required to explain credit decisions. However, as data becomes increasingly complex and voluminous, these models struggle to keep pace, highlighting the need for more advanced techniques such as AI and ML.

1.2. Credit Risk Assessment Models Using Artificial Intelligence and Machine Learning

Recently, AI and ML have emerged as powerful tools for credit risk assessment. These methods offer several advantages over traditional models, mainly due to their ability to process large and complex datasets and uncover nonlinear relationships. Examples of such models include Random Forests, Support Vector Machines (SVMs), and Neural Networks, which can learn from vast amounts of data and recognize patterns that traditional models might miss.

Random Forest, an ensemble learning model, builds multiple decision trees and then combines their outputs for better accuracy. It is highly effective in credit risk assessment because the relationship between borrower attributes and default probability is often complex and nonlinear. Research has shown that Random Forests outperform conventional models in predictive accuracy, especially when the model is enhanced with data sources such as social media activities and transaction history.

Neural Networks have also been shown to reveal complex nonlinear relationships among variables. They are instrumental when managing vast amounts of unstructured data, such as transaction histories or mobile phone usage data (Zhang & Yu, 2024). However, interpretability is considered one of the main disadvantages of neural networks, referred to as the "black box" problem-which makes it difficult for financial institutions to justify their credit decisions.

SVMs also show exciting potential for credit risk modeling, especially when dealing with unbalanced datasets, as is often the case where the number of defaulters is less than that of non-defaulters. Support Vector Machines seek an optimal hyperplane that separates the two classes: defaulters versus non-defaulters. In situations where generalizing from traditional models is difficult, SVMs are particularly robust.

1.3. Incorporation of Alternative Information Providers

One of the most crucial advantages of AI and ML models in evaluating credit risk is their ability to incorporate alternative data. Traditional models have relied heavily on structured data such as credit scores and income levels, which are inaccessible to all borrowers, particularly in emerging markets or those with incomplete credit histories (Breeden, 2020). In contrast, AI models can use unstructured data from social media activity, mobile phone usage, and transaction history to better understand a borrower's behavior.

For instance, Kou et al. (Kou et al., 2019) highlighted the contributions of social media data in enhancing credit risk assessment. AI models can better predict

financial reliability based on factors like the size of borrowers' social networks, posting frequencies, and sentiment analysis from their online activities. Mobile usage data, such as call logs and location information, can provide in-depth insights into a borrower's stability and predict their likelihood of default.

These alternative data sources are particularly useful for assessing borrowers with minimal traditional credit records, making them an integral part of modern credit risk models. However, the use of alternative data raises ethical concerns regarding data privacy and potential discrimination. Financial organizations should ensure that their alternative data practices comply with data protection regulations, such as the GDPR in Europe, and that their models are not discriminatory across different demographic groups.

1.4. Explainability and Transparency in Artificial Intelligence Models

One of the biggest challenges posed by AI and ML models in credit risk assessment is their lack of transparency. Traditional models, such as logistic regression, are comparatively easier to explain, allowing lending and regulatory bodies to understand how decisions are made. In contrast, many AI models, particularly neural networks, function as an impenetrable "black box", making it difficult to explain how they derive their predictions.

This lack of transparency creates serious problems in regulated industries like finance, where institutions must demonstrate the reasonableness of their lending decisions. Consequently, researchers have developed techniques to make AI models more explainable. For example, Shapley values provide a method for quantifying the contribution of each feature to model predictions, thereby helping financial institutions more easily explain their decisions to regulatory bodies and customers.

Another explainability technique is LIME, which generates interpretable models to approximate complex AI models for individual predictions. These methods are paramount for enhancing trust in AI credit assessment models, especially when a borrower is denied credit and an explanation is required (Angelini et al., 2008).

1.5. Bias and Fairness of Artificial Intelligence Models

While credit risk estimates have improved, biases inherent in AI models remain a concern. AI models are trained on historical data, which can be biased and lead to discriminatory outcomes. For example, a model trained on data from discriminatory lending practices may still be biased against certain demographic groups, even if sensitive attributes like race or gender are excluded.

Some methods incorporate fairness constraints during model training to prevent sensitive attributes from affecting predictions. Other biases may be reduced by approaches like "fairness through unawareness", which removes sensitive variables from the model. However, Kou et al. (Kou et al., 2019) demonstrated that

information like geographical location or employment status can introduce biases even without sensitive variables.

Financial institutions must ensure that their AI models comply with regulatory standards that prevent discrimination, such as the FCRA in the United States and the Equality Act in the United Kingdom. According to Fernandez (Fernandez, 2019), failure to meet these regulations can generate serious legal and reputational risks for financial institutions.

1.6. Performance Metrics for Artificial Intelligence Models in Credit Risk

Beyond accuracy, various performance metrics are used for AI models in credit risk assessment. In imbalanced datasets-with few defaulters compared to non-defaulters-metrics like precision, recall, their harmonic mean (F1 score), and the area under the ROC curve (AUC-ROC) are crucial.

Accuracy is defined as the ratio of correctly classified instances to all instances. While accuracy is useful, it can be misleading in imbalanced datasets where one class (the majority class) dominates. Precision is the ratio of true positive predictions (correctly predicted defaulters) to all positive predictions made by the model. Precision is relevant in credit risk assessment because a false positive-classifying a borrower as a defaulter when they are not-can lead to lost lending opportunities.

Recall measures the proportion of true positives out of all actual positive instances (defaulters). A high recall ensures that a high proportion of high-risk borrowers are correctly identified. The F1 score, the harmonic mean of precision and recall, provides a balanced metric that considers both false positives and false negatives.

The AUC-ROC is a measure indicating the model's capability to differentiate between defaulters and non-defaulters. A higher AUC indicates better model performance. Utilizing these performance metrics helps in understanding the strengths and weaknesses of different AI models in credit risk assessment.

2. Methodology

2.1. Data Collection and Preprocessing

This research is based on a comprehensive dataset from various sources that integrates both structured and unstructured data to enhance the validity of models when assessing credit risk. The structured data includes conventional financial variables related to borrowers' income, debt-to-income ratio, credit history, and loan amounts-definite elements in any traditional credit risk assessment model (Breeden, 2020; Addo et al., 2018). Additionally, it leverages data from social media activities, mobile transactions, and geolocation to capture more depth in borrower behavior than conventional data may miss.

The dataset consists of over 200,000 records of individual borrowers spanning from 2017 to 2022. This large sample size provides high statistical power and

makes the model dependable. All personal information is anonymized in accordance with strict standards, such as the European Union's General Data Protection Regulation (GDPR), which demands that organizations maintain stringent data handling and protection.

Data preprocessing involves several important steps. Missing values are imputed using mean and mode strategies for numerical and categorical variables, respectively. Outliers are detected and removed to avoid bias in the results. For unstructured data, natural language processing techniques such as sentiment analysis and keyword extraction are employed to transform text data into analyzable features. Sentiment analysis, also known as sentiment analysis or opinion mining, is the process of analyzing, processing, summarizing, and inferring subjective texts with emotional connotations. By utilizing the ability of sentiment analysis, it is possible to automatically determine the positive and negative emotional tendencies of natural language texts with subjective descriptions and provide corresponding results. Keyword extraction is the process of extracting the most relevant words from the text that are most relevant to the meaning of the article. It has important applications in literature search, automatic summarization, text clustering, and text classification. Transaction data is summarized to reveal financial behaviors, including transaction frequency and volume.

Referring to the commonly used classification standards for training and testing sets in deep learning networks, this article divides the preprocessed data into an 80% training set and a 20% testing set. To avoid overfitting, 10-fold cross-validation is used to tune the models and measure their performance.

2.2. Model Selection and Architecture

In this study, we will evaluate the effectiveness of several AI and ML models in credit risk assessment and compare them with the baseline logistic regression model:

Random Forests: As an integrated learning model, random forests improve prediction accuracy by constructing multiple decision trees and merging their outputs. They are excellent at preventing overfitting and can provide an ordering of the importance of features that affect prediction.

Support Vector Machines (SVMs): SVMs are excellent at handling unbalanced datasets and are able to find optimal hyperplanes in high-dimensional spaces to separate different classes. This makes SVMs particularly well suited to deal with class imbalance.

Neural networks: These models can handle complex nonlinear relationships between input variables. Deep learning models with a multi-layer structure make neural networks particularly effective at handling structured and unstructured data. They are especially good at recognizing patterns that may not be detected by traditional models.

Logistic regression: used as a benchmark model for performance comparisons, although logistic regression is widely used for credit risk modeling and its linear

nature has limitations when dealing with interactions between variables.

2.3. Feature Engineering and Selection

The predictive power of a model can be significantly enhanced through efficient feature engineering. New features derived from raw data can more accurately reflect borrower behavior and risk profile, such as calculating debt-to-income ratios, default rates, and loan maturities.

Sentiment analysis of social media posts can reveal borrower sentiment and risk tolerance, information hidden in unstructured data. Variables such as transaction frequency, average transaction amount, and stability derived from geolocation data further enrich the database for risk assessment models. These features demonstrate details of financial behavior that are difficult to capture in traditional data.

2.4. Evaluation Metrics

Therefore, a range of metrics was considered to comprehensively evaluate the model performance: accuracy, precision, recall, F1 score, and area under the receiver operator characteristic curve (AUC-ROC). Since credit risk datasets are often classically unbalanced, relying on accuracy alone may be misleading.

Accuracy rate: The ratio of correct classifications to the total number of instances.

Precision rate: These are the number of instances correctly predicted as defaulters to all instances predicted as defaulters, with the importance of avoiding opportunity costs due to incorrect lending decisions.

Recall rate: the ratio of correctly identified true defaulters to highlight high-risk borrowers accordingly.

F1-score: Harmonic mean of the precision and recall rates; it allows for a balanced view to find the model's score, considering the effect of false positives and false negatives.

AUC-ROC: It describes the model performance in terms of the capability to distinguish between defaulter and non-defaulter classes where higher values are better.

2.5. Cross-Validation and Hyperparameter Tuning

In this thesis, 10-fold cross-validation will be applied to make sure that the model is generalizable for unseen data. One dataset for this purpose is further divided into ten subsets and is, in turn, trained on nine subsets and evaluated on the remaining one. In this manner, overfitting can be avoided, and the model can be generalized well to various subsets of data. The hyperparameter tuning was done through a grid search in which different combinations of hyperparameters that yield an optimum configuration were systematically tried. It ranges from modifying the number of trees in a random forest and the maximum depth of each tree to adjusting the number of hidden layers, the number of neurons per layer, and

the learning rate in a neural network, making fits of model performance.

2.6. Explainability and Interpretability

Given the complexity of AI models, especially neural networks, interpretability is an important challenge for financial institutions. Shapley values and LIME are two techniques that help to solve the "black box" problem of AI models by clarifying the contribution of each feature to the model prediction.

Shapley value: inspired by cooperative game theory, it is used to quantify the marginal contribution of each feature to the model output.

LIME: approximates the behavior of complex models by constructing simplified interpretable models that revolve around individual predictions.

The integration of these interpretable techniques not only improves the accuracy of the model, but also ensures that the model is transparent and interpretable, enabling GSEs to make lending decisions that are both informed and interpretable, with results that are easily understood by regulators and customers.

3. Results and Analysis

3.1. Model Performance Comparison

This section introduces the comparative performance of the Random Forest, Support Vector Machine, Neural Networks, and baseline Logistic Regression models for credit risk assessment. Each model was evaluated for accuracy, precision, recall, F1 score, and AUC-ROC, which refers to the Area Under the Receiver Operating Characteristic Curve.

During the analysis, the Random Forest model yielded the best performance with 93% accuracy and a mean AUC-ROC score of 0.94. Its high capability to deal with nonlinear relationships in big, complicated datasets and its robustness when structured data, such as social media activities or transaction history, give it a special place in credit default prediction.

The neural network models fared equally well by achieving the accuracy of 91% with an AUC-ROC score of 0.92. The strong points of the methods are treating high volume data to capture nonlinear interactions among features and allowing both structured and unstructured data inclusion.

Support Vector Machine showed worse results when the accuracy was 89%, and the AUC-ROC score equaled 0.88. It could be useful in a case if there is a small portion of defaulters compared with the overall population, and its strong classification capabilities will be helpful to identify a high-risk class of borrowers (Goodell et al., 2021).

While the traditional credit risk assessments applied the LR baseline model, it resulted in the poorest performance among the models, where its accuracy was 84% and AUC-ROC was 0.79. Some of the limitations of LRs include reliance on linear assumptions, making it not that effective to capture the complexity in borrower behaviors.

Table 1 is the summary of key performance metrics in a tabular form.

Table 1. Model performance comparison table.

Model	Accuracy	AUC-ROC	Precision	Recall	F1 Score
Random Forest	93%	0.94	0.91	0.88	0.90
Neural Networks	91%	0.92	0.89	0.86	0.87
Support Vector Machine	89%	0.88	0.85	0.82	0.84
Logistic Regression	84%	0.79	0.78	0.74	0.76

Performance of Visualization:

ROC Curves: The ROC curve of each model has been plotted to visually compare the strength of each model in classifying between defaulters and non-defaulters. Of these, the Random Forest had the maximum AUC, thus becoming the most powerful predictive model.

Precision-Recall Curves: Since credit risk datasets are usually imbalanced, precision-recall curves will give insight into how each model will perform to minimize false positives and false negatives. Random Forests and Neural Networks performed particularly well in maintaining high precision while finding the right balance in terms of recall.

Confusion Matrix: The confusion matrix for each model provides different values of true positives, those respective defaulters who were correctly identified as true negatives, false positives, and false negatives that give detailed insight into model performance in a real-world setting.

These analyses pinpoint the potential of AI models, especially Random Forest and Neural Networks, which, by treating complex borrower behavior and using sources of structured and unstructured data, can outperform traditional methods such as Logistic Regression.

3.2. Alternative Data Sources—Effectiveness

The key insight from this research is the exceptional performance gain achieved by incorporating additional data sources, such as social media activity, transaction history, and geolocation data. These alternative data sources gave a better context on how borrowers behave; this was especially true for those who did not have any formal credit history in the first place (Fernandez, 2019). In the case of social media analysis, for instance, good sentiments reflected a minimal risk of default. Other indicative factors showed that borrowers with stable high-frequency transactions have a lower risk of default, again proving the predictive power of alternative data.

The Random Forest and Neural Network models benefited the most from all the alternative data sources (Zhang & Yu, 2024). Because both can manage structured and unstructured data, they can notice complex patterns in borrower behavior, which the Logistic Regression model couldn't (Addo et al., 2018). In contrast, the RF model used the frequency and volume of transactions as a significant measure of stability, while NN models leveraged social media activities and transaction behavior to build far more accurate default predictions.

3.3. Explainability and Transparency of Artificial Intelligence Models

While the models Random Forest and Neural Network outperformed Logistic Regression on predictive performance, their opaqueness is yet a big concern in highly regulated sectors like finance. According to Fernandez (Fernandez, 2019), financial institutions should be able to explain the rationale behind credit decisions; therefore, explainability is a crucial factor when considering AI model adoption. This has been accomplished through Shapley values and LIME in this work, Local Interpretable Model-Agnostic Explanations, by Goodell et al. (Goodell et al., 2021).

Shapley values provided insights into the contributions of individual features, such as debt-to-income ratio, payment history, and transaction frequency, to model predictions. For instance, debt-to-income ratio was one of the most influential variables in predicting default with continuity in most applications. LIME was used in specific instances to explain, for instance, why a particular borrower had been classified as high-risk. This at least allows more insight into the decision-making process.

While these explanation techniques brought much-needed transparency, they added even more complexity. The trade-off between model complexity and explainability remains among the most important challenges that financial institutions need to address to ensure that regulatory bodies are satisfied and trust is instilled in AI-driven credit assessments.

3.4. Bias and Fairness in Artificial Intelligence Models

However, their use in credit risk management has raised several concerns regarding algorithmic bias and fairness. In general, AI models are subject to learning potential biases from the data used for training to disadvantage one demographic group or another. For instance, even though discriminatory on lending issues, AI models could practice it if they had been biased in the beginning against some groups, even if sensitive attributes like race or gender are excluded from the dataset.

The principle of FTU compliance in this work eliminated the bias by not considering sensitive variables in the models. However, even with FTU, biases are still possible, like proxy variables of location or employment sector. In line with this, the calculations of fairness metrics were performed for demographic parity and equal opportunity, which are used to determine whether models make equitable predictions across different demographic groups. Fairness constraints increased equity but slightly reduced performance in the Neural Network model (Thakkar, & Chaudhari, 2021). This reveals that fairness is a competing factor against accuracy.

3.5. Comparison of Artificial INTELLIGENCE and Traditional Model

Overall, the findings of this study confirm that AI models outperform traditional

logistic regression in credit risk evaluation. The performance of the Random Forest and Neural Network models was consistently better than logistic regression in all the individual metrics, especially when exploiting nonlinear relationships and alternative data sources.

While logistic regression is still appreciated for its simplicity and transparency, its reliance on linear relationships limits its performance on more complex, real-world credit risk scenarios. The Support Vector Machine model balances accuracy and interpretability well without being as complex as Neural Networks, thus making it implementable for financial institutions that want better performance without going all the way to black-box models.

The findings reveal that it is high time financial institutions adopt more advanced AI-driven systems; otherwise, they would likely be excluded from contemporary finance. Nevertheless, the problems of model transparency and fairness, together with assurance of conformation with regulatory requirements, must be addressed so that AI can deploy its duties responsibly in credit risk assessment.

4. Discussion

4.1. Implication of the Findings for Credit Risk Assessment

These results indicate benefits of using AI and ML models in evaluating credit risk compared to traditional approaches. The two AI models, Forests and Neural Networks, are significantly better and can improve the accuracy of the predictions in identifying complex nonlinear relationships that are easily ignored by traditional techniques, such as those based on logistic regression. Indeed, prior studies have established that these AI models can use alternative data, such as how active a credit-seeker has been on social media, his transaction history, and geolocation, to arrive at more accurate risk assessments (Fernandez, 2019).

One key takeaway is that those financial institutions using AI-based credit risk models gain a much deeper understanding of the behavior of borrowers. Integrating unstructured data offers better predictions of creditworthiness for people who either have a thin or no formal credit history. This is particularly important in developing markets and among gig economy workers, who typically do not have access to traditional credit systems because of a lack of structured financial data.

4.2. Challenges and Limitations of Artificial Intelligence Models

Conversely, several challenges dominate deploying AI and ML models to mainstream adoption in the financial industry. The most important of these is the inherent complexity of deep learning models, which requires several computational resources and technical skills for modeling, deployment, and maintenance substantial enough to make smaller institutions struggle to compete with large organizations able to invest more energy in state-of-the-art AI solutions.

Another critical issue with these AI models is their lack of transparency. That is why traditional models, like Logistic Regression, have remained preferable because they are easy to interpret and explain the decisions, both to a regulator and

a customer who has been rejected. It is true that deep learning models have emerged as "black boxes", and it becomes challenging to justify credit decisions, especially in cases of denial to applicants. This causes severe complications regarding regulatory issues, such as adhering to a framework that needs to be set by the General Data Protection Regulation for explainability when a decision is made.

This paper tried to overcome these challenges by incorporating Shapley values to explain model predictions in a transparent manner without sacrificing predictive power. Goodell et al. (Goodell et al., 2021) have proposed LIME as a method to derive locally interpretable models for individual predictions, increasing the potential for transparency such that institutions may give clear justifications for their credit decisions, as stated by (Bussmann et al., 2020). In any case, implementing these explainability techniques requires additional computational effort, adding to the complexity of adopting AI.

4.3. Some Ethics to Consider: Bias and Fairness in Artificial Intelligence Models

Credit risk assessment always faces the algorithmic bias issue from deployed AI models. The model can inherit biases leading to disparity in outcomes, as most models are trained with historical data. For instance, a model that has been trained on some discriminatory practices of the past might keep discriminating against certain groups, even when sensitive attributes like race or gender are removed. This will be especially detrimental in regulated industries where equity and fairness are core ends.

The researchers in this work have excluded sensitive variables from the dataset, using a principle called Fairness Through Unawareness. However, there is still a chance that proxy features might encode sensitive information; therefore, fairness constraints during model training were necessary to make the predictions equal among demographic groups.

Fairness constraints imposed to enforce equity came at some cost in predictive accuracy, especially for the Neural Network model. This depicts the challenge of trading off fairness against performance, an issue now at the forefront of research. Organizations must ensure that their AI models comply with the legislation enacted to eradicate discrimination, such as the U.S. Fair Credit Reporting Act and the Equality Act in the UK.

4.4. Future Research Directions

Results from this work have also pointed out some directions for future research and development concerning the applications of AI models in credit risk assessment. Hybrid model development will be one very promising avenue that combines the predictive power of AI with the interpretability of traditional models. By integrating AI techniques with Logistic Regression or decision trees, financial institutions can have the best of both worlds: accuracy from AI and transparency

for regulatory compliance. Hybrid systems could allow real practicality for financial institutions that might be uneasy about thoroughly implementing black-box models like Neural Networks.

Another exciting direction is improving fair-aware AI models. Current approaches, such as Fairness Through Unawareness, are a good starting point for which much future research needs to develop sophisticated techniques that tend to reduce bias without performance compromise. Techniques such as adversarial de-biasing or representation learning for fairness might hold promise for mitigating bias with at least loss in model accuracy.

Then, there is a valid ethical basis upon which this research into the effects of alternative data sources in credit risk assessment can be pursued. Although social media activities and transaction histories present quite valuable advantages concerning improvement in credit risk predictions, data privacy and regulatory compliance issues implicate essential considerations, especially under rigid data protection laws such as the GDPR, as identified by (Goodell et al., 2021). Projects soon must be considered considering the impact of more privacy-preserving AI techniques, federated learning, and differential private ways to train AI models from decentralized data without touching borrower privacy.

Finally, future studies need to be directed at enhancing the explainability of AI models, since they are still opaque despite the high accuracy of deep learning models. New developments in explainability techniques which can make complex models more interpretable would thus be crucial in developing AI-driven credit assessments that are transparent, trustworthy, and adherent to regulatory standards.

5. Conclusions

The present study has explored the transformative power of AI and ML in credit risk assessment by comparing various state-of-the-art techniques with conventional models like Logistic Regression. The results show more significant improvements in predictive accuracy and identification of risk features for AI models, particularly Random Forest and Neural Networks. These models outperform conventional ones by processing and integrating a large volume of structured and unstructured information to attain much deeper insights into the behavior of borrowers (Breeden, 2020; Addo et al., 2018).

Another key takeaway is that incorporating alternative data sources, such as social media activity, transaction histories, and geolocation data, improves the performance of AI models. Moreover, a better breakthrough of predictive power is achieved that assists institutions in making more accurate estimates of credit risk, especially for people who have limited formal credit records. These AI models synthesize these sources of data in a way that would not otherwise be available for evaluating borrower risk, hence making the models valuable for underrepresented or emerging market borrowers.

However, there are some issues that AI models need to go through before they

can earn inconspicuous acceptance. First, transparency in the decision-making process, such as Neural Networks, is imperative due to its lack. Their black-box nature antagonizes explanations of how the model derived a particular decision, which is inappropriate for regulated industries. This research used Explainable AI techniques, including Shapley values and LIME, for enhanced interpretability according to feature importance and the building of overall trust in AI-driven decisions.

It is also important to consider bias and equity. Models built on historical data run the risk of perpetuating bias, leading to discriminatory outcomes. Fairness Through Unawareness and fairness constraints are among the various techniques applied in this study, but more research needs to be done to develop methods that reduce bias with minimal impact on performance. Showing compliance with regulatory standards goes hand in hand with responsible AI adoption.

5.1. Future of AI in Credit Risk Assessment

While the future of AI in credit risk assessment has a promising direction, several key considerations must be met to ensure that this technology is used responsibly and effectively. One auspicious direction is the development of hybrid models that take advantage of the predictive power of AI but combine such powers with the interpretability of traditional models. Adding AI techniques to the models with Logistic Regression or decision trees may give some accuracy of AI but retain the transparency to satisfy regulatory compliance. Hybrid models can thus provide a practical alternative for financial institutions that may be quite apprehensive about their use.

Another interesting path might be investigating the development of eventual fairness-aware AI models that can reduce bias without sacrificing accuracy. Techniques like adversarial de-biasing and fair representation learning are promising methods for mitigating bias with minimal model performance compromise. Future research should investigate such techniques within the context of credit risk assessment, especially regarding their effectiveness in reducing bias in real-world applications.

Another important direction of research is related to the ethical impact of using alternative data sources in credit risk assessment. While alternative data offers advantages in making credit risk predictions, it also raises major concerns about data privacy and regulatory issues, especially with strict data protection laws such as GDPR. Further research is warranted, which would apply newly developed AI privacy-preserving techniques, including federated learning and differential privacy, to enable the training of an AI model on decentralized data without violation of the borrower's privacy.

Explainability in AI models is a key element of any responsible use of AI within credit risk evaluation. While the Shapley values and LIME are indeed especially useful tools to explain the decisions of a model, much remains to be undertaken regarding the development of more sophisticated techniques of explainability that

may render complex models, such as Neural Networks, more transparent. As AI continues to develop, it will be relevant to ensure that models remain interpretable, trustworthy, and comply with regulatory standards (Breeden, 2020).

5.2. Realistic Recommendations to Financial Institutions

Based on the findings of this study, the following are some practical recommendations that any financial institution willing to adopt AI in its credit risk assessment processes can consider. First, institutions should consider the adoption of hybrid models that provide a balancing act between the accuracy of AI and the interpretability of traditional models. These models can be a workable solution for organizations that have to comply with regulatory requirements while leveraging the predictive power of AI.

Second, fairness and equity should be inherent concerns of any AI model within financial institutions. This means periodic auditing of models for bias elimination, constraining models at development for fairness, and adherence to relevant regulations on the subject. Institutions should also be open towards borrowers regarding data considered during credit assessment and give meaningful reasons if adverse decisions are being passed.

Thirdly, institutions should consider the use of alternative data sources to populate their credit risk profiles. This will be particularly important in the case of borrowers with limited or no formal credit history. Nonetheless, this must be pursued while giving full respect to borrowers' privacy and keeping in mind the needs of data protection regulations such as the GDPR. Privacy-preserving AI techniques, such as federated learning, should be considered to mitigate the risks associated with using alternative data.

The second is that it must make investments in explainable AI techniques, which shall help and work towards more transparency of the AI models that one uses. These techniques can include Shapley values and LIME in enabling institutions to explain their credit decisions to regulators and customers and therefore further build trust and ensure that regulatory standards are met. With increasing complexity, AI models would require the clear and interpretable development of institutions that can be trusted by all stakeholders.

5.3. Future Directions and Considerations

AI and ML represent the future of credit risk assessment: more enabling correct decisions, lower default rates, and increased credit accessibility (Milojević & Redzepagic, 2021). Minimizing transparency, bias, and fairness challenges, along with compliance issues, paves the path for successful adoption. Successful implementation of hybrid models, fairness-aware approaches, and techniques for explainable AI will enable institutions to harness all the benefits of AI while never compromising on ethical grounds or regulatory compliance.

While much fair, private, and explainable AI models are developed today, with the rapid development of AI, future research in this area should be channeled to

make AI models fairer, more private, more explainable for a truly more equitable financial system that exists between borrowers and lenders.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Addo, P., Guegan, D., & Hassani, B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. *Risks, 6,* Article 38. https://doi.org/10.3390/risks6020038

Angelini, E., di Tollo, G., & Roli, A. (2008). A Neural Network Approach for Credit Risk Evaluation. *The Quarterly Review of Economics and Finance, 48,* 733-755. https://doi.org/10.1016/j.qref.2007.04.001

Breeden, J. L. (2020). Adjusting Loss Forecasts for the Impacts of Government Assistance and Loan Forbearance during the COVID-19 Recession. *Journal of Risk Management in Financial Institutions, 14,* 25-32. https://doi.org/10.69554/dzwx6781

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable Machine Learning in Credit Risk Management. *Computational Economics, 57,* 203-216. https://doi.org/10.1007/s10614-020-10042-0

Fernandez, A. (2019). *Artificial Intelligence in Financial Services. Banco de Espana Article 3/19*. https://doi.org/10.2139/ssrn.3366846

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial Intelligence and Machine Learning in Finance: Identifying Foundations, Themes, and Research Clusters from Bibliometric Analysis. *Journal of Behavioral and Experimental Finance, 32,* Article 100577. https://doi.org/10.1016/j.jbef.2021.100577

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., & Herrera-Viedma, E. (2019). Machine Learning Methods for Systemic Risk Analysis in Financial Sectors. *Technological and Economic Development of Economy, 25,* 716-742. https://doi.org/10.3846/tede.2019.8740

Milojević, N., & Redzepagic, S. (2021). Prospects of Artificial Intelligence and Machine Learning Application in Banking Risk Management. *Journal of Central Banking Theory and Practice, 10,* 41-57. https://doi.org/10.2478/jcbtp-2021-0023

Thakkar, A., & Chaudhari, K. (2021). A Comprehensive Survey on Deep Neural Networks for Stock Market: The Need, Challenges, and Future Directions. *Expert Systems with Applications, 177,* Article 114800. https://doi.org/10.1016/j.eswa.2021.114800

Zhang, X., & Yu, L. (2024). Consumer Credit Risk Assessment: A Review from the State-Of-the-Art Classification Algorithms, Data Traits, and Learning Methods. *Expert Systems with Applications, 237,* Article 121484. https://doi.org/10.1016/j.eswa.2023.121484