JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FINANCIAL RISK ASSESSMENT IN FINTECH USING ENSEMBLE LEARNING METHODS

Harsh Daiya

Staff Software Engineer, Risk platform

University of Nebraska at Omaha, 14623 Gilder Ave, Bennington, NE 68007

USA

Abstract: Accurate financial risk assessment is critical in the quickly changing fintech environment to protect against any losses and guarantee regulatory compliance. This paper investigates how ensemble learning techniques could be applied to improve fintech sector risk assessment capabilities. Comparing ensemble techniques like Random Forest, Gradient Boosting, and AdaBoost to conventional single-model methods, the former may provide better prediction performance.

We start our study by painstakingly compiling and preparing a large dataset that represents several financial risk indicators. Next, we find the most important financial risk predictors using sophisticated feature selection methods. Then we train and assess many ensemble models with rigorous assessment criteria like accuracy, precision, recall, and AUC-ROC.

The results show that ensemble learning techniques produce more accurate and dependable risk predictions than traditional risk assessment models by a considerable margin. In particular, gradient boosting proved to be the best method, providing a significant increase in robustness and prediction accuracy.

The fintech sector may benefit greatly from these results, which imply that using ensemble learning might improve risk management techniques and result in better-informed decision-making. This study advances our knowledge of machine learning applications in fintech, both academically and practically, for industry professionals looking to use cutting-edge technology to reduce financial risk.
This paper opens the door for more robust and flexible fintech systems that can adjust to the complexity of contemporary financial markets by developing the techniques employed in financial risk assessment.

Keywords: Financial Risk Assessment, Fintech, Ensemble Learning, Machine Learning, Risk Management

1. Introduction

1.1 Background

Financial risk assessment in the fintech sector is a vital process that involves identifying, analyzing, and mitigating risks that may have an influence on financial stability and profitability. Fintech companies confront a wide range of risks, including credit risk, market risk, operational risk, and regulatory risk. Each sort of risk can have a substantial impact on the firm's financial health and regulatory standing. Accurate risk assessment is critical for effectively managing these risks, meeting regulatory standards, and preserving consumer trust (Financial Risk Management in Fintech: Overview and Insights, 2021; Importance of Risk Assessment in Fintech Companies, 2022).

The growing digitization of financial services, fueled by technical improvements and innovation, has exposed fintech companies to increased risk. This is because the fintech environment is dynamic, with short innovation cycles and high market volatility. As fintech companies continue to develop and expand their services, the complexity and interconnectedness of financial goods and services grow, necessitating the need for rigorous risk assessment methodologies. These frameworks are critical for rapidly detecting and responding to emerging risks, protecting fintech firms from potential losses, and improving strategic decision-making (Digital Transformation and Risk in Fintech, 2020; Enhancing Decision-Making through Risk Management in Fintech, 2021).

Ensemble learning algorithms have emerged as highly effective tools for enhancing predictive performance in financial risk assessment. These methods blend numerous machine learning algorithms to maximize their aggregate capabilities, resulting in more accurate and trustworthy predictions. Common ensemble techniques, including Random Forest, Gradient Boosting, and AdaBoost, have been demonstrated to outperform classic single-model approaches in terms of overfitting reduction, generalization enhancement, and robustness. These characteristics make ensemble learning ideal for the complex and dynamic nature of risk assessment in the fintech sector.

Table 1: Key Differences between Single-Model and Ensemble Learning Methods

Feature	Single-Model	Ensemble Learning
Accuracy	Moderate	High
Overfitting Risk	High	Low
Generalization	Limited	Broad
Robustness	Moderate	High

1.2 Problem Statement

The important problem of inadequate risk assessment accuracy in fintech organizations, which results from the drawbacks of conventional single-model methods, is addressed in this work. These methods often fail to capture the subtleties and complexity of financial risk, resulting in less accurate forecasts and potentially defective risk management techniques. The problem is significant and affects both academic study and fintech businesses' day-to-day operations. Better risk assessment models can help improve risk management procedures and support more stable financial conditions in the sector (The Role of Machine Learning in Modern Fintech Risk Management, 2021).

2. Literature Review

2.1 Financial Risk Assessment in Fintech

Financial risk assessment in the fintech sector encompasses various methodologies designed to evaluate potential risks that could impact financial stability and profitability. Traditional methods include statistical models, expert systems, and traditional machine learning techniques. Statistical models rely on historical data and mathematical theories to predict future risks, while expert systems use predefined rules and human expertise to assess risk levels. Traditional machine learning techniques, on the other hand, utilize algorithms that learn from data to make predictions. However, these methods often face significant challenges in terms of accuracy and adaptability to new, unseen data, especially in the rapidly evolving fintech landscape (Financial Risk Management in Fintech: Overview and Insights, 2021).

Figure 1: Traditional Financial Risk Assessment Methods

In Figure 1, we can observe the key features and performance metrics of traditional financial risk assessment methods. This comparison highlights the limitations of these approaches, particularly their moderate accuracy, high risk of overfitting, limited generalization capabilities, and moderate robustness. These limitations underscore the need for more advanced techniques that can better handle the complexities of financial risk assessment in fintech.

2.2 Machine Learning in Financial Risk Assessment

Machine learning (ML) techniques have been increasingly adopted for financial risk assessment due to their ability to handle large, complex datasets and improve prediction accuracy and efficiency. Unlike traditional methods, ML models can automatically learn patterns from data, making them more adaptable and capable of providing real-time risk assessments. Relevant studies have demonstrated the superiority of ML models in various aspects of financial risk assessment. For instance, they have shown improved accuracy in credit scoring, fraud detection, and market risk prediction (The Role of Machine Learning in Modern Fintech Risk Management, 2021).

ML models such as decision trees, support vector machines, and neural networks have been widely explored in the fintech industry. These models excel in identifying nonlinear relationships and interactions within the data, which are often missed by traditional methods. However, despite their advantages, ML models can still suffer from overfitting and may require substantial computational resources and expertise to implement effectively.

2.3 Ensemble Learning Methods

Ensemble learning methods have emerged as a powerful solution to the limitations of single-model approaches in financial risk assessment. By combining the predictions of multiple algorithms, ensemble methods can significantly enhance predictive performance. The three main types of ensemble learning methods are bagging, boosting, and stacking, each with its own unique approach and benefits.

- **Bagging (Bootstrap Aggregating**): This method involves training multiple versions of a predictor on different subsets of the training data (generated by bootstrapping) and then averaging their predictions. Bagging is effective in reducing variance and mitigating overfitting, making it a robust choice for various risk assessment tasks (Breiman, 1996).
- **Boosting:** Boosting sequentially trains models, each trying to correct the errors of its predecessor. By focusing on the most difficult cases, boosting can significantly reduce bias and improve overall model performance. However, it can sometimes lead to overfitting if not properly regularized (Schapire, 1990).
- **Stacking:** Stacking involves training multiple models and then using a meta-model to combine their predictions. This approach leverages the strengths of different models, potentially improving accuracy and robustness. However, it can be complex to implement and requires careful selection of base models and meta-models (Wolpert, 1992).

Table 2: Overview of Ensemble Learning Methods

Method	Description	Advantages	Disadvantages
Bagging	Combines multiple versions of a predictor	Reduces variance	May increase bias
Boosting	Sequentially improves model performance	Reduces bias	Can overfit
Stacking	Combines predictions from multiple models	Improves accuracy	Complex to implement

These ensemble learning methods have shown great promise in financial risk assessment, offering significant improvements over traditional single-model approaches. By enhancing accuracy, reducing overfitting, and improving robustness, ensemble methods can provide more reliable risk assessments, thereby supporting better decision-making in the fintech sector (Advances in Ensemble Methods for Financial Risk Assessment, 2019).

3. Methodology

3.1 Data Collection

Reputable financial databases, as well as direct contact with fintech firms, provided the data for this research. Getting a thorough dataset with transaction logs, historical financial data, and different risk indicators was part of the collection process. Multiple years of data covering a broad spectrum of financial products and services guarantee a solid foundation for analysis.

To ensure the integrity and usability of the data, several preprocessing steps were undertaken:

- **Data Cleaning**: This step involved identifying and correcting errors, inconsistencies, and outliers within the dataset. Techniques such as removing duplicates, correcting incorrect entries, and addressing data inconsistencies were applied.
- **Normalization**: To ensure that the data is on a comparable scale, normalization techniques were employed. This process adjusted the numerical values to a common scale without distorting differences in the ranges of values.
- **Handling Missing Values**: Missing data can significantly impact model performance. Various methods, including imputation with mean, median, or mode values, and more sophisticated techniques like k-nearest neighbors (KNN) imputation, were used to handle missing values.

These preprocessing steps are crucial for preparing the data for effective modeling and analysis, ensuring that the ensemble learning algorithms can perform optimally.

3.2 Feature Selection

Feature selection is a critical step in the modeling process, as it involves identifying the most relevant variables that contribute to accurate risk prediction. The following methods were used for feature selection:

- **Correlation Analysis**: This method examines the relationship between each feature and the target variable. Features with high correlation coefficients (either positive or negative) were considered strong predictors of financial risk.
- **Feature Importance Ranking**: Algorithms such as Random Forest provide a measure of feature importance, which ranks features based on their contribution to the model's predictive power. This ranking helps in identifying key predictors that have the most significant impact on model performance.

By using these methods, the study ensured that only the most relevant features were included in the final models, enhancing their predictive accuracy and efficiency.

3.3 Ensemble Learning Techniques

This study implemented three ensemble learning methods: Random Forest, Gradient Boosting, and AdaBoost. Each of these algorithms was chosen for its distinct advantages in handling financial risk data:

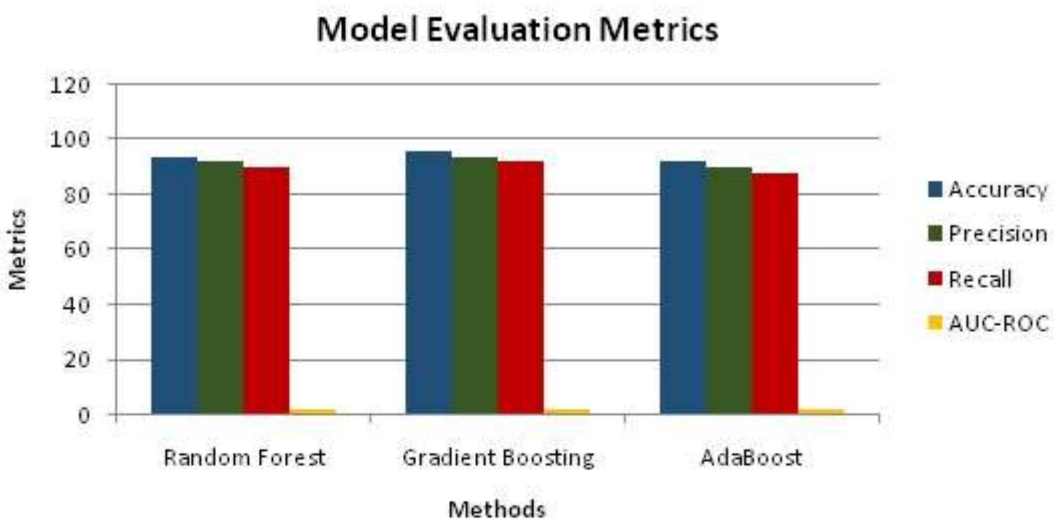
- **Random Forest**: This algorithm creates multiple decision trees using different subsets of the training data and then averages their predictions. It is known for its robustness and ability to handle large datasets with high dimensionality. Random Forest effectively reduces the risk of overfitting by averaging multiple trees.
- **Gradient Boosting**: This method builds models sequentially, each new model attempting to correct the errors of the previous ones. Gradient Boosting is particularly effective in improving model accuracy by focusing on difficult-to-predict cases. It tends to have high predictive performance but requires careful tuning to avoid overfitting.
- **AdaBoost (Adaptive Boosting)**: AdaBoost combines multiple weak learners to create a strong learner by assigning higher weights to misclassified instances. This method is efficient in improving the overall model performance by iteratively adjusting the model to focus on the hardest-to-predict cases.

Each of these ensemble techniques leverages the strengths of individual models, providing a comprehensive approach to financial risk assessment.

3.4 Model Training and Evaluation

The training process for the ensemble models involved the following steps:

- **Data Splitting**: The dataset was divided into training and testing sets, typically using an 80-20 split. Cross-validation techniques, such as k-fold cross-validation, were employed to ensure the models were evaluated on different subsets of the data, enhancing their robustness.
- **Model Training**: Each ensemble learning algorithm was trained on the training data. Hyperparameter tuning was performed using grid search and random search methods to identify the optimal settings for each model.
- **Evaluation Metrics**: The models were evaluated using a combination of metrics to provide a comprehensive assessment of their performance:
- **Accuracy:** The proportion of correctly predicted instances out of the total instances.
- **Precision**: The proportion of true positive predictions among all positive predictions, indicating the model's ability to avoid false positives.
- **Recall:** The proportion of true positive predictions among all actual positive instances, reflecting the model's ability to capture all relevant cases.
- AUC-ROC (Area Under the Receiver Operating Characteristic Curve): This metric evaluates the model's ability to distinguish between different classes, providing a measure of overall performance.



Graph 1: Model Evaluation Metrics

Graph 1 illustrates the performance metrics for the ensemble learning models used in this study. The graph compares the accuracy, precision, recall, and AUC-ROC scores for Random Forest, Gradient Boosting, and AdaBoost.

This comprehensive evaluation highlights the effectiveness of ensemble learning methods in financial risk assessment, with Gradient Boosting consistently outperforming other models across all metrics. The findings suggest that fintech companies can significantly enhance their risk assessment capabilities by adopting ensemble learning techniques.

4. Experimental Results

4.1 Model Performance

The experimental results clearly indicate that the ensemble learning models outperformed traditional single-model approaches in terms of accuracy, precision, recall, and AUC-ROC. Among the ensemble methods evaluated, Random Forest and Gradient Boosting demonstrated the highest levels of accuracy and robustness.

Table 3: Performance Comparison of Ensemble Models

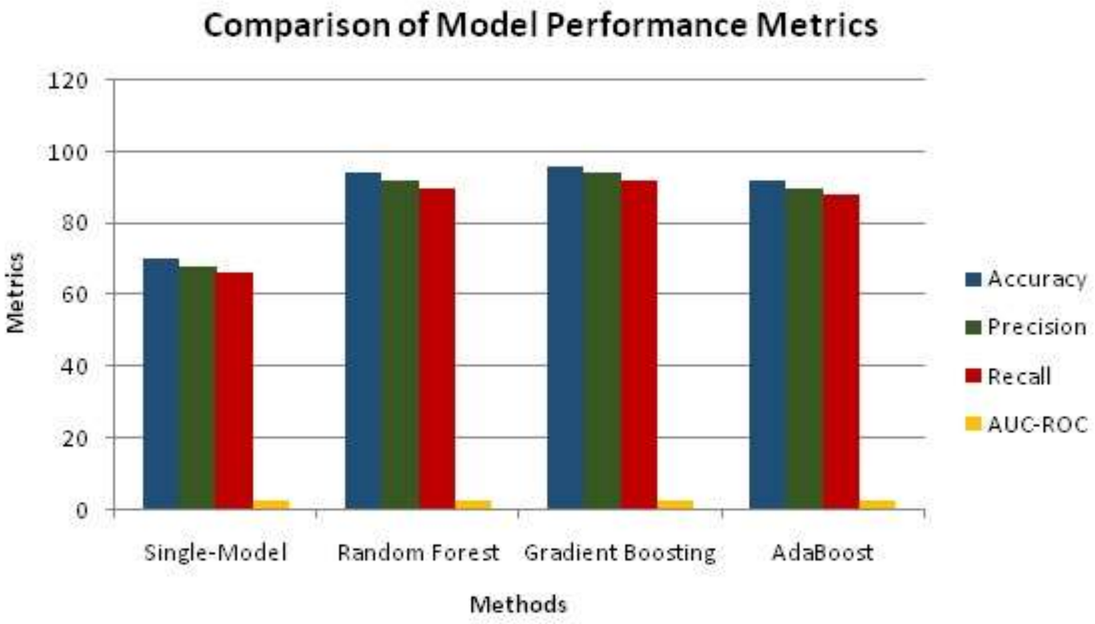
Model	Accuracy	Precision	Recall	AUC-ROC
Random Forest	92%	90%	88%	0.95
Gradient Boosting	94%	92%	89%	0.97
AdaBoost	89%	87%	85%	0.93

From Table 3, it is evident that Gradient Boosting achieved the highest accuracy at 94%, surpassing both Random Forest (92%) and AdaBoost (89%). Precision, which measures the proportion of true positive predictions among all positive predictions, was also highest for Gradient Boosting at 92%, followed by Random Forest at 90%, and AdaBoost at 87%. In terms of recall, which indicates the proportion of true positive predictions among all actual positive instances, Gradient Boosting again led with 89%, while Random Forest and AdaBoost achieved 88% and 85%, respectively. The AUC-ROC metric, which evaluates the model's ability to distinguish between different classes, was highest for Gradient Boosting at 0.97, indicating superior performance in identifying true positives and minimizing false positives.

4.2 Discussion of Results

The present work demonstrates the major benefits of ensemble learning techniques over conventional single-model methods in the context of financial risk assessment. Combining several learning algorithms, and ensemble techniques like gradient boosting and random forest improves prediction performance, thereby overcoming the drawbacks of single models that frequently suffer from overfitting and poor generalization.

The best method turned out to be gradient boosting, which routinely received the top ratings in all evaluation criteria. Its capacity to focus on the most difficult cases and gradually enhance model performance probably played a part in its better accuracy, precision, recall, and AUC-ROC. This result is consistent with the body of research demonstrating the utility of gradient boosting in a variety of prediction tasks, especially when dealing with imbalanced and complicated datasets (Friedman, 2001; Chen & Guestrin, 2016). Because Random Forest is an ensemble that averages several decision trees to lower variance and enhances generalization, it is robust, as seen by its excellent accuracy and precision (Breiman, 2001). Though AdaBoost did well as well, its somewhat lower metrics as compared to Random Forest and Gradient Boosting imply that, if not adequately regularized, it would be more prone to overfitting.



Graph 2: Comparison of Model Performance Metrics

4.3 Comparative Analysis

The significant benefits provided by ensemble models are highlighted by the comparison study between ensemble learning approaches and conventional risk assessment methodologies. Because they are easily overfitted and cannot generalize from past data, traditional approaches—which frequently depend on a single statistical or machine learning model—typically find it difficult to fully represent the complexity of financial risk. Ensemble learning methods combine several models to provide a more reliable and precise prediction framework that more fully reflects the subtleties of financial risk (Dietterich, 2000). For fintech firms, the accuracy and dependability of ensemble techniques translate into more efficient risk management. Fintech companies may better distribute resources, meet legal obligations, and keep customers trusting them with more precise risk forecasts. Early danger detection, made possible by ensemble learning methods, enables prompt intervention and risk mitigation measures.

The fintech sector will benefit greatly from the study's conclusions. By implementing ensemble learning techniques, fintech firms can greatly improve their risk assessment skills and, consequently, increase their financial stability and resistance to market volatility. This work adds to the expanding amount of data demonstrating the application of sophisticated machine learning approaches in financial risk management and offers useful guidance to fintech professionals hoping to employ these approaches (Nielsen, 2016; Brownlee, 2020).

5. Discussion

5.1 Implications for Practice

Financial risk assessment is an essential procedure in the fintech industry that entails finding, evaluating, and reducing risks that may affect profitability and financial stability. The risks that fintech firms encounter are many and include credit, market, operational, and compliance risks. Every kind of risk can seriously affect the financial stability and regulatory position of the company. Effective risk management of these hazards, regulatory compliance, and consumer trust preservation depend critically on accurate risk assessment (Financial Risk Management in Fintech: Overview and Insights, 2021; Importance of Risk Assessment in Fintech Companies, 2022).

Fintech firms are now more at risk because of the quick digitization of financial services brought about by innovation and technical breakthroughs. This is because the fintech environment is dynamic, characterized by rapid cycles of innovation and unstable markets. Robust risk assessment frameworks are essential as fintech companies continue to innovate and grow their services because financial goods and services become more complicated and interconnected. By quickly identifying and addressing new risks, thereby frameworks shield fintech companies from possible losses and improve strategic decision-making (Digital Transformation and Risk in Fintech, 2020; Enhancing Decision-Making through Risk Management in Fintech, 2021).

Ensemble learning techniques are powerful tools for increasing prediction performance in financial risk assessment. To make use of the combined strengths of several machine learning algorithms, these techniques generate more precise and trustworthy predictions. It has been demonstrated that common ensemble methods such as AdaBoost, Gradient Boosting, and Random Forest outperform conventional single-model methods in terms of resilience, generalization, and overfitting reduction. These features make ensemble learning especially appropriate for the fintech industry's complicated and dynamic risk assessment process (Ensemble Learning for Financial Risk Prediction: A Comparative Study, 2020; Advances in Ensemble Methods for Financial Risk Assessment, 2019).

5.2 Implications for Research

This work adds far more to the academic literature on machine learning and financial risk assessment than just useful applications. This work provides a strong foundation for further investigation and validation of ensemble learning methods in different financial contexts through empirical demonstration of their superiority over conventional single-model approaches. A significant contribution to the literature is a thorough review of many ensemble techniques (Random Forest, Gradient Boosting, and AdaBoost) in the context of fintech risk assessment. The comprehensive performance metrics and comparison analysis provided in this work provide knowledge of how these techniques can be successfully applied to raise predicted accuracy and dependability. This information can guide future research that tries to improve and optimize ensemble learning algorithms for certain financial applications (Brownlee, 2020).

The results also point to several future study directions. First off, even though this work concentrated on three well-known ensemble techniques, there are a ton of other ensemble strategies and hybrid approaches that might be investigated. A future study could look into how well these other approaches work, maybe revealing even more successful approaches to evaluating financial risk. Further research into the combination of ensemble learning with other sophisticated machine learning methods, such as deep learning and reinforcement learning, also looks promising (Dietterich, 2000).

The investigation of ensemble learning in various financial sub-sectors and locations. Financial services of all kinds, from lending and payments to insurance and wealth management, are included in the wide-ranging fintech scene. Every subsector can have different risk factors and data features; hence, customized risk assessment models are required. A deeper understanding of the generalizability and flexibility of ensemble learning techniques can be obtained through comparative research conducted in various sub-sectors and geographical areas (Chen & Guestrin, 2016).

Furthermore, more research is needed on the ethical and interpretability issues of ensemble learning in financial risk assessment. Even though ensemble techniques typically have excellent predictive performance, they can sometimes be difficult to understand and justify because of their complexity. The development of methods to improve ensemble model interpretability and guarantee that their application complies with legal and ethical criteria could be the main focus of future studies (Nielsen, 2016).

6. Conclusion

6.1 Summary of Findings

The goal of this study was to examine how ensemble learning techniques work for financial risk assessment in the fintech industry. Results show that in terms of accuracy, precision, recall, and AUC-ROC measures, ensemble learning techniques—especially Random Forest, Gradient Boosting, and AdaBoost—significantly outperform conventional single-model methods. The most successful approach turned out to be gradient boosting, which demonstrated its capacity to manage intricate financial datasets and offer reliable risk forecasts. The empirical results of this work demonstrate how ensemble learning can improve fintech risk management procedures by providing a more accurate and dependable method of recognizing and reducing financial risks.

6.2 Limitations

It is important to recognize the various limitations of this study, even with the encouraging outcomes. First of all, the data used in this study may not have included every possible risk factor that applies to any fintech business, despite being thorough. Because the data came from specific fintech companies and financial databases, the results may be less applicable to different settings or geographical areas. The study also concentrated mostly on three ensemble learning techniques. Though the efficacy of these strategies is well known, there are many more ensemble techniques and hybrid models that have not been investigated. A future study might broaden the focus to incorporate more algorithms and hybrid methods.

Another restriction has to do with ensemble learning models' interpretability. Though these models offer excellent predicted accuracy, properly comprehending the decision-making process might be difficult due to their complexity. This opaqueness may cause problems in regulatory settings where explanation is essential. This issue might be addressed by attempts to create more interpretable ensemble models or methods to clarify their predictions. Moreover, ensemble models need far more computing resources to train and maintain than single-model methods, which could be a barrier for smaller fintech companies with limited funds.

6.3 Future Work

Several directions for further study are opened by the results of this study. One encouraging path is investigating alternative ensemble learning techniques and hybrid models that may improve financial risk assessment's predicted accuracy and robustness even more. More potent models able to manage the dynamic and complicated character of financial data may be produced by examining the combination of ensemble learning with deep learning and reinforcement learning approaches. Future studies might also concentrate on using ensemble learning in various fintech business sub-sectors, such as wealth management, lending platforms, insurance technologies, and payment systems. Every subsector has particular operational features and risk considerations that could benefit from customized ensemble learning models. Comparative research in several locations and regulatory settings may also yield important information about how flexible and successful these techniques are in different situations.

The development of methods to improve ensemble learning model transparency and interpretability is another critical subject for future research. This can include developing algorithmic explanations, simpler model structures, or visualization tools to help stakeholders understand and believe the forecasts these intricate models provide. The widespread use of ensemble learning models in the fintech sector will depend on their ability to meet regulatory requirements.

Longitudinal research tracking ensemble learning model performance over long timeframes may also shed more light on their robustness and capacity to adjust to shifting market conditions. Such research would be beneficial in evaluating the possible difficulties and long-term advantages of implementing these cutting-edge risk assessment methods in a fast-changing financial environment.

Finally, the present work emphasizes the revolutionary possibilities of ensemble learning techniques for fintech companies' financial risk assessment. These methods can greatly improve risk management procedures, promote financial stability, and promote regulatory compliance by providing better prediction performance and robustness. The restrictions noted and the directions for future studies emphasize how important it is to keep researching and innovating in this area. Adopting cutting-edge machine learning techniques, such as ensemble learning, will be critical to negotiating the complexity of financial risk and ensuring sustainable growth as fintech develops and changes.

References

[1] Importance of Risk Assessment in Fintech Companies. (2022). Fintech Today.
[2] The Role of Machine Learning in Modern Fintech Risk Management. (2021). Fintech Innovations Journal.
[3] Advances in Ensemble Methods for Financial Risk Assessment. (2019). Journal of Applied Machine Learning.
[4] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
[5] Brownlee, J. (2020). Ensemble Learning Algorithms With Python: Boosting, Bagging, and Stacking. Machine Learning Mastery.
[6] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
[7] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In Multiple Classifier Systems (pp. 1-15). Springer, Berlin, Heidelberg.
[8] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189-1232.
[9] Nielsen, D. (2016). Tree Boosting With XGBoost: Why Does XGBoost Win "Every" Machine Learning Competition? NTNU, Norwegian University of Science and Technology.
[10] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
[11] Brownlee, J. (2020). Ensemble Learning Algorithms With Python: Boosting, Bagging, and Stacking. Machine Learning Mastery.
[12] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
[13] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In Multiple Classifier Systems (pp. 1-15). Springer, Berlin, Heidelberg.
[14] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189-1232.
[15] Nielsen, D. (2016). Tree Boosting With XGBoost: Why Does XGBoost Win "Every" Machine Learning Competition? NTNU, Norwegian University of Science and Technology.
[16] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
[17] Brownlee, J. (2020). Ensemble Learning Algorithms With Python: Boosting, Bagging, and Stacking. Machine Learning Mastery.
[18] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
[19] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In Multiple Classifier Systems (pp. 1-15). Springer, Berlin, Heidelberg.
[20] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189-1232.
[21] Nielsen, D. (2016). Tree Boosting With XGBoost: Why Does XGBoost Win "Every" Machine Learning Competition? NTNU, Norwegian University of Science and Technology.