Big Data Analytics and Data Science E-ISSN: 3123-9986

Research Article

Evaluating Explainable Artificial Intelligence Methods for Interpretable Machine Learning Models in Large scale Enterprise Data Analytics Systems

Indra Ava Dianta [1], Greget Widhiati [2], Andreas Tigor Oktaga [3],

[1] Universitas Sains dan Teknologi Komputer 1; e-mail : indra@stekom.ac.id .
[2] Universitas Sains dan Teknologi Komputer 2; e-mail : greget@stekom.ac.id
[3] Institut Teknologi dan Bisnis Semarang 3; e-mail : tigoroktaga@gmail.com
* Corresponding Author : name

Abstract: Explainable Artificial Intelligence (XAI) has become a critical area of research within artificial intelligence, focusing on improving the transparency and interpretability of machine learning (ML) models, often referred to as "black-box" models. The need for XAI techniques arises from the inherent complexity of ML models, which can make their decision making processes difficult for users to understand. This study investigates various XAI techniques, including LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations), to assess their impact on model interpretability without significantly compromising predictive performance. A comparative experimental design was used, applying these XAI methods to different ML models, including deep neural networks and ensemble methods, within large scale enterprise data analytics systems. The results indicate that XAI methods significantly enhance model transparency and decision traceability, allowing users to understand the influence of individual features on predictions. While a slight reduction in predictive accuracy was observed, especially with simpler models, the trade-off between interpretability and performance was deemed acceptable, particularly in fields requiring transparency, such as healthcare, finance, and autonomous systems. The use of XAI in enterprise data systems has practical implications for fostering trust and enabling informed decision making among stakeholders. Furthermore, the study discusses the challenges and limitations of applying XAI techniques, such as complexity, scalability, and model-specific limitations. Future research is suggested to focus on developing more scalable and efficient XAI methods, enhancing their applicability across various model types, and addressing the challenges of real-time applications. This will be crucial in ensuring the widespread adoption of XAI in critical domains, promoting the ethical use of AI while maintaining predictive accuracy.

Keywords: Explainable AI; Machine learning; LIME method; SHAP values; Enterprise systems.

Received: 21, November 2025
Revised: 10, December 2025
Accepted: 29, December 2025
Published: 15, January 2026
Curr. Ver.: 19, January 2026

1. Introduction

Enterprise data analytics systems have become an integral part of modern business operations, helping organizations leverage vast amounts of data to drive decision making and optimize processes [1]. The integration of machine learning (ML) models within enterprise information systems (EIS) has particularly grown in sectors like digital marketing, where data-driven strategies are essential for understanding customer behavior and developing targeted campaigns [2]. These models are also employed in various business analytics applications, such as customer segmentation, market prediction, and resource optimization, making them essential tools for modern enterprises [3], [4].

The increasing reliance on ML models is primarily driven by their ability to handle complex, unstructured data-something traditional analytical methods struggle with [1]. Machine learning models, including Random Forests, Support Vector Machines, and Neural Networks, offer superior performance in predictive analytics and decision support [3], [4]. As enterprise data analytics systems demand real-time data processing and predictive insights, ML models provide organizations with a competitive advantage and enhance operational efficiency [5].

Despite these advantages, ML models often suffer from low interpretability, commonly referred to as the "black-box" problem [6]. This lack of transparency poses significant challenges, especially in critical domains such as healthcare, finance, and education, where understanding the decision making process is essential for trust and accountability [6], [7]. Specifically, the opaque nature of complex ML models can undermine user trust, hinder adoption, and complicate the accountability of systems [6]. Additionally, such models may perpetuate biases present in the training data, further exacerbating issues related to fairness and transparency [8].

To address these challenges, the field of Explainable Artificial Intelligence (XAI) has emerged, with a focus on developing methods to enhance the interpretability of ML models without sacrificing performance [9]. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), and Counterfactual Explanations (CF) are being increasingly applied to provide clearer insights into model behavior [3], [10]. These methods, along with model-specific approaches and visualization tools, are essential in making ML models more transparent, thereby improving their trustworthiness and applicability in enterprise settings [1], [11].

Explainable Artificial Intelligence (XAI) is an emerging field within artificial intelligence (AI) that focuses on enhancing the transparency and interpretability of machine learning (ML) models [12]. Traditional machine learning models, especially complex ones like deep learning networks, are often regarded as "black boxes" due to their lack of transparency, which makes it difficult to understand how these models arrive at specific conclusions or predictions [13], [14]. XAI aims to address this limitation by providing human-readable explanations for AI model outputs, thus fostering trust, accountability, and usability across various applications [15], [16].

One of the primary goals of XAI is to improve model interpretability by making decisions more understandable and trustworthy. Through techniques that highlight important variables or provide clearer insights into model behaviors, XAI aims to enhance both the transparency and accuracy of ML models [12], [17]. Moreover, XAI has significant potential in sensitive areas such as healthcare, where understanding the rationale behind AI decisions is crucial for ethical decision making and ensuring patient safety [18], [19]. In these fields, being able to interpret and justify AI-generated decisions can lead to better outcomes and greater trust from users and stakeholders [15], [16].

The importance of XAI is further emphasized in the context of improving ethical and equitable decision making. For instance, in sectors like finance and healthcare, where fairness and bias mitigation are critical, transparent AI models allow for the identification and correction of potential biases, thereby promoting fairness in decision making [18], [19]. Additionally, by providing clearer explanations of decisions, XAI allows users to trust that the models are making decisions based on relevant and equitable factors [20].

The primary objective of this study is to evaluate the effectiveness of XAI techniques in improving the interpretability of machine learning models while maintaining their predictive performance in large scale analytics systems. This study will focus on assessing how various XAI techniques can improve transparency and understanding of model behavior [13], [20]. Moreover, it will investigate whether these techniques can maintain or even enhance the predictive accuracy of the models [12], [21]. Finally, the study will explore the practical applications of XAI across various domains, including healthcare, finance, and education, highlighting its potential to enhance decision making and operational efficiency [14], [18].

2. Literature Review

Exploration of Existing Machine Learning Models Used in Enterprise Data Analytics Systems

Machine learning (ML) models have become an essential part of enterprise data analytics systems, offering powerful capabilities to analyze large datasets and make predictions in real-time. Among the most commonly used ML models in enterprise settings is *ensemble learning*, which combines multiple models to improve prediction accuracy and stability. This technique is particularly effective in enterprise risk prediction, where reducing bias and overfitting is critical for making robust predictions in complex business environments [22]. Ensemble learning is beneficial for tasks where various algorithms can complement each other, providing a more accurate and reliable prediction.

Support Vector Machines (SVM) and Decision Trees (DT) are also widely applied in business analytics, such as customer segmentation, market forecasting, and data classification [2]. These models are particularly useful for handling intricate datasets, including text and image data, which are common in business applications. Additionally, Artificial Neural Networks (ANN) are used for forecasting tasks, such as predicting electricity generation based on variables like population growth rates [23]. These models are often employed in enterprise systems to manage both structured and unstructured data.

Other popular models include Random Forest, Naïve Bayes, k-Nearest Neighbors, and Logistic Regression, which are commonly used in educational data analytics to predict student success and classify learning data [24]. These models are also applicable in business intelligence applications, where they help organizations derive insights from large volumes of data and make informed decisions.

Overview of the Current State of Model Interpretability and Challenges Associated with Complex, Black-Box Models

While machine learning models offer substantial advantages in terms of predictive performance, one of the significant challenges is their lack of interpretability. Many ML models, particularly deep learning models and ensemble methods, are often regarded as "black boxes" due to their complex decision making processes that are difficult for humans to understand [25]. This opacity is especially problematic in critical applications, such as healthcare and finance, where understanding the rationale behind decisions is essential for trust, accountability, and ethical considerations [26].

The challenge of model interpretability is compounded by the trade-off between performance and explainability. While complex models like deep learning networks may provide higher predictive accuracy, their lack of transparency makes them unsuitable for high-stakes decision making, where model explanations are critical [26]. This trade-off raises concerns about the trustworthiness of AI systems, particularly when used in sensitive areas like patient care or financial decision making.

To address the interpretability challenges of black-box models, Explainable Artificial Intelligence (XAI) has emerged as a solution. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are widely used to enhance the transparency of these models [11]. These methods offer human-readable explanations for the predictions made by complex models, making it easier for users to understand how a model arrived at a particular decision. Model-specific techniques, like decision trees, are also employed because they inherently provide better interpretability compared to more complex models like deep neural networks [12].

Despite these advancements, there are still significant challenges in achieving complete interpretability, particularly when working with very complex models. For instance, the need for transparency in AI systems raises ethical concerns, especially in regulated industries where compliance and fairness are paramount [27]. Moreover, ensuring that explanations are both accurate and understandable remains a challenge, as the explanations provided by some XAI techniques can sometimes oversimplify the decision making process, leading to potentially misleading interpretations [22].

As a result, ongoing research into XAI is focused on developing more effective interpretability methods that balance the need for model performance with the need for transparent, ethical, and accountable AI systems [28]. Hybrid models, which combine symbolic knowledge with machine learning techniques, are one promising direction in this area [26]. These models aim to make machine learning more transparent while maintaining high performance, offering a potential solution to the challenges posed by traditional black-box models.

Review of XAI Techniques, Including Their Purpose, Methods, and Applications in Various Fields

Explainable Artificial Intelligence (XAI) focuses on enhancing the interpretability and transparency of complex machine learning (ML) models. The main objective of XAI is to provide human-readable explanations for the decisions made by these models, which are often regarded as "black-box" due to their lack of transparency [29]. The purpose of XAI techniques is to bridge the gap between model accuracy and understandability, ensuring that users can trust and validate the predictions made by ML models, particularly in high-stakes applications like healthcare, finance, and autonomous systems [30].

XAI techniques are commonly categorized by their purpose (pre-model, in-model, post-model), scope (local vs. global), and usability (*model-agnostic vs. model-specific*). These techniques are designed to make models more interpretable by providing clear insights into the decision making process. Some methods are model-agnostic, meaning they can be applied to any type of ML model, while others are model-specific, designed to work with particular model architectures [31].

LIME (Local Interpretable Model-Agnostic Explanations) is a popular XAI method that provides local explanations by approximating the black-box model with an interpretable model around the prediction of interest. This method helps users understand how specific features contributed to a particular prediction [29]. SHAP (SHapley Additive exPlanations), based on cooperative game theory, assigns a value to each feature to indicate its contribution to the prediction. SHAP values provide a unified measure of feature importance, making it easier to identify which features are most influential in the model's decision making process [32].

Other XAI methods include Grad-CAM (*Gradient-weighted Class Activation Mapping*) for explaining deep learning models in computer vision, as well as decision trees and rule-based systems, which are inherently more interpretable compared to more complex models like neural networks [33]. These methods are particularly useful for tasks where model interpretability is crucial, such as in medical diagnoses, fraud detection, and autonomous vehicles.

XAI techniques are applied across various domains. In healthcare, XAI helps in interpreting medical diagnoses and treatment recommendations, ensuring that medical professionals can trust AI-driven insights and make informed decisions [34]. In finance, XAI assists in risk assessment and regulatory compliance by providing transparent explanations for credit scoring and investment decisions [30]. Autonomous vehicles benefit from XAI by explaining decision making processes, improving safety and trust in these systems [35]. Additionally, XAI is used in agriculture, manufacturing, energy efficiency, education, and content recommendation, demonstrating its wide applicability in enhancing transparency and user trust [13].

Examination of Previous Studies Comparing XAI Methods with Traditional Machine Learning Models Regarding Transparency and Performance

A major challenge with traditional machine learning models, particularly deep neural networks and ensemble methods, is their lack of transparency. Unlike models like decision trees, which are inherently interpretable, complex models often operate as "black boxes," making it difficult to understand how decisions are made. This opacity creates issues in domains that require trust and accountability, such as healthcare and finance, where understanding the rationale behind predictions is essential [26].

XAI methods like LIME and SHAP have been shown to significantly improve the interpretability of machine learning models by providing insights into how predictions are made and which features are most influential [31]. These techniques enhance model transparency by offering clear, human-readable explanations for each decision made by the model. In comparison to traditional ML models, which offer no explanation, XAI methods allow users to understand the decision making process, making it easier to identify potential errors or biases in the model [29].

However, there is often a trade-off between interpretability and performance. While simpler models like decision trees are more interpretable, they may not offer the same level of predictive accuracy as more complex models like deep neural networks [26]. This trade-off is a key consideration when choosing between traditional ML models and XAI-enhanced models. In some cases, the need for interpretability may outweigh the performance benefits of more complex models, especially in high-stakes fields such as healthcare [32].

Despite these challenges, combining multiple XAI techniques can provide a more comprehensive understanding of the model's behavior and mitigate some of the limitations of individual methods [36]. For example, combining LIME and SHAP allows for a more detailed explanation of the model's decisions, improving both transparency and performance [32]. This combination helps address the inherent trade-offs between model complexity and interpretability, offering a balanced approach to understanding and trusting AI-driven decisions.

3. Proposed Method

The study employs a comparative experimental design to evaluate multiple Explainable Artificial Intelligence (XAI) techniques in large scale enterprise data analytics systems, focusing on their impact on model interpretability and predictive performance. XAI methods such as LIME, SHAP, and decision trees are applied to various machine learning models, with metrics like user comprehension, decision traceability, and predictive accuracy used to assess interpretability and transparency. The experimental setup involves applying XAI techniques to enterprise systems, gathering feedback from stakeholders, and comparing performance metrics to understand the trade-offs between model complexity and interpretability, with a focus on healthcare and finance applications.

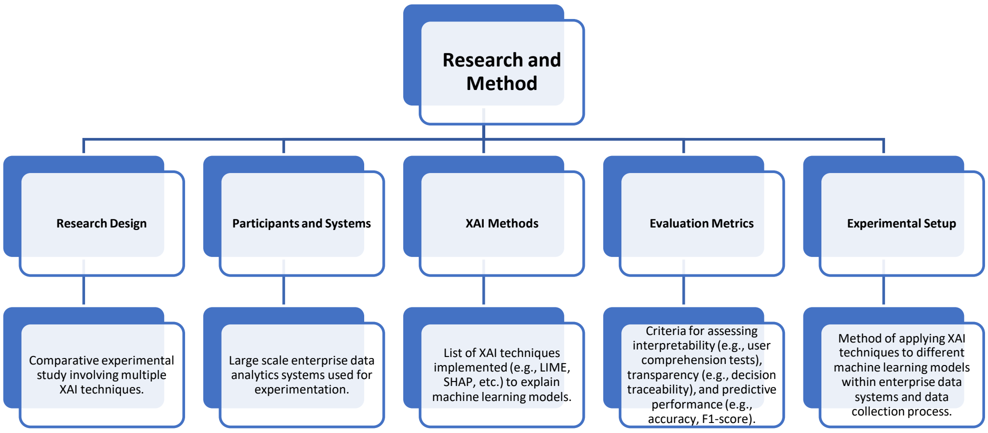

Figure 1. Flowchart structure.

Research Design

This study follows a comparative experimental design aimed at evaluating the effectiveness of multiple Explainable Artificial Intelligence (XAI) techniques in enhancing the interpretability of machine learning (ML) models while preserving their predictive performance. The design involves applying various XAI methods to large scale enterprise data analytics systems to assess their impact on transparency, interpretability, and model performance. This approach allows for a detailed comparison of how each XAI method contributes to making the decision making process of ML models more understandable and trustworthy.

Participants and Systems

The study utilizes large scale enterprise data analytics systems that are commonly deployed in industries such as healthcare, finance, and manufacturing. These systems typically process vast amounts of structured and unstructured data, making them ideal for experimenting with different ML models and XAI techniques. The experimental participants include various stakeholders from these industries, such as data scientists, domain experts, and decision-makers who interact with the outputs of the models and interpret their predictions. The systems used for experimentation incorporate enterprise-grade data platforms and analytics tools that support the deployment and continuous validation of machine learning models.

XAI Methods

The study applies several widely used XAI techniques to improve the interpretability of machine learning models in enterprise data systems. These include LIME, which provides local explanations by approximating black-box models with interpretable models around specific predictions; SHAP, based on cooperative game theory, which quantifies the contribution of each feature to the model's output; Grad-CAM, used in deep learning models for visualizing the most influential regions in an image for decision making; and decision trees, which offer a simple, interpretable structure for understanding model decisions. These techniques are implemented to assess their effectiveness in enhancing model transparency without compromising predictive performance.

Evaluation Metrics

The study evaluates the effectiveness of XAI techniques using three key criteria: interpretability, transparency, and predictive performance. Interpretability is assessed through user comprehension tests, where data scientists and domain experts rate their understanding of the explanations provided by the XAI methods, alongside subjective evaluations of the clarity and relevance of these explanations. Transparency is measured by examining decision traceability, which tracks how well users can understand the decision making process of the models. Predictive performance is assessed using metrics like accuracy and F1-score to determine how well the XAI techniques maintain or improve the model's ability to make accurate predictions while enhancing interpretability and transparency.

Experimental Setup

In the experimental setup, multiple machine learning models are applied to the enterprise data analytics systems, and different XAI techniques are integrated into these models. The models tested include both simple, interpretable models such as decision trees and complex models like deep neural networks and ensemble methods, commonly used in large scale enterprise systems. XAI techniques like LIME and SHAP are applied to these models to generate explanations for individual predictions.

The data collection process involves tracking how the XAI methods influence the interpretability and performance of the models. This includes gathering feedback from stakeholders (e.g., data scientists, business analysts) through surveys and interviews to assess the clarity and usefulness of the explanations provided by the XAI methods. Additionally, performance metrics such as accuracy and F1-score are recorded to measure any potential trade-offs between interpretability and model performance. The experimental procedure is repeated across various domains, such as healthcare and finance, to evaluate the generalizability of the results.

4. Results and Discussion

The application of XAI techniques, such as LIME and SHAP, significantly improved the interpretability of machine learning models, particularly complex ones like ensemble methods and deep neural networks, by providing clear, human-readable explanations. While these methods enhanced transparency and helped stakeholders understand the decision making process, there was a trade-off with predictive performance, as simpler models like decision trees, which are more interpretable, often had lower accuracy. The balance between transparency and accuracy remains a challenge in XAI, with ongoing research focused on finding hybrid solutions that combine the strengths of both interpretable and high-performance models to ensure both trust and high predictive power.

Results

The implementation of multiple Explainable Artificial Intelligence (XAI) techniques on machine learning models provided varied results in terms of interpretability and predictive performance. LIME and SHAP were particularly effective in enhancing the interpretability of complex models, such as ensemble methods and neural networks. LIME, for example, provided local explanations by approximating the complex model with simpler, interpretable models around individual predictions. This helped stakeholders understand how specific features influenced predictions, especially in models that are traditionally considered black boxes. SHAP, on the other hand, offered a more global view by providing feature importance scores, making it easier to understand the overall contribution of each feature to the model's decision making process.

Table 1. Comparison of XAI Techniques and Their Impact on Interpretability and Performance

XAI Technique	Model Type	Impact on Interpretability	Predictive Performance	Key Advantages
LIME	Model-agnostic, local	High	Moderate	Local explanations for individual predictions

SHAP	Model-agnostic, global	Very High	Slight decrease in performance	Provides feature importance scores, good for feature analysis
Grad-CAM	Deep learning (image data)	Moderate	High	Visualizes important regions for image-based models
Decision Trees	Model-specific	Very High	Low	Easily interpretable, fast predictions
Ensemble Methods	Complex models (e.g., random forest)	Moderate	Very High	High accuracy, but harder to interpret

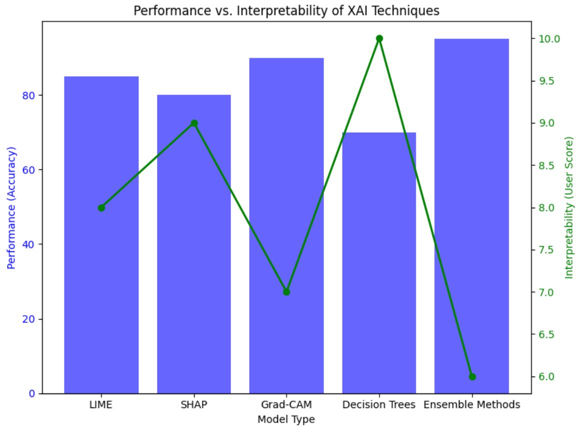

Figure 2. Performance Vs. Interpretability Of XAI Techniques.

The study applies various Explainable Artificial Intelligence (XAI) techniques, including LIME, SHAP, Grad-CAM, and decision trees, to enhance the interpretability of machine learning models used in large scale enterprise systems. The results show that XAI methods like LIME and SHAP significantly improve transparency by providing clear, human-readable explanations for model predictions, with SHAP offering a more global view of feature importance. However, there is a trade-off between interpretability and performance, as simpler models like decision trees are highly interpretable but less accurate compared to more complex models like deep neural networks. Despite this, combining multiple XAI methods can offer a comprehensive understanding of model behavior, enhancing both transparency and performance while addressing the inherent challenges of balancing interpretability with predictive power.

However, the application of XAI methods did not come without trade-offs in terms of performance. While techniques like LIME and SHAP enhanced transparency and interpretability, there was a slight reduction in the models' predictive accuracy, especially for more complex models. This is particularly evident when comparing decision trees, which are inherently interpretable, with deep learning models that offer higher predictive accuracy but are less transparent. Nonetheless, combining multiple XAI techniques allowed for a more comprehensive understanding of the models' behavior, helping to mitigate some of the performance limitations.

Discussion

The results from this study highlight the significant impact of XAI techniques on improving the interpretability of machine learning models. LIME and SHAP are two of the most effective methods in enhancing transparency by providing clear, human-readable explanations. LIME's ability to approximate black-box models with interpretable models around individual predictions makes it particularly useful in real-time decision making environments. Similarly, SHAP's application of cooperative game theory offers a unified measure of feature importance, which is valuable for understanding the contributions of individual features across various models. These techniques play a crucial role in ensuring that models, especially complex ones, are understood by the users, thus improving trust and accountability.

However, there remains a noticeable trade-off between interpretability and predictive performance. Simpler models, such as decision trees, are inherently more interpretable but often suffer from lower accuracy compared to more complex models, like deep neural networks. The trade-off between model transparency and accuracy is an ongoing challenge in the field of XAI, as users must often choose between a model that is easy to interpret and one that provides higher predictive accuracy. The results of this study show that while interpretability is critical for user trust and ethical decision making, the complexity of certain tasks may require the use of more sophisticated, less interpretable models.

Furthermore, while XAI methods such as LIME and SHAP significantly enhance model interpretability, they do not fully eliminate the challenges associated with complex models. For instance, Grad-CAM, though effective for computer vision models, did not provide as clear an explanation in tasks involving structured data, such as finance or healthcare. The need for a balance between model complexity and explainability remains, with the field of XAI still grappling with how to provide more interpretable explanations without sacrificing the predictive power of deep models. As the research continues, a key focus will be to develop hybrid approaches that combine the strengths of both interpretable models and complex, high-performance models.

5. Comparison

The performance of XAI-enhanced models was compared against traditional black-box models, revealing a notable trade-off between interpretability and predictive accuracy. Traditional models, particularly deep learning networks and ensemble methods, often provide high predictive performance but lack transparency, making it difficult for stakeholders to understand how decisions are made. In contrast, XAI techniques such as LIME and SHAP significantly enhanced the interpretability of these models by providing clear, human-readable explanations for individual predictions and overall feature importance. While these XAI methods improved model transparency, there was a slight reduction in predictive accuracy, particularly for models that rely on simple interpretability techniques like decision trees. This comparison highlights the challenge of balancing transparency and performance in AI models, where achieving high levels of interpretability may sometimes come at the cost of predictive accuracy.

The use of XAI in large scale enterprise data systems presents several advantages, particularly in enhancing stakeholder trust, decision making, and transparency. By providing explanations for complex machine learning models, XAI techniques help users better understand the underlying decision making processes, which is crucial in sectors such as healthcare, finance, and autonomous systems. The ability to trace decisions back to specific features and data points enables more informed decision making and fosters greater trust among stakeholders. Furthermore, XAI improves the transparency of models, ensuring that stakeholders can scrutinize and validate the predictions made by AI systems. This transparency is essential for ethical AI use, especially in regulated industries where accountability and fairness are paramount.

However, the application of XAI techniques also comes with several challenges and limitations. One key challenge is the complexity and scalability of certain XAI methods. For example, techniques like LIME and SHAP can be computationally expensive, particularly when applied to large scale data sets or real-time decision making processes. Additionally, while XAI techniques improve interpretability, they are not universally applicable to all model types. For instance, methods like Grad-CAM, which are effective for computer vision models, may not offer the same level of clarity when applied to structured data models commonly used in industries like finance or healthcare. These limitations point to the need for further refinement and optimization of XAI techniques to enhance their scalability and applicability across diverse domains and model types.

6. Conclusions

The key findings from this study indicate that Explainable Artificial Intelligence (XAI) techniques, such as LIME and SHAP, significantly improve the transparency and interpretability of machine learning models without causing significant reductions in predictive performance. These techniques provide clear explanations of model decisions, which enhance decision traceability and help stakeholders understand how predictions are made, especially in complex, black-box models. While there is a slight trade-off between interpretability and model accuracy, the use of XAI methods offers a practical balance that benefits industries where transparency is critical.

The practical implications for enterprise data analytics systems are profound. Adopting explainable models is crucial for fostering stakeholder trust and ensuring informed decision making. By integrating XAI into enterprise systems, organizations can enhance the understanding of AI-driven decisions, which is especially important in regulated fields such as healthcare and finance. This adoption not only helps build trust but also improves the ethical deployment of AI, as it allows for the examination of how decisions are made and provides accountability in AI systems.

Future research should focus on developing new XAI methods that are more scalable and efficient for larger datasets and real-time applications. Additionally, there is a need to refine current XAI techniques to improve their applicability across diverse model types and industries. Addressing the challenges of complexity, scalability, and interpretability limitations will be key in making XAI methods more accessible and effective for a wider range of applications. As XAI continues to evolve, it will play an increasingly important role in ensuring that machine learning models are both powerful and transparent, meeting the growing demand for ethical and trustworthy AI systems.

References

[1] J. Henriques, T. Rocha, P. de Carvalho, C. Silva, and S. Paredes, "Interpretability and Explainability of Machine Learning Models: Achievements and Challenges," in *IFMBE Proceedings*, 2024, pp. 81 – 94. doi: 10.1007/978-3-031-59216-4_9.

[2] A. Orlov, M. Ozhgibesov, and A. Agapov, "Enterprise machine learning and data platform," in *Saint Petersburg 2020 - Geosciences: Converting Knowledge into Resources*, 2020. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099534612&partnerID=40&md5=82882a49a08361aa105108f7c008560c

[3] M. A. Wajeed, K. M. Quadry, M. A. Mallik, and K. R. Khanna, "EXPLAINABLE MACHINE LEARNING FOR TEXT CLASSIFICATION: A NOVEL APPROACH TO TRANSPARENCY AND INTERPRETABILITY," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 3, pp. 907 – 915, 2025, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-86000130472&partnerID=40&md5=41626c819996be79427f950e73334fe6

[4] M. Bakator *et al.*, "Transforming Industry 5.0 Through Advanced Analytics and Machine Learning," *J. Soft Comput. Decis. Anal.*, vol. 3, no. 1, pp. 122 – 128, 2025, doi: 10.31181/jscda31202564.

[5] K. Hemachandran, S. Khanra, R. V Rodriguez, and J. R. Jaramillo, *Machine Learning for Business Analytics: Real-Time Data Analysis for Decision making*. 2022. doi: 10.4324/9781003206316.

[6] N. Kanchepu, *Unveiling the Black Box: The Crucial Role of Interpretability in Machine Learning Models*. 2025. doi: 10.2174/9789815305548125010004.

[7] R. Elshawi, Y. Sherif, M. Al-Mallah, and S. Sakr, "Interpretability in healthcare a comparative study of local machine learning interpretability techniques," in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2019, pp. 275 – 280. doi: 10.1109/CBMS.2019.00065.

[8] X. Wang *et al.*, *Assessing learning outcomes with interpretable learner models*. 2024. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209599448&partnerID=40&md5=f9d9dcecfb3713ac1965d5f145407cb1

[9] M. Saqr and S. López-Pernas, *Explainable Artificial Intelligence in Education: A Tutorial for Identifying the Variables that Matter*. 2025.

doi: 10.1007/978-3-031-95365-1_6.

[10] M. A. Shakir *et al.*, "Developing Interpretable Models for Complex Decision making," in *Conference of Open Innovation Association, FRUCT*, 2024, pp. 66 – 75. doi: 10.23919/FRUCT64283.2024.10749922.

[11] D. Gaspar, P. Silva, and C. Silva, "Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron," *IEEE Access*, vol. 12, pp. 30164 – 30175, 2024, doi: 10.1109/ACCESS.2024.3368377.

[12] Y. Swathi and M. Challa, "A Comparative Analysis of Explainable AI Techniques for Enhanced Model Interpretability," in *Proceedings - 2023 3rd International Conference on Pervasive Computing and Social Networking, ICPCSN 2023*, 2023, pp. 229 – 234. doi: 10.1109/ICPCSN58827.2023.00043.

[13] B. Mohammed, "A Review on Explainable Artificial Intelligence Methods, Applications, and Challenges," *Indones. J. Electr. Eng. Informatics*, vol. 11, no. 4, pp. 1007 – 1024, 2023, doi: 10.52549/ijeei.v11i4.5151.

[14] F. H. Yagin and A. Pinar, *Explainable artificial intelligence in medical research: A synopsis for clinical practitioners—Comprehensive XAI methodologies*. 2025. doi: 10.1016/B978-0-443-23979-3.00012-9.

[15] J. Jung, H. Lee, H. Jung, and H. Kim, "Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review," *Heliyon*, vol. 9, no. 5, 2023, doi: 10.1016/j.heliyon.2023.e16110.

[16] M. Thamer and Z. N. Sultani, "Explainable AI in the Medical Field: A Survey on Machine Learning Interpretability and Use Cases," *Al-Nahrain J. Sci.*, vol. 28, no. 4, pp. 188 – 206, 2025, doi: 10.22401/ANJS.28.4.15.

[17] B. P. Sheela and H. Girisha, "An Explainable Artificial Intelligence (XAI) Framework for Deep Learning Based Classification to Generate Textual Explanations on Predicted Images," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 6, pp. 651 – 662, 2024, doi: 10.22266/ijies2024.1231.50.

[18] S. Verma, S. Avasthi, P. Mishra, M. Gupta, and G. Agarwal, *XAI in healthcare: Black box to interpretable models*. 2024. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85192864780&partnerID=40&md5=efda84ca5dd8a898ea0c2ab68c9ff01a

[19] U. Khakurel and D. B. Rawat, "Evaluating Explainable Artificial Intelligence (XAI): Algorithmic Explanations for Transparency and Trustworthiness of ML Algorithms and AI Systems," in *Proceedings of SPIE - The International Society for Optical Engineering*, 2022. doi: 10.1117/12.2620598.

[20] S. Roychowdhury, V. Lanfranchi, and S. Mazumdar, "Evaluating explanation performance for clinical decision support systems for non-imaging data: A systematic literature review," *Comput. Biol. Med.*, vol. 197, 2025, doi: 10.1016/j.compbiomed.2025.110944.

[21] L. Weber, S. Lapuschkin, A. Binder, and W. Samek, "Beyond explaining: Opportunities and challenges of XAI-based model improvement," *Inf. Fusion*, vol. 92, pp. 154 – 176, 2023, doi: 10.1016/j.inffus.2022.11.013.

[22] A. Srinivasulu, C. V Ravikumar, S. Mastan, R. S. Vali, and S. Rajeyyagari, *Unveiling the Transformative Infuence of Machine Learning Techniques on Business Data Analytics: A Comprehensive Exploration*. 2025. doi: 10.1201/9781003640776-1.

[23] J. Narmadha, G. N. Kaavya, and S. D. Preethii, "Analysis on electricity generation forecasting system," in *Proceedings - 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering, ICEICE 2017*, 2017, pp. 1 – 5. doi: 10.1109/ICEICE.2017.8191901.

[24] A. Aleksieva-Petrova, V. Gancheva, and M. Petrov, "APTITUDE framework for learning data classification based on machine learning," *Int. J. Circuits, Syst. Signal Process.*, vol. 14, pp. 379 – 385, 2020, doi: 10.46300/9106.2020.14.51.

[25] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," *Cognit. Comput.*, vol. 16, no. 1, pp. 45 – 74, 2024, doi: 10.1007/s12559-023-10179-8.

[26] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206 – 215, 2019, doi: 10.1038/s42256-019-0048-x.

[27] M. Garouani, J. Mothe, A. Barhrhouj, and J. Aligon, "Investigating the Duality of Interpretability and Explainability in Machine Learning," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2024, pp. 861 – 867. doi: 10.1109/ICTAI62512.2024.00125.

[28] Y.-W. Chen, S.-Y. Chien, and F. Yu, "An overview of XAI Algorithms," in *2023 International Automatic Control Conference, CACS 2023*, 2023. doi: 10.1109/CACS60074.2023.10326174.

[29] S. Bhatnagar and R. Agrawal, "Understanding explainable artificial intelligence techniques: a comparative analysis for practical application," *Bull. Electr. Eng. Informatics*, vol. 13, no. 6, pp. 4451 – 4455, 2024, doi: 10.11591/eei.v13i6.8378.

[30] M. Kopzhasarova and D. Kozhamzharova, "Explainable AI (XAI): Techniques, applications, and challenges," in *CEUR Workshop Proceedings*, 2025. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105006886912&partnerID=40&md5=4007adb38600ba88e07d459c1ef6a021

[31] M. H. Azam, M. H. Hasan, N. Y. Murad, and E. A. Bt Patah, "Transparency in AI: A Review of Explainable Artificial Intelligence Techniques," in *2024 8th International Conference on Computing, Communication, Control and Automation, ICCUBEA 2024*, 2024. doi: 10.1109/ICCUBEA61740.2024.10774981.

[32] A. M. Salih *et al.*, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," *Adv. Intell. Syst.*, vol. 7, no. 1, 2025, doi: 10.1002/aisy.202400304.

[33] M. H. Albydhany and I. Al-Darraji, "Explainable AI for Visual Inspection: A Comparative Analysis and Review," *J. Vis. Exp.*, vol. 2025-October, no. 224, 2025, doi: 10.3791/69440.

[34] M. Panda and S. R. Mahanta, *Explainable Artificial Intelligence for Healthcare Applications Using Random Forest Classifier with LIME and SHAP*. 2024. doi: 10.1201/9781003442509-6.

[35] J. Duell, M. Seisenberger, and X. Fan, "A Comparison of Global Explanations Given on Electronic Health Records," *Lect. Notes Networks Syst.*, vol. 794, pp. 413 – 424, 2024, doi: 10.1007/978-3-031-44981-9_34.

[36] S. Kumar, N. Dhanda, and K. K. Gupta, "Exploring Explainable AI Methods for Decoding Machine Learning Model Decisions," in *Proceedings of the 2025 3rd International Conference on Inventive Computing and Informatics, ICICI 2025*, 2025, pp. 1083 – 1089. doi: 10.1109/ICICI65870.2025.11069883.