BALANCING ACCURACY AND INTERPRETATION: AN EMPIRICAL STUDY OF EXPLAINABLE AI TECHNIQUES IN BUSINESS CRITICAL PREDICTIVE MODELS

Sungho Kim[1], Sanjar Azizov[2], Sadeq Almanaseer[3], Edgar Plasas Mueses[4], Sumaia Arif[5], Abdul Alim[6], S M Jobayer Al Amin[5], Hossain Ahmed[5], FNU Anjali[6], Jaime Cantillo[4], Amar Shrestha[4], Sajeet Raj Aryal[6]

[1] Department of Computer Science, Korea University, Seoul, South Korea
[2] Department of Business Analytics, Pacific States University, Los Angeles, USA
[3] Department of Accounting (ACP – Accounting Certificate Program), Pacific States University, Los Angeles, USA
[4] Department of Computer Science, Pacific States University, Los Angeles, USA
[5] Department of Information Systems, Pacific States University, Los Angeles, USA
[6] Department of Cybersecurity, Pacific States University, Los Angeles, USA

ABSTRACT

The increasing use of artificial intelligence in business has led to widespread adoption of predictive models in critical decision-making processes. While highly accurate models offer strong performance, their lack of interpretability raises concerns related to trust, accountability, and regulatory compliance. This study examines the trade-off between predictive accuracy and interpretability in business-critical AI applications. It empirically analyzes explainable artificial intelligence techniques to evaluate their impact on model transparency and performance. The findings aim to support informed model selection and responsible AI deployment in real-world business contexts.

KEYWORDS

explainable artificial intelligence, predictive accuracy, model interpretability, business-critical models, decision support.

1. BACKGROUND AND PROBLEM DEFINITION

Artificial intelligence is increasingly used in business environments to support predictive decision-making in areas such as credit approval, fraud detection, customer churn analysis, and demand forecasting. In these applications, predictive models are often embedded directly into operational and strategic processes, making their outputs highly influential on business outcomes (Burrell, 2016).

2. PREDICTIVE ACCURACY IN BUSINESS–CRITICAL MODELS

Predictive accuracy is commonly treated as the primary performance metric for machine learning models in business settings. Highly accurate models enable organizations to reduce financial risk, optimize operations, and improve decision efficiency. As a result, businesses frequently adopt

complex algorithms that maximize predictive performance, especially in competitive or high-volume environments (Breiman, 2001).

3. LIMITATIONS OF BLACK-BOX MODELS

Despite their strong performance, many advanced machine learning models function as black-box systems. Their internal decision logic is difficult to interpret, which limits the ability of managers and stakeholders to understand how specific predictions are generated. This lack of transparency can undermine trust, complicate decision justification, and restrict the practical adoption of AI systems in business-critical contexts (Burrell, 2016; Rudin, 2019).

4. ROLE OF EXPLAINABILITY

Explainable Artificial Intelligence has emerged as a response to the interpretability limitations of black-box models. Explainability aims to provide insights into model behavior, either by describing general decision patterns or by explaining individual predictions. These explanations support transparency, oversight, and informed decision-making, particularly in regulated or high-stakes business environments (Doshi-Velez & Kim, 2017; Molnar, 2023).

5. PROBLEM STATEMENT

The central challenge addressed in this research is the trade-off between predictive accuracy and interpretability in business-critical AI applications. While accuracy is essential for reliable predictions, interpretability is necessary for trust, accountability, and compliance. Understanding how explainable AI techniques affect this balance is critical for selecting models that are both effective and suitable for real-world business use (Rudin, 2019).

6. EXPLAINABLE AI (XAI) TECHNIQUES

Business-critical predictive models often prioritize high predictive accuracy by employing complex machine learning algorithms. Although these models deliver superior performance, they typically lack transparency, making their decisions difficult to interpret. Explainable Artificial Intelligence (XAI) techniques address this challenge by providing interpretable explanations of model behavior while preserving predictive accuracy, thereby improving trust, accountability, and regulatory compliance (Molnar, 2023).

7. OVERVIEW OF COMMON XAI TECHNIQUES

Feature Importance: Feature importance methods identify the relative contribution of each input variable to the model's predictions. They provide global explanations that help stakeholders understand overall model behavior. While computationally efficient and easy to interpret, these methods do not explain individual predictions and may overlook feature interactions (Molnar, 2023).

SHAP (SHapley Additive exPlanations): SHAP is a model-agnostic technique grounded in game theory. It assigns each feature a value representing its contribution to a specific prediction relative to a baseline. SHAP supports both local and global explanations and is widely used in business-critical applications, despite its higher computational cost (Lundberg & Lee, 2017).

LIME (Local Interpretable Model-agnostic Explanations): LIME explains individual predictions by approximating a complex model locally with a simpler interpretable model. It highlights the most influential features for a specific decision but may produce unstable results due to random sampling (Ribeiro et al., 2016).

Example: Customer Churn Prediction

In customer churn prediction, SHAP values can show that high subscription prices, low service usage, and frequent customer complaints significantly increase churn risk. For individual customers, SHAP quantifies each factor's contribution, while LIME provides concise local explanations. These insights enable targeted retention strategies without sacrificing predictive accuracy (Lundberg & Lee, 2017; Ribeiro et al., 2016).

Summary

Explainable AI techniques play a crucial role in balancing accuracy and interpretability in business-critical predictive models. By integrating XAI methods, organizations can maintain high-performing models while ensuring transparency, trust, and responsible AI deployment (Doshi-Velez & Kim, 2017; Molnar, 2023).

8. ACCURACY VS. INTERPRETABILITY TRADE-OFF

8.1. Model Complexity and Predictive Accuracy

In predictive modeling, model complexity plays a central role in determining both accuracy and interpretability. Simple models such as linear regression, logistic regression, and shallow decision trees rely on transparent mathematical relationships between input variables and outcomes. Their structure allows stakeholders to clearly understand how predictions are generated, but this simplicity often limits their ability to capture complex, nonlinear patterns in large or high-dimensional datasets (Breiman, 2001; Molnar, 2023).

In contrast, complex models such as random forests, gradient boosting machines, and neural networks are designed to capture intricate relationships and interactions among variables. These models typically achieve higher predictive accuracy, especially in tasks such as fraud detection, credit risk assessment, and demand forecasting. However, their internal mechanisms are difficult to interpret, making them less transparent and harder to justify in business-critical decision-making contexts (Burrell, 2016; Rudin, 2019).

This creates a fundamental trade-off: as model complexity increases, predictive accuracy often improves, but interpretability decreases (Breiman, 2001).

8.2. Interpretability of Simple vs. Complex Models

Simple models are often preferred in regulated or high-risk environments because their predictions can be directly explained. For example, a logistic regression model used in credit scoring allows decision-makers to identify how variables such as income, debt ratio, and credit history influence approval decisions. This transparency supports regulatory compliance and helps organizations explain outcomes to customers (Rudin, 2019).

Complex models, while more accurate, operate as black boxes. A neural network used for fraud detection may outperform logistic regression in identifying fraudulent transactions, but it does not

naturally provide human-readable explanations for individual decisions. Without additional explanation tools, stakeholders may struggle to trust or validate these predictions, especially when errors occur (Burrell, 2016). Explainable AI techniques aim to reduce this gap by adding interpretability to complex models without significantly sacrificing performance (Doshi-Velez & Kim, 2017).

8.3. Empirical Evidence from Prior Studies

Empirical studies consistently show that simple models tend to underperform complex models in terms of raw accuracy, particularly when data relationships are nonlinear or highly interactive. However, these same studies highlight that interpretable models often lead to better human decision-making outcomes (Rudin, 2019).

For example, research in financial risk modeling shows that while neural networks achieve higher classification accuracy, decision-makers are more likely to rely on and correctly use predictions from interpretable models or explainable versions of complex models. In many cases, slightly lower accuracy is offset by improved trust, oversight, and error detection (Rudin, 2019).

Similarly, studies in fraud detection and credit approval demonstrate that combining complex models with XAI techniques such as SHAP improves acceptance and usability without causing a significant drop in predictive performance (Lundberg & Lee, 2017).

These findings may vary depending on data quality, industry context, and organizational decision-making environments.

8.4. Analytical Comparison

The accuracy–interpretability trade-off can be summarized as follows:

- Simple models provide high interpretability but may lack sufficient accuracy for complex tasks (Breiman, 2001).
- Complex models offer superior predictive performance but suffer from low transparency (Burrell, 2016).
- Explainable AI techniques enable complex models to retain most of their accuracy while providing meaningful explanations (Lundberg & Lee, 2017; Ribeiro et al., 2016).

Although black-box models often outperform interpretable models in controlled or benchmark settings, this advantage may diminish in real operational environments where human oversight and accountability are required (Rudin, 2019).

Empirical evidence suggests that the optimal solution in business-critical settings is not choosing between accuracy and interpretability, but strategically balancing both using explainable modeling approaches (Doshi-Velez & Kim, 2017; Molnar, 2023).

Summary

Model complexity directly influences the trade-off between accuracy and interpretability in predictive analytics. While complex models generally outperform simpler ones in accuracy, their lack of transparency limits their use in high-stakes business environments. Empirical studies indicate that integrating explainable AI techniques allows organizations to benefit from advanced predictive performance while maintaining accountability, trust, and regulatory compliance. This

balance is essential for responsible deployment of predictive models in business-critical applications (Rudin, 2019).

9. EMPIRICAL EVALUATION / CASE STUDIES (EXPLAINABLE AI)

Focus

This section analyzes real-world studies and examples where Explainable Artificial Intelligence (XAI) has been applied. The goal is to evaluate how this type of AI affects model accuracy and the quality of human decision-making (Doshi-Velez & Kim, 2017). Unlike traditional black-box models, XAI allows users to understand how and why decisions are made (Molnar, 2023).

9.1. Real-World Case Studies

9.1.1. Case1: Pricing Optimization in Businesses

Context. Many companies use artificial intelligence to determine the prices of their products or services. These systems analyze large amounts of data, such as sales history, market demand, seasonal trends, and customer behavior. Traditional AI models are often very accurate, but they operate as black boxes by providing a final price without explaining how the decision was reached. This creates challenges because managers cannot easily justify price changes to customers, employees, or executives (Bertsimas & Kallus, 2020).

What Changes with XAI. Explainable Artificial Intelligence shows the main reasons behind each pricing recommendation (e.g., increased demand, limited product availability, and changes in customer behavior). With these explanations, managers can understand the system's reasoning and decide whether the recommendation fits the business context (Molnar, 2023).

Results. Accuracy was slightly lower compared to black-box models, managers trusted the system more, and decisions were easier to explain and justify.

Conclusion. Even though explainability may reduce a small amount of mathematical accuracy, it can improve the quality of business decisions because humans can understand, evaluate, and justify the actions taken (Doshi-Velez & Kim, 2017).

9.1.2. Case2: Medical Diagnosis Support

In healthcare, artificial intelligence is used to assist doctors in diagnosing diseases from medical images such as X-rays or scans. Traditional AI models can detect patterns with high accuracy but do not explain how they reached their conclusions, which can be risky in critical medical decisions (Ghassemi et al., 2021).

What Changes with XAI. XAI systems visually highlight the areas of the image that influenced the model's decision (e.g., specific regions of the lungs and areas showing possible abnormalities). This allows doctors to verify whether the AI's decision makes sense, combine medical expertise with AI recommendations, and detect possible model errors (Ghassemi et al., 2021).

Results. Doctors made fewer mistakes, trust in AI recommendations increased, and medical decisions were easier to review.

Conclusion. In this case, XAI not only improves understanding of the system but can also increase safety and accuracy in medical decision-making (Ghassemi et al., 2021).

Performance vs. Interpretability Outcomes

Area	Accuracy	DecisionQuality	Interpretability
Pricingsystems	Slightlylower	Higher	High
Medicaldiagnosis	Higher	Higher	High
Black-boxAI	Higher	Higher	Low

10. BUSINESS IMPLICATIONS AND CONCLUSIONS

Highly accurate models are widely used in critical business areas, but their lack of transparency creates a trade-off between performance and explainability, especially under regulatory and stakeholder scrutiny (Burrell, 2016; Rudin, 2019). From a managerial perspective, this trade-off becomes especially relevant when AI-driven decisions directly affect customers, patients, or financial outcomes.

10.1. Managerial Implications

Managers must balance model accuracy and interpretability based on decision risk. High-stakes decisions require explainable models for accountability, oversight, and regulatory justification, while low-risk routine decisions can prioritize accuracy over transparency. Thus, the choice of the model should be context-driven, with explainability seen as a strategic consideration (Rudin, 2019).

10.2. Regulatory Implications

Regulations are increasingly demanding transparency, fairness, and accountability in automated decisions, especially in finance and healthcare. Non-compliance can lead to legal, reputational, and operational risks. Explainable AI helps organizations show how decisions are made, even with complex models, allowing businesses to stay ahead of regulatory changes and reduce compliance risks (European Commission, 2021).

10.3. Ethical Implications

Using opaque models in critical business applications raises ethical concerns about bias, fairness, and trust. Explainable models help identify and correct unfair outcomes, ensure alignment with ethical values, and build stakeholder confidence (Burrell, 2016; Rudin, 2019).

When to Prioritize Accuracy versus Explainability

Deciding whether to prioritize accuracy or explainability should depend on business risk and potential impact on stakeholders. In high-stakes, regulated areas like healthcare and finance, explainability should come first, even if it slightly reduces predictive accuracy. In contrast, for internal efficiency projects or low-impact applications, accuracy can be prioritized, while explainability is mainly used for model oversight and monitoring (Rudin, 2019).

11. RECOMMENDATIONS FOR BUSINESSES

Integrate Explainability Methods

Use explainability techniques such as SHAP and LIME alongside high-accuracy models to ensure transparency and build trust, particularly in customer-facing decisions (Lundberg & Lee, 2017; Ribeiro et al., 2016).

Adopt Hybrid Models

Combine complex, high-performance models with simpler, interpretable ones to balance accuracy and clarity (Rudin, 2019).

Educate Stake holders

Ensure employees, managers, and regulators understand the importance of both accuracy and interpretability for smoother AI adoption (Doshi-Velez & Kim, 2017).

Implement Ethical Oversight

Regularly audit AI systems for fairness and bias, using explainable methods to ensure ethical, transparent, and compliant decision-making (Burrell, 2016; European Commission, 2021).

12. FUTURE RESEARCH DIRECTIONS

Evaluation Frameworks

Develop comprehensive frameworks that assess both the accuracy and interpretability of AI models in real-world business settings. These frameworks should help businesses navigate the trade-offs between performance and explainability (Doshi-Velez & Kim, 2017).

Industry-Specific Best Practices

Future research could focus on creating industry-specific guidelines for balancing accuracy and explainability, taking into account each sector's unique regulatory and ethical needs (European Commission, 2021).

Human–AI Collaboration

Investigate how AI can work alongside human decision-makers, improving both the accuracy and interpretability of results. In industries like healthcare, AI might assist clinicians rather than replace them, providing insights that are easy to interpret and act upon (Ghassemi et al., 2021).

Bias Mitigation in Explainable AI

Research should explore how explainable AI techniques can help detect and correct bias in machine learning models, especially in high-risk areas like hiring, lending, and criminal justice (Rudin, 2019).

13. KEY CONCLUSIONS

In regulated sectors like healthcare, finance, and insurance, explainability is essential for compliance and managing legal risks. While accuracy matters, justifying AI decisions is key to building trust and accountability (European Commission, 2021; Rudin, 2019).

There is a trade-off between accuracy and interpretability. In high-risk areas, businesses may prioritize explainability over slight accuracy loss, but in operational tasks, accuracy is more critical (Breiman, 2001).

Transparent, explainable AI models are more likely to be trusted, leading to broader adoption and reduced resistance, while minimizing risks by ensuring fairness and accountability (Burrell, 2016; Doshi-Velez & Kim, 2017).

REFERENCES

[1] Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. Management Science,66(3), 1025–1044. https://doi.org/10.1287/mnsc.2018.3253

[2] Breiman, L. (2001). Statistical modeling: The two cultures. Statistical Science ,16(3), 199–231. https://doi.org/10.1214/ss/1009213726

[3] Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. Big Data & Society ,3(1), 1–12. https://doi.org/10.1177/2053951715622512

[4] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning [arXiv preprint arXiv:1702.08608]. https://arxiv.org/abs/1702.08608

[5] European Commission. (2021). Ethics guidelines for trustworthy AI [Accessed 2026-02-03]. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[6] Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health,3(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

[7] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems ,30, 4765–4774.

[8] Molnar, C. (2023). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). https://christophm.github.io/interpretable-ml-book/

[9] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “why should I trust you?”: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining , 1135–1144. https://doi.org/10.1145/2939672.2939778

[10] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence ,1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x