INTERNATIONAL REVIEW OF
ARTIFICIAL INTELLIGENCE
IN MARKETING

NAWALA

Volume 1, Number 1, 2022

Explainable AI in Marketing Analytics: Balancing Predictive Power and Managerial Trust

Rizki Nur Fatimah [1]

[1] Universitas Sarjanawiyata Tamansiswa, Yogyakarta, Indonesia

Abstract

Article history:
Received: January 15, 2022
Revised: February 27, 2022
Accepted: April 3, 2022
Published: June 30, 2022

Keywords:
Algorithmic Transparency, Explainable Artificial Intelligence, Managerial Trust, Marketing Analytics, Predictive Models.

Identifier:
Nawala
Page: 15-28
https://nawala.io/index.php/iraim

This article examines how explainable artificial intelligence can reconcile the tension between predictive performance and managerial trust in contemporary marketing analytics. Drawing on a synthesis of prior studies, it maps how artificial intelligence and machine learning are deployed for targeting, personalization, resource allocation, and customer journey optimization, while highlighting the persistent opacity of complex black box models. The review identifies key families of explanation techniques, including global and local model explanations, feature importance analysis, and counterfactual reasoning, and discusses how these tools can make algorithmic decision paths intelligible for managers. The findings show that explainable artificial intelligence can enhance error diagnosis, perceived fairness, and confidence in model outputs, yet also reveal that technical transparency alone is insufficient to overcome algorithm aversion. The article argues that effective use of explainable artificial intelligence in marketing requires integrated governance arrangements that combine powerful predictive models with human oversight, user control, and clear accountability structures. These conditions support more responsible data driven marketing decisions.

*Corresponding author:
(Rizki Nur Fatimah)

1. Introduction

Artificial intelligence has become a core engine of contemporary marketing analytics, powering applications such as customer segmentation, personalization, churn prediction, and dynamic pricing. Recent work shows that machine learning systems can process massive, unstructured data streams and generate highly accurate predictions that reshape how firms design and execute marketing strategies (Davenport et al., 2020; Huang & Rust, 2021). In advertising and digital customer journeys, artificial intelligence is already embedded in programmatic media buying, recommendation engines, and real-time message optimization, amplifying the strategic role of data driven decision making in marketing practice (Kietzmann et al., 2018; Paschen et al., 2019).

However, the models that offer the greatest predictive power are often complex "black boxes" whose internal logic is opaque to managers. Surveys of black box explanation methods emphasize that state of the art algorithms, especially deep learning, trade interpretability for accuracy, making it difficult for non technical decision makers to understand why a particular customer was targeted, scored as high risk, or offered a specific price (Guidotti et al., 2018; Arrieta et al., 2020). In high stakes marketing contexts such as credit-based offers, insurance products, or vulnerable consumer segments, this opacity raises concerns about fairness, accountability, and alignment with brand values, and can limit the willingness of managers to fully rely on algorithmic recommendations (Davenport et al., 2020).

Explainable artificial intelligence has emerged as a response to this tension, aiming to make complex predictive models more transparent by generating human understandable explanations of their outputs. Conceptual work on explainable artificial intelligence highlights that explanations can be global, describing overall model behavior, or local, clarifying why a specific prediction was made, and that both forms of explanation are important for building trust and enabling effective human oversight (Guidotti et al., 2018; Arrieta et al., 2020). Empirical studies in business analytics contexts show that explainable artificial intelligence can enhance users' ability to diagnose model errors and foster more calibrated trust by revealing the features and decision paths that drive algorithmic outcomes (Mahbooba et al., 2021).

At the same time, behavioral research on algorithm aversion indicates that managers may reject algorithmic advice when it appears inscrutable or fallible, even when it is objectively more accurate than their own judgment. Providing transparency, controllability, or opportunities to adjust model recommendations has been shown to mitigate this aversion and increase acceptance of analytic tools (Dietvorst et al., 2018; Jung et al., 2021). In the context of marketing analytics, these insights suggest that the value of explainable artificial intelligence lies not only in regulatory compliance or technical transparency, but also in supporting managerial sensemaking, responsibility, and confidence when acting on model outputs.

The discussion builds on these streams of research to examine how explainable artificial intelligence can be integrated into marketing analytics in ways that balance predictive performance with managerial trust. It addresses the conceptual foundations of explainable artificial intelligence, the specific challenges of opacity in marketing decision environments, and the behavioral mechanisms through which explanations shape managers' perceptions of usefulness, fairness, and risk. By articulating design principles and governance considerations for explainable marketing analytics, the discussion aims to offer guidance on how firms can harness powerful predictive models while preserving the human judgment, accountability, and stakeholder trust that remain essential to effective marketing management.

2. Literature Review

The literature on artificial intelligence in marketing analytics highlights both the transformative potential of predictive models and the managerial concerns that arise when these models are opaque. At the broad level, studies show that artificial intelligence and machine learning increasingly underpin key marketing tasks such as targeting, personalization, and resource allocation, with firms using large-scale behavioral and contextual data to improve prediction and decision quality (Ma & Sun, 2020). Systematic reviews emphasize that artificial intelligence in marketing is moving from isolated tools to a strategic capability that reshapes customer relationship management, channel management, and value creation, while simultaneously raising questions around governance and responsible deployment (Libai et al., 2020; Verma et al., 2021).

At the same time, a parallel stream of work in computer science and information systems stresses that the models delivering the highest predictive accuracy are often black-box systems whose internal workings are difficult to interpret. Surveys of explainable artificial intelligence document a fast growing body of techniques such as feature importance measures, local surrogate models, and counterfactual explanations that seek to make complex models more transparent without fully sacrificing performance (Adadi & Berrada, 2018). This literature clarifies the distinction between global explanations, which describe overall model behavior, and local explanations, which focus on individual predictions, and argues that both are essential for meaningful human oversight in applied decision environments.

Research on customer relationship management further shows that the diffusion of artificial intelligence changes how firms prioritize and treat customers, often based on predicted lifetime value and risk scores that are not directly visible to managers or consumers (Libai et al., 2020; Ma & Sun, 2020). These developments heighten concerns about discrimination, unintended bias, and misalignment with brand values when algorithmic treatment rules are not well understood by decision makers. From a behavioral perspective, studies of algorithmic affordance demonstrate that perceived fairness, accountability, and transparency are central antecedents of trust in algorithmic systems and shape whether managers are willing to adopt and rely on algorithmic recommendations (Shin & Park, 2019). When stakeholders lack insight into why a model produced a given recommendation, they may either reject superior advice or over rely on outputs they do not fully understand, both of which undermine the potential value of analytics.

Overall, existing work suggests that explainable artificial intelligence sits at the intersection of technical model design and human judgment in marketing analytics. Prior studies establish that artificial intelligence can significantly enhance prediction and decision making in marketing, and that explainability techniques can make complex models more interpretable. However, there remains a need for domain specific evidence on how different forms of explanation affect managers' perceptions of usefulness, fairness, and risk in concrete marketing decisions, and on how firms can design governance arrangements that balance predictive power with managerial trust (Libai et al., 2020; Verma et al., 2021).

3. Methods

The study employs a systematic literature review method to synthesize current knowledge on explainable artificial intelligence in marketing analytics and its role in balancing predictive power with managerial trust. The review began by defining the research questions and developing a review protocol that specified the concepts of interest, namely artificial intelligence and machine learning applications in marketing analytics, explainable artificial intelligence techniques, and managerial perceptions of usefulness, fairness, risk, and trust. Academic databases such as Scopus, Web of Science, ScienceDirect, and Google Scholar were searched using combinations of keywords including "artificial intelligence," "machine learning," "marketing analytics," "explainable artificial intelligence," "algorithmic transparency," "interpretability," and "managerial trust." Only peer reviewed journal articles written in English and focusing on marketing, customer analytics, or closely related business contexts were included, while conference papers, non scholarly reports, and purely technical studies without managerial implications were excluded.

The remaining articles were screened through title, abstract, and full-text evaluation against the inclusion criteria, followed by a quality assessment based on clarity of research design, transparency of methods, and relevance to the research questions. For each selected study, data were extracted on the type of artificial intelligence models used, the explainability techniques applied, the marketing application area, the outcomes related to predictive performance, and the reported effects on managerial or user trust. These data were then analyzed through thematic synthesis to identify recurring patterns, tensions, and gaps, enabling the review to map how explainable artificial intelligence is currently conceptualized and implemented in marketing analytics and where further research is needed to support responsible and trustworthy deployment.

4. Results and Discussion

The systematic review shows that artificial intelligence has already become deeply embedded in core marketing analytics activities, but that this technical progress has not been matched by equivalent advances in transparency and managerial understanding. Studies on customer targeting, personalization, and resource allocation indicate that firms increasingly rely on machine learning models trained on large-scale behavioral and contextual data to improve prediction accuracy and campaign performance (Davenport et al., 2020; Ma & Sun, 2020; Huang & Rust, 2021). In line with this, empirical work on advertising and digital customer journeys finds that artificial intelligence supports programmatic media buying, recommendation engines, and real-time message optimization, effectively turning data driven decision making into a strategic capability that reshapes customer relationship management and channel management (Kietzmann et al., 2018; Paschen et al., 2019; Libai et al., 2020; Verma et al., 2021). These findings confirm that predictive power and automation are now central to contemporary marketing practice.

However, the review also reveals a persistent tension between predictive performance and managerial trust, which is largely driven by the opacity of state of the art models. Evidence from computer science and information systems highlights that the most accurate algorithms, in particular deep learning models, often behave as black boxes whose internal logic remains inaccessible to non technical decision makers (Adadi & Berrada, 2018; Guidotti et al., 2018; Arrieta et al., 2020). In marketing applications this opacity becomes particularly problematic where predictions affect sensitive outcomes such as eligibility for credit-based offers, pricing, or the prioritization of vulnerable customer segments (Davenport et al., 2020; Libai et al., 2020). Research on customer relationship management shows that algorithmic rules based on predicted lifetime value or risk scores can change how firms classify and treat customers, while the underlying criteria remain invisible to managers and consumers (Libai et al., 2020; Ma & Sun, 2020). These patterns raise concerns about discrimination, unintended bias, and misalignment with brand values, and they create a trust gap between what models can technically achieve and what managers feel comfortable implementing in practice.

The findings further indicate that explainable artificial intelligence provides a promising, although still partial, response to this gap. Surveys of explainable artificial intelligence techniques document a growing toolkit that includes feature importance measures, local surrogate models, and counterfactual explanations designed to make complex models more interpretable without fully sacrificing accuracy (Adadi & Berrada, 2018; Guidotti et al., 2018; Arrieta et al., 2020). Conceptual work clarifies that global explanations, which describe aggregate model behavior, and local explanations, which justify specific predictions, serve complementary roles and are both necessary for meaningful human oversight (Guidotti et al., 2018). Empirical studies in business analytics contexts suggest that such techniques can help users diagnose model errors, understand key drivers of predictions, and develop more calibrated trust in algorithmic systems by revealing the features and decision paths that shape outcomes (Mahbooba et al., 2021). In the context of marketing analytics this implies that well designed explanation routines can transform artificial intelligence from an opaque oracle into a more collaborative decision aid that supports managerial sensemaking.

At the same time, the review underscores that technical explainability alone is not sufficient to guarantee acceptance of artificial intelligence in marketing. Behavioral research on algorithm aversion shows that managers may still resist algorithmic recommendations when systems are perceived as inscrutable, rigid, or fallible, even when objective performance exceeds human judgment (Dietvorst et al., 2018). Evidence that providing users with limited control, such as the ability to adjust or override model outputs, can reduce aversion and increase willingness to rely on algorithms points to the importance of perceived autonomy and accountability in human algorithm interaction (Jung et al., 2021). Complementary work on algorithmic affordance highlights that perceptions of fairness, accountability, and transparency are central antecedents of trust, and shape whether managers view artificial intelligence as a legitimate basis for decisions that affect customers (Shin & Park, 2019). Taken together, these insights suggest that explainable artificial intelligence must be embedded in broader governance arrangements that combine technical transparency with clear roles, responsibilities, and escalation paths for human decision makers.

Overall, the review indicates that explainable artificial intelligence sits at the intersection of technical model design, organizational capabilities, and managerial cognition in marketing analytics. Prior studies confirm that artificial intelligence can significantly enhance predictive performance and that explainability techniques can increase interpretability (Adadi & Berrada, 2018; Davenport et al., 2020; Ma & Sun, 2020; Arrieta et al., 2020). However, there remains limited domain specific evidence on which types of explanations are most effective for different marketing tasks, how explanations should be tailored to the needs of managers with varying levels of analytical expertise, and how explanation practices interact with organizational norms, brand positioning, and regulatory pressures (Libai et al., 2020; Verma et al., 2021). These gaps point to a future research agenda that moves beyond generic discussions of transparency toward experimentally and qualitatively grounded studies of explainable artificial intelligence in real marketing decision environments, with particular attention to fairness, accountability, and the conditions under which predictive power can be reconciled with sustained managerial trust.

5. Conclusion

This study concludes that artificial intelligence has become a strategic pillar of contemporary marketing analytics, enabling firms to leverage large scale behavioral and contextual data for more precise targeting, personalization, and resource allocation. The evidence shows that machine learning and related technologies substantially enhance predictive performance and support data driven decision making across customer journeys, advertising, and customer relationship management. At the same time, the review reveals a persistent tension between the pursuit of accuracy and the need for managerial trust. Highly complex, black box models intensify concerns about fairness, accountability, and alignment with brand values, particularly in high stakes contexts where algorithmic decisions shape access to offers, prices, and attention. These concerns underscore that technical sophistication alone is not sufficient; explainability and governance are equally central to the responsible use of artificial intelligence in marketing.

Explainable artificial intelligence emerges from the review as a promising, yet incomplete, bridge between predictive power and managerial trust. Techniques such as global and local explanations, feature importance, and counterfactual reasoning can increase transparency, improve error diagnosis, and foster more calibrated trust by making decision paths more intelligible. However, the findings also make clear that explanation mechanisms must be embedded within broader organizational arrangements that recognize the behavioral dynamics of algorithm aversion, the importance of perceived fairness and accountability, and the continued role of human judgment. For practitioners, this implies that designing marketing analytics systems requires simultaneous attention to model performance, interpretability, user control, and clear lines of responsibility. For future research, the gaps identified point to the need for more domain specific, empirically grounded studies on which explanation types work best for different marketing tasks, how managers with diverse analytical capabilities interpret and use explanations, and how governance frameworks can be structured to ensure that powerful predictive models reinforce rather than erode stakeholder trust.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24–42.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1–42.

Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50.

Jung, M., & Seiter, M. (2021). Towards a better understanding on mitigating algorithm aversion in forecasting: An experimental study. *Journal of Management Control, 32*(4), 495–516.

Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. *Journal of Advertising Research, 58*(3), 263–267.

Libai, B., Bart, Y., Gensler, S., Hofacker, C. F., Kaplan, A., Kötterheinrich, K., & Kroll, E. B. (2020). Brave new world? On AI and the management of customer relationships. *Journal of Interactive Marketing, 51*(1), 44–56.

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing: Connecting computing power to human insights. *International Journal of Research in Marketing, 37*(3), 481–504.

Mahbooba, B., Timilsina, M., Sahal, R., & Serrano, M. (2021). Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model. *Complexity, 2021*(1), 6634811.

Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons, 63*(3), 403–414.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277–284.

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights, 1*(1), 100002.