INTERNATIONAL REVIEW OF
ARTIFICIAL INTELLIGENCE
IN MARKETING

NAWALA

Volume 2, Number 2, 2023

Customer Engagement Scoring with Machine Learning: Integrating Clickstream, Social, and CRM Data

Ainun Mardia Syamsir[1*]

[1] Universitas Muhammadiyah Makassar, Makassar, Indonesia

Abstract

Article history:
Received: August 30, 2023
Revised: September 17, 2023
Accepted: October 28, 2023
Published: December 30, 2023

Keywords:
Clickstream Data, CRM Data, Customer Engagement Scoring, Machine Learning, Social Media Data.

Identifier:
Nawala
Page: 105-118
https://nawala.io/index.php/iraim

This study investigates how machine learning is used to construct customer engagement scores from clickstream, social media, and CRM data in recent marketing research. It asks how engagement is conceptualized and operationalized, which data sources and feature representations are employed, and to what extent multi source integration is implemented in existing models. Using a systematic literature review of peer reviewed studies from 2017 to 2022, the article maps contexts, data characteristics, modelling approaches, and evaluation practices. The synthesis shows that most studies model engagement as a single channel behavioral outcome and rely on task specific indicators, with deep learning and tree-based methods frequently outperforming traditional models. Empirical work that jointly exploits clickstream, social, and CRM data in unified engagement scoring architectures remains limited. The review develops an integrative view of current design patterns, identifies conceptual and methodological gaps, and outlines priorities for future research on cross channel, value-oriented engagement scoring with machine learning.

*Corresponding author:
(Ainun Mardia Syamsir)

1. Introduction

Customer engagement has become a central construct in contemporary marketing, capturing the cognitive, emotional, and behavioral manifestations of how customers relate to brands beyond mere transactions. In data-rich digital environments, engagement is increasingly conceptualized as a dynamic, multi-touch process that unfolds across channels and throughout the customer journey (Kunz et al., 2017). Empirical studies, particularly in social media and online platforms, show that higher engagement is associated with outcomes such as loyalty, advocacy, and firm performance, yet its conceptualization and operationalization remain heterogeneous and fragmented (Dissanayake et al., 2019; Trunfio & Rossi, 2021; Vinerean & Opreana, 2021). This inconsistency complicates attempts to design robust, comparable engagement scores that can effectively guide managerial decision-making.

In parallel, firms now routinely collect detailed clickstream traces from websites and mobile applications, interaction data from social media, and rich customer relationship management (CRM) records covering purchases, complaints, service contacts, and campaign responses. Big-data-enabled CRM frameworks emphasize that integrating these heterogeneous data sources is essential to construct a holistic, relationship-oriented view of the customer and enable personalized, real-time marketing strategies (Zerbino et al., 2018; Anshari et al., 2019). Research on advanced customer analytics further shows that strategic value arises when organizations build infrastructures that combine relationship-oriented data across silos to support predictive modelling of acquisition, retention, and customer equity

(Kitchens et al., 2018). Nevertheless, most of this work focuses on general analytics capabilities rather than on constructing unified, multi-source customer engagement scores that span clickstream, social, and CRM data.

Machine learning has rapidly become the dominant analytical paradigm in marketing analytics, offering powerful tools to exploit high-dimensional, longitudinal behavioural data. Recent reviews indicate that machine learning is widely applied to churn prediction, customer lifetime value estimation, recommendation, and response modelling tasks (Ngai & Wu, 2022). In online settings, deep learning models that ingest raw clickstream sequences have achieved substantial gains in predicting purchase behaviour and conversion probabilities (Koehn et al., 2020), while machine-learning-based retention models relying on CRM attributes frequently outperform traditional statistical approaches (Schaeffer & Sanchez, 2020). Yet these applications generally treat clickstream, social media, and CRM data as separate domains, and engagement measurement studies continue to focus on single-channel indicators, particularly social media metrics, rather than truly integrated, cross-channel engagement scoring (Trunfio & Rossi, 2021; Vinerean & Opreana, 2021).

This article addresses these gaps by conducting a systematic literature review of peer-reviewed studies published between 2017 and 2022 that employ machine learning to model or score customer engagement using clickstream, social, and CRM data, either individually or in combination. Building on customer engagement and big-data CRM perspectives, the review synthesises how engagement is conceptualised and operationalised, which machine learning techniques and feature representations are adopted, and the extent to which multiple data sources are integrated into unified scoring models (Kunz et al., 2017; Zerbino et al., 2018; Kitchens et al., 2018). The study then develops an integrative framework for customer engagement scoring with machine learning, identifies methodological and organizational challenges related to multi-source data integration, and proposes a research agenda to advance engagement-centric analytics capable of supporting more nuanced, cross-channel customer strategies.

2. Literature Review

Research on customer engagement in digital environments has developed along two main lines: conceptual clarification and empirical operationalization. Early work in the Big Data context describes engagement as a multidimensional construct that emerges across channels and touchpoints, combining cognitive, emotional, and behavioral manifestations (Kunz et al., 2017). Subsequent reviews show that most empirical studies rely on observable behavioral indicators such as clicks, likes, comments, shares, time spent, and depth of navigation, usually within single platforms or channels (Dissanayake et al., 2019; Trunfio & Rossi, 2021; Vinerean & Opreana, 2021). More recent studies highlight that content related factors, such as authenticity and entertainment value, shape how customers respond and engage in online settings, reinforcing the idea that engagement is both behavioral and experiential (Eigenraam et al., 2021).

In parallel, machine learning has become central to marketing analytics, including applications related to engagement. In social media contexts, supervised learning models have been used to predict engagement reactions to brand posts by combining content features with interaction histories to forecast likes, comments, or shares (Dai & Wang, 2021). Clickstream based studies exploit sequential navigational data to trace stages of user engagement along the online journey, showing that path structures and temporal patterns are informative about engagement intensity and progression (Kumar et al., 2019; Koehn et al., 2020). CRM oriented research, supported by big data enabled architectures, uses classification and regression models to predict outcomes such as churn, retention, and cross sell, often treating engagement either as an intermediate predictor or as a narrow outcome rather than as a comprehensive scoring construct (Kitchens et al., 2018; Zerbino et al., 2018; Ngai & Wu, 2022).

Despite these advances, there remains a notable fragmentation in how engagement is modelled across clickstream, social, and CRM domains. Studies grounded in big data CRM frameworks argue that strategic value arises when firms integrate relationship-oriented data across silos, including transactional records, service interactions, and external digital traces (Zerbino et al., 2018; Anshari et al., 2019). However, empirical work that builds unified customer engagement scores with machine learning using combined clickstream, social, and CRM data is still scarce. Existing research typically focuses on one dominant data source, which limits comparability across studies and underutilizes the potential of multi-source integration for engagement scoring (Koehn et al., 2020; Schaeffer & Sanchez, 2020). This fragmentation provides the core motivation for a systematic literature review that maps how different data sources and machine learning techniques have been used so far, and identifies design patterns and gaps for integrated customer engagement scoring.

3. Methods

The study adopts a systematic literature review design to synthesize how machine learning has been used to construct customer engagement scores from clickstream, social, and CRM data. The review is guided by research questions concerning 1) how customer engagement is conceptualized and operationalized when modelled with machine learning, 2) which data sources and feature representations are used, and 3) how far existing work integrates multiple data sources in unified scoring models. Relevant studies were identified through structured searches in major databases, including Scopus, ScienceDirect, IEEE Xplore, ACM Digital Library, and Google Scholar. Search strings combined terms related to customer engagement, machine learning, and data sources (for example, "customer engagement score", "engagement prediction", "clickstream", "social media", "CRM", "machine learning", "deep learning"). The search was restricted to peer reviewed journal articles and conference papers published in English between 2017 and 2022. Duplicates were removed, then titles and abstracts were screened, followed by full text assessment based on predefined inclusion and exclusion criteria.

Studies were included if they reported empirical applications of machine learning models that predict, classify, or score customer engagement at the individual or segment level using clickstream, social media, CRM, or combinations of these data sources. Conceptual papers, purely methodological contributions without empirical data, non-marketing applications, and studies that did not focus on engagement related outcomes were excluded. For each included study, data were extracted on bibliographic information, context and industry, engagement definitions and indicators, data sources, feature engineering strategies, machine learning methods, evaluation metrics, and approaches to multi source integration. A simple quality appraisal was conducted, focusing on clarity of research design, transparency of data and model description, appropriateness of evaluation procedures, and relevance to the review questions. The extracted information was then synthesized narratively and organized around engagement conceptualization, data integration patterns, and machine learning techniques.

4. Results and Discussion

4.1 How Do Current Machine Learning Approaches Conceptualize and Model Customer Engagement Across Channels?

The reviewed studies show that customer engagement scoring with machine learning is still at an early and fragmented stage, particularly when viewed from a multi-channel perspective. Conceptually, most articles adopt engagement as a behavioral construct captured through observable actions such as clicks, views, likes, comments, shares, time on site, recency of visits, and response to campaigns, often with only brief reference to the richer cognitive and emotional dimensions highlighted in broader engagement frameworks (Ng et al., 2020; Santini et al., 2020). Operationalizations are typically task specific. Social media studies define engagement scores in terms of post level interactions or aggregated engagement classes for brand accounts (Jung & Jeong, 2020; Dai & Wang, 2021), whereas clickstream research focuses on navigation depth, path types, or stage-based engagement indices within the online journey (Kumar et al., 2019). CRM centric work tends to treat engagement as propensity to respond, purchase, or remain active, which is then inferred from transactional and interaction histories (Kitchens et al., 2018; Zerbino et al., 2018). As a result, the field lacks a consistent, cross channel definition of what a customer engagement score represents and how it should relate to downstream outcomes such as customer lifetime value or loyalty.

From a data and modelling perspective, the corpus confirms strong progress at the single source level. Social media studies rely on platform APIs to collect large scale interaction logs and often use supervised learning to predict engagement categories or intensities from content features, user characteristics, and historical interactions (Jung & Jeong, 2020; Dai & Wang, 2021). Results generally show that tree-based ensembles and deep learning architectures outperform linear baselines for predicting likes, comments, and shares. In clickstream contexts, sequence models such as recurrent neural networks and other deep architectures leverage ordered page views and time stamps to predict conversion or to classify users into engagement stages, demonstrating the predictive value of path structure and temporal patterns (Kumar et al., 2019; Koehn et al., 2020). CRM focused research, in turn, integrates structured attributes about purchases, tenure, and service contacts into classification or regression models for churn, retention, and campaign response, often achieving substantial gains over traditional scoring rules (Kitchens et al., 2018; Ngai & Wu, 2022). However, only a small subset of studies explicitly frame their outputs as engagement scores rather than as task specific probabilities, and even fewer specify how these scores should be calibrated and compared across models or contexts.

4.2 To What Extent Do Existing Studies Integrate Multi-Source Data to Build Holistic Customer Engagement Scoring Models?

With respect to multi source integration, the findings highlight a clear gap between conceptual calls for holistic engagement and actual empirical practice. Big data CRM frameworks and engagement reviews argue that value is created when firms combine transactional, interactional, and digital trace data to obtain a unified view of the customer (Kunz et al., 2017; Zerbino et al., 2018; Ng et al., 2020). A handful of empirical works illustrate this potential. For example, Aluri et al. (2019) integrate loyalty program transactions with behavioral and demographic information in a hospitality setting to build a dynamic engagement model that guides personalized promotions, while maintaining a clear link to value outcomes. Other studies implicitly approximate integration by enriching social media or clickstream models with CRM inspired features such as prior spend, tenure, or previous campaign exposure, but still focus their discussion on single channel engagement metrics (Jung & Jeong, 2020; Dai & Wang, 2021). Overall, explicit architectures that jointly exploit clickstream, social, and CRM data in a single machine learning pipeline for engagement scoring remain rare. Most work treats each data source as a separate domain with its own indicators, models, and evaluation metrics, which limits comparability and hinders the development of transferable engagement scoring practices.

Taken together, these patterns have several theoretical and managerial implications. Conceptually, the fragmentation of engagement definitions suggests the need to link machine learning based scores more tightly to established engagement frameworks and to clarify whether a given score reflects momentary interaction, relationship depth, or value potential across the customer journey (Ng et al., 2020; Santini et al., 2020). Methodologically, there is an opportunity to adopt more systematic feature engineering and representation learning strategies that explicitly capture cross channel behavior, such as joint embeddings of clickstream paths, social interactions, and CRM events, as well as to adopt consistent evaluation criteria that relate engagement scores to downstream business outcomes like loyalty, advocacy, and profitability (Aluri et al., 2019; Ngai & Wu, 2022). For practitioners, the review suggests that current models often excel at optimizing local engagement metrics within individual platforms but fall short of supporting strategic, engagement centric decision making across channels and touchpoints. Future research should therefore explore architectures that integrate heterogeneous data sources, experiment with alternative target formulations such as multi task or sequence to engagement models, and investigate issues of interpretability and fairness in engagement scoring. By addressing these gaps, customer engagement scoring with machine learning can evolve from isolated predictive tools into a coherent, cross channel analytics capability that better supports relationship management and long-term customer value creation.

5. Conclusion

This review has examined how machine learning is currently used to construct customer engagement scores based on clickstream, social, and CRM data. Overall, the literature shows clear progress in single channel modelling, where deep and tree-based algorithms deliver strong predictive performance for specific tasks such as social media reactions, clickstream-based conversion, or CRM driven churn and retention. However, engagement is most often treated as a behavioral, task specific outcome rather than as a coherent, cross channel construct linked to broader relationship quality and value. As a result, there is still no widely accepted definition or operational template for what a customer engagement score should represent across platforms and touchpoints.

At the same time, the review highlights substantial fragmentation and several limitations that reduce the cumulative validity of existing work. Most studies rely on data from a single source or platform, often within one industry and geography, which restricts generalizability and makes it difficult to compare engagement scores across contexts. Many articles provide limited detail about feature engineering choices, data preprocessing, and model calibration, which hampers replication and critical assessment of their findings. Studies rarely consider issues such as sampling bias, the stability of engagement scores over time, or potential fairness concerns when engagement-based models are used to allocate marketing resources. These shortcomings raise important questions about how far current results can be transferred to different markets, data environments, or strategic objectives.

Building on these insights, future research should focus on more explicit integration of clickstream, social, and CRM data within unified engagement scoring architectures, supported by clear conceptual links to established engagement frameworks. Methodologically, there is room for richer representation learning across channels, systematic sensitivity analyses, and more transparent reporting of modelling choices and evaluation procedures. Substantively, researchers could investigate how engagement scores relate to long term outcomes such as loyalty, advocacy, and customer lifetime value, and explore their role in dynamic decision systems such as real time personalization or omnichannel journey orchestration. By addressing these gaps and limitations, subsequent projects can move the field from isolated predictive exercises toward a more robust, theoretically grounded, and practically useful approach to customer engagement scoring with machine learning.

References

Aluri, A., Price, B. S., & McIntyre, N. H. (2019). Using machine learning to cocreate value through dynamic customer engagement in a brand loyalty program. *Journal of Hospitality & Tourism Research, 43*(1), 78-100.

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization and customization of services. *Applied Computing and Informatics, 15*(2), 94-101.

Dai, Y., & Wang, T. (2021). Prediction of customer engagement behaviour response to marketing posts based on machine learning. *Connection Science, 33*(4), 891-910.

Dissanayake, D. M. R., Siriwardana, A., & Ismail, N. (2019). Social media marketing and customer engagement: A review on concepts and empirical contributions. *Kelaniya Journal of Management, 8*(1), 71-85.

Eigenraam, A. W., Eelen, J., & Verlegh, P. W. (2021). Let me entertain you? The importance of authenticity in online customer engagement. *Journal of Interactive Marketing, 54*(1), 53-68.

Jung, S. H., & Jeong, Y. J. (2020). Twitter data analytical methodology development for prediction of start-up firms' social media marketing level. *Technology in Society, 63*, 101409.

Kitchens, B., Dobolyi, D. G., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship oriented big data. *Journal of Management Information Systems, 35*(2), 540-574.

Koehn, D., Lessmann, S., & Schaal, M. (2020). Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications, 150*, 113342.

Kumar, A., Salo, J., & Li, H. (2019). Stages of user engagement on social commerce platforms: Analysis with the navigational clickstream data. International *Journal of Electronic Commerce, 23*(2), 179-211.

Kunz, W., Aksoy, L., Bart, Y., Heinonen, K., Kabadayi, S., Villarroel Ordenes, F., Sigala, M., Diaz, D., & Theodoulidis, B. (2017). Customer engagement in a big data world. *Journal of Services Marketing, 31*(2), 161-171.

Ng, S. C., Sweeney, J. C., & Plewa, C. (2020). Customer engagement: A systematic review and future research priorities. *Australasian Marketing Journal, 28*(4), 235-252.

Ngai, E. W. T., & Wu, Y. (2022). Machine learning in marketing: A literature review, conceptual framework, and research agenda. *Journal of Business Research, 145*, 35-48.

Santini, F. D. O., Ladeira, W. J., Pinto, D. C., Herter, M. M., Sampaio, C. H., & Babin, B. J. (2020). Customer engagement in social media: A framework and meta-analysis. *Journal of the Academy of Marketing Science, 48*(6), 1211-1228.

Schaeffer, S. E., & Sanchez, S. V. R. (2020). Forecasting client retention: A machine learning approach. *Journal of Retailing and Consumer Services, 52*, 101918.

Trunfio, M., & Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing, 2021*(3), 267-292.

Vinerean, S., & Opreana, A. (2021). Measuring customer engagement in social media marketing: A higher order model. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(7), 2633-2654.

Zerbino, P., Aloini, D., Dulmin, R., & Mininno, V. (2018). Big data-enabled customer relationship management: A holistic approach. *Information Processing & Management, 54*(5), 818-846.