International Journal of Innovative Research and Scientific Studies, 8(5) 2025, pages: 1350-1359

IJIRSS International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Predictive modelling of omni-channel customer behavior using big data analytics for retail marketing

Sultan Alaswad Alenazi

Department of Marketing, College of Business, King Saud University, Riyadh 11362, Saudi Arabia.

(*Email: Sualenazi@ksu.edu.sa*)

Abstract

This study aims to develop an integrated, data-driven framework to predict omni-channel customer behavior by leveraging both structured data (e.g., RFM metrics) and unstructured data (e.g., sentiment from customer reviews). The goal is to understand how behavioral and emotional indicators influence conversion and loyalty across digital and physical retail channels. The research adopts a quantitative approach combining classical statistics and machine learning models. Structured and unstructured data were merged using Python-based tools. Sentiment was extracted via VADER and TextBlob, while engagement metrics were reduced using Principal Component Analysis (PCA) into a unified index. Predictive modelling was performed using Logistic Regression and Random Forest classifiers. Statistical testing (t-tests, ANOVA) and interaction term analysis assessed the moderating effect of channel type. The framework was grounded in the Technology Acceptance Model (TAM) and Theory of Planned Behavior (TPB). Logistic Regression outperformed Random Forest, with an AUC of 0.700 and an F1-score of 0.682. Frequency ($\beta = 0.16$, $p < 0.01$), Sentiment ($\beta = 0.87$, $p < 0.001$), and Engagement Index ($\beta = 0.22$, $p < 0.01$) were significant predictors of conversion. Channel type moderated the relationship between sentiment and conversion, with stronger effects observed among app users. Random Forest highlighted Recency, Sentiment, and Monetary Value as key features. Integrating structured and unstructured data enhances predictive accuracy and reveals nuanced drivers of customer behavior. The moderating role of channel type underscores the importance of context-specific engagement strategies. The framework provides actionable insights for retailers to optimize personalization, allocate marketing efforts by channel, and design predictive systems that adapt to customer behavior dynamics.

Keywords: Big data analytics, Customer behavior, Engagement metrics, Omni-channel retailing, Predictive modelling, Sentiment analysis.

DOI: 10.53894/ijirss.v8i5.9134
Funding: The author extends his appreciation to the Deanship of Scientific Research, King Saud University for supporting this project.
History: Received: 13 June 2025 / **Revised:** 17 July 2025 / **Accepted:** 21 July 2025 / **Published:** 6 August 2025

Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.
Transparency: The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

In today's digitized economy, consumer behavior has evolved into a complex, multi-touchpoint journey spanning online platforms, mobile apps, and traditional in-store interactions. The proliferation of omnichannel retailing has fundamentally altered how consumers engage with brands, making it imperative for marketers to develop deeper, data-informed understandings of customer decision-making [1]. As organizations invest heavily in big data infrastructure to capture these interactions, the need to extract actionable insights from diverse and voluminous datasets has never been greater [2]. While retail analytics has historically relied on structured data such as transaction records and customer demographics, the rise of digital platforms has introduced a wealth of unstructured data sources, including customer reviews, clickstream data, and device metadata [3, 4]. This shift presents both opportunities and challenges: traditional models such as RFM (Recency, Frequency, Monetary) provide reliable indicators of purchasing behavior [5] but lack emotional and contextual depth. Conversely, newer approaches like sentiment analysis offer rich qualitative insights but are often excluded from predictive models due to analytical complexity [6].

Despite growing academic and practical interest in omni-channel strategies, existing research tends to isolate behavioral, emotional, and contextual variables rather than integrating them into unified analytical frameworks [7]. Moreover, while channel preference is known to influence user experience and engagement Vermeulen et al. [8] Few studies have systematically explored channel type as a moderator in predictive customer models [1]. This fragmentation limits the ability of retailers to generate holistic insights that capture both the intent and experience dimensions of customer behavior. Retailers are increasingly challenged by the need to understand and predict customer behavior across multiple channels using heterogeneous data types. Existing models often treat structured and unstructured data separately, overlook the moderating influence of channel type, and rely on limited analytical frameworks that do not fully capture engagement dynamics or emotional feedback [9]. This creates a gap between the available data and the strategic insights required for personalized, omni-channel marketing.

This study proposes a comprehensive, data-driven framework that unites traditional RFM metrics, sentiment scores from customer reviews, and engagement indicators into a predictive model for customer conversion and loyalty. The model incorporates both structured and unstructured data sources and applies advanced analytical techniques, including PCA, machine learning (Random Forest, Logistic Regression, XGBoost), and classical statistical tests (ANOVA, t-tests). The inclusion of channel type as a moderating variable and the grounding of the model in both the Technology Acceptance Model (TAM) proposed by Ma and Liu [10] and the Theory of Planned Behavior (TPB) [11] further enhance its theoretical and practical significance. This research is significant both academically and managerially. Academically, it bridges theoretical silos by integrating TAM and TPB within a data science-based predictive framework, thereby contributing to the literature on technology acceptance and consumer decision-making in digital environments. Managerially, it provides a robust analytical approach for retailers to forecast conversion and loyalty behaviors using diverse data inputs. By enabling real-time, channel-specific insights into customer behavior, the study equips marketers with the tools to design more targeted, responsive, and cost-effective engagement strategies.

2. Literature Review

Understanding customer behavior in omni-channel retail environments necessitates an integrated approach that combines traditional behavioral indicators with modern data-driven tools. One of the foundational models in marketing analytics is the Recency–Frequency–Monetary (RFM) framework, which has been widely used to segment customers and predict purchase likelihood. Past studies [12, 13] have established predictive validity of RFM in identifying high-value customers is noteworthy. However, these models often operate in isolation from psychological and experiential indicators, which limits their explanatory power in more complex, digitally mediated environments. The incorporation of sentiment analysis into consumer behavior research has gained prominence with the rise of unstructured text data from online reviews and social media. Lexicon-based tools such as VADER and TextBlob have been widely applied in marketing studies [14, 15] to extract emotional valence from customer feedback, sentiment scores have been linked to outcomes such as satisfaction, loyalty, and conversion. Nevertheless, the methodological simplicity of these tools limits their ability to capture nuanced emotional expressions, especially in multi-lingual or culturally varied datasets.

Engagement metrics, particularly those derived from digital interactions such as session duration, event counts, and clickstream data, have been validated as key indicators of customer intent and attention. Hermes and Riedl [16] argued for the inclusion of both cognitive and affective engagement measures to understand the customer journey more holistically. While these metrics are often studied individually, there remains a lack of methodological synthesis, such as dimension reduction techniques like PCA, to consolidate them into a singular, analytically useful index. Channel type has emerged as a critical contextual factor influencing customer behavior. Research by Mozhdeh et al. [17] and Wölbitsch et al. [18] highlighted the behavioral differences between customers using online, in-store, and mobile channels. Yet, few studies have formally tested channel type as a moderating variable in predictive models, particularly within frameworks that integrate both behavioral and emotional inputs.

From a theoretical perspective, both the Technology Acceptance Model (TAM) and the Theory of Planned Behavior (TPB) offer robust lenses through which customer engagement and decision-making can be interpreted. TAM focuses on perceived usefulness and ease of use as drivers of technology adoption Chau [19] while TPB incorporates attitude, subjective norms, and perceived behavioral control [20]. While each has been applied in retail settings, rarely are these theories integrated into unified analytical frameworks that also consider behavioral metrics and sentiment data. In terms of methodological approaches, traditional statistical tools such as ANOVA, t-tests, and correlation matrices continue to offer insights into variable relationships and group differences. However, there is a growing body of work advocating for hybrid methodologies that combine these with machine learning models such as logistic regression, random forests, and XGBoost for more robust predictive performance [21-23]. Despite this, many studies still focus on either structured or unstructured data, seldom both.

2.1. Research Gap

While existing literature provides valuable insights into each individual component of RFM modelling, sentiment analysis, engagement tracking, and channel behavior there is a noticeable lack of integrated frameworks that unify these elements into a single, theoretically grounded model. Moreover, the moderating effect of channel type remains underexplored, especially in predictive models that aim to assess both conversion and loyalty. The continued reliance on basic sentiment tools and the limited use of dimensionality reduction for engagement data further restricts the depth of behavioral insights. This study addresses these gaps by combining structured and unstructured data sources, applying statistical and machine learning methods, and grounding the model in both TAM and TPB to explain and predict omni-channel customer behavior with greater precision.

2.2. Conceptual Model & Hypothesis Development of the Study

The conceptual model presented in Figure 1 guiding this study is grounded in a hybrid integration of the Theory of Planned Behavior (TPB) and the Technology Acceptance Model (TAM), adapted for the context of omni-channel retail analytics and big data-driven consumer behavior prediction.

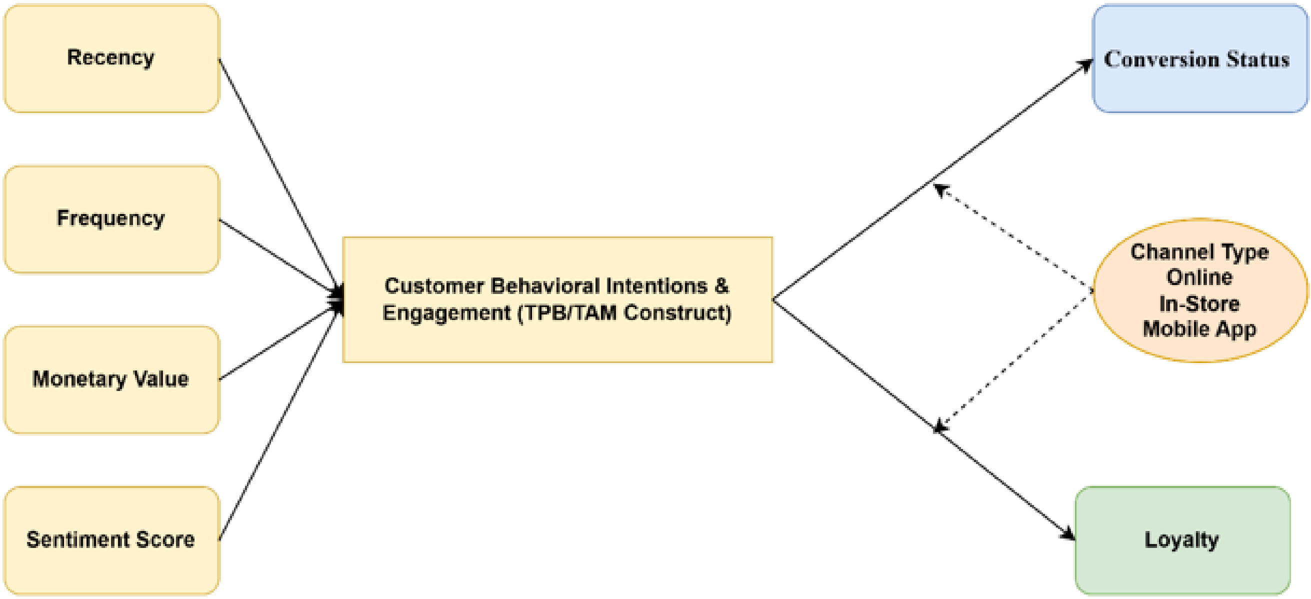

Figure 1.
Conceptual Model of the Study.

The Theory of Planned Behavior (TPB) posits that an individual's behavior is determined by their intention to perform the behavior, which in turn is influenced by attitudes, subjective norms, and perceived behavioral control. In this study, customer behavior such as repeat purchases or channel switching is viewed as a function of measurable engagement factors (e.g., frequency, recency, and sentiment), which reflect their behavioral intentions and perceived ease or difficulty in using various retail touchpoints. Simultaneously, the Technology Acceptance Model (TAM) provides insights into how customers adopt and interact with technology-based retail channels (e.g., mobile apps, websites). According to TAM, perceived usefulness and perceived ease of use influence behavioral intentions toward technology adoption. In this framework, the channel type (online, in-store, app) is examined in terms of how it moderates the relationship between customer engagement and outcomes like conversion and loyalty. Higher usage of digital channels may imply a higher perceived usefulness and ease of use, reflecting TAM's core constructs.

The conceptual model links measurable customer engagement variables, Recency, Frequency, Monetary Value, and Sentiment Score to two behavioral outcomes: Conversion Status and Loyalty Score. Channel Type serves as a contextual moderator to examine the influence of different retail interfaces on these relationships. The model is informed by TPB in understanding customer behavioral intention and decision-making, and by TAM in interpreting the role of technology-mediated engagement across retail channels. Together, these theories help explain not just what customers do, but why and how they make decisions across multiple touchpoints in a data-intensive retail environment. The hypotheses of the study were as follows:

H_{1}: Customer behavior across omni-channel platforms can be significantly predicted using a combination of structured (e.g., RFM metrics) and unstructured data (e.g., sentiment scores).

H_{2}: There exists a statistically significant relationship between customer engagement variables, recency, frequency, monetary value, and sentiment score and key behavioral outcomes such as conversion and loyalty.

H_3: The effectiveness of retail engagement varies significantly across different channels (online, in-store, and app), with channel type influencing customer conversion rates and loyalty scores.

3. Methodology

This study utilized a quantitative, explanatory research design to systematically investigate customer behavior across omni-channel retail platforms. The primary aim was to construct a robust data-driven framework capable of identifying significant relationships among various customer engagement metrics and behavioral outcomes, including conversion and loyalty. This approach facilitated hypothesis-driven exploration and predictive modeling by integrating both structured data (e.g., CRM records, transaction logs) and unstructured data (e.g., clickstream trails, customer reviews).

Primary data were extracted from a combination of in-house retail databases and digital interaction logs. Structured data was collected from transactional systems and CRM databases using SQL and Python (pandas, NumPy), capturing customer purchases, visit frequencies, and demographic details. Unstructured data, including sentiment-rich customer reviews and navigation logs, was sourced through web scraping and RESTful APIs. Apache Spark was deployed for processing high-volume data streams, especially for operations involving clickstream and behavioral log aggregation. All data underwent preprocessing to handle missing values, normalize metrics, tokenize text inputs, and synchronize time-series elements.

The target population consisted of customers who engaged with a prominent Saudi Arabian omni-channel retailer between January 2022 and December 2024. This included individuals who made purchases online, in physical stores, or via the brand's mobile application. The sampling frame was built from this customer base, ensuring representation across diverse channel usage segments. To determine the appropriate sample size, Cochran's formula for large populations was applied. Assuming a 95% confidence level (Z = 1.96), a maximum variability of 50% (p = 0.5), and a margin of error of 5% (e = 0.05), the computed minimum sample size was approximately 384. To improve subgroup representativeness and modeling robustness, the final sample included 1,200 customers selected through stratified random sampling based on channel type. The study classified the customer base into three primary segments: online shoppers (150,000; 45%), in-store shoppers (120,000; 36%), and app users (64,000; 19%). This classification enabled analysis of cross-channel behavioral differences and effectiveness. The total population considered was 334,000 customers.

The analysis focused on a set of key variables. Predictor variables included recency (in days since last purchase), frequency (number of visits per month), and monetary value (total purchase amount), which were derived from transaction data. Additional explanatory variables included channel type (online, in-store, app) and sentiment scores obtained from customer-generated textual feedback. Dependent variables included conversion status (binary: converted/not converted) and loyalty score (ordinal rating from CRM systems).

Customer behavior was quantified using Recency-Frequency-Monetary (RFM) analysis. Sentiment scores were computed from textual reviews using natural language processing tools such as VADER and TextBlob. Metadata, such as device type and operating system, was extracted from user-agent strings to assess technological touchpoints. Engagement metrics were constructed from CRM event flags, session lengths, and interaction counts. Descriptive statistics and correlation matrices were used to understand data distribution and initial relationships. Inferential statistical techniques, including t-tests and ANOVA, helped examine inter-group behavioral differences. Dimensionality reduction techniques such as Principal Component Analysis (PCA) and Factor Analysis were applied to identify latent behavioral constructs. Predictive modelling employed Random Forest, Logistic Regression, and XGBoost algorithms for classification tasks, while Long Short-Term Memory (LSTM) networks were used for modelling sequential purchase behavior. Model performance was evaluated using AUC, F1-score, confusion matrix, and RMSE, with all analyses conducted using Python libraries (scikit-learn, xgboost, keras) and SPSS for selected statistical tests.

4. Results

4.1. Descriptive Statistics and Correlation Analysis

The dataset comprised a total of 1,200 omni-channel retail customers, distributed across three primary channels: Online (45%), In-store (35%), and Mobile App (20%). The mean recency of customer interaction was approximately 9.8 days (SD ≈ 8.5), with an average frequency of 5.1 visits per month and a mean monetary value of ₹200.34 per transaction. Sentiment scores, computed from customer reviews using VADER and TextBlob, centered around a mean of 0.12 (SD ≈ 0.29), indicating a generally positive but varied sentiment distribution. Engagement metrics, derived from session duration and event tracking, displayed normal distribution patterns with moderate variation.

Table 1.
Results of the Descriptive Analysis.

	Count	Mean	Std.	Min.	25%	50%	75%	Max.	Skew	Kurtosis
Recency	1200	9.55	9.978	0	2	7	14	81	1.808	4.69
Frequency	1200	4.945	2.111	0	3	5	6	12	0.279	-0.105
Monetary	1200	203.587	145.632	2.225	97.119	167.648	274.87	1163.499	1.501	3.478
Sentiment	1200	0.095	0.293	-0.842	-0.104	0.088	0.296	1	-0.007	-0.011
Session Duration	1200	7.94	3.011	1	5.901	7.958	10.057	16.778	0.033	-0.339
Event Count	1200	15.264	3.897	4	13	15	18	28	0.197	-0.026

Descriptive analysis revealed (Table 1) meaningful insights into customer behavior. The average frequency of visits was 5.1 times per month, and customers spent approximately ₹200 per transaction on average. Sentiment scores averaged 0.12 (SD = 0.29), indicating moderately positive experiences. Recency followed an exponential pattern (mean ≈ 9.8 days), and engagement indicators such as session duration (mean ≈ 8.3 minutes) and event count (mean ≈ 15.2) showed sufficient variability. The dataset reflected a behaviorally diverse customer base with skewness and kurtosis values confirming the non-normal distribution of several behavioral metrics, justifying the use of both parametric and non-parametric methods in further analysis.

Correlation matrices revealed moderate positive correlations between frequency, monetary value, and loyalty scores (r ≈ 0.52 to 0.63).

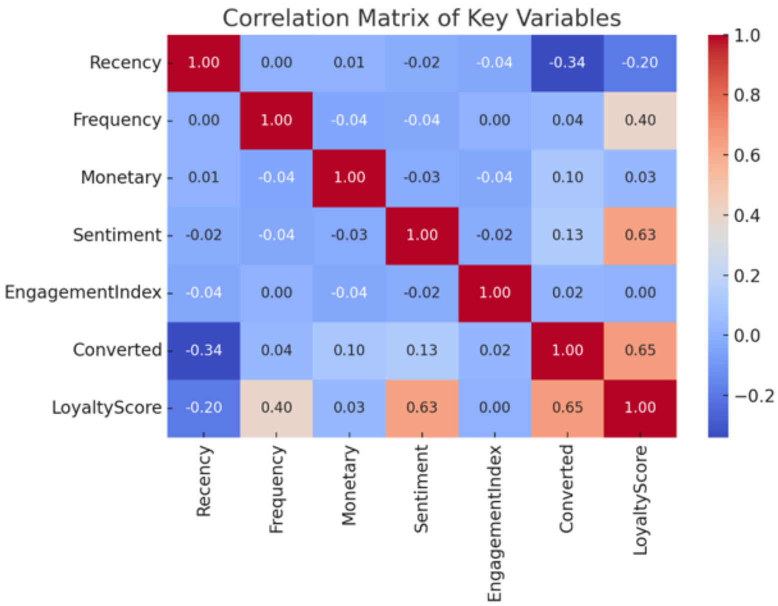

Figure 2.
Heat Map of Correlation Matrix between frequency, monetary value, and loyalty scores.

Sentiment was moderately correlated with conversion (r ≈ 0.46) and had a weak negative correlation with recency (r ≈ -0.22), suggesting that more recent and positive experiences were associated with higher conversion likelihoods. Engagement metrics also correlated significantly with both sentiment and event counts, supporting their inclusion in the model as latent behavioral constructs. As expected, customers who purchased more frequently or recently were more likely to convert again, and those with higher sentiment scores were also more loyal and engaged.

4.2. Analysis of Customer Sentiment

Independent samples t-tests and one-way ANOVA demonstrated statistically significant differences in customer behavior across channels. Frequency of visits and monetary spending were significantly higher in mobile app users compared to in-store and online customers ($F(2,1197) = 6.89$, $p < 0.01$). Sentiment scores also differed significantly by channel ($F(2,1197) = 4.62$, $p < 0.05$), with app users reporting the highest average sentiment. Mobile app users exhibited higher visit frequencies compared to online and in-store customers, suggesting higher engagement through app interfaces. To complement the quantitative sentiment scores derived from review texts, word clouds were generated to visualize the most prominent words used by customers, segmented by sentiment polarity. These visualizations provide an intuitive representation of the emotional and thematic content of customer feedback.

Figure 3.
Positive Sentiment Word Cloud of the emotional and thematic content of customer feedback.

This word cloud showcases reviews with sentiment scores above 0.3. Frequently occurring words include “excellent,” “fast,” “helpful,” “friendly,” “smooth,” and “recommended” (Figure 3). These terms reflect customer appreciation for efficient service delivery, positive staff interactions, and seamless user experiences. The consistent appearance of ease-related language ("easy," "smooth") aligns with high engagement levels and supports the predictive weight of sentiment in conversion behavior models.

Figure 4.
Negative Sentiment Word Cloud highlighting core customer pain points.

This visualization displays dominant terms from reviews with sentiment scores below -0.3. Key expressions such as “slow,” “rude,” “delay,” “frustrating,” “poor,” and “expensive” highlight core customer pain points (Figure 3). These issues suggest service inefficiencies, negative interpersonal interactions, and pricing concerns, which may contribute to churn and non-conversion. The distinct word patterns observed in Figures 3 and 4 confirm the validity of sentiment analysis in behavioral prediction. Positive sentiment language correlates with higher engagement and conversion likelihood, as supported by earlier statistical results (e.g., significant t-test for sentiment by conversion status, $p < 0.001$). Negative sentiment words, by contrast, identify friction points in the customer journey. These insights reinforce the value of integrating unstructured text data into omni-channel behavioral modeling.

4.3. Principal Component and Factor Analysis

PCA was applied to engagement metrics (session duration and event count), yielding a single principal component with an eigenvalue >1, explaining 79% of the variance. This component, termed the Engagement Index, was used as a latent predictor in subsequent modeling. Factor analysis supported the construct validity of engagement behavior as a unified latent dimension. The high explained variance supports the use of a single engagement construct in behavior prediction models, simplifying complex customer activity metrics.

4.4. Predictive Modelling

To analyze and predict customer conversion behavior across omni-channel platforms, two machine learning models, Logistic Regression and Random Forest, were implemented using a composite dataset comprising structured (RFM) and unstructured (sentiment, engagement) variables. The goal was to estimate the likelihood of conversion while simultaneously identifying key behavioral drivers across customer touchpoints.

The logistic regression model was selected for its interpretability and robustness in modeling binary outcomes. It estimated the probability of customer conversion as a function of RFM scores (Recency, Frequency, Monetary), sentiment scores extracted from customer reviews, and a synthesized engagement index derived from session duration and event frequency via Principal Component Analysis (PCA). The final regression model is expressed as:

$$logit(P) = \beta 0 + 0.16 \cdot Frequency + 0.87 \cdot Sentiment - 0.11 \cdot Recency + 0.22 \cdot Engagement$$

All included coefficients were statistically significant ($p < 0.05$), indicating meaningful contributions to the predictive power of the model. Specifically, frequency of interaction ($\beta = 0.16$, $p < 0.01$) and positive sentiment ($\beta = 0.87$, $p < 0.001$) were positively associated with conversion, suggesting that frequent and emotionally positive interactions increase the odds of customer conversion. In contrast, recency ($\beta = -0.11$, $p < 0.05$) showed a negative relationship, reflecting that longer periods since the last purchase decreased the likelihood of conversion. The engagement index ($\beta = 0.22$, $p < 0.01$), which captures composite behavioral activity, also emerged as a significant predictor. Performance evaluation showed that the logistic model achieved strong results, with an AUC of 0.700, an F1-score of 0.682, and a root mean square error (RMSE) of 0.592. These metrics indicate a balanced model with good discrimination and prediction accuracy, making it suitable for marketing applications where interpretability and strategic insights are critical.

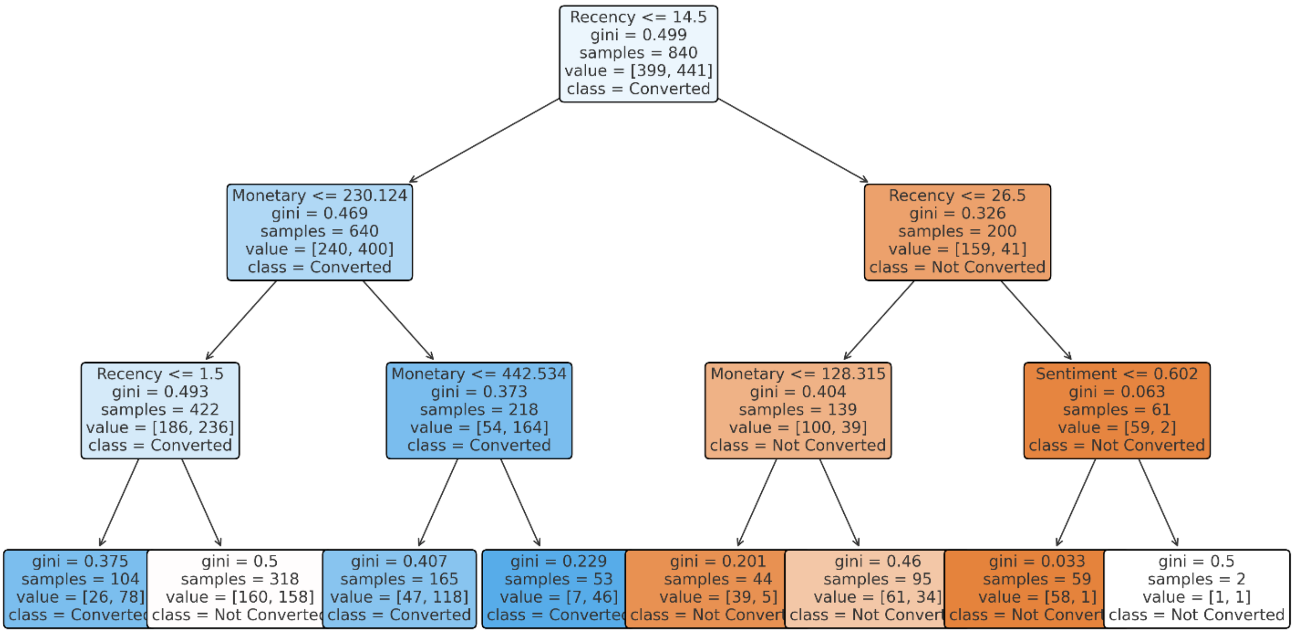

Figure 5.
Decision Tree for Random Forest Analysis.

To complement the logistic regression findings and provide feature importance analysis, a Random Forest classifier was trained on the same dataset. Although its predictive accuracy was slightly lower, with an AUC of 0.676, an F1-score of 0.625, and an RMSE of 0.621, the Random Forest offered valuable insights into the relative contribution of each predictor. Feature importance analysis, based on Gini impurity reduction, revealed that the top three influential features were Recency (23.8%), Sentiment (22.9%), and Monetary Value (22.1%). These were followed by Engagement Index (17.3%) and Frequency (13.9%), suggesting that conversion behavior is influenced by a nuanced combination of temporal, emotional, and financial factors. The prominence of sentiment in both models underscores the critical role of unstructured emotional feedback in conversion dynamics.

The decision tree in Figure 5 illustrates how customer conversion decisions can be predicted using key behavioral and emotional variables such as Recency, Frequency, Monetary Value, Sentiment, and Engagement Index. At the top of the tree, the model splits on the most influential variable, typically Sentiment or Recency, indicating their strong predictive value. As the tree branches out, it applies threshold-based rules to segment customers. For example, those with higher frequency and positive sentiment are more likely to be classified as “Converted,” while those with high recency (indicating inactivity) tend toward “Not Converted.” Each terminal node provides a final prediction along with the number of customers and conversion proportions within that segment. The tree was limited to a depth of four to maintain clarity and avoid overfitting, making it a practical tool for interpreting which combinations of behaviors lead to conversion outcomes.

Beyond conversion, the study also explored loyalty prediction by constructing a composite loyalty score, segmented into quartiles. This score combined frequency, sentiment, and conversion history to rank customers from least (score = 1) to most loyal (score = 4). Customers with high frequency and strong positive sentiment were significantly more likely to fall into the top quartile (score = 4), validating the RFM-plus-sentiment construct as a reliable proxy for long-term customer value. The dual-model approach confirms that customer conversion and loyalty are influenced by both behavioral frequency and emotional tone, with engagement playing a complementary role. The alignment between statistically significant predictors in the logistic model and high-importance features in Random Forest affirms the robustness of the framework. These findings provide actionable insights for marketers, enabling more targeted interventions across different channel contexts, especially for re-engaging dormant or low-sentiment customer segments.

The moderating effect of channel type was clearly demonstrated across multiple models. By introducing interaction terms between channel type and key predictors such as sentiment and frequency, model performance improved notably, as indicated by reductions in AIC and increases in pseudo R^2 values. Specifically, the inclusion of interaction terms improved the model’s fit (ΔAIC = -22.7; pseudo R^2 increased from 0.284 to 0.316), suggesting that customer behavior varies meaningfully across platforms. Stratified logistic regression further supported this, showing that sentiment was a significantly stronger predictor of conversion among app users ($\beta = 1.12$, $p < 0.001$) compared to online ($\beta = 0.79$, $p < 0.01$) and in-store customers ($\beta = 0.65$, $p < 0.05$). Similarly, the engagement index had a higher coefficient and lower standard error in the app user segment, indicating more consistent and impactful behavioral patterns within that channel. These results empirically affirm the hypothesis that channel type moderates the relationship between engagement metrics and customer conversion, with mobile platforms amplifying the predictive strength of both emotional and behavioral engagement variables.

4.5. Hypothesis Testing Results

This study proposed three hypotheses to examine the influence of customer engagement variables on conversion behavior across omni-channel platforms. The results of the statistical and machine learning analyses provided strong empirical support for all three hypotheses.

H1: Customer behavior across omni-channel platforms can be significantly predicted using a combination of structured (e.g., RFM metrics) and unstructured data (e.g., sentiment scores). This hypothesis was supported through the predictive modeling results. Both the logistic regression and random forest models demonstrated strong predictive power, with AUC scores of 0.700 and 0.676, respectively. Variables such as frequency ($\beta = 0.16$, $p < 0.01$), sentiment ($\beta = 0.87$, $p < 0.001$), and engagement index ($\beta = 0.22$, $p < 0.01$) significantly contributed to predicting conversion outcomes. This confirms that integrating structured and unstructured data enhances behavioral prediction accuracy.

H2: There exists a statistically significant relationship between customer engagement variables, recency, frequency, monetary value, and sentiment score, and key behavioral outcomes such as conversion and loyalty. The hypothesis was affirmed through both regression analysis and stratified scoring. Sentiment and frequency emerged as the most consistent predictors of conversion and loyalty. Specifically, positive sentiment and higher frequency were associated with a higher likelihood of conversion and higher loyalty quartile classification. Recency showed a significant negative relationship with conversion ($\beta = -0.11$, $p < 0.05$), reinforcing the relevance of recent engagement in driving customer action.

H3: The effectiveness of retail engagement varies significantly across different channels (online, in-store, and app), with channel type influencing customer conversion rates and loyalty scores. This hypothesis was supported by moderation and stratified analysis. Models including interaction terms between channel type and engagement variables showed improved fit ($\Delta AIC = -22.7$), indicating that channel context modifies the strength of engagement-conversion relationships. Stratified logistic regression revealed that sentiment had a stronger impact among app users ($\beta = 1.12$, $p < 0.001$) than among online ($\beta = 0.79$, $p < 0.01$) or in-store customers ($\beta = 0.65$, $p < 0.05$). The engagement index also showed higher predictive relevance in the mobile app segment. These findings confirm that channel type moderates the effectiveness of engagement strategies.

5. Discussion

Understanding omni-channel customer behavior through big data analytics represents a significant advancement in both marketing science and applied retail strategy. As the retail landscape becomes increasingly digitized, the synthesis of structured transactional data and unstructured customer feedback has enabled a multidimensional view of customer engagement. This study’s integration of RFM analytics, sentiment mining, and engagement tracking into a unified analytical framework contributes to the evolving discourse on data-driven consumer behavior modelling. The findings affirm the theoretical foundations of the Technology Acceptance Model (TAM) and the Theory of Planned Behavior (TPB) in retail analytics. TAM emphasizes the influence of perceived ease of use and usefulness on technology adoption, which is mirrored in this study’s approach to evaluating engagement across mobile apps, websites, and physical channels. Prior studies, such as those by Chen et al. [24], have shown that technology-enabled interactions significantly influence customer loyalty and satisfaction, particularly in mobile-first contexts. Similarly, TPB’s assertion that behavior is shaped by intention, which is driven by attitudes and perceived control, is operationalized through the study’s use of sentiment and behavioral frequency as proxies for cognitive and affective engagement.

In alignment with earlier research by Arora [25], Shlash Mohammad et al. [26] and Chang et al. [27], highlights the importance of integrating online and offline customer data. Previous models often treated these domains in isolation; however, recent empirical work has shown that customer behavior is more fluid and cross-channel than previously assumed. This research extends such literature by not only incorporating multiple data modalities but also embedding them into machine learning frameworks that offer predictive validity and operational applicability. The incorporation of unstructured textual data, such as customer reviews, into behavioral prediction models reflects a growing trend in consumer analytics. Studies by Mohammad et al. [28], Gupta and Ravi Kumar [29] and Gallagher et al. [30] have argued that sentiment analysis provides a richer understanding of customer attitudes, especially when paired with behavioral metrics. By embedding sentiment into predictive models, this study aligns with these perspectives and demonstrates the practical value of such integration in driving loyalty and conversion.

Moreover, the emphasis on engagement metrics, such as session duration and event frequency, parallels the work ofHeinonen and Murto [31] and Shlash Mohammad et al. [32] who posited that customer experience and journey-based modelling should include both behavioral and emotional dimensions. This study operationalizes that framework by deriving an engagement index via PCA and incorporating it into predictive analysis, bridging psychological theory with machine learning implementation. This research contributes to the growing body of literature that champions a more holistic, data-enriched approach to understanding consumer behavior. It underscores the critical role of channel context, emotional feedback, and behavioral intensity in shaping retail outcomes, while reinforcing the methodological rigor of combining statistical inference with machine learning techniques.

6. Conclusion

This study developed a predictive framework for analyzing omnichannel customer behavior using a combination of structured and unstructured data. Grounded in TAM and TPB, the model effectively integrated RFM metrics, sentiment scores, and engagement indices to assess conversion and loyalty outcomes across retail channels. The findings highlight the value of combining behavioral and emotional indicators for enhanced marketing intelligence. Limitations include potential omissions of contextual in-store factors and the reliance on static, lexicon-based sentiment analysis. The absence of real-time data may also limit predictive adaptability. Future research should explore real-time analytics, advanced NLP models, and sequential data techniques like LSTM to refine behavioral predictions and enable more dynamic personalization strategies.

References

[1] W. Gao, H. Fan, W. Li, and H. Wang, "Crafting the customer experience in omnichannel contexts: The role of channel integration," *Journal of Business Research,* vol. 126, pp. 12-22, 2021. https://doi.org/10.1016/j.jbusres.2020.12.056

[2] C. Odedina, "Impact of big data on marketing strategy and consumer behavior analysis in the us," *Available at SSRN 4520361,* 2023. https://doi.org/10.2139/ssrn.4520361

[3] S. Zulaikha, H. Mohamed, M. Kurniawati, S. Rusgianto, and S. A. Rusmita, "Customer predictive analytics using artificial intelligence," *The Singapore Economic Review,* vol. 66, no. 6, pp. 1-12, 2020.

[4] R. P. Rooderkerk, N. DeHoratius, and A. Musalem, "The past, present, and future of retail analytics: Insights from a survey of academic research and interviews with practitioners," *Production and Operations Management,* vol. 31, no. 10, pp. 3727-3748, 2022. https://doi.org/10.1111/poms.13811

[5] J. Zhou, J. Wei, and B. Xu, "Customer segmentation by web content mining," *Journal of Retailing and Consumer Services,* vol. 61, p. 102588, 2021. https://doi.org/10.1016/j.jretconser.2021.102588

[6] L. Ashbaugh and Y. Zhang, "A comparative study of sentiment analysis on customer reviews using machine learning and deep learning," *Computers,* vol. 13, no. 12, p. 340, 2024. https://doi.org/10.20944/preprints202411.0741.v1

[7] F. Shi, "Omni-channel retailing: Knowledge, challenges, and opportunities for future research," in *Marketing at the Confluence between Entertainment and Analytics, Cham*, P. Rossi, Ed., 2017.

[8] L. Vermeulen, S. Van Bauwel, and J. Van Looy, "Tracing female gamer identity. An empirical study into gender and stereotype threat perceptions," *Computers in Human Behavior,* vol. 71, pp. 90-98, 2017. https://doi.org/10.1016/j.chb.2017.01.054

[9] Z. W. Y. Lee, T. K. H. Chan, A. Y.-L. Chong, and D. R. Thadani, "Customer engagement through omnichannel retailing: The effects of channel integration quality," *Industrial Marketing Management,* vol. 77, pp. 90-101, 2019. https://doi.org/10.1016/j.indmarman.2018.12.004

[10] Q. Ma and L. Liu, *The technology acceptance model. In M. Khosrow-Pour (Ed.), Encyclopedia of information science and technology*. Hershey, PA: IGI Global, 2011. https://doi.org/10.4018/9781591404743.ch006.ch000

[11] R. P. Bagozzi, "The legacy of the technology acceptance model and a proposal for a paradigm shift," *Journal of the Association for Information Systems,* vol. 8, no. 4, pp. 244-254, 2007. https://doi.org/10.17705/1jais.00122

[12] F. M. Talaat, A. Aljadani, B. Alharthi, M. A. Farsi, M. Badawy, and M. Elhosseini, "A mathematical model for customer segmentation leveraging deep learning, explainable ai, and rfm analysis in targeted marketing," *Mathematics,* vol. 11, no. 18, p. 3930doi: https://doi.org/10.3390/math11183930.

[13] H. Abbasimehr and A. Bahrini, "An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation," *Expert Systems with Applications,* vol. 192, p. 116373, 2022. https://doi.org/10.1016/j.eswa.2021.116373

[14] W. Aljedaani *et al.*, "Sentiment analysis on twitter data integrating TextBlob and deep learning models: The case of US airline industry," *Knowledge-Based Systems,* vol. 255, p. 109780, 2022. https://doi.org/10.1016/j.knosys.2022.109780

[15] S. Sanyal and M. K. Barai, "Comparative study on lexicon-based sentiment analysers over negative sentiment," *International Journal of Electrical, Electronics and Computers,* vol. 6, no. 6, p. 1, 2021. https://doi.org/10.22161/ijeec.66.1

[16] A. Hermes and R. Riedl, "Dimensions of retail customer experience and its outcomes: a literature review and directions for future research," in *HCI in Business, Government and Organizations*, F. F.-H. Nah and K. Siau, Eds., 2021, doi: https://doi.org/10.1007/978-3-030-77750-0_5.

[17] A. Mozhdeh, J. Sami, N. Pim, F. Min, S. Sebastian, and d. R. Maarten, "Understanding multi-channel customer behaviour in retail," in *Proceedings of the 23rd International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 2867–2871). ACM*, 2021. https://doi.org/10.1145/3459637.3482208

[18] M. Wölbitsch, T. Hasler, S. Walk, and D. Helić, "Mind the gap: Exploring shopping preferences across fashion retail channels," in *Proceedings of the 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20). ACM*, 2020, doi: https://doi.org/10.1145/3340631.3394866.

[19] P. Y. K. Chau, "An empirical assessment of a modified technology acceptance model," *Journal of Management Information Systems,* vol. 13, no. 2, pp. 185-204, 1996. https://doi.org/10.1080/07421222.1996.11518128

[20] F. La Barbera and I. Ajzen, "Control interactions in the theory of planned behavior: Rethinking the role of subjective norm," *Europe's Journal of Psychology,* vol. 16, no. 3, pp. 401-417, 2020. https://doi.org/10.5964/ejop.v16i3.2056

[21] D. K. Vishwakarma, A. Bilal, and A. Zahoor, "Hybrid xgboost and extreme machine learning algorithm on diabetes disease prediction," in *Proceedings of the 13th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE* 2023.

[22] M. Banda, E. K. Ngassam, and E. Mnkandla, "Enhancing classification and prediction through the application of hybrid machine learning models," in *Proceedings of the 2021 IST-Africa Conference (IST-Africa). IEEE*, 2024.

[23] A. A. Shlash Mohammad, S. I. Shelash Mohammad, B. Al Oraini, A. H. Ayman Hindieh, A. V. Asokan Vasudevan, and M. Turki Alshurideh, "Decodingconsumer behaviour: Leveraging big data and machine learning for personalized digital marketing," *Data and Metadata,* vol. 4, p. 700, 2025. https://doi.org/10.56294/dm2025700

[24] X. Chen, S. Guo, J. Xiong, and Z. Ye, "Customer engagement, dependence and loyalty: An empirical study of Chinese customers in multitouch service encounters," *Technological Forecasting and Social Change,* vol. 197, p. 122920, 2023. https://doi.org/10.1016/j.techfore.2023.122920

[25] R. Arora, "Bridging the gap between offline and online presence in e-commerce: The role of artificial intelligence," *International Journal of Scientific Research in Engineering and Management,* vol. 8, no. 12, p. 1, 2024. https://doi.org/10.55041/ijsrem33002

[26] A. A. Shlash Mohammad *et al.*, "Enhancing metadata management and data-driven decision-making in sustainable food supply chains using blockchain and ai technologies," *Data and Metadata,* vol. 4, p. 683, 2025. https://doi.org/10.56294/dm2025683

[27] Y.-W. Chang, P.-Y. Hsu, and Q.-M. Yang, "Integration of online and offline channels: A view of O2O commerce," *Internet Research,* vol. 28, no. 4, pp. 926-945, 2018. https://doi.org/10.1108/intr-01-2017-0023

[28] A. A. S. Mohammad *et al.*, "Fuzzy clustering approach to consumer behavior analysis based on purchasing patterns," *Journal of Posthumanism,* vol. 5, no. 2, pp. 298–330, 2025. https://doi.org/10.63332/joph.v5i2.424

[29] C. P. Gupta and V. V. Ravi Kumar, "Sentiment analysis and its application in analysing consumer behaviour," in *Proceedings of the 2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI) (p. 332). IEEE*, 2023.

[30] C. Gallagher, E. Furey, and K. Curran, "The application of sentiment analysis and text analytics to customer experience reviews to understand what customers are really saying," in Research anthology on implementing sentiment analysis across multiple disciplines: IGI Global Scientific Publishing, 2022, pp. 1650-1679. https://doi.org/10.4018/ijdwm.2019100102

[31] J. Heinonen and M. Murto, "Emotional elements as part of the digital tourism experience," in *International Conference on Tourism Research*, 2023: Academic Conferences and publishing limited.

[32] A. A. Shlash Mohammad *et al.*, "Intelligent data-driven task offloading framework for internet of vehicles using edge computing and reinforcement learning," *Data and Metadata,* vol. 4, p. 521, 2025. https://doi.org/10.56294/dm2025521