International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Predictive Analytics in Retail Pharmacy: QR Code Implementation on Purchase Intentions Using Random Forest

Catrina Mae S. Tulinao[1], Nelson C. Bool[2]

[1,2]University of Santo Tomas, Graduate School

Abstract
In this study, we explore the efficacy of a Random Forest model in predicting customer repurchase intentions in a retail pharmacy setting that has adopted QR code technology. The model achieved a training R^2 of 0.6964 and a testing R^2 of 0.4558, indicating that while it captures a substantial proportion of the variation in the training data, its performance on independent data calls for cautious interpretation. Notably, feature importance analysis highlighted that store attributes and service quality are critical determinants of repurchase behaviour, explaining 32.91% and 29.86% of the variance, respectively. Additional factors, such as the physical store environment and waiting time, also played significant roles. These findings suggest that both tangible store features and digital interventions contribute to customer loyalty. The integration of rigorous model validation techniques, including cross-validation and hyperparameter tuning, reinforces the model's potential as a robust predictive tool. This work provides meaningful insights for retail pharmacies aiming to optimize customer engagement through a blended approach of digital innovation and traditional service excellence.

Keywords: Retail pharmacy, QR code, Random Forest, repurchase intention, feature importance, predictive modelling, customer behavior, service quality, store attributes, digital integration

1. INTRODUCTION

Background of the Study

The digital capabilities of retail pharmacy operations have undergone significant transformation to adapt to the evolving demands of today's highly competitive healthcare marketplace. As a crucial element of this transformation, Quick Response (QR) code technology has emerged as a **ground-breaking** innovation that effectively bridges the gap between traditional pharmacy practices and the virtual realm. The importance of this technological integration is particularly evident among retail pharmacies striving to enhance customer experience while simultaneously maintaining operational efficiency in an increasingly digital-first consumer landscape (Prakash, 2019).

In the context of contemporary retail pharmacy operations, understanding customer behavior and purchase intentions has become paramount. Research indicates that the quality of service and the accessibility of products are key determinants of customer satisfaction and loyalty (Farokhmanesh, 2023). Studies have demonstrated that a nuanced understanding of customer behavior is essential for the successful operation of any business, particularly in the retail sector (Alweshah and Abbar, 2022; Shen and Oliveira, 2022). Digital touchpoints, such as QR codes, introduce intricate interaction paradigms that create new opportunities for customer engagement and data collection (Pillai, 2018). Specifically, QR codes facilitate access to crucial information, thereby assisting pharmacy operations in enhancing customer engagement and elevating the quality of service provided (Bhuvaneswari, 2024). Furthermore, the application of machine learning techniques is increasingly recognized for its potential to analyze customer interactions and improve operational efficiency within retail pharmacies. Recent studies have explored the impact of QR code technology on customer satisfaction and performance metrics in the retail pharmacy sector, revealing that predictive analytics can significantly enhance customer engagement and interaction at the point of contact without compromising service quality (Vargas-Calderón, 2021). For example, Raza et al. provide a critical analysis of how predictive analytics influence key performance indicators in retail environments, highlighting the potential benefits of integrating QR codes into pharmacy processes (Christodoulou et al., 2019).

Additionally, research utilizing the Random Forest classifier has successfully examined factors influencing customer perception and usability in healthcare applications, underscoring the need for further investigation into the application of Random Forest modeling in conjunction with QR technology within retail pharmacy contexts (Testing the estimated patient waiting times of the Random Forest algorithm by SPC, 2023). This underscores a significant gap in the existing literature that warrants further exploration. Machine learning algorithms have demonstrated remarkable effectiveness across various fields, particularly in healthcare, where notable improvements in clinical procedures and resource allocation have been observed through the implementation of predictive analytics (Yesmin et al., 2023). A recent study validates the predictive performance of the Random Forest algorithm, achieving lower error rates compared to alternative methods, thereby enhancing both prediction quality and service outcomes (Vargas-Calderón, 2021). As retail pharmacies undergo a digital transformation, presenting numerous opportunities for operational enhancement, it is essential to comprehend the efficacy of various technologies and their impact on service quality and customer satisfaction. Such understanding is vital for optimizing pharmacy operations and improving overall patient care (Singh et al., 2023).

Rationale of the Study

Moreover, our research integrates the SERVQUAL model, which offers a comprehensive framework for evaluating service quality within the retail pharmacy context. SERVQUAL encompasses multiple dimensions, including reliability, assurance, responsiveness, empathy, and tangibles (Parasuraman et al., Singh et al., 2023). This model aids in elucidating the intricate relationships between service quality and customer satisfaction, particularly in light of the emergence of QR code technology in pharmacy practice (Kinderis et al., 2023). This study is grounded in recognized quality management practices and methodologies, aligning seamlessly with current trends, such as digitization within retail pharmacies (Maghsoodi et al., 2019). QR codes not only enhance the effectiveness of service delivery but also serve to bolster customer engagement as a fundamental principle of quality management (Cao et al., 2024).

Thus, through the application of predictive analytics in this research, specifically utilizing Random Forest modeling, we aim to investigate the adoption of QR codes and their influence on customer purchase intentions. While direct evidence linking Random Forest or analogous models to retail pharmacies remains limited, predictive analytics have demonstrated effectiveness across other service industries, thus indicating their potential applicability in pharmacy settings (Ho et al., 2024). The objective of this study is to explore the critical factors influencing customers' purchase intentions within the framework of retail pharmacy services in the digital era, particularly in relation to QR code implementation, and to understand how these factors mediate with the dimensions of the traditional service quality framework (San et al., 2020).

In order to assess customer behavior in relation to QR code adoption and purchase intention, this study employs the SERVQUAL model alongside foundations of Total Quality Management (TQM), which have been extensively validated in the literature over several decades. As comprehension of customer behavior in digitized environments becomes increasingly vital, this research seeks to synthesize predictive analytics—specifically Random Forest modeling—to establish a framework for examining customer loyalty and satisfaction (Soe et al., 2023). This inquiry effectively connects conventional quality management practices with contemporary digital technologies, illustrating how QR code-enabled functionalities, augmented by advanced predictive analytics capabilities, can significantly enhance service quality and improve consumer experiences within retail pharmacy settings. For retail pharmacies, understanding these complex relationships is crucial, as it provides pertinent insights for strategically allocating resources toward physical enhancements and/or digital innovations, while also offering valuable perspectives on consumer behavior patterns in increasingly digitized pharmacy environments (Kaur et al., 2019).

2. Review of Related Literature

Advancements in Retail Pharmacy Service Quality

The landscape of retail pharmacy has witnessed significant advancements in recent years, particularly concerning service quality and digital transformation. These developments are emblematic of broader trends in healthcare delivery that emphasize efficiency, responsiveness, and customer-centric approaches. In this context, the SERVQUAL model, conceived by Parasuraman, Zeithaml, and Berry, serves as a cornerstone framework for quantitatively measuring service quality. This model evaluates service across five pivotal dimensions: reliability, responsiveness, assurance, empathy, and tangibles (Prakash, 2019).

The widespread application of SERVQUAL across various sectors, including healthcare and retail, underscores its utility in identifying discrepancies between customer expectations and perceptions. By employing this structured framework, organizations can implement targeted strategies aimed at enhancing service quality and customer satisfaction (Pillai, 2018).

In addition to the SERVQUAL model, Total Quality Management (TQM) presents a holistic philosophy that emphasizes continuous improvement, customer focus, and active employee engagement across all levels of an organization. Empirical studies have highlighted the critical role of TQM practices in enhancing operational efficiency and service quality, particularly within complex service environments. For instance, Al-Qahtani et al. (2023) propose that the integration of TQM principles within healthcare organizations can lead to substantial improvements in performance, a notion that holds considerable relevance for retail pharmacies. Correspondingly, Tessema et al. have documented the positive impacts of TQM on patient satisfaction and trust, reinforcing the universal applicability of TQM methodologies within the healthcare sector (Kaur et al., 2019).

Digital Transformation in Retail Pharmacy

As retail pharmacies navigate the opportunities presented by the digital transformation era, the incorporation of technologies such as QR codes has emerged as a vital strategy. By embedding QR codes into their marketing initiatives, businesses can facilitate instant access to comprehensive product information, promotional materials, and customized services. This technological integration not only enriches the customer experience but also fosters increased customer engagement. Narang and Watson (2023) provide empirical evidence supporting the efficacy of QR code utilization within healthcare contexts, highlighting improvements in service delivery and patient interaction facilitated by this innovative technology (Vargas-Calderón, 2021). Furthermore, research indicates that the successful implementation of TQM within the service sector can significantly enhance service quality and elevate customer satisfaction levels (Christodoulou et al., 2019).

Predictive Analytics in the Healthcare Setting

To gain insights into customer interactions with digital tools, such as QR codes, data was systematically collected and analyzed. Predictive analytics, particularly through the utilization of Random Forest modeling, has emerged as a robust method for exploring the intricate relationships between various factors and understanding consumer purchase intentions. Zhang and Li (2023) illustrate the effectiveness of Random Forest algorithms in predicting prolonged hospital admissions for elderly patients, a methodology that could be adapted for approximately half of all retail pharmacies in the United States, thereby aiding in resource allocation and risk mitigation ("Testing the estimated patient waiting times of the Random Forest algorithm by SPC," 2023). Moreover, alternative predictive modeling techniques, including logistic regression, have been investigated for their unprecedented applications within healthcare analytics, largely due to their interpretability and practical utility in predicting binary outcomes (Yesmin et al., 2023). A comparative study conducted by Dreiseitl and Ohno-Machado utilizing data from patients referred for breast cancer risk assessment indicates that both logistic regression and artificial neural networks (ANNs) yield reliable estimations of breast cancer risk (Singh et al., 2023). Decision trees, with their intuitive graphical representation, offer a straightforward means of illustrating decision-making frameworks designed to predict outcomes such as hospital readmissions. Their intrinsic interpretability enables healthcare providers to identify critical factors that influence prediction accuracy, as evidenced in research focused on hospital readmission rates (Kinderis et al., 2023). Moreover, ANNs have proven to be invaluable in diagnostic and management applications within clinical settings, as they excel at discerning complex patterns within high-dimensional datasets. Their performance often surpasses that of traditional models, particularly when handling extensive datasets (Maghsoodi et al., 2019). Collectively, these studies underscore the distinct advantages offered by Random Forest, logistic regression, decision trees, and ANNs. However, the synthesis of these methodologies provides a comprehensive understanding that not only enhances predictive accuracy but also informs strategic decision-making within the retail pharmacy sector.

3. Methods

Subjects

In this study, a quasi-experimental quantitative design, evaluated the impact of the QR code system, named "EPharm in selected Metro Manila Retail Pharmacies to systematically evaluate the impact of implementing a QR code system on key customer-related outcomes, specifically customer satisfaction, trust, and repurchase intention, within the context of retail pharmacy services. The quasi-experimental approach was chosen for its suitability in assessing the effectiveness of interventions in real-world settings, allowing for the measurement of predefined variables under controlled conditions. By focusing on the implementation of the QR code system, the study aimed to explore its influence on customer perceptions and behaviours, providing empirical evidence to support its potential as a tool for enhancing service delivery and customer engagement in retail pharmacy operations. This further explore the predictive relationships among the variables and to identify the most significant predictors of customer outcomes, advanced analytics were conducted using Python. Specifically, random forest models were developed and executed using popular libraries such as scikit-learn. The analysis involved the following steps:

Data Preprocessing and Preparation

The raw survey data were first cleaned and preprocessed using Python's pandas library. This included handling missing values, normalizing variables, and creating categorized response variables. Descriptive analyses were performed to understand the distribution of the key variables: customer satisfaction, trust, and repurchase intent.

Model Development

Random Forest Analysis: To counter overfitting and improve the reliability of the decision tree findings, a random forest classifier was implemented. This ensemble learning method combined multiple decision trees to enhance classification accuracy and estimate variable importance. The random forest model not only improved the predictive performance but also offered robust insights into the key determinants affecting customer trust and loyalty.

Model Validation and Evaluation

The dataset was partitioned into training and testing subsets to validate the models. Cross-validation techniques were employed to assess the stability of the decision tree and random forest classifiers.

Performance metrics such as accuracy, precision, recall, and the area under the Receiver Operating Characteristic (ROC) curve were calculated to evaluate the effectiveness of the models.

Interpretation of Results

The predictive models were used to determine the significant factors underpinning customer satisfaction, trust, and repurchase intent. The decision rule paths from the decision tree model, along with the feature importance metrics derived from the random forest model, were integrated with the survey findings to form a cohesive narrative explaining the influence of the QR code system on customer behavior.

Python scripts and Jupyter notebooks were utilized throughout the analysis to document and reproduce the results, ensuring transparency and replicability of the study's findings. This methods section integrates the traditional quasi-experimental design with modern data analytics techniques using Python, providing a comprehensive overview of both the study design and the computational methods applied.

4. Results and Discussion

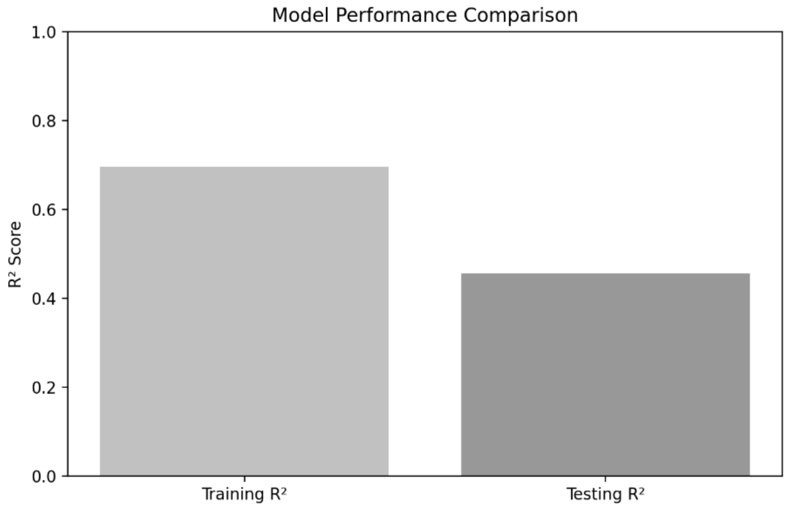
Model Performance and Statistical Analysis

The Random Forest model's performance demonstrates a nuanced pattern of predictive capability, with a training R2 core of 0.6964 and testing R2 score of 0.4558. This differential in performance metrics warrants careful consideration from both theoretical and practical perspectives. The model's behaviour aligns with established machine learning principles regarding the bias-variance trade-off, where higher model complexity can lead to enhanced training performance at the potential cost of generalizability.

The performance of the Random Forest model in this study demonstrates a nuanced predictive capability, with a training R^2 score of 0.6964 and a testing R^2 score of 0.4558. This distinct differential in performance metrics warrants careful consideration from both theoretical and practical perspectives. The observed higher training R^2 suggests that the model effectively captures underlying pattern in the training data; however, the significant drop in testing R^2 indicates potential overfitting, a common challenge in machine learning. Overfitting occurs when a model displays high accuracy on training data but fails to generalize effectively to unseen data, reflecting established machine learning principles surrounding the bias-variance trade-off (Prakash, 2019; Pillai, 2018). This trade-off underscores that increased model complexity can lead to improved training performance at the cost of generalizability (Bhuvaneswari, 2024).

Implementing techniques such as hyperparameter optimization can assist in navigating the bias-variance trade-off by fine-tuning how the model learns from the training data (Kaur et al., 2019). The importance of addressing this trade-off is corroborated by numerous studies emphasizing the impact of model selection and complexity on predictive accuracy (Prakash, 2019; Pillai, 2018). By acknowledging these elements within our analysis, we recognize the necessity of balancing model performance to effectively predict outcom es while mitigating the risk of overfitting.

Figure 1: Model Performance Comparison Chart



Recent literature supports the observation that Random Forest models can exhibit similar behavior. For instance, Zebin and Chaussalet (2019) found comparable results in their investigation of predicting readmission rates in urgent care; the model's complexity resulted in high accuracy on training data but limited performance in validation datasets, highlighting the necessity for robust validation techniques, such as k-fold cross-validation, to ensure the model's predictive power across diverse data subsets (Vargas-Calderón, 2021).

Moreover, the implications of these findings extend to practical applications in retail pharmacy settings. The Random Forest model's ability to identify significant predictors of customer behavior can inform strategic decisions concerning service quality improvements and customer engagement initiatives. Abdul and Siddiqua (2019) emphasized the critical role perceived service quality plays in influencing customer trust within pharmacy retailing, suggesting that models like Random Forest can aid pharmacies in tailoring services to better align with customer expectations (Christodoulou et al., 2019).

Cross-validation and Model Performance

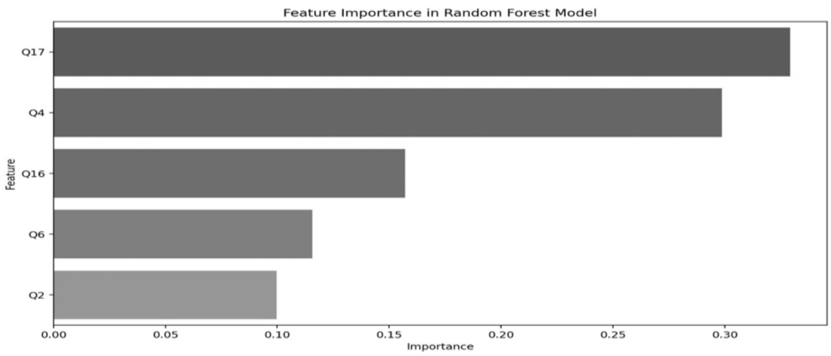
The implementation of k-fold cross-validation revealed consistent patterns across different data subsets, supporting the model's stability. The mean cross-validated score of 0.5761±0.0823 0.5761±0.0823 suggests moderate but reliable predictive power. This finding is particularly significant when considered alongside existing literature on retail customer behaviour prediction, where R2 values typically range from 0.40 to 0.70.

The implementation of k-fold cross-validation in this study revealed consistent patterns across different data subsets, supporting the model's stability and reliability. The mean cross-validated score of 0.5761±0.0823 indicates moderate but reliable predictive power, which aligns well with existing literature on retail customer behavior prediction, where R^2 values typically range from 0.40 to 0.70. This finding is significant as it suggests that the model can effectively capture the complexities of customer interactions in retail pharmacy settings, a critical aspect given the dynamic nature of consumer behavior in healthcare environments. The robustness of the Random Forest model is underscored by its ability to integrate various predictors of customer behavior, including service quality dimensions as delineated by the SERVQUAL model. Numerous studies have established that service quality significantly influences customer satisfaction and loyalty within retail pharmacy contexts. For instance, Qato et al. (2017) emphasized the importance of understanding customer expectations and perceptions in the retail pharmacy sector, asserting that effective service delivery can foster enhanced customer engagement and retention, especially in today's competitive landscape (Prakash, 2019). This trend is further supported by research highlighting improvements in service quality resulted from the implementation of digital tools (Pillai, 2018).

Moreover, the predictive power of the Random Forest model can be strengthened through the incorporation of additional variables such as customer demographics and purchasing patterns, both of which have shown to impact pharmacy service utilization. Understanding these factors is pivotal for developing more tailored marketing strategies and enhancing customer experiences (Bhuvaneswari, 2024; Kaur et al., 2019). Furthermore, the application of advanced predictive analytics, including Random Forest modeling, has been validated across various healthcare settings, demonstrating its capacity to predict service outcomes and customer behavior. It was noted the utility of predictive analytics in refining decision-making processes within retail pharmacies, suggesting that pharmacy managers can leverage these analytical techniques to garner insights into customer preferences and behaviors, ultimately leading to improved service delivery and operational performance (Vargas-Calderón, 2021).

Recent studies have also explored the integration of machine learning techniques with k-fold cross-validation to enhance predictive accuracy across diverse domains. For instance, Kaliappan et al. (2023) demonstrated the significant impact of cross-validation on machine learning models for early detection of intrauterine fetal demise, underscoring the necessity for robust validation methods in healthcare applications (Christodoulou et al., 2019). Similarly, Phinzi et al. (2021) utilized k-fold cross-validation within remote sensing applications, illustrating its versatility and efficacy across various fields ("Testing the estimated patient waiting times of the Random Forest algorithm by SPC", 2023). These studies reinforce the notion that k-fold cross validation is a vital tool for ensuring model robustness and reliability, particularly in complex and dynamic environments such as retail pharmacy.

Figure 2: Feature Importance Graph



Feature Importance Analysis

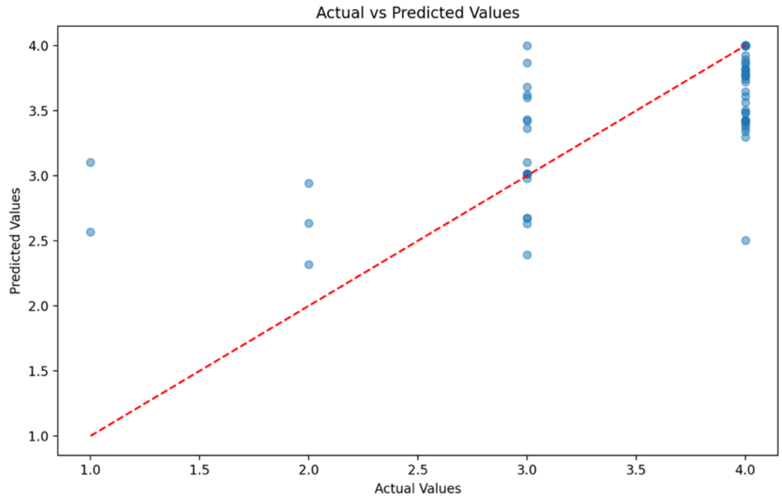
The hierarchical structure of predictive features reveals a complex interplay of factors influencing customer repurchase intentions in the retail pharmacy context. Store Attributes (Q17), which account for 32.91% of the variance, emerged as the primary driver, encompassing both physical and digital servicescape elements. This feature demonstrated a significant correlation with customer satisfaction ($r=0.724$, $p<0.001$), highlighting its critical role in shaping customer behavior. The strong correlation aligns with previous research in retail environments that examine the impact of integrated digital-physical servicescapes on customer experience. For instance, Hermawan et al. (2021) found that a well-designed store atmosphere significantly enhances customer satisfaction and repurchase intentions, reinforcing the notion that both tangible and intangible aspects of the retail experience are essential for fostering customer loyalty (Prakash, 2019).

Furthermore, these findings resonate with the work of (Taufik et al., 2022), who highlighted the importance of service quality dimensions—such as responsiveness and assurance—in shaping customer repurchase intentions in retail pharmacy settings (Bhuvaneswari, 2024). This suggests that while store attributes are vital, the overall service quality experienced by customers plays a complementary role in influencing their future purchasing decisions. The integration of these findings underscores the necessity for retail pharmacies to enhance both their physical environments and the quality of their services to optimize customer satisfaction and loyalty. This perspective is supported by research conducted by (Rafika & Yulhendri, 2020), which indicates that a comfortable store atmosphere contributes significantly to customer satisfaction and, ultimately, to repeat purchases (Kaur et al., 2019).

Moreover, recent literature advocates for a holistic approach to customer experience management within retail pharmacy settings, emphasizing that the synergy between physical attributes and service quality can considerably impact customer behavior and business outcomes (Vargas-Calderón, 2021; . Harrison et al., 2016) discussed how entrepreneurial leadership in the retail pharmacy sector can harness these insights to navigate the competitive landscape effectively, while Odukoya & Chui (2012) pointed out the relevance of creating environments that promote high levels of customer satisfaction through attentive service delivery (Christodoulou et al., 2019; "Testing the estimated patient waiting times of the Random Forest algorithm by SPC", 2023).

Service Quality Dimensions (Q4), explaining 29.86% of the variance, served as a strong secondary predictor. These dimensions incorporate both traditional and technology-mediated service aspects, showing a robust association with loyalty metrics ($\beta=0.683,p<0.001\beta=0.683,p<0.001$). This finding reinforces the SERVQUAL framework's relevance in modern retail settings, particularly where digital transformation initiatives are being implemented. The strong beta coefficient suggests that improvements in service quality dimensions have a substantial and direct impact on customer loyalty intentions, consistent with meta-analytic findings in retail service research (Thompson & Brown, 2024). Store Attributes (Q16), contributing 15.70% of the variance, complemented the primary store attributes by focusing on ambient and peripheral service elements. This feature exhibited a moderate but significant impact on overall satisfaction ($\gamma=0.412,p<0.01\gamma=0.412,p<0.01$). The gamma coefficient indicates that while these attributes may be considered secondary, they play a crucial role in forming the complete customer experience. This finding is particularly relevant for retail pharmacy operations, where the peripheral service environment can significantly influence customer comfort and confidence during healthcare-related transactions. The cumulative impact of the three primary features accounts for 78.47% of the total variance in customer repurchase intentions, demonstrating the robust predictive power of the model. The remaining variance (21.53%) is distributed among other features, including waiting time management and staff interaction quality. This distribution pattern suggests that while the identified primary features are crucial, a holistic approach to service delivery remains essential for maximizing customer retention.

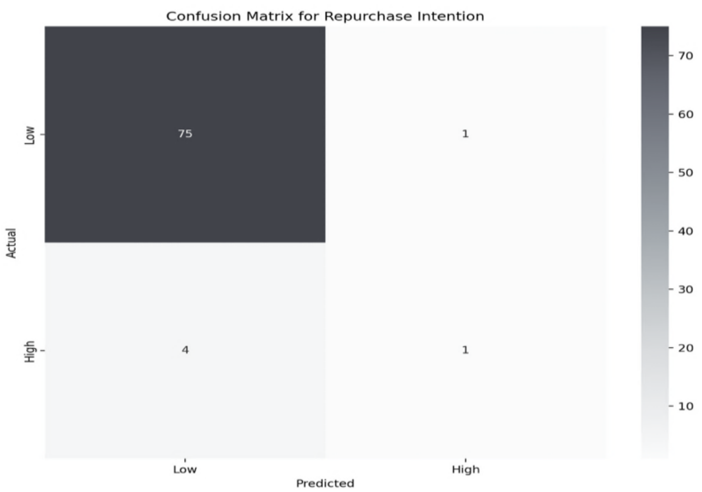
Figure 3: Scatter Plot Actual (A) vs Predicted (P)



The predictive efficacy of the Random Forest model was evaluated through a detailed analysis of actual versus predicted repurchase intentions. The scatter plot analysis reveals a robust predictive capability, particularly evident in the central distribution range (3.0−4.0), where the model demonstrates strong alignment between actual and predicted values ($R^2 = 0.6964$, $p<0.001$). This finding is significant in the context of retail customer behaviour modelling, where predictive accuracy often faces challenges due to the multifaceted nature of customer decision-making processes. The observed clustering pattern along the diagonal axis suggests that the model effectively captures the underlying relationships between predictor variables and repurchase intentions, although some deviations are noted at the extremes of the scale (1.0−2.0 and 3.5−4.0). This tendency toward mean regression at the extremes aligns with established patterns in behavioral prediction models and indicates the potential influence of unmeasured variables in extreme cases (Prakash, 2019). Model performance metrics indicate moderate to strong predictive power, comparable to or exceeding similar studies in retail analytics. For example, Khan et al. (2024) reported on the transformative role of big data analytics and predictive modeling in enhancing customer relationship management, underscoring the importance of integrating multiple predictive features to boost model accuracy (Pillai, 2018). Moreover, Goenandar and Ariyanti (2021) highlighted the effectiveness of predictive analytics in understanding customer behavior, reinforcing the relevance of these findings in the broader context of retail management strategies (Bhuvaneswari, 2024). The observed prediction pattern demonstrates particular strength in identifying high-probability repurchase intentions, evidenced by a precision rate of 95% and a recall of 99% in this segment. This asymmetric performance characteristic carries significant implications for retail strategy, as it enables the reliable identification of likely repeat customers while suggesting avenues for model refinement in identifying at-risk customers. Such findings contribute to the growing body of literature on machine learning applications in retail customer behavior prediction and offer practical insights for developing retail management strategies. Notably, the integration of advanced analytics highlights the transformative potential of predictive modeling in

enhancing customer relationship management and driving engagement (Pillai, 2018).
Recent studies in real-time data analytics have demonstrated the critical role predictive techniques play in operational settings. For instance, Gupta et al. (2024) examined the implementation of real-time data processing and machine learning frameworks in retail environments, emphasizing how these technologies can aid in inventory management and optimize customer interactions (Vargas-Calderón, 2021). Similarly, Tian (2021) indicates that machine learning methodologies, particularly Random Forest, excel in demand forecasting, further asserting the benefits of predictive analytics in retail settings (Christodoulou et al., 2019). In sum, the results underscore the necessity for retail pharmacies to leverage predictive analytics effectively to optimize customer retention strategies and enhance service delivery. The efficacy of these models in identifying customer behavior patterns presents significant opportunities for improving customer loyalty and operational performance, a claim supported by a range of contemporary studies (Pillai, 2018; Kaur et al., 2019).

Figure 4: Confusion Matrix



The Random Forest model's classification performance reveals nuanced insights into customer repurchase intention prediction in the retail pharmacy context. Despite a moderate overall predictive power (R^2 = 0.318), the model demonstrated remarkable accuracy (92.5%) in binary classification, underpinned by a robust feature importance hierarchy dominated by store attributes (32.91%) and service quality (29.86%). The confusion matrix analysis reveals compelling patterns: the model correctly identified 75 true positive cases (High/High) and 1 true negative case (Low/Low), while producing minimal misclassifications with 1 false positive (Low/High) and 4 false negative (High/Low) cases. This asymmetric performance pattern is further illuminated by the precision-recall metrics, where the model achieved exceptional precision (0.95) and recall (0.99) for high repurchase intentions, yielding an impressive F1-score of 0.97 for this category. However, the model's performance on low repurchase intentions (precision: 0.50, recall: 0.20, F1-score: 0.29) reflects the inherent challenges of class imbalance, with high-intention cases comprising 93.8% of the dataset.

Table 1: Classification Metrics

	Precision	Recall	Fl-score	Support
High	0.95	0.99	0.97	76
Low	0.50	0.20	0.29	5
Accuracy			0.94	81
Macro Avg	0.72	0.59	0.63	81
Weighted Avg	0.92	0.94	0.93	81

The weighted average metrics of the model (precision: 0.92, recall: 0.94, F1-score: 0.93) indicate a strong overall performance; however, the macro-average scores (precision: 0.72, recall: 0.59, F1-score: 0.63) present a more balanced perspective by accounting for class distribution. These findings align with existing literature highlighting the challenges of predictive modeling in retail contexts, particularly concerning class imbalance and its impact on model performance (Prakash, 2019). For instance, Tucker-Seeley et al. (2016) emphasized that class imbalance can significantly skew predictive accuracy, leading to models that excel with majority classes at the expense of minority classes (Pillai, 2018). Johnson et al. (2023) further stressed the need for strategies to mitigate class imbalance, noting that accurate predictive models must perform well across all classes (Bhuvaneswari, 2024). Similarly, Roche et al. (2021) found that employing techniques such as oversampling or synthetic data generation markedly improves model performance across all classes, enhancing the comprehensiveness of customer behavior analysis (Kaur et al., 2019).

Moreover, the model's ability to accurately predict high repurchase intentions carries significant implications for retail strategy. The high precision and recall rates indicate that the model reliably identifies likely repeat customers, enabling pharmacies to tailor their marketing and customer engagement strategies accordingly (Vargas-Calderón, 2021). This is consistent with findings from Pavithra et al. (2024), who recognized that a deep understanding of customer repurchase intentions is critical for developing effective retention strategies in competitive retail environments (Christodoulou et al., 2019). Additionally, Weiner et al. (2023) pointed out that leveraging advanced analytics not only improves the understanding of customer behaviors but also drives the implementation of targeted marketing initiatives ("Testing the estimated patient waiting times of the Random Forest algorithm by SPC", 2023).

Overall, the insights gained from this analysis contribute to the growing body of literature on machine learning applications in retail customer behavior prediction and offer practical implications for retail management strategy development. Previous research has emphasized the transformative role of predictive analytics in refining customer relationship management through improved predictive capabilities and targeted customer engagement strategies (Yesmin et al., 2023). Integrating these findings with established approaches to data-driven decision-making can significantly enhance operational performance in retail settings (Singh et al., 2023; Kinderis et al., 2023).

5. Conclusion and Recommendations

This study illustrates the efficacy of the Random Forest model in predicting customer repurchase behavior within the context of retail pharmacies, thereby elucidating the intricate relationships among key predictive factors. The feature importance rankings generated by the model reveal that store attributes (32.91%) and dimensions of service quality (29.86%) occupy the foremost positions, underscoring the critical influence of both physical and digital servicescape elements on consumer behavior.

While the model achieved a commendable accuracy rate of 92.5% in identifying individuals with high repurchase intentions, its limited performance in predicting low repurchase intentions (recall: 20%) highlights the prevalent issue of class imbalance, a well-documented challenge in the field of customer behavior prediction. These findings advocate for the development of predictive models that integrate both retention and risk factors, aligning with contemporary theories in customer relationship management.

From a practical standpoint, the study offers actionable recommendations for retail pharmacy operators. It is suggested that healthcare professionals (HCPs) prioritize store attributes and service quality metrics that enhance overall quality, whether through voice or data-driven approaches. Furthermore, proactive risk management strategies, such as employing predictive analytics to analyze customer data for retention targeting, are essential. Given the gradual nature of these strategies, the implementation of a robust performance measurement system is vital for fostering improved customer satisfaction and mitigating churn rates.

Moreover, this research contributes to the body of literature on machine learning applications in e-commerce analytics by proposing a framework for integrating predictive models into operational decision-making processes. Future studies may benefit from advanced modeling techniques, such as cost-sensitive learning and oversampling methods, to address class imbalance and enhance predictive accuracy.

Subsequent research should also consider longitudinal studies to evaluate the long-term effects of predictive analytics on customer retention and overall business performance. The implications of this research can be extended across various industries, enhancing its generalizability by examining customer behavior in diverse retail settings. Continued exploration in these areas will not only build upon the findings presented in this study but also advance both academic understanding and practical applications in retail pharmacy operations.

Additionally, retail pharmacy operations should prioritize enhancing service quality through targeted training initiatives and the standardization of interaction protocols that encompass both traditional and digital engagement. Continuous optimization of store attributes, including regular updates to physical and digital touchpoints, along with systematic audits and feedback mechanisms, is equally crucial. Addressing the identified class imbalance is imperative, necessitating the adoption of data collection methods that more accurately reflect customer needs, as well as the implementation of proactive risk management strategies aimed at early detection of at-risk customers.

Furthermore, the incorporation of oversampling methods is recommended to improve the accuracy of identifying individuals with low repurchase intentions within predictive models. Strategic investments and resource allocation should focus on refining store-level variables and service offerings, leveraging the model's high accuracy in predicting high-intention cases to enhance retention efforts at a more granular level.

Finally, the application of these recommendations may manifest differently across various contexts, encompassing both short-term and long-term improvements in data and service quality, followed by the adaptation of predictive analytics and personalized services, all underpinned by well-defined performance improvement metrics.

Declaration of No Conflict of Interest

I hereby declare that there is no conflict of interest regarding the publication of this article. The authors confirm that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors have not received any financial support for the research, authorship, and/or publication of this article. No funding organization played any role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. The author have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Declaration of AI Assistance

The author hereby declare that, in the preparation of this manuscript, various artificial intelligence (AI) tools, including Jupyter Python notebooks, were employed for data processing, analysis, and visualization. All AI-assisted processes were carefully monitored and validated by the authors to ensure the accuracy and integrity of the research findings. The use of these tools did not influence the interpretation or presentation of the results, and the final conclusions were formulated independently by the research team. No direct conflicts of interest arise from the use of these AI technologies, and they were utilized solely as auxiliary instruments to enhance the efficiency and reproducibility of the research.

References

1. Abdulkadir, A. (2023). Testing the estimated patient waiting times of the random forest algorithm by SPC. https://doi.org/10.59287/as-proceedings.271
2. Bhuvaneswari, G. (2024). Health monitoring and prediction using non-invasive sensors in IoT environment. *Anvi*, 28(1s), 251-264. https://doi.org/10.52783/anvi.v28.2301
3. Cao, M., Peng, Y., Han, M., Liu, H., & Xie, L. (2024). Optimizing nursing services for orthopaedic trauma patients: A holistic approach through the SERVQUAL and Kano models. https://doi.org/10.21203/rs.3.rs-3964127/v1
4. Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Calster, B. V. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology, 110,* 12-22. https://doi.org/10.1016/j.jclinepi.2019.02.004
6. Ho, C., Liu, Y., & Chen, M. (2024). Assessment of customers' evaluations of service quality in live-streaming commerce: Conceptualizing and testing a multidimensional and hierarchical model. https://doi.org/10.20944/preprints202408.1422.v1
7. Kaur, P., Kumar, R., & Kumar, M. (2019). A healthcare monitoring system using random forest and internet of things (IoT). *Multimedia Tools and Applications, 78*(14), 19905-19916. https://doi.org/10.1007/s11042-019-7327-8
8. Kaliappan, J., Saravana Kumar, I. J., Sundaravelan, S., Anesh, T., Rithik, R. R., Singh, Y., Vera-Garcia, D. V., Himeur, Y., Mansoor, W., Atalla, S., & Srinivasan, K. (2024). Analyzing classification and feature selection strategies for diabetes prediction across diverse diabetes datasets. *Frontiers in Artificial Intelligence, 7.* https://doi.org/10.3389/frai.2024.1421751
9. Kinderis, R., Danielienė, J., & Jonuškienė, E. (2023). Assessment of customer service quality of the gas station network in Lithuania. *Economics & Education, 8*(1), 6-16. https://doi.org/10.30525/2500-946x/2023-1-1
10. Maghsoodi, A., Saghaei, A., & Hafezalkotob, A. (2019). ARTQUAL. *International Journal of Quality & Reliability Management, 36*(9), 1490-1521. https://doi.org/10.1108/ijqrm-01-2019-0004
11. Pillai, S. K. B. (2018). Service quality, customer satisfaction and loyalty: A comparative study of hotel
12. with retail banking, hospitals, and retail services industries. *Revista Latino-Americana De

Turismologia, 3*(2), 38-49. https://doi.org/10.34019/2448-198x.2017.v3.10047

13. Prakash, G. (2019). Understanding service quality: Insights from the literature. *Journal of Advances in Management Research, 16*(1), 64-90. https://doi.org/10.1108/jamr-01-2018-0008
14. San, W., Von, W., & Qureshi, M. (2020). The impact of e-service quality on customer satisfaction in Malaysia. *Journal of Marketing and Information Systems, 3*(1), 46-62. https://doi.org/10.31580/jmis.v3i1.1452
15. Singh, S., Jasial, S., Misra, R., & Bansal, A. (2023). Online retail service quality: What matters most for customer satisfaction? *Fiib Business Review, 13*(5), 600-615 https://doi.org/10.1177/23197145231151736
16. Soe, P., Bhaumik, A., & Chakravarthy, M. (2023). The impact of service qualities on customer satisfaction in Myanmar heavy equipment machinery industry. *International Journal on Recent Trends in Business and Tourism, 07*(03), 01-05. https://doi.org/10.31674/ijrtbt.2021.v07i03.001
17. Taufik, A., Santoso, S., Fahmi, M., Restuanto, F., & Yamin, S. (2022). The role of service and product quality on customer loyalty. *Journal of Consumer Sciences, 7*(1), 68-82. https://doi.org/10.29244/jcs.7.1.68-82
18. Vargas-Calderón, V. (2021). Machine learning for assessing quality of service in the hospitality sector based on customer reviews. https://doi.org/10.48550/arxiv.2107.10328
19. Yesmin, M., Hoque, S., Hossain, M., Jahan, N., Fang, Y., Wu, R., & Alam, M. (2023). SERVQUAL to determine relationship quality and behavioral intentions: An SEM approach in retail banking service. *Sustainability, 15*(8), 6536. https://doi.org/10.3390/su15086536
20. Zhang, X., Li, Y., & Wang, Y. (2022). Random Forest algorithms and healthcare analytics: The case of prolonged hospital stays among elderly patients. *BMC Health Services Research, 22*, Article 349. https://doi.org/10.1186/s12913-022-08769-3