Review of Applied Science and Technology

Volume: 2; **Issue:** 3
Pages: 67– 114

Neural Network–Based Customer Retention Forecasting in Mobile Wallet Services Using 200k Historical User Profiles

Iftekhar Ahmed[1]; Binayan Dey[2];

[1]. Assistant Manager, Pathao Limited, Bangladesh; Email: iftekhar.nabil@gmail.com

[2]. Assistant Manager, Systems & IT, Chittagong Stock Exchange Ltd, Bangladesh; Email: binayan.dey@gmail.com

Doi: 10.63125/ee5eas98

Received: 12 June 2023; **Revised:** 20 July 2023; **Accepted:** 14 August 2023; **Published:** 11 September 2023

Abstract

This study developed and evaluated a neural network–based retention forecasting model using 200,000 historical mobile wallet user profiles. Retention was operationalized as continued transactional activity within a defined post-observation horizon, with 68% of users classified as retained and 32% classified as churned. Descriptive results indicated substantial behavioral heterogeneity, with mean transaction frequency of 12.8 transactions (SD = 21.3), median of 6.0, and mean recency gap of 18.6 days (SD = 27.4). Retained users demonstrated higher average merchant diversity (M = 6.8) compared with churned users (M = 2.9), and shorter inter-transaction gaps (M = 11.2 days vs. 34.7 days). Logistic regression results showed significant effects for recency (β = -0.042, $p < .001$), transaction frequency (β = 0.087, $p < .001$), and engagement trend (β = 0.054, $p < .001$). The logistic baseline achieved an AUC of 0.78 and PR-AUC of 0.64, while regularized regression improved performance to an AUC of 0.80 and PR-AUC of 0.67. The neural network model achieved superior discrimination with an AUC of 0.86 and PR-AUC of 0.74, and produced a lift of 3.5 in the top 10% risk segment. Segment-level results remained stable across new users (AUC = 0.84) and mature users (AUC = 0.88). Findings demonstrated that combining intensity, diversity, and stability constructs within a neural architecture significantly enhanced retention forecasting accuracy in large-scale mobile wallet environments.

Keywords

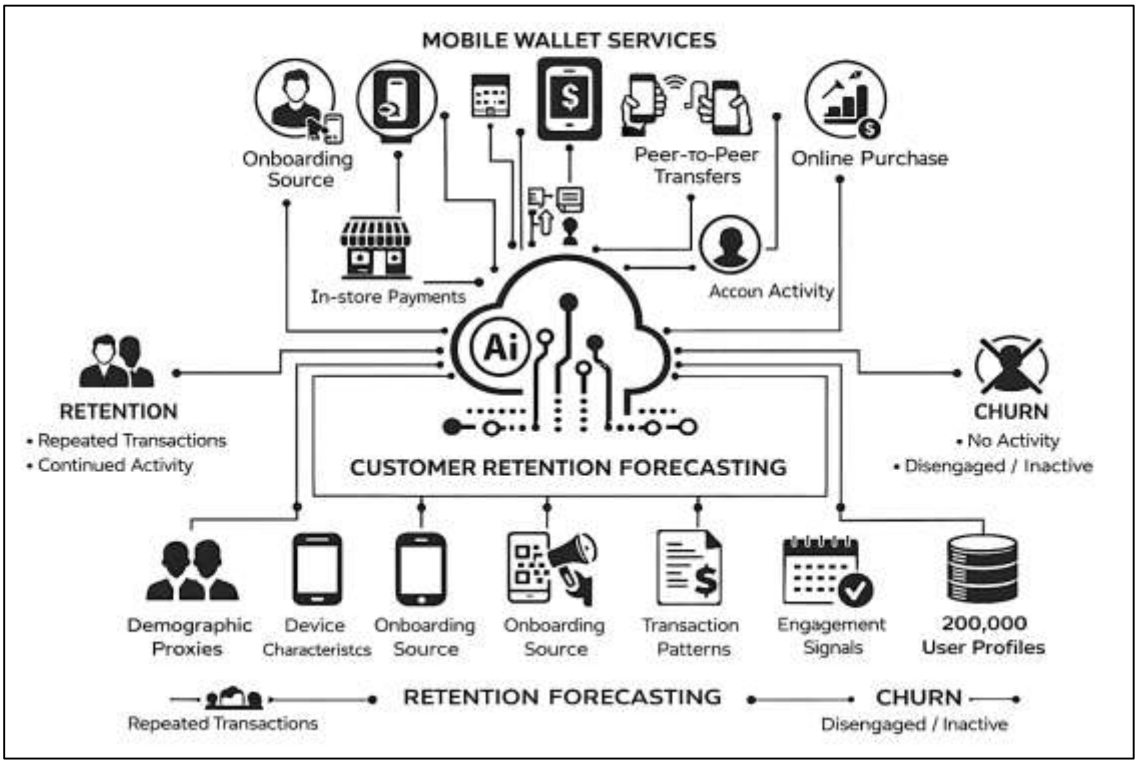
Neural Networks, Customer Retention, Mobile Wallets, Predictive Modeling, Behavioral Analytics.

INTRODUCTION

Customer retention forecasting in mobile wallet services begins with precise definitions of the core concepts that guide quantitative investigation. A mobile wallet can be defined as a digital payment application that stores monetary value or payment credentials and enables users to conduct financial transactions through a smartphone or connected device (Khoa, 2020). These transactions may include peer-to-peer transfers, merchant payments, online purchases, bill payments, airtime top-ups, and account funding activities. Customer retention, in this context, refers to the sustained continuation of user activity within the mobile wallet ecosystem over a defined time horizon. Retention is commonly measured through repeated transactions, continued logins, ongoing balance activity, or persistent engagement with wallet features during consecutive periods. Churn is the opposite behavioral outcome, representing inactivity or discontinuation, typically operationalized as the absence of any meaningful activity for a predefined number of days or billing cycles. Retention forecasting is therefore the statistical task of predicting whether a user will remain active in the future, given their historical behavior and profile characteristics. In quantitative terms, this becomes a supervised learning problem where the dependent variable is binary (retained vs. churned) or probabilistic (likelihood of retention). The independent variables are derived from historical user profiles, including demographic proxies, device characteristics, onboarding source, transaction patterns, and engagement signals (Gao & Waechter, 2017). When the dataset contains 200,000 historical user profiles, retention forecasting becomes particularly suited to modern machine learning methods because the volume allows stable estimation, subgroup coverage, and strong statistical power. However, it also demands careful construction of time-consistent predictors to prevent leakage, meaning the model must only use information that would have been available before the forecasting point. The definitional foundation is essential because mobile wallet engagement is not uniform: some users transact daily, others transact monthly, and some use the wallet only for specific events such as salary receipt, family remittances, or seasonal purchases (Gimpel et al., 2018). A rigorous quantitative introduction must therefore treat retention as a measurable behavioral phenomenon rather than a vague marketing concept, while also recognizing that retention is embedded within financial routines, trust perceptions, and service utility. This definitional clarity provides the conceptual base for neural network–based retention forecasting, where large-scale historical user profiles become structured inputs for predicting future continuity of mobile wallet usage.

The international significance of customer retention forecasting in mobile wallet services is grounded in the central role mobile payments now play in everyday commerce and digital financial participation across the world. Mobile wallets have expanded rapidly in both developed and emerging economies, becoming an essential layer of the payment infrastructure for consumers, merchants, and service providers (Yang et al., 2021). In many countries, mobile wallets are not simply optional convenience tools; they are critical channels for salary disbursement, government payments, remittances, micro-merchant transactions, and digital commerce. Retention matters internationally because the sustainability of mobile wallet ecosystems depends on active user bases. A wallet platform may acquire large numbers of users through marketing campaigns, referral bonuses, and onboarding promotions, yet these users do not generate stable economic value unless they continue transacting over time. Retained users are more likely to increase transaction frequency, adopt additional wallet services, and participate in network-driven features such as peer transfers and merchant acceptance expansion. At the ecosystem level, retention also affects merchant confidence (Su et al., 2021). Merchants are more likely to invest in QR code infrastructure, payment acceptance training, and integration with digital settlement processes when customer usage is persistent rather than sporadic. On the provider side, retention forecasting has operational significance because it influences customer support load, fraud monitoring volume, settlement traffic, and revenue predictability. The global scale of mobile wallet services amplifies these effects, as millions of micro-behaviors aggregate into macroeconomic payment flows. Retention forecasting is also important because mobile wallets operate under varying cultural and regulatory environments. Some markets exhibit strong cash preferences and low digital trust, while others show high smartphone penetration and established digital payment norms. These differences produce distinct engagement patterns that must be captured through data-driven modeling. A retention forecasting model built on 200K historical profiles provides a valuable quantitative lens for understanding how user behaviors translate into continued usage under real-world conditions. The international significance is therefore not limited to business profitability; it extends to digital payment stability, consumer reliance on mobile financial services, and the broader continuity of digital commerce participation. Because mobile wallets often act as gateways to financial inclusion, particularly for underbanked populations, retention can also reflect whether users sustain access to convenient payment tools or revert to informal alternatives (Liébana-Cabanillas et al., 2017). In this way, neural network–based retention forecasting becomes a globally relevant analytical task: it aims to predict continuity within a service category that increasingly underpins economic interaction across borders, income levels, and transaction contexts.

Figure 1: Neural Network Mobile Wallet Retention



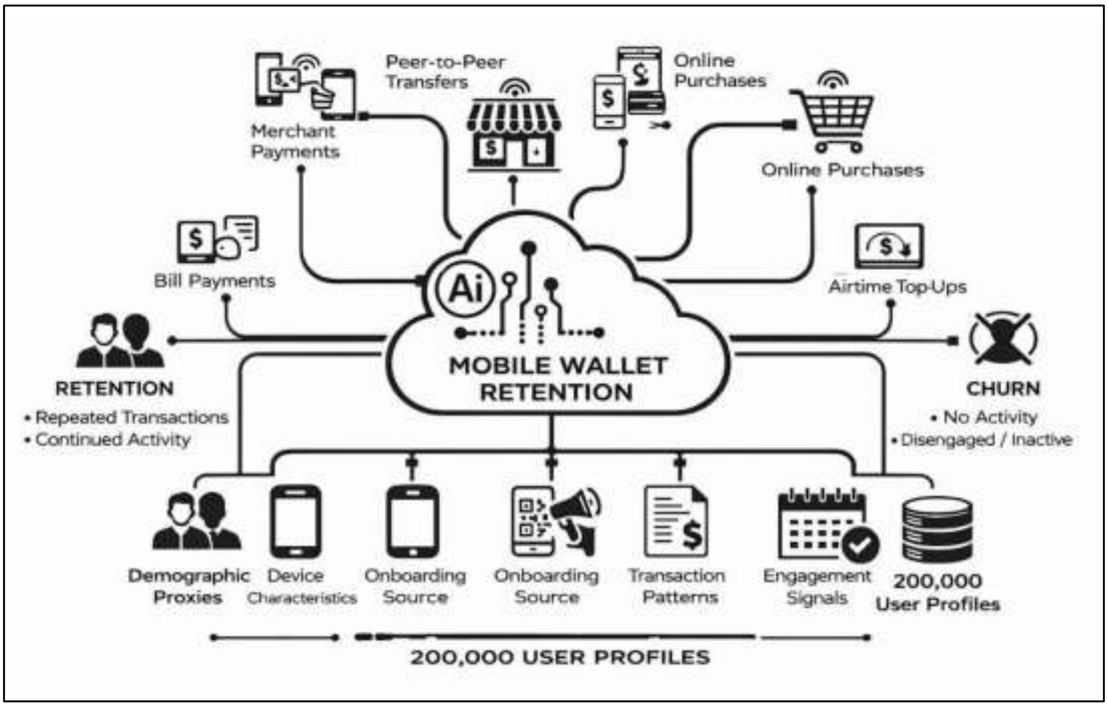
The quantitative study of retention forecasting draws on several well-established perspectives from service research, consumer behavior, and information systems. One foundational idea is that continued usage of a service is driven by perceived value, satisfaction, trust, and habitual integration into daily routines (Gremler et al., 2020). In mobile wallet contexts, perceived value may be reflected in the convenience of payments, the speed of transfers, the breadth of merchant acceptance, and the availability of rewards or cashback. Satisfaction can be influenced by transaction success rates, user interface usability, dispute resolution quality, and customer support responsiveness. Trust is especially central because mobile wallets involve financial risk perceptions, privacy concerns, and fears of fraud. Habitual usage emerges when the wallet becomes a default tool for recurring activities such as transportation payments, utility bills, subscription renewals, or peer transfers. These constructs, although psychological in nature, can be approximated quantitatively through behavioral proxies. For example, stable and increasing transaction frequency may indicate rising utility and habit formation. High diversity in transaction categories may indicate deeper integration into multiple life domains. Sudden decreases in activity may signal dissatisfaction, reduced trust, or competing service substitution. A quantitative model built on historical user profiles aims to learn the mapping between these behavioral proxies and future retention outcomes (Kumar & Ayodeji, 2021). This mapping is rarely linear. Users do not necessarily retain because they transact more; some high-transaction users may churn after promotional periods, while some low-transaction users may remain stable long-term due to occasional but recurring needs. Retention forecasting therefore requires analytical approaches that can capture interactions among variables. For instance, a small increase in transaction frequency may strongly predict retention among new users, while the same increase may have minimal predictive value among mature users who already transact frequently. Similarly, incentives may influence retention differently depending on user segment, transaction size, or merchant type. These interaction effects become more visible in large datasets such as 200K profiles, where enough observations exist to represent varied user pathways (Liébana-Cabanillas et al., 2021). Quantitative retention forecasting also involves careful outcome definition. A user may appear inactive for weeks yet return later, so defining churn requires selecting an inactivity threshold that represents meaningful discontinuation. The forecasting horizon must align with service usage cycles, which may depend on pay periods, bill schedules, or seasonal events. These definitional choices affect label stability and predictive performance. Within this theoretical frame, retention forecasting becomes a measurable behavioral classification problem, grounded in service continuity principles and operationalized through large-scale transaction and engagement records. This foundation supports the rationale for applying neural networks, which can learn complex relationships from such data without requiring rigid assumptions about linearity or independence.

Neural network–based forecasting represents a powerful quantitative approach for predicting customer retention because neural architectures are designed to model non-linear relationships and learn layered representations from high-dimensional inputs. Artificial neural networks consist of interconnected computational units that transform inputs through weighted combinations and non-linear activation functions, enabling models to approximate complex decision boundaries. In retention forecasting, the input space often includes dozens or hundreds of engineered features derived from historical profiles (Pal et al., 2021). These features may include recency, frequency, monetary aggregates, transaction category diversity, session engagement measures, failed transaction counts, device switching frequency, geographic stability, and response to incentives. Traditional models such as logistic regression can provide baseline predictions, yet they may struggle to capture non-linear interactions unless explicitly engineered. Neural networks, in contrast, can learn interactions automatically by combining features across hidden layers. This is particularly relevant in mobile wallet services, where user behavior is shaped by multiple simultaneous factors such as convenience, trust, merchant availability, and promotion schedules. Neural networks can also incorporate embeddings for categorical variables such as device type, onboarding channel, or region, allowing the model to learn similarity relationships among categories rather than treating them as unrelated one-hot indicators. Another advantage is the ability to learn from sequential data. If the dataset includes ordered transaction histories, sequence-based neural models such as recurrent networks or long short-term memory units can encode temporal dependencies, capturing how the timing and order of events influence retention (Buttle & Maklan, 2019). For example, a user who makes three transactions in the first week and then becomes inactive may have a different retention probability than a user who spreads transactions evenly across months, even if both have the same total transaction count. Neural sequence models can learn such patterns by processing events chronologically. The scale of 200K profiles supports the training requirements of neural models, as deep learning generally benefits from large datasets to learn robust patterns and reduce overfitting. Yet scale alone does not guarantee performance. Neural networks can memorize artifacts if features contain leakage or if the data split mixes future and past behaviors incorrectly. For that reason, neural network–based retention forecasting requires disciplined data preparation, time-consistent splitting, and careful regularization strategies. Regularization techniques such as dropout, weight decay, and early stopping help prevent overfitting, especially when feature sets are large. Optimization algorithms such as adaptive gradient methods can improve training stability (Hair Jr et al., 2019). These methodological considerations position neural networks as suitable for mobile wallet retention forecasting because they can learn the complex, non-linear, and temporally structured relationships that characterize digital payment behavior at scale.

The dataset focus of 200,000 historical user profiles introduces methodological richness and statistical strength, while also raising important quantitative design considerations. A user profile in mobile wallet services typically combines static information and dynamic behavioral histories. Static information may include onboarding channel, account verification status, device characteristics, language settings, and region indicators. Dynamic information includes transaction timestamps, amounts, merchant identifiers, merchant categories, peer transfer patterns, top-up behaviors, and session engagement metrics (Lopes et al., 2021). In predictive modeling, these dynamic signals are usually transformed into fixed-length feature vectors by aggregating behavior over defined observation windows. Common aggregation strategies include computing recency measures, frequency counts, monetary sums, averages, standard deviations, and trend indicators across multiple time intervals. A multi-window approach is often used, such as features for the last 7 days, last 30 days, and last 90 days, enabling the model to capture short-term momentum and long-term stability simultaneously. Feature scaling and transformation become essential because transaction amounts often have heavy-tailed distributions, and activity counts can vary widely between low-use and high-use users. Large datasets also highlight the problem of class imbalance, where the retained class may dominate the churn class depending on the horizon definition. Class imbalance affects both training and evaluation, making it necessary to consider metrics beyond simple accuracy. Evaluation may emphasize ranking ability and probabilistic calibration, ensuring that predicted retention probabilities correspond meaningfully to observed outcomes. Time-aware validation is critical. A retention model must be tested on later cohorts to reflect real-world forecasting, since random splits can inflate performance by allowing the model to learn patterns that are not stable over time (Koghut & AI-Tabbaa, 2021). Large datasets also make it possible to analyze heterogeneity across segments, such as differences by onboarding month, device type, transaction type dominance, or incentive responsiveness. This is particularly relevant for mobile wallets because users can be driven by distinct motivations: some primarily use the wallet for peer transfers, others for merchant payments, others for bill pay, and others for occasional online purchases. These different usage modes may correspond to different retention dynamics. The availability of 200K profiles provides the statistical power to model these differences without collapsing users into overly broad categories. At the same time, large-scale datasets often contain noise, missingness, and inconsistencies. Some users may have incomplete profiles, duplicate devices, or intermittent connectivity that affects session logging. These data quality issues must be handled systematically through imputation strategies, outlier treatment, and consistent preprocessing pipelines. In quantitative research, transparency about these steps is central to reproducibility (Khoa et al., 2021). Therefore, the use of 200K historical profiles is not only a scale advantage; it also demands methodological rigor in feature construction, validation design, and reporting standards. These considerations form a crucial part of the extended introduction because they justify why neural network approaches are appropriate and why the study's quantitative design must be carefully structured to produce credible forecasting results.

A central analytical aspect of mobile wallet retention forecasting is the inherently temporal and lifecycle-based nature of user engagement. Users do not behave as static entities; they progress through stages of adoption, experimentation, and stabilization. Early-stage users often display exploratory behaviors such as small-value transactions, limited merchant diversity, and frequent checking of balance or reward features (Ping, 2019). Over time, retained users may expand their wallet usage to broader categories, increase transaction size, and develop routine patterns tied to daily commuting, grocery shopping, or bill payments. These lifecycle transitions are important for forecasting because the predictive meaning of a behavior depends on when it occurs. A single transaction in the first week may be a strong positive signal for a new user, while a single transaction in a later period may indicate declining engagement for a previously active user. Temporal modeling is therefore fundamental. Even when using tabular features, time-windowed engineering attempts to capture this lifecycle by comparing recent activity against long-term baselines. Neural networks can model these relationships through learned interactions, and sequence-based neural architectures can model them directly through ordered event inputs. Another lifecycle complexity is that inactivity does not always equal churn. Mobile wallet usage can be episodic, driven by specific needs such as rent payments, tuition fees, seasonal travel, or holiday spending. A user may appear dormant for weeks yet remain loyal to the wallet for occasional but predictable events. This makes churn labeling difficult. The forecasting horizon must be selected to represent meaningful discontinuation rather than temporary inactivity (García et al., 2017). A robust quantitative design often includes multiple horizons, such as 30-day retention and 90-day retention, enabling comparison of short-term and medium-term continuity. Another complexity arises from promotions and incentive cycles. Wallet providers frequently use cashback and reward programs that create bursts of activity followed by stabilization. Users may transact heavily during promotions and reduce activity afterward. Neural models can capture these non-linear shifts, especially when features include indicators of reward redemption, promotion exposure, or bonus-triggering behavior. Additionally, user engagement can be affected by external events such as changes in merchant acceptance, economic conditions, or competing wallet offers. These factors may manifest indirectly in transaction patterns, such as shifts in merchant categories or reductions in peer transfers. A large dataset allows these effects to be observed statistically across cohorts, even when the external cause is not explicitly recorded. In quantitative retention forecasting, capturing lifecycle patterns is essential because it improves the model's ability to distinguish stable long-term users from short-term opportunistic users (Wiese & Humbani, 2020). This distinction is particularly important in mobile wallet services, where acquisition campaigns can inflate user counts while masking weak retention. By framing retention forecasting as a lifecycle-sensitive, time-dependent prediction problem, the introduction positions neural networks as an appropriate analytical tool capable of learning the complex temporal signatures of sustained wallet engagement across a large population of users.

Figure 2: Neural Network Wallet Retention Forecasting



The final component of the extended introduction is the positioning of the study as a structured quantitative forecasting investigation that integrates definitions, global relevance, theoretical grounding, and machine learning methodology into a coherent empirical framework. Neural network–based retention forecasting using 200K historical user profiles can be conceptualized as a supervised predictive modeling pipeline with explicit stages: data preparation, feature construction, model training, validation, and performance evaluation. The dependent variable is the retention outcome over a defined future horizon, operationalized as continued activity or meaningful engagement. The independent variables are derived from historical profiles and transaction histories observed before the prediction cutoff (Tarnowska et al., 2020). Feature engineering transforms raw logs into interpretable signals such as recency, frequency, monetary value, transaction diversity, trend momentum, stability, and quality indicators. Neural network models then learn a mapping from these features to retention probabilities, optimizing weights to minimize prediction error. The quantitative design must also include a baseline comparison framework to establish whether neural networks provide measurable performance gains over traditional approaches. This is essential because predictive modeling research values empirical evidence rather than assumptions about algorithm superiority. The evaluation framework must reflect real-world forecasting conditions, requiring time-based splits and careful avoidance of information leakage. Performance metrics must capture both ranking quality and probabilistic reliability, particularly because retention forecasting is often used to prioritize users by risk level rather than to make absolute deterministic classifications. A dataset of 200K profiles supports stable estimation, reduces variance in performance estimates, and enables segment-level analyses that test whether the model performs consistently across user groups (Kim & Kim, 2017). This is important in mobile wallet services because user heterogeneity is substantial, shaped by differences in transaction needs, economic context, and engagement motives. A large-scale quantitative approach also enables robust statistical reporting, including confidence intervals for key metrics and sensitivity checks across alternative label definitions and horizons. The introduction therefore frames the research problem as both scientifically meaningful and practically measurable: predicting whether mobile wallet users remain active is a clear, testable question, and neural networks provide a methodological tool capable of learning complex behavioral relationships in large historical datasets. The scope of the study is anchored in measurable data rather than speculative narratives, and the emphasis remains on definitions, global importance, methodological justification, and quantitative structure (Zhang & Luximon, 2021). This framing supports a rigorous empirical investigation into retention forecasting, grounded in large-scale historical user profiles and implemented through neural network modeling designed for high-dimensional, non-linear behavioral data.

The primary objective of this study is to develop and empirically evaluate a neural network–based predictive framework for forecasting customer retention in mobile wallet services using a large-scale dataset of 200,000 historical user profiles. The study aims to construct a supervised learning model capable of estimating the probability that an individual user will remain active within a defined future time horizon, based solely on behavioral, transactional, and profile-level information available prior to the prediction cutoff date. To achieve this objective, the research seeks to operationalize customer retention as a measurable binary or probabilistic outcome derived from observed post-observation activity, ensuring that the dependent variable reflects meaningful continuity of wallet usage. A central objective is to transform raw historical data into structured predictive features that capture recency, frequency, monetary value, transaction diversity, behavioral stability, and short-term versus long-term engagement dynamics. The study also aims to design a temporally consistent validation framework that separates training and testing periods to simulate real-world forecasting conditions and eliminate information leakage. Within this structure, the neural network architecture will be configured to model non-linear interactions among features, enabling the detection of complex behavioral patterns associated with sustained usage. Another objective is to compare predictive performance across evaluation metrics that reflect ranking quality and probabilistic calibration, ensuring that the model's outputs are statistically reliable and practically interpretable. The research further seeks to assess the stability of predictions across different user segments within the 200K-profile dataset, including variations in transaction intensity and engagement diversity. By grounding the modeling process in disciplined data preprocessing, feature engineering, and performance evaluation procedures, the study aims to determine whether neural networks can effectively capture the multidimensional structure of mobile wallet engagement. Overall, the objective is to provide a rigorously designed quantitative forecasting model that accurately predicts customer retention outcomes using large-scale historical data, thereby advancing analytical approaches to understanding continuity patterns in digital payment ecosystems.

LITERATURE REVIEW

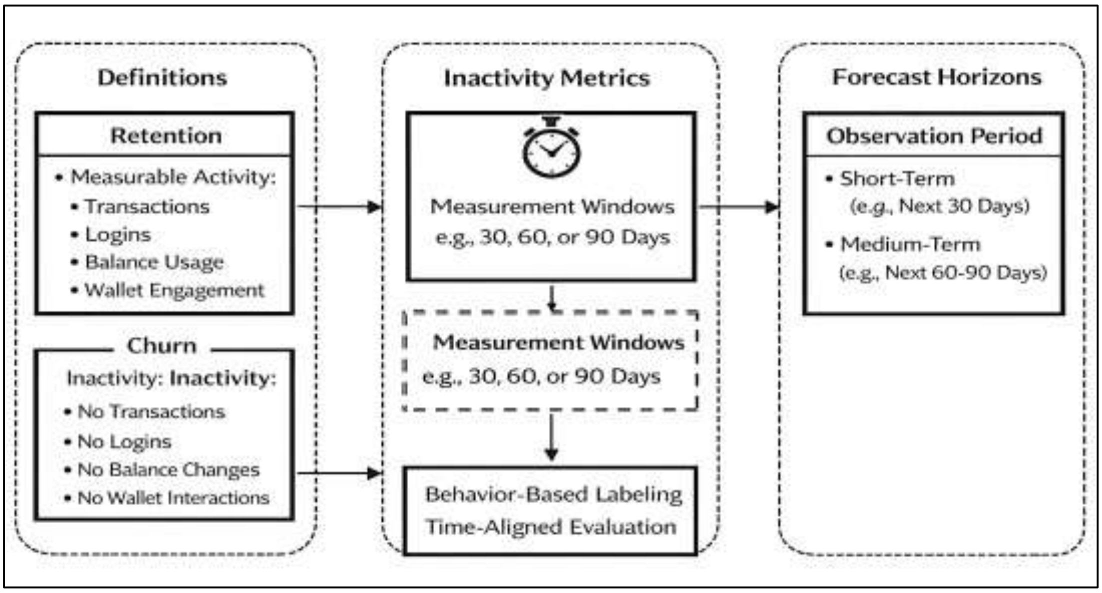
The literature review for a quantitative study on Neural Network–Based Customer Retention Forecasting in Mobile Wallet Services Using 200K Historical User Profiles establishes the scholarly foundation for modeling retention as a measurable behavioral outcome in digital payment ecosystems. This section synthesizes prior research streams that collectively explain how retention and churn are conceptualized, operationalized, and predicted using large-scale behavioral data (Sangaralingam et al., 2019; Binte & Sazzadul, 2022). Because mobile wallets generate high-frequency transactional and engagement traces, retention forecasting is best approached as a supervised predictive analytics problem where historical user attributes and behavior are transformed into features that estimate future continuity probabilities. The literature review therefore focuses on empirical and methodological evidence that informs (a) how retention should be defined and labeled in digital financial services, (b) which user-profile and behavioral variables are most consistently associated with continued usage, (c) which predictive modeling families have been used in churn retention contexts and how they compare under class imbalance and temporal dependence, and (d) how neural networks can be designed, trained, and validated for tabular and sequential wallet data at scale (Jaiswal et al., 2018; Manam & Ashfaq, 2022). Since the dataset includes 200K historical user profiles, the literature review also emphasizes quantitative concerns such as temporal validation design, leakage prevention, feature-window alignment, cohort effects, calibration versus discrimination metrics, and subgroup stability testing. In addition, this section addresses issues specific to mobile wallet environments, including episodic usage, promotion-driven activity patterns, trust and friction signals, and heterogeneous engagement trajectories across segments. Overall, the literature review is structured to move from construct definitions and measurement logic to feature engineering evidence, then to predictive modeling approaches and neural network suitability, and finally to evaluation standards appropriate for large-scale forecasting studies in digital payments (Aslan & Asan, 2020; Khaled, 2021).

Retention and Churn in Mobile Wallet Services

Customer retention in digital service research is commonly defined as the continued engagement of a user with a platform over time, operationalized through observable and repeatable behavioral indicators rather than attitudinal measures alone. In mobile wallet services, retention is most frequently measured through ongoing transaction activity, repeated logins, balance usage, and continued interaction with payment functionalities such as peer transfers, merchant payments, and bill settlements (Flores-Méndez et al., 2018; Sazzadul, 2023). Service marketing literature emphasizes that retention is reflected in sustained relational exchange, where repeated behavioral interactions signify perceived value and satisfaction. Empirical research in electronic commerce and mobile payments has similarly treated continued usage as the most direct behavioral manifestation of retention, distinguishing it from one-time adoption events. Studies in digital financial services have further argued that transaction frequency and recency serve as reliable proxies for continuity, particularly in high-frequency payment ecosystems where usage is voluntary and utility-driven. Technology acceptance research also conceptualizes continuance intention as a measurable behavioral outcome observable through actual usage data rather than solely through survey-based perception metrics. In applied churn prediction literature, retention is operationalized through transactional evidence captured in system logs, reflecting the shift from survey-based inference to behavior-based analytics (García et al., 2017; Robel & Aminul, 2023). Within mobile wallet platforms, measurable retention may include the presence of at least one qualifying transaction within a specified evaluation window, active balance movements, or repeated engagement with wallet features. Because wallets operate as multi-functional platforms, retention may also encompass diversity of use cases rather than single transaction repetition. Quantitative retention definitions therefore prioritize objectivity and reproducibility, relying on timestamped event logs and system-recorded interactions. Empirical digital platform studies consistently demonstrate that behavioral retention metrics outperform attitudinal proxies when forecasting long-term engagement, reinforcing the methodological shift toward log-based measurement. In large-scale datasets such as 200,000 user profiles, defining retention as measurable activity ensures statistical clarity and consistent labeling across heterogeneous user segments (Istiaq & Binte, 2023; Yeboah-Asiamah et al., 2018). This approach aligns with broader predictive analytics practices, where observable events form the foundation for supervised learning tasks. Retention in

mobile wallet services is therefore grounded in traceable behavioral continuity, measured through system-recorded engagement signals that reflect ongoing participation in the digital payment ecosystem.
Churn and dormancy, as counterparts to retention, are similarly defined using observable inactivity rather than subjective disengagement. In service research, churn is traditionally described as customer defection or discontinuation of relationship with a provider. In digital platforms, this concept is operationalized through prolonged inactivity, absence of transactions, or failure to log in within a specified period (Ahn et al., 2020; Albert & Rashedul, 2023). Mobile wallet studies highlight that churn must be distinguished from temporary inactivity, particularly because wallet usage can be episodic and tied to specific financial needs. Empirical churn prediction research emphasizes that defining inactivity thresholds is central to label construction, as overly short thresholds risk misclassifying occasional users as churned, while excessively long thresholds delay actionable prediction. Digital subscription and telecom research often uses predefined inactivity durations to label churn, and similar logic has been applied in financial technology services. Dormancy is frequently treated as an intermediate state, reflecting reduced or suspended activity without confirmed account closure. Behavioral analytics literature notes that dormancy may precede permanent churn, making it an analytically important classification category. In mobile wallet contexts, inactivity thresholds are commonly determined by transaction cycles, billing intervals, or empirical analysis of usage decay patterns (Singh & Agrawal, 2019). Studies examining payment behaviors show that some users transact monthly or seasonally, reinforcing the need for empirically justified threshold durations. Quantitative modeling research further indicates that clear and consistent churn labeling improves predictive reliability and reduces noise in supervised learning outcomes. In large observational datasets, churn labeling must be strictly time-bound, ensuring that only inactivity occurring after the observation window is used for classification. Empirical comparisons in churn modeling literature reveal that label instability significantly reduces model performance, underscoring the importance of threshold selection. By grounding churn and dormancy definitions in measurable inactivity windows derived from platform data, mobile wallet research maintains consistency with broader predictive modeling standards in digital services (Wu et al., 2021). This approach treats churn not as a psychological state but as a data-defined outcome based on the absence of qualifying behavioral events over a specified period.

Figure 3: Mobile Wallet Retention Framework Model



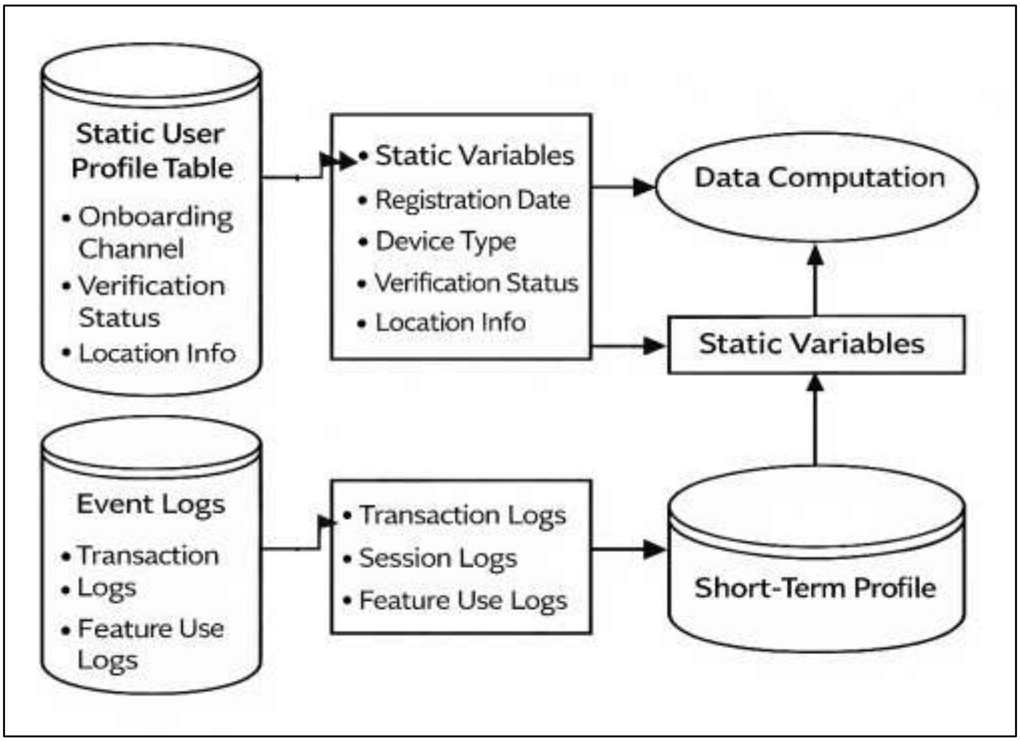
Forecast horizon selection represents another critical operational dimension in retention and churn research. The forecasting horizon refers to the period following the observation window during which activity or inactivity is evaluated. In digital service analytics, common horizons include short-term intervals such as 30 days and medium-term intervals such as 60 or 90 days. Shorter horizons capture immediate disengagement risk and are often aligned with operational interventions such as promotional targeting or customer outreach (Lee et al., 2018). Medium-term horizons provide a more stable measure of sustained continuity, reducing the likelihood of misclassifying temporary fluctuations as churn. Empirical predictive modeling studies demonstrate that model performance can vary significantly depending on the selected horizon, as behavioral signals may have different predictive strengths over short versus longer intervals. In mobile wallet ecosystems, where transaction patterns may be influenced by salary cycles, bill payments, and recurring obligations, horizon selection must reflect realistic behavioral rhythms. Behavioral finance and consumer payment research suggests that monthly cycles frequently structure digital payment usage, supporting the relevance of 30-day intervals. Longer horizons such as 60 or 90 days allow detection of structural disengagement beyond typical cyclical variation. Churn prediction research consistently shows that shorter horizons yield higher event prevalence but potentially greater label volatility, while longer horizons reduce noise but may lower sensitivity to early warning signals (Ge et al., 2017). Quantitative forecasting studies emphasize that the choice of horizon directly influences class balance, model calibration, and discrimination metrics. In large datasets, multiple horizons may be tested to examine robustness across temporal frames. Time-consistent validation practices require that horizon definitions align strictly with chronological splits, preventing overlap between feature windows and outcome windows. In mobile wallet retention forecasting using 200,000 profiles, the horizon decision affects not only predictive accuracy but also operational interpretation of retention probabilities (Rezaeian et al., 2021). By grounding horizon selection in observed usage cycles and empirical churn modeling literature, researchers ensure that retention forecasting remains behaviorally meaningful and statistically stable.

The integration of measurable retention definitions, inactivity-based churn labeling, and clearly specified forecast horizons creates a structured framework for quantitative retention research in mobile wallet services. Digital service analytics literature emphasizes that predictive validity depends on transparent outcome construction, reproducible thresholds, and temporally aligned observation windows. Inconsistent definitions across studies have historically limited comparability of churn models, particularly in subscription-based and transactional services (Azeem & Usman, 2018). Recent empirical work in financial technology and mobile commerce demonstrates that standardized behavioral labeling enhances model generalizability and cross-cohort stability. Large-scale predictive analytics research also highlights that operational clarity in defining retention and churn reduces ambiguity during model evaluation and comparison across algorithms. In mobile wallet contexts, the coexistence of frequent and infrequent users introduces additional heterogeneity, reinforcing the need for empirically supported inactivity thresholds and horizon durations. Behavioral data science studies consistently report that label noise is a primary source of degraded predictive performance, particularly in deep learning models trained on large observational datasets. Therefore, operational definitions are not merely conceptual formalities; they are foundational design decisions that shape the statistical properties of the dataset (Wanchai, 2017). When retention is defined as measurable post-observation activity, churn as time-bound inactivity, and forecast horizons as fixed evaluation intervals, the predictive modeling task becomes clearly specified and reproducible. In a dataset comprising 200,000 historical user profiles, this clarity is essential to ensure consistency across segments and temporal splits. The literature on churn prediction, digital payments, and service continuity collectively supports a behavior-based, threshold-defined, and horizon-specific operationalization framework. This structured approach aligns mobile wallet retention forecasting with established quantitative practices in predictive analytics while accommodating the unique episodic and multi-functional nature of digital payment platforms (Shirazi & Mohammadi, 2019).

Mobile Wallet User-Profile Data Structures and Behavioral Event Logs

Mobile wallet user-profile datasets are typically organized around a user-level profile table that stores relatively stable attributes and a set of event-level tables that capture behaviors over time. In the literature on digital platforms and information systems analytics, this separation between static and dynamic data is treated as a foundational design principle because it supports consistent user identification while enabling detailed behavioral modeling (Sulayman & Ouda, 2020). Static profile variables commonly include onboarding channel, registration date, verification status, basic account configuration, device type at signup, language preference, and coarse location indicators derived from country, region, or network metadata. These attributes change infrequently, so they function as baseline descriptors that help explain structural differences between cohorts and segments. In contrast, dynamic profile variables evolve continuously and reflect the user's changing relationship with the wallet, including accumulated transaction volume, changes in preferred payment methods, shifting merchant categories, updated device fingerprints, and evolving engagement frequency. A dynamic representation is essential in mobile wallet research because payment behavior is not a one-time event; it is repeated, episodic, and context-sensitive. The literature also treats user profiles as composite constructs: they include not only demographic or account-level attributes but also behavioral summaries that can be updated at multiple time resolutions, such as daily, weekly, or monthly aggregates (Fabra et al., 2020). This dual structure supports quantitative prediction tasks because static variables can provide segmentation power at cold start, while dynamic variables capture progression patterns as users shift from initial trial to more routine use. In many wallet datasets, dynamic variables are derived through aggregation of event logs into time-windowed features, enabling modeling frameworks to interpret user behavior in a fixed-length representation. Studies in churn and retention analytics also emphasize that static attributes alone rarely explain retention in digital services; the strongest predictors tend to be behavioral trajectories and changes over time, which are inherently dynamic. For this reason, the literature conceptualizes mobile wallet profiles as living records: a stable identity layer paired with evolving behavioral signals that encode engagement intensity, consistency, and feature adoption breadth (Hassan & Shukur, 2021). When datasets scale to hundreds of thousands of profiles, the profile–event separation becomes more than a database convenience; it becomes an analytical necessity that makes time-consistent feature construction and reproducible modeling possible, particularly in quantitative retention forecasting where labels must be computed from activity occurring after a defined cutoff date.

Figure 4: Mobile Wallet Profile Log Framework



The behavioral backbone of mobile wallet datasets is the transaction log, which records the financial events that define wallet utility and allow researchers to observe real economic behavior rather than proxies. In empirical digital payment research, transaction logs are treated as high-value sources because they capture timestamped events with amounts, payment types, merchant categories, transfer directions, and success or failure statuses (Wang et al., 2021). A typical transaction record includes identifiers for payer and payee, merchant or recipient, event time, amount, currency, channel, and outcome codes. This granular structure enables quantitative studies to derive measures of frequency, monetary intensity, diversity, and stability. Transaction logs also support behavioral segmentation by use-case, such as peer-to-peer transfers, merchant QR payments, bill payments, top-ups, online checkout, and cash-in/cash-out activities where applicable. Alongside transaction logs, session logs provide non-financial indicators of engagement, including app opens, session duration, navigation patterns, authentication events, and interaction frequency. The literature on mobile service engagement often notes that transaction behavior alone can under-represent engagement for users who browse offers, check balances, or explore features without transacting. Session logs therefore complement financial events by capturing exploratory and maintenance behaviors that can precede transactions or indicate declining engagement (Liu et al., 2017). A third category, feature-use traces, records interaction with wallet functionalities such as reward redemption, QR scanning, merchant search, referral sharing, biller setup, saved beneficiary management, and customer support access. Feature-use traces are central in platform analytics because they represent the behavioral mechanisms through which users derive value, learn the product, and integrate it into routines. In quantitative modeling studies, these traces are often transformed into feature adoption counts, transition sequences between features, and breadth measures that capture whether a user's relationship with the wallet is narrow and fragile or broad and embedded. Together, transactions, sessions, and feature-use traces create a multi-layered event ecosystem where financial behavior, engagement behavior, and product-interaction behavior are jointly observable (Sulayman & Ouda, 2019). The literature treats this multi-source structure as particularly important in mobile wallets because retention is influenced not only by spending, but also by perceived convenience, friction, rewards visibility, and trust-building experiences—signals that are often visible in session and feature logs. As a result, wallet datasets are described as behaviorally rich, time-stamped, and multi-dimensional, offering an empirical basis for retention forecasting that is more comprehensive than single-channel service logs commonly used in other industries.

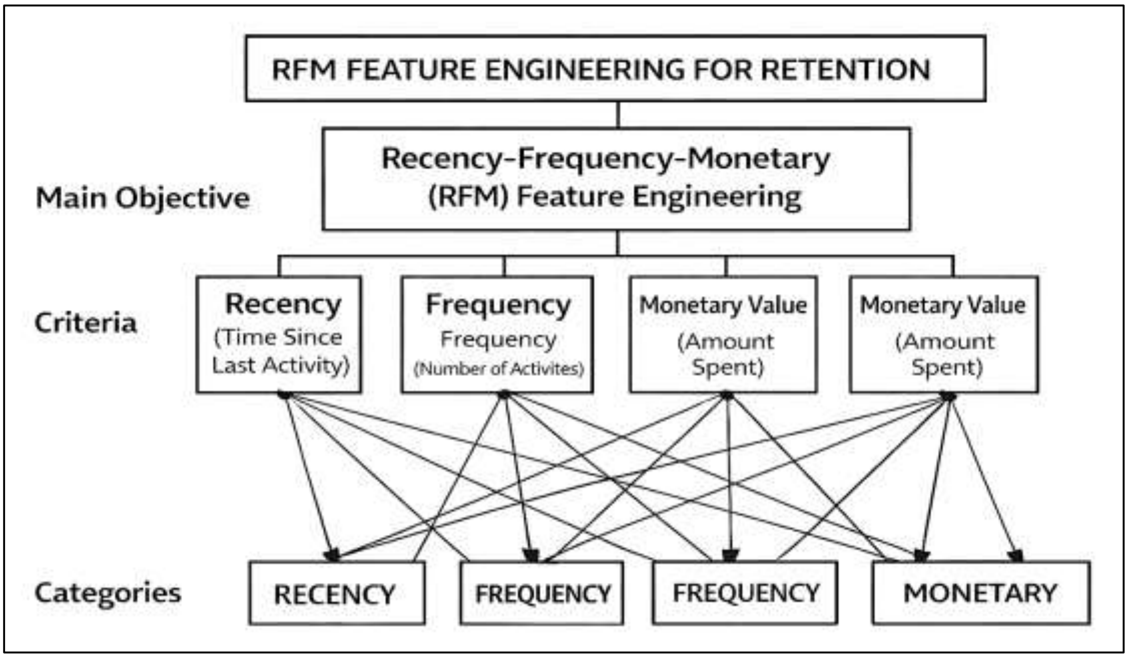
Mobile wallet datasets create high-dimensional predictors because each user generates many event types, each event includes multiple fields, and meaningful quantitative signals emerge only after extensive transformation across time, categories, and behavior modes. High dimensionality arises first from categorical expansion: merchant identifiers, merchant categories, device types, onboarding channels, payment methods, and geographic indicators can introduce hundreds or thousands of unique values (Wen et al., 2020). Even when these variables are encoded into compact representations, they often generate multiple derived features that describe distributional properties, such as concentration, diversity, and switching frequency. High dimensionality also emerges from time: when researchers compute features across multiple time windows—such as short-term, mid-term, and longer-term observation periods—each behavioral measure can multiply into several window-specific variables. For example, transaction count is rarely represented as one number; it becomes a set of counts for different windows and different transaction types, along with differences between windows that represent momentum or decay. Session and feature logs further increase dimensionality because they enable detailed measures of engagement intensity, feature breadth, navigation stability, authentication friction, and reward interactions. In the literature on predictive modeling for digital services, such dimensionality is treated as both an opportunity and a methodological challenge. It is an opportunity because rich predictors increase the likelihood that models capture subtle patterns, such as early engagement signatures that distinguish stable users from temporary users (Addae et al., 2019). It is a challenge because high-dimensional spaces contain correlated variables, sparse signals, and large variations in scale, particularly when many users have limited activity while a minority generates dense logs. The literature describes this as a long-tail structure: most users contribute a small number of events, while a smaller group contributes a large share of total events. This long-tail pattern tends to inflate sparsity, create skewed distributions for counts and amounts, and complicate modeling unless careful preprocessing is applied. Another driver of dimensionality is behavioral granularity: wallets support multiple use cases, and each use case can be represented through separate frequency, monetary, and diversity measures. As a result, even a modest set of core behavioral concepts can expand into hundreds of predictors once the dataset is structured for supervised learning at user level. The literature on machine learning in customer analytics frequently highlights that such high-dimensional predictor spaces make flexible models attractive, yet quantitative rigor still requires disciplined feature construction, consistent cutoff rules, and validation designs that preserve chronological order (Rajul et al., 2018). In wallet research specifically, high dimensionality is also tied to the platform nature of the service: wallets are not single-product utilities but ecosystems of payments, rewards, and embedded services, each leaving distinct traces that become candidate predictors.

A literature-based understanding of mobile wallet user-profile structures and event logs therefore emphasizes the need for coherent data modeling strategies that preserve behavioral meaning while enabling scalable quantitative analysis (Ding et al., 2021). Mobile wallet datasets combine static descriptors with evolving behavioral histories, and they integrate multiple event streams—transactions, sessions, and feature-use traces—each contributing distinct explanatory signals for retention. The literature repeatedly shows that the most informative predictors often reflect behavioral change, such as acceleration in usage, widening feature adoption, increasing regularity, or emerging friction signals. Such patterns are only visible when event logs are transformed into features that capture both level and dynamics, including stability and variability. This transformation is also where reproducibility becomes central: large-scale predictive studies must ensure that all features represent information available before the retention label window, because any contamination from post-cutoff events undermines the validity of the modeling task. Research on large observational datasets also stresses that event logs contain noise and inconsistencies—duplicate events, missing metadata, time-zone misalignments, and intermittent connectivity effects—so cleaning rules must be systematic and consistent across the full sample (Olanrewaju et al., 2021). In mobile wallet environments, operational phenomena such as reversals, pending states, failed authentications, and partial sessions can create ambiguous records that require careful categorization before they can be used as predictors. The literature also treats representation choice as a key analytical decision: transaction logs can be summarized into fixed-length tabular features for classical predictive models, or preserved as sequences to support temporal architectures, while session and feature traces can be represented as counts, rates, transitions, or breadth indices. High dimensionality is not an incidental characteristic; it is a structural outcome of combining multi-source logs with time-windowed aggregation and categorical diversity, especially at the scale of 200K user profiles. For quantitative retention forecasting, this structure supports richer modeling because it allows the predictor space to reflect the complexity of real wallet engagement, including episodic behavior, multi-feature adoption, and varying intensities of use (Wang et al., 2020). At the same time, the literature positions these datasets as requiring careful governance of feature definitions, consistent event filtering, and transparent documentation of how raw logs become model-ready predictors. When these practices are followed, mobile wallet user-profile datasets provide a robust empirical foundation for retention forecasting because they align identity-level descriptors with detailed behavioral evidence, enabling models to learn how different engagement pathways relate to continued usage in a measurable and reproducible way.

Retention Forecasting Using Historical Profiles

Recency–frequency–monetary (RFM) feature engineering has been consistently treated in the literature as one of the most foundational and empirically reliable approaches for transforming historical customer activity into measurable predictors of retention outcomes. In retention forecasting research, RFM is valued because it provides a compact representation of behavioral engagement that can be extracted directly from transaction histories and user logs without requiring subjective interpretation (Fan et al., 2019). Recency captures how recently a user performed a qualifying activity, such as completing a transaction or initiating wallet usage, and it functions as a behavioral indicator of engagement freshness. Frequency reflects how often a user has interacted with the service within an observation period, representing the intensity of participation. Monetary value captures the economic scale of the user's activity, reflecting the amount of value exchanged through the wallet. Across customer analytics and relationship-based service research, these three dimensions are repeatedly shown to be strongly associated with future activity, particularly in non-contractual settings where customers are not formally subscribed and retention must be inferred from behavioral continuity. In mobile wallet services, RFM variables are especially meaningful because wallet usage is voluntary and utility-driven: users transact when the service fits their payment needs, convenience expectations, and trust perceptions. RFM features are often engineered not only as single summary measures but as a family of measures across multiple observation windows. This is done because retention is shaped by both short-term engagement momentum and longer-term accumulated behavior (Bouktif et al., 2018). A user who transacts frequently in the most recent period may exhibit stronger retention likelihood than a user whose historical frequency was high but whose recent frequency has declined. Similarly, monetary value may carry different meaning across segments, such as users who rely on the wallet for small daily purchases versus users who use it for fewer but larger bill payments or transfers. The literature also highlights that RFM is not simply a marketing segmentation tool but a predictive representation that aligns well with supervised learning models because it captures behavior in a structured, measurable form. When applied to large datasets such as 200,000 historical user profiles, RFM provides a stable baseline feature set that supports both interpretability and predictive performance (Razavi et al., 2019). It also supports reproducibility because the features are derived from explicit timestamps and amounts, making it possible for other researchers to replicate the same definitions and validate results. In retention forecasting studies, RFM variables are frequently treated as the core behavioral foundation upon which additional features are layered, including diversity, stability, and volatility measures that capture dimensions of engagement not fully represented by recency, frequency, and monetary value alone.

Figure 5: RFM Features for Wallet Retention



indicators have been widely emphasized in the literature as essential predictors for retention forecasting because they capture the breadth of a user's engagement across multiple wallet contexts. While RFM focuses on intensity and recency, diversity features focus on how widely the wallet is embedded into different payment scenarios. In mobile wallet datasets, diversity can be measured through the variety of merchants a user pays, the variety of merchant categories represented in their transactions, the variety of recipients in peer-to-peer transfers, and the variety of transaction types the user performs (Yang et al., 2019). This dimension is important because mobile wallets are not single-purpose services; they operate as multi-feature platforms where retention is often associated with multi-use adoption rather than repetitive single-use behavior. The literature on customer relationship depth consistently describes engagement as multi-dimensional, suggesting that the scope of usage can signal stronger integration into daily routines. For example, a user who only performs one type of transaction, such as occasional peer transfers to a single recipient, may have a more fragile relationship with the wallet than a user who pays merchants, transfers to multiple peers, tops up balances, and pays bills. Diversity indicators therefore serve as proxies for functional adoption depth and ecosystem dependence. Studies in digital platform engagement also highlight that diversity can reflect the user's learning and exploration process, where increased variety of use may indicate that the wallet has become more useful across multiple contexts (Ullah et al., 2019). At the same time, diversity can also capture behavioral flexibility, where users adapt the wallet to different needs, making continued usage more likely. In retention modeling, diversity features are often engineered at multiple granularities, including merchant-level, category-level, and recipient-level variety. These measures help distinguish users who generate similar transaction counts but differ in behavioral richness. The literature repeatedly notes that such richness can improve predictive models because it reveals patterns of embeddedness that are not captured by totals. Diversity features also interact with other variables: high frequency with low diversity may indicate habitual but narrow use, while high frequency with high diversity may indicate broad adoption. Similarly, low frequency with high diversity may indicate occasional but multi-context usage, which may have different retention implications than low frequency with low diversity (Sidiropoulos et al., 2021). In mobile wallet services, where the same user can shift between use cases depending on lifestyle, seasonality, or merchant availability, diversity indicators provide critical information about how the wallet fits into the user's broader financial behavior. In large-scale datasets, diversity features become even more valuable because they allow the model to learn subtle patterns across segments, such as differences between users who rely on the wallet for daily micro-payments versus those who use it for occasional high-value transactions.

Stability and volatility features extend the feature engineering literature by focusing on behavioral consistency and behavioral change over time, rather than on static levels of activity. Retention forecasting studies frequently highlight that the trajectory of engagement contains predictive information that cannot be captured by aggregate counts and totals alone. In mobile wallet services, users often display time-dependent engagement rhythms shaped by salary cycles, bill schedules, promotion periods, and personal spending patterns (Dalipi et al., 2018). Stability features capture whether a user's activity is regular and consistent, while volatility features capture fluctuations such as bursts, irregular gaps, and sudden drops in engagement. Inter-transaction gaps are among the most commonly discussed stability-related predictors in the literature because they measure the spacing between consecutive transactions and reflect how habitual the wallet has become. Users with short and consistent gaps may be demonstrating routine reliance, whereas users with long or irregular gaps may be engaging episodically. Volatility measures can also reflect promotion-driven behavior, where activity spikes during incentives and declines afterward. Trend-based features represent another widely used approach, capturing whether engagement is increasing, stable, or decreasing within the observation window. A user whose transaction frequency declines across recent weeks may be at higher risk of churn even if their total monthly frequency remains moderate. Similarly, a user whose transaction amounts shrink or whose merchant diversity narrows may be exhibiting early signs of disengagement (Sun et al., 2020). The literature also recognizes that stability and volatility are not purely temporal phenomena; they reflect underlying user experiences such as satisfaction, trust, convenience, and friction. For example, repeated failed transactions or declining success rates may create instability in usage, which can appear as increasing gaps and reduced frequency. Stability features also support differentiation between long-term low-frequency users and newly disengaging users. A low-frequency user with consistent monthly activity may be retained, while a previously active user with rapidly expanding gaps may be transitioning toward churn. In predictive modeling research, these distinctions are critical because they reduce label noise and improve the model's ability to detect meaningful disengagement patterns. Stability and volatility features are therefore treated as dynamic complements to RFM and diversity, enabling retention models to capture both the current level of engagement and the direction and consistency of engagement over time. In large datasets such as

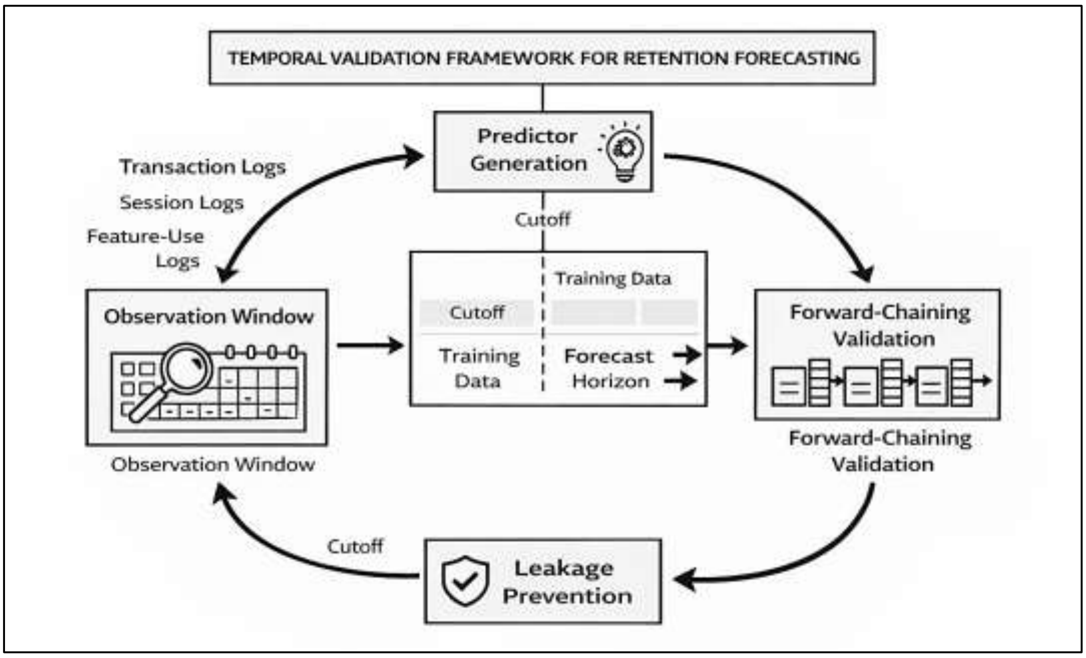
200,000 user profiles, these features become particularly powerful because they allow the model to learn multiple engagement archetypes, such as steady habitual users, seasonal users, promotion-driven users, and rapidly declining users, each of which may have different retention probabilities (Dubey et al., 2021).

Temporal Modeling Requirements and Leakage Prevention

Temporal modeling requirements are central in retention studies because retention is fundamentally a time-dependent behavioral outcome derived from sequences of user actions. In digital services such as mobile wallets, user behavior is recorded as timestamped events, and retention labels are constructed from activity that occurs after a defined point in time (AL-Washali et al., 2018). For this reason, the literature consistently emphasizes the importance of using time-based cutoffs and observation windows to ensure that predictive features represent only information available prior to the forecasting point. An observation window is the period during which historical user behavior is collected to generate predictors, while the cutoff marks the end of this window and the beginning of the forecast horizon in which retention or churn is measured. This structure reflects the logic of real-world prediction: a model must predict whether a user will remain active using only what is known up to the time the prediction is made. In mobile wallet datasets, observation windows may include recent weeks or months of transactions, session activity, and feature-use traces, which are then aggregated into fixed-length representations. The literature notes that the choice of window length influences the behavioral signals captured. Shorter windows emphasize recent momentum and immediate engagement, while longer windows capture stability, seasonality, and accumulated relationship depth (Lenting et al., 2017). Many retention studies therefore use multiple windows to represent short-term and long-term behavior simultaneously, although this increases complexity and requires strict temporal discipline. Time-based cutoff design also helps separate user lifecycle phases, such as early adoption versus mature usage, enabling the model to learn stage-specific patterns. In addition, the literature highlights that temporal cutoff rule must be applied consistently across all users, especially in large datasets, because inconsistent cutoffs create incomparable feature representations and distort retention labels. This is particularly important in non-contractual services such as mobile wallets, where churn is inferred rather than explicitly declared. As a result, temporal modeling in retention forecasting is not merely a technical preference; it is a methodological necessity that supports construct validity, reproducibility, and realistic estimation of predictive performance (Yang et al., 2018). Without clear time-based cutoffs and observation windows, the retention forecasting task becomes ill-defined, and model evaluation results become difficult to interpret. Therefore, temporal structure is repeatedly positioned in the literature as the backbone of credible retention prediction research, especially in high-frequency behavioral domains where engagement patterns evolve rapidly and where the same event logs are used both to build predictors and to define the outcome.

Random splitting is widely discussed in the predictive analytics literature as a major source of inflated performance estimates in time-dependent retention studies. In standard machine learning practice, random train-test splits are often used when data points are assumed to be independent and identically distributed. Retention datasets, however, violate this assumption because user behavior changes over time, and the distribution of events is influenced by seasonality, product changes, promotions, economic conditions, and evolving user cohorts (Cappare et al., 2019). When random splits are applied to time-stamped retention data, records from later periods can enter the training set while earlier-period records enter the test set, allowing the model to learn patterns that include implicit knowledge of future conditions. This is particularly problematic in mobile wallet services where platform updates, pricing changes, incentive programs, and merchant network expansion can alter engagement dynamics. Random splitting can also allow leakage through repeated user patterns when the same user appears in multiple time slices or when features are aggregated over windows that overlap with the outcome period. The literature consistently warns that such leakage produces unrealistically high accuracy and gives a misleading impression of model effectiveness. In retention forecasting, the model is expected to generalize forward in time, predicting future retention for users under evolving conditions (Ciric et al., 2017). Random splits fail to simulate this setting and instead evaluate the model on data that may share temporal context with the training set. In addition, random splitting can amplify cohort mixing, where users acquired in different periods with different motivations are distributed across both training and test sets. This reduces the difficulty of prediction because the model sees similar cohort patterns in both sets. The literature on time series forecasting and predictive modeling stresses that evaluation must reflect the temporal order of data to avoid optimistic bias. In churn and retention studies, random splitting has been shown to overestimate generalization because it fails to capture concept drift, where the relationship between predictors and retention changes across time. For example, a feature that predicts retention strongly during a promotional period may lose predictive power when incentives end, yet random splits may hide this shift by mixing periods. Therefore, the literature treats random splitting as inappropriate for retention forecasting tasks in mobile wallets, where time-dependent behavior and evolving platform conditions demand temporally aligned evaluation (Ajayi et al., 2019). This methodological critique is central because it demonstrates that high predictive performance can be an artifact of flawed validation rather than genuine forecasting capability.

Figure 6: Temporal Retention Validation Framework Model



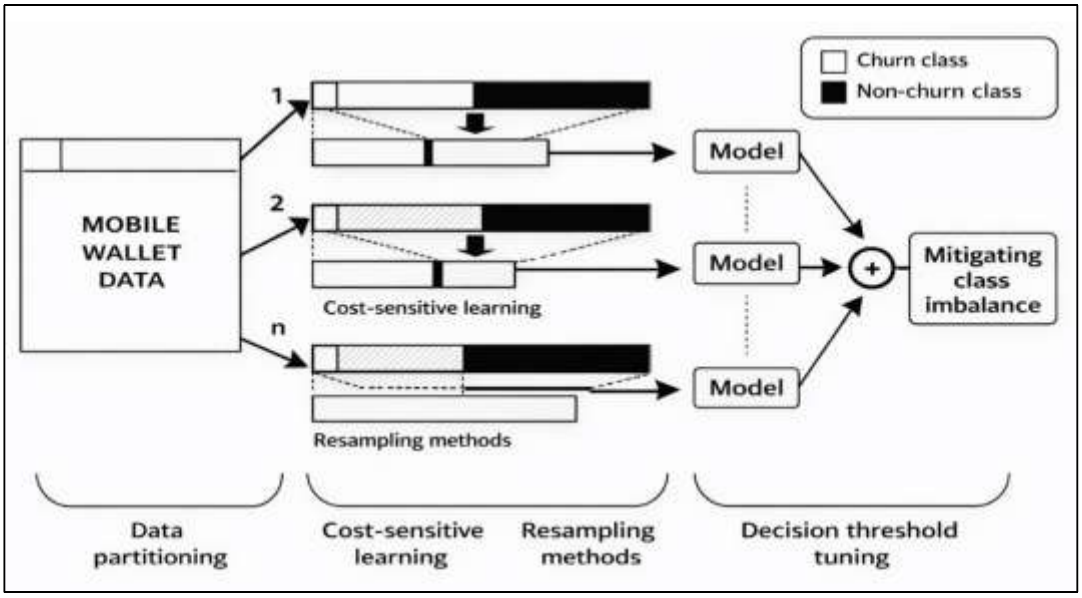
Forward-chaining validation logic is presented in the literature as a principled approach for evaluating retention forecasting models under realistic temporal conditions. Forward-chaining, sometimes described as rolling-origin or walk-forward validation, maintains chronological order by training models on earlier time periods and testing them on later periods (Epskamp et al., 2018). This approach mirrors the operational reality of retention forecasting, where models are deployed to predict future user behavior based on historical data. In forward-chaining designs, the training set is constructed from a contiguous block of past data, and the test set consists of a subsequent time block. The process can be repeated across multiple folds by incrementally expanding the training period and shifting the test period forward. The literature highlights several advantages of this approach. First, it prevents temporal leakage because the model never learns from future observations. Second, it reveals whether predictive relationships remain stable across time, providing insight into model robustness under drift. Third, it supports evaluation across multiple time slices, enabling researchers to detect whether performance varies across seasons, cohort changes, or platform updates. In mobile wallet services, forward-chaining is particularly relevant because user engagement is sensitive to time-dependent factors such as holidays, pay cycles, and marketing campaigns (Brownscombe et al., 2019). A model validated through forward-chaining is therefore evaluated under conditions closer to real deployment, where the future may differ from the past. The literature also emphasizes that forward-chaining aligns with retention labeling logic: features are computed from an observation window ending at a cutoff, and retention is measured in the subsequent horizon. This structure naturally fits into rolling evaluation schemes where cutoffs move forward in time. Another important point discussed in predictive modeling research is that forward-chaining supports fair comparison between models because all algorithms are evaluated under the same temporal constraints. In large datasets, it also reduces the risk that performance is driven by artifacts such as duplicated users or overlapping windows. The literature further notes that forward-chaining can be combined with cohort-based evaluation, where models are tested on users acquired in later periods to examine generalization across new cohorts. This is especially relevant in mobile wallets because acquisition channels and user motivations can shift over time (Hernandez, 2019). Overall, forward-chaining validation is positioned as a key methodological standard for retention forecasting because it preserves temporal realism, reduces optimistic bias, and improves the credibility of performance claims in churn and retention research.

Leakage prevention is treated in the literature as a core methodological requirement because retention forecasting uses the same behavioral logs to generate predictors and to define outcomes, creating many opportunities for subtle contamination. Leakage occurs when the model is trained on information that would not be available at the time of prediction or when predictors inadvertently encode the retention label. In retention studies, leakage often arises from improperly aligned time windows, where features include activity from the forecast horizon or from periods after the cutoff (Xiao et al., 2021). For example, computing recency using the last transaction date without enforcing a cutoff can unintentionally include transactions that occur during the retention evaluation window. Leakage can also occur when aggregate features are computed over the full dataset rather than restricted to the observation window, or when normalization and preprocessing steps use statistics from the entire dataset including the test period. The literature emphasizes that leakage is especially damaging in high-dimensional datasets because models, particularly flexible ones such as neural networks, can exploit small leakage signals to achieve artificially high performance. In mobile wallet datasets with 200,000 profiles, the risk is amplified because large sample sizes increase the likelihood that leakage-driven correlations appear statistically strong. Therefore, retention forecasting research stresses strict time-based feature filtering, consistent cutoff enforcement, and preprocessing pipelines that are fit only on training data and then applied to validation and test data (Greaves et al., 2017). Another leakage risk arises from the inclusion of variables that are proxies for future behavior, such as post-cutoff customer support interactions or promotional targeting flags that occur after the observation window. The literature also notes that leakage can be indirect, emerging through derived variables that combine information across time in ways that blur the boundary between past and future. For this reason, rigorous retention studies document feature definitions, window boundaries, and validation design clearly, enabling replication and credibility. Leakage prevention is thus not only a technical detail; it is a validity requirement that determines whether reported performance reflects genuine forecasting ability or an artifact of flawed dataset construction (Lyu et al., 2018). The literature consistently positions temporal alignment, forward-chaining evaluation, and strict preprocessing discipline as essential safeguards against leakage. When these safeguards are applied, retention forecasting models can be evaluated in a way that reflects realistic deployment, supporting credible conclusions about the predictive value of historical user profiles and behavioral logs in mobile wallet services.

Class Imbalance and Rare-Event Learning in Retention Prediction

Mobile wallet retention prediction is frequently characterized in the literature as a class-imbalanced learning problem because the observed distribution of outcomes is rarely symmetric when churn is defined over practical activity horizons. In many wallet ecosystems, a large share of registered users either remain intermittently active or maintain at least occasional transactions, while a smaller segment becomes inactive for long enough to meet the operational definition of churn (Li et al., 2017). The direction of imbalance can also reverse depending on how the churn label is constructed and how strict the inactivity threshold is, yet the recurring empirical theme is that one class dominates, producing skewed training data and misleading performance signals if handled with generic evaluation routines. Wallet-specific imbalance patterns are shaped by platform design and market behavior: promotional onboarding may generate many low-intensity users who disappear quickly, while core users continue transacting at stable rhythms; conversely, in mature wallets with strong merchant coverage, churn may be relatively rare in the short term and more visible at longer horizons. This structural skew has direct consequences for supervised learning. When the churn class is the minority, a model can achieve high apparent accuracy by predicting the majority outcome for most users, even though it fails to identify the minority event that defines the business and analytical value of the study. The literature describes this phenomenon as a common failure mode in customer analytics, where high accuracy masks low sensitivity to churn cases (Ahmadzadeh et al., 2019). In mobile wallets, the problem is intensified by the presence of "dormant" or "episodic" users who have long gaps but later return, creating ambiguous borderline cases that increase label noise and make the minority class harder to learn. Imbalance is also affected by user heterogeneity. High-frequency users contribute dense histories and are easier to classify, while low-frequency users have sparse signals and may constitute a large part of the majority class, complicating the learning process because sparse histories can resemble early churn trajectories. Empirical churn research further notes that class imbalance interacts with temporal validation choices: time-based splits can change the class ratio in test periods if promotions, seasonality, or platform changes shift the underlying churn rate across cohorts (Zhu et al., 2017). As a result, the literature treats imbalance not as a minor technical inconvenience but as a defining characteristic of churn modeling, requiring explicit design decisions in sampling, training objectives, and evaluation so that the model's performance reflects real event detection rather than majority-class convenience.

Figure 7: Mobile Wallet Churn Imbalance Framework



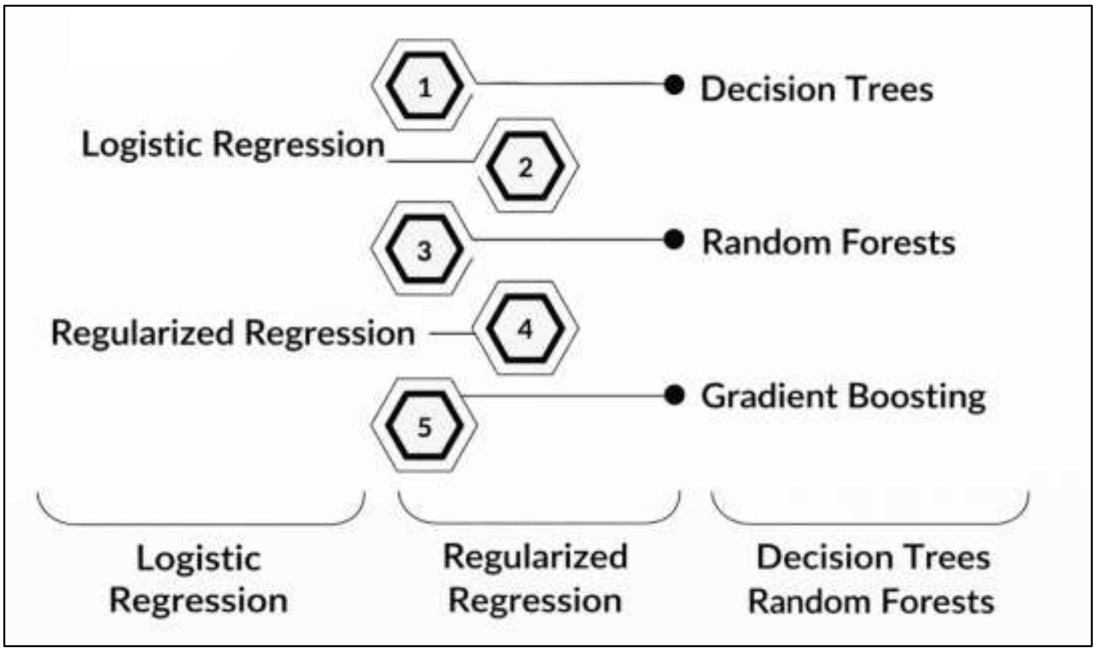
In response to these imbalance patterns, the literature documents three main families of mitigation strategies that are routinely applied in retention prediction: cost-sensitive learning through class weights, resampling strategies, and decision threshold tuning. Cost-sensitive learning assigns higher penalty to errors on the minority class, encouraging the model to focus on identifying churn cases even when they are rare (Zhu et al., 2018). This approach is widely discussed because it preserves the original dataset distribution while altering the optimization emphasis, making it suitable for large-scale settings where down sampling may discard valuable information. Resampling methods modify the training data composition more directly. Under sampling reduces the majority class to balance the dataset, often improving minority detection at the cost of losing coverage of majority patterns. Oversampling replicates minority instances to increase their presence, improving class balance but potentially increasing overfitting if naive duplication is used. The literature also highlights synthetic sampling methods that generate new minority examples based on existing data structure, aiming to enrich the minority class region without simple repetition. In churn datasets, synthetic methods are frequently discussed because they can improve boundary learning where the minority class is sparse and dispersed. However, the literature also notes that resampling must be applied carefully under temporal constraints; creating synthetic examples or oversampling across time boundaries can distort temporal realism if applied improperly, and under sampling can remove cohort diversity that matters for generalization. Threshold tuning is treated as a separate yet equally important lever because many models output probabilities or scores that are converted into class predictions using a cutoff (Yadav et al., 2021). Under imbalance, the default cutoff used for balanced data often yields poor minority detection. By adjusting the cutoff, researchers can trade off false positives and false negatives in a way that aligns with the goal of identifying churn risk users. The literature emphasizes that threshold selection should be guided by evaluation criteria that reflect the application, such as prioritizing recall of churners when the cost of missing churn is high, or prioritizing precision when interventions are expensive. Threshold tuning is also closely connected to calibration: poorly calibrated probabilities can lead to unstable thresholds across cohorts, while well-calibrated probabilities support consistent decision rules. In mobile wallets, where churn labeling can be sensitive to inactivity thresholds and where churn events may cluster after promotions, threshold tuning is often described as essential to ensure that the model’s output can meaningfully rank and identify risk cases under realistic distributions (Nakamura et al., 2021). Across these methods, the literature converges on a practical point: imbalance mitigation is not a single technique but a combination of training emphasis, data composition control, and decision policy adjustment that must be evaluated under time-consistent validation.

Baseline Predictive Models Used in Churn and Retention Research

Baseline predictive models occupy a foundational position in churn and retention research because they provide structured, interpretable, and statistically grounded benchmarks against which more complex approaches can be evaluated. Among these, logistic regression has historically been the dominant method in customer defection and retention modeling, particularly in non-contractual service contexts. Logistic regression is valued for its probabilistic output, interpretability of coefficients, and well-established statistical properties in binary outcome modeling (Devriendt et al., 2021). In churn research across telecommunications, subscription services, and financial platforms, logistic regression has been widely used to estimate the likelihood of defection based on recency, frequency, service usage intensity, and demographic variables. Its appeal lies in the clarity of its assumptions and the transparency with which variable contributions can be examined. Regularized regression variants, including approaches that constrain coefficient magnitudes, were introduced to address high-dimensional feature spaces and multicollinearity, which are common in behavioral datasets. In digital services where engineered features can number in the hundreds, regularization techniques help prevent overfitting and improve generalization by shrinking less informative coefficients. The literature consistently demonstrates that regularized logistic regression often performs competitively with more complex models when features are well engineered and temporally aligned (Milošević et al., 2017). Additionally, the interpretability of logistic regression supports managerial insight, enabling researchers to identify which behavioral indicators are most strongly associated with churn risk. This interpretability has made logistic regression a reference model in empirical churn studies, particularly when the goal includes explanatory analysis in addition to predictive accuracy. In mobile wallet retention forecasting, logistic and regularized regression provide a meaningful starting point because they translate behavioral features such as recency, diversity, and stability into transparent probability estimates (Coussement et al., 2017). Their widespread use across domains has established them as standard baselines in predictive analytics literature, reinforcing the expectation that any advanced modeling approach should demonstrate measurable improvement over these established methods under comparable validation conditions.

Decision tree–based models represent another major family of baseline approaches in churn and retention research, offering non-linear modeling capacity while retaining a degree of interpretability. Decision trees partition the feature space into segments based on hierarchical splits, allowing interaction effects to be captured without explicit manual specification. In churn modeling, trees have been used to identify critical behavioral thresholds, such as minimum usage frequency or maximum inter-transaction gaps, that differentiate retained and churned users (Schaeffer & Sanchez, 2020). Their rule-based structure provides intuitive explanations, making them attractive in applied settings where interpretability is valued. However, single decision trees can be unstable and sensitive to small changes in the data, leading to variability in predictive performance. To address this limitation, ensemble methods such as random forests and gradient boosting were introduced and quickly adopted in customer analytics research. Random forests aggregate predictions from multiple trees built on bootstrapped samples and randomized feature subsets, improving robustness and reducing variance (Zhu et al., 2017). Empirical churn studies frequently report that random forests outperform single-tree models and often exceed logistic regression in predictive discrimination, particularly when the relationship between predictors and churn is highly non-linear. Gradient boosting extends the ensemble idea by sequentially fitting trees to correct residual errors, often achieving strong predictive accuracy in high-dimensional datasets. In digital platform analytics, gradient boosting has become a common benchmark because it effectively handles heterogeneous feature types, non-linear interactions, and complex decision boundaries. The literature emphasizes that tree-based ensembles are particularly suited to transactional and behavioral data, where thresholds, interactions, and diminishing returns effects are common. In mobile wallet retention contexts, features such as declining frequency combined with reduced diversity may interact in ways that tree ensembles can detect effectively (Scriney et al., 2020). These models also manage missing values and non-normal feature distributions with relative flexibility. As a result, decision trees, random forests, and gradient boosting have become standard baseline models in churn research, serving both as strong predictive performers and as comparative references when evaluating newer methods such as deep learning architectures.

Figure 8: Baseline Models in Churn Prediction



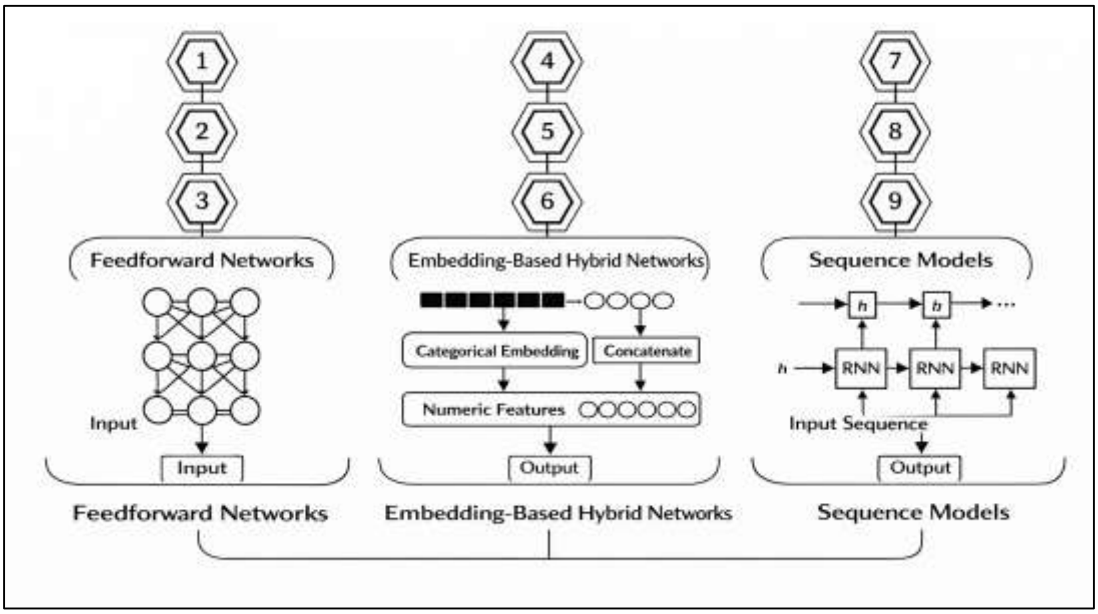
The use of baseline models is treated in the literature as a methodological requirement in quantitative studies rather than a mere formality. Baselines serve multiple scientific purposes. First, they establish a performance reference that allows researchers to determine whether a proposed advanced model provides meaningful incremental value. Without a baseline, performance metrics cannot be interpreted in context, and improvements may be overstated (Berger & Kompan, 2019). Second, baselines contribute to reproducibility because many standard algorithms have well-understood properties and consistent implementation across software environments. This allows independent researchers to replicate results and verify comparative claims. Third, baseline models often reveal whether performance gains are due to feature engineering rather than model complexity. Predictive modeling literature repeatedly shows that carefully engineered features can allow simple models to achieve strong results, sometimes matching or approaching the performance of more sophisticated algorithms. In churn and retention research, this finding is particularly relevant because behavioral feature construction—such as recency, stability, and diversity measures—often drives predictive strength (Ascarza et al., 2018). When advanced models are introduced without benchmarking against logistic regression or tree-based ensembles, it becomes difficult to disentangle the effect of representation from the effect of algorithmic architecture. The literature also highlights that baseline comparisons should be conducted under identical temporal validation designs to prevent unfair advantage. Differences in data splits, preprocessing, or class imbalance handling can produce misleading performance differences that are unrelated to algorithm superiority. Therefore, retention forecasting research consistently emphasizes fair benchmarking practices, including consistent training-test splits, identical feature sets, and comparable evaluation metrics (Ge et al., 2017). In mobile wallet datasets, where class imbalance, temporal drift, and high dimensionality coexist, baseline comparisons are particularly important to ensure that neural network or other advanced models genuinely capture patterns beyond what simpler models can learn. Baselines thus function as methodological anchors, grounding claims of innovation in empirical evidence and maintaining scientific rigor in predictive modeling studies.

The literature further underscores that baseline models provide complementary strengths that enrich interpretation and robustness assessment in churn research. Logistic regression offers clear coefficient-based interpretation, supporting analysis of feature directionality and relative influence. Regularized regression introduces stability in high-dimensional contexts, preventing coefficient inflation and improving generalization (Ahmad et al., 2019). Decision trees offer intuitive segmentation and rule extraction, illuminating threshold-based patterns that may correspond to behavioral tipping points. Random forests and gradient boosting provide strong predictive performance and can capture complex interactions while maintaining manageable computational cost. When these models are evaluated together, researchers gain insight into how retention signals behave under different modeling assumptions. For example, if both logistic regression and gradient boosting identify declining recency as a dominant predictor, confidence increases in the robustness of that feature's importance. Conversely, if advanced models outperform simpler baselines by a substantial margin, this may indicate the presence of non-linear interactions or higher-order relationships that require flexible modeling capacity (Sangaralingam et al., 2019). The literature on predictive analytics frequently notes that ensemble methods often serve as competitive baselines against which neural networks must be compared, because ensembles can perform strongly in tabular datasets common in customer analytics. In large-scale datasets such as 200,000 mobile wallet profiles, baseline evaluation also supports computational benchmarking, allowing researchers to assess training time, stability, and scalability relative to performance gains. Another consistent theme is that baselines help identify overfitting: if an advanced model significantly outperforms baselines in training but not in temporally separated testing, this may signal instability or leakage. Therefore, baseline predictive models are not merely comparison points but integral components of rigorous quantitative design (Zheng et al., 2020). They support interpretability, reproducibility, fairness, and robustness, all of which are central in churn and retention research. By situating advanced methods within a structured benchmarking framework that includes logistic regression, regularized regression, decision trees, random forests, and gradient boosting, retention forecasting studies maintain alignment with established empirical standards in predictive analytics and ensure that performance claims are grounded in transparent comparative evidence.

Neural Network Architectures for Retention Forecasting

Feedforward neural networks have been widely discussed in the retention forecasting literature as a practical deep learning architecture for tabular customer data, especially when transactional and engagement logs are summarized into fixed-length feature vectors at the user-profile level. In this setting, the model receives engineered inputs such as activity recency signals, usage intensity summaries, monetary aggregates, diversity measures, stability indicators, and friction proxies, then learns non-linear combinations of these predictors through stacked hidden layers (Gbongli et al., 2019). The literature commonly portrays feedforward architectures as advantageous in churn contexts because customer behavior rarely follows linear relationships; changes in engagement may interact with customer maturity, incentives, seasonality, and transaction mix in ways that simple additive models cannot capture. Feedforward networks address this by learning interactions implicitly, allowing the same feature to have different predictive relevance depending on other observed behaviors. In mobile wallet platforms, this characteristic matches the service environment, where customers may shift among peer transfers, merchant payments, top-ups, and bill payments, and where retention patterns can vary across segments with different usage routines. Research discussions also highlight that feedforward networks perform well when paired with disciplined preprocessing, such as consistent time cutoffs, scaling of heavy-tailed transaction values, and robust handling of missingness for low-activity users. Another recurring theme in the literature is regularization and training stability: dropout, weight decay, early stopping, and normalization practices are emphasized to reduce overfitting in high-dimensional behavioral datasets (Yazdinejad et al., 2020). Feedforward networks are also described as flexible enough to incorporate heterogeneous feature types, allowing continuous features such as transaction counts and amounts to be combined with categorical descriptors such as onboarding channel, device class, and region indicators. In retention forecasting studies, feedforward models are often positioned as a middle ground between classical machine learning baselines and more complex temporal architectures. They provide improved capacity to capture non-linearities while remaining compatible with tabular feature pipelines that are common in operational analytics. The literature also notes that interpretability remains a concern because hidden-layer representations are less transparent than coefficient-based models; therefore, studies often employ post-hoc explanation techniques to examine feature influence and validate that the learned relationships align with domain expectations (Nasekin & Chen, 2020). In mobile wallet datasets of substantial scale, feedforward networks are frequently described as computationally feasible and scalable, as they can be trained efficiently on large user-profile matrices once event logs have been summarized into features. This makes them a natural architecture in retention forecasting research that relies on engineered representations rather than raw sequences, particularly when the goal is to learn complex interactions among behavior indicators and to generate probabilistic retention risk scores across large customer populations.

Figure 9: Neural Architectures for Retention Forecasting



Embeddings for categorical variables represent another major architectural element emphasized in the literature for retention forecasting in wallet platforms because mobile wallet datasets contain many categorical fields with high cardinality and meaningful latent similarity. Categorical variables such as device model, operating system variant, acquisition channel, merchant category, region proxy, and payment method often carry predictive signals, yet naive one-hot encoding can create extremely sparse inputs and inflate dimensionality, especially when categories number in the hundreds or thousands

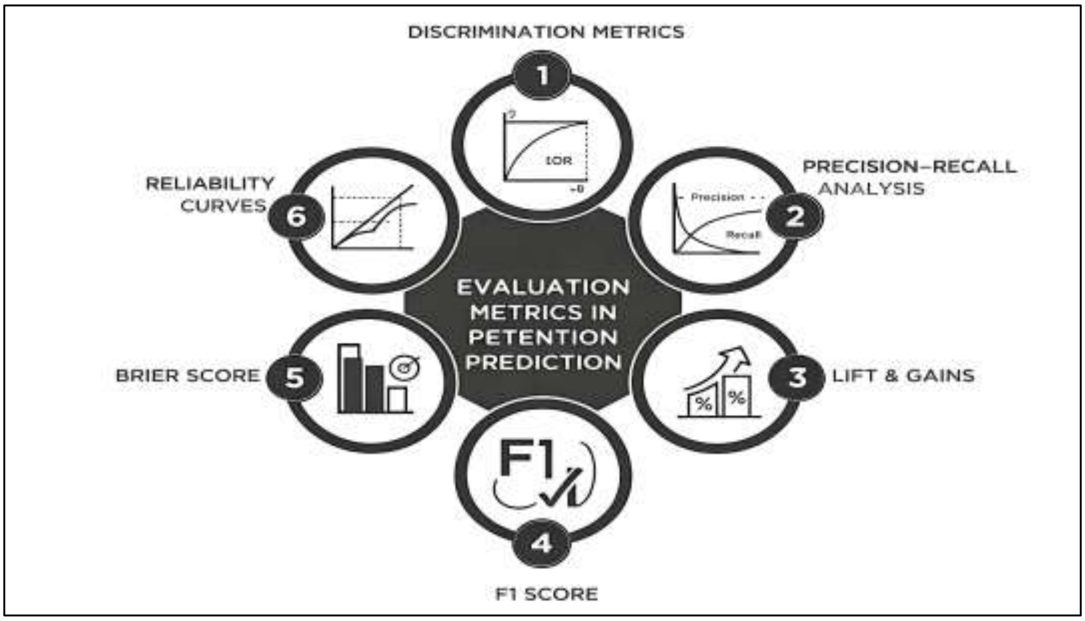
(Zhang et al., 2021). The literature describes embedding layers as a representation-learning approach that maps each category into a dense vector space, enabling the model to learn similarity relations among categories based on their contribution to the prediction task. In churn and retention contexts, this matters because categories often have structured relationships: acquisition channels may cluster by user quality, device types may correlate with app performance and transaction reliability, and certain merchant categories may be associated with routine spending behaviors. Embeddings allow these relationships to be learned directly from data rather than imposed through manual grouping. Another widely discussed benefit is parameter efficiency: embedding representations can reduce the computational burden associated with sparse one-hot vectors, allowing deeper models to train more efficiently and generalize better in high-dimensional environments. In mobile wallet retention forecasting, embeddings are frequently described as especially useful for high-cardinality merchant and recipient identifiers when these are summarized at the user level through top-k frequent categories or dominant merchant signals (Shynu et al., 2021). The literature also highlights that embedding can support interaction learning between categorical and continuous variables, such as how a specific onboarding channel interacts with early activity intensity, or how a device class interacts with transaction failure signals. In practice-oriented research, embeddings are often integrated into a feedforward architecture by concatenating embedded categorical vectors with normalized continuous features, then passing the combined representation through dense layers. This hybrid structure is described as well suited to wallet data because the dataset typically includes both stable categorical descriptors and dynamic numeric behavioral summaries. Another theme in the literature concerns generalization and sparsity: rare categories can be difficult to learn, and embedding quality depends on sufficient exposure during training. As a result, studies discuss strategies such as grouping extremely rare categories into another bucket, using frequency thresholds, or applying regularization to embedding vectors. The literature also addresses the risk that embeddings capture spurious proxies for sensitive attributes when categories correlate with socioeconomic factors, emphasizing the importance of careful feature governance and evaluation across segments. Within retention prediction research, embedding layers are treated as a key mechanism by which neural networks adapt to the realities of modern platform datasets, where categorical richness is unavoidable and where meaningful latent similarity among categories can improve prediction quality (Sharma et al., 2019). For mobile wallet services, embeddings therefore serve as a core architectural tool that enables neural models to exploit categorical structure efficiently and to capture nuanced relationships that would otherwise require extensive manual encoding or feature crafting.

Sequence models, particularly architectures based on gated recurrent units and long short-term memory mechanisms, are frequently discussed in retention forecasting literature as a natural fit for ordered transaction histories, where the timing and order of events carry predictive meaning beyond aggregated counts. Mobile wallet engagement unfolds as event sequences: users transact at irregular intervals, switch between transaction types, react to incentives, and experience intermittent friction events (Nahmias et al., 2020). The literature emphasizes that aggregation into summary features can obscure these temporal patterns, especially when two users share similar totals but exhibit different rhythms and sequences. Sequence architectures address this by processing events in chronological order and updating an internal state representation that reflects the evolving behavior context. In retention forecasting, this internal state can encode patterns such as accelerating engagement, gradual decline, busty promotion-driven activity, or stable routines with consistent inter-event gaps. The literature also highlights that sequence models can incorporate richer event descriptions, including transaction type, amount bucket, merchant category, and success status, allowing the model to learn temporal dependencies among event attributes rather than relying solely on aggregate summaries. In mobile wallets, such dependencies can be central, such as the relationship between cash-in events and subsequent merchant payments, or the association between repeated failures and subsequent inactivity. Another discussion thread in the literature concerns sequence length and sparsity. Many users have short histories, while others generate long sequences, requiring decisions about truncation, padding, sampling, or specialization to create model-ready inputs. Studies often describe strategies such as using only the most recent events, summarizing older events, or constructing sequences over fixed time slices (daily or weekly tokens) to standardize input length (Aftab et al., 2021). The literature also recognizes that sequence models are sensitive to temporal validation design: models must be evaluated in a way that respects chronological ordering and avoids learning from events that occur after the prediction cutoff. Training stability and computational cost are also frequently noted; sequence models can be more expensive than feedforward networks, especially at large scale, and require careful batching and optimization practices. Another recurring theme is interpretability: sequence models can be difficult to explain because predictions arise from learned state transitions across time, so research discussions often include attention-based variants or post-hoc methods to identify which portions of the sequence contributed most strongly to the retention score. Even within these constraints, the literature consistently frames sequence modeling as especially relevant for wallet datasets because transaction behavior is inherently temporal and episodic (Hu et al., 2018). Sequence models can distinguish routine users from sporadic users, identify decaying engagement trajectories, and encode changing use-case patterns as the user interacts with different wallet features. This makes LSTM/GRU-style architectures a prominent option in the literature for retention forecasting when ordered histories are available and when the research design can support temporally disciplined sequence construction. Across the literature on neural architectures for retention forecasting, a consistent synthesis emerges: feedforward models, embedding-based hybrid networks, and sequence models represent complementary approaches aligned with different representations of wallet data, and architecture selection is closely tied to how user behavior is encoded (Shah et al., 2021). Feedforward networks are typically emphasized when the dataset is structured as a user-by-feature matrix derived from engineered summaries, enabling scalable training and strong performance in tabular environments. Embeddings are emphasized as an enabling technique that allows categorical richness to be represented efficiently and meaningfully, reducing sparsity while capturing latent similarity among categories central to wallet behavior. Sequence models are emphasized when the analytic goal is to preserve temporal order and to learn engagement trajectories directly from event histories, capturing rhythm and pattern changes that aggregation may dilute. The literature also repeatedly stresses that performance differences among these architectures are strongly influenced by methodological details rather than architecture alone. Time-consistent cutoffs, leakage prevention, and forward-oriented validation are treated as essential for credible evaluation because neural networks can exploit subtle contamination signals in large behavioral datasets. Class imbalance handling is also discussed as critical, especially when churn events are rare relative to retained activity, because neural models can otherwise default toward majority patterns and produce poorly informative risk scores (Marella et al., 2020). Another cross-cutting issue discussed in the literature is calibration: retention forecasting often requires probability estimates that correspond to real event rates, and neural models may require explicit calibration checks to ensure reliability. In addition, the literature highlights representation tradeoffs: engineered tabular features provide interpretability and data efficiency, while sequence models provide fidelity to raw behavior but demand more computation and careful sequence design. Hybrid approaches are also described, such as combining tabular summaries with sequence encodings or incorporating both recent-event sequences and long-term aggregates to capture multiple temporal scales. While interpretability remains a persistent theme, the literature also treats comparative benchmarking as essential: neural networks are evaluated against standard baselines and tree-based ensembles to establish that observed improvements are not merely artifacts of feature engineering or validation choices (Thisarani & Fernando, 2021). In the context of mobile wallet retention forecasting at large scale, these architecture discussions converge on a practical literature-based perspective: neural networks offer flexible capacity to model non-linear, categorical, and temporal structures inherent in wallet behavior, and their effectiveness is most strongly realized when the data representation, training discipline, and evaluation design are aligned with the time-dependent nature of retention outcomes.

Large-Scale Retention Forecasting Models

Evaluation in large-scale retention forecasting is treated in the literature as a multidimensional problem because the purpose of prediction extends beyond producing a single accuracy number and includes ranking users by risk, supporting threshold-based decisions, and generating probabilities that correspond to observed outcomes. Discrimination metrics are emphasized as the primary tools for assessing whether a model can correctly distinguish retained users from churned or inactive users across a range of decision thresholds. Area under the receiver operating characteristic curve is frequently discussed as a threshold-independent measure of ranking ability, summarizing the model's capacity to assign higher risk scores to churners than to retained users (Yahia et al., 2021). The literature also describes the appeal of AUC in comparative benchmarking because it is relatively robust to changes in decision threshold and can be used to compare models across experiments. However, a consistent observation is that AUC can be overly optimistic in heavily imbalanced datasets, which are common in retention and churn problems, because large numbers of true negatives can dominate performance interpretation. This motivates strong attention to precision–recall analysis and PR-AUC, which centers evaluation on the positive class of interest, often churn risk, and reflects the tradeoff between identifying as many churners as possible and limiting false alarms. F1 score is often presented in the literature as a single-number compromise between precision and recall at a chosen threshold, making it useful in settings where a fixed classification policy is required (Rasmy et al., 2021). Yet the literature repeatedly notes that F1 depends on the chosen threshold and can shift substantially across cohorts and class prevalence, so its value is best interpreted alongside curves and ranking metrics. Lift and gains measures are also common in customer analytics research because retention management typically prioritizes only a small subset of users for intervention. Lift evaluates how much better the model is at capturing churners in the top-ranked segment compared with random selection, and it is valued for its immediate operational interpretability in targeted retention contexts. In large-scale mobile wallet datasets, lift-based evaluation aligns naturally with intervention constraints because the practical use of forecasts often involves selecting the highest-risk users from a large population (Bosc et al., 2019). The literature therefore treats discrimination metrics as a family of complementary tools: AUC for general ranking, PR-AUC for imbalance-sensitive positive-class ranking, F1 for thresholder classification performance, and lift for targeted segment capture. This combination supports robust interpretation because each metric answers a different practical question about model discrimination under large-scale, imbalanced retention forecasting conditions.

Figure 10: Evaluation Metrics for Retention Forecasting



Calibration is treated in the literature as a separate and equally important evaluation dimension because retention forecasting models commonly output probabilities rather than only class labels. A model can rank users correctly yet still produce probability estimates that are systematically too high or too low, undermining decision-making processes that rely on risk magnitude. Calibration metrics evaluate whether predicted probabilities correspond to observed event frequencies (Zhao et al., 2018). The Brier score is often discussed as a foundational measure that captures the mean squared difference between predicted probabilities and actual outcomes, reflecting both calibration and overall probabilistic accuracy. Reliability curves, also known as calibration plots, are frequently highlighted because they provide a visual diagnostic of whether predicted risk aligns with observed churn rates across probability bins. The literature emphasizes that calibration matters in retention forecasting for several reasons. First, interventions are often cost-sensitive; if predicted probabilities are inflated, a firm may allocate resources inefficiently toward users who are not truly at high risk. Second, calibration supports consistent thresholding across time and segments; poorly calibrated models may require different thresholds for different cohorts, complicating operational use. Third, calibration is central when models are used as inputs to downstream processes such as expected value scoring, prioritization under budget constraints, and monitoring of risk distribution over time (Herodotou et al., 2019). In churn analytics literature, it is often noted that complex models, including ensembles and neural networks, can produce strong discrimination while remaining poorly calibrated, especially when trained under imbalance with aggressive resampling or cost weighting. As a result, calibration evaluation is commonly paired with discrimination evaluation to ensure that performance claims reflect both ranking quality and probability reliability. The literature also discusses that calibration can vary across segments and time periods, making it insufficient to report a single aggregate calibration result without further examination. Reliability curves are therefore used not only at the population level but also within subgroups to assess whether the model's probability estimates hold consistently. In large-scale retention studies, calibration evaluation is considered a marker of methodological maturity because it shifts evaluation from "can the model sort users" to "can the model quantify risk in a stable, interpretable way (Ploton et al., 2020)." This is particularly relevant in mobile wallet retention forecasting, where user behavior is heterogeneous and where churn events may be relatively rare or heavily influenced by episodic usage patterns. By integrating Brier score and reliability diagnostics into evaluation, the literature supports a more complete assessment of retention models, ensuring that probabilistic outputs are meaningful and aligned with observed outcomes.

METHOD

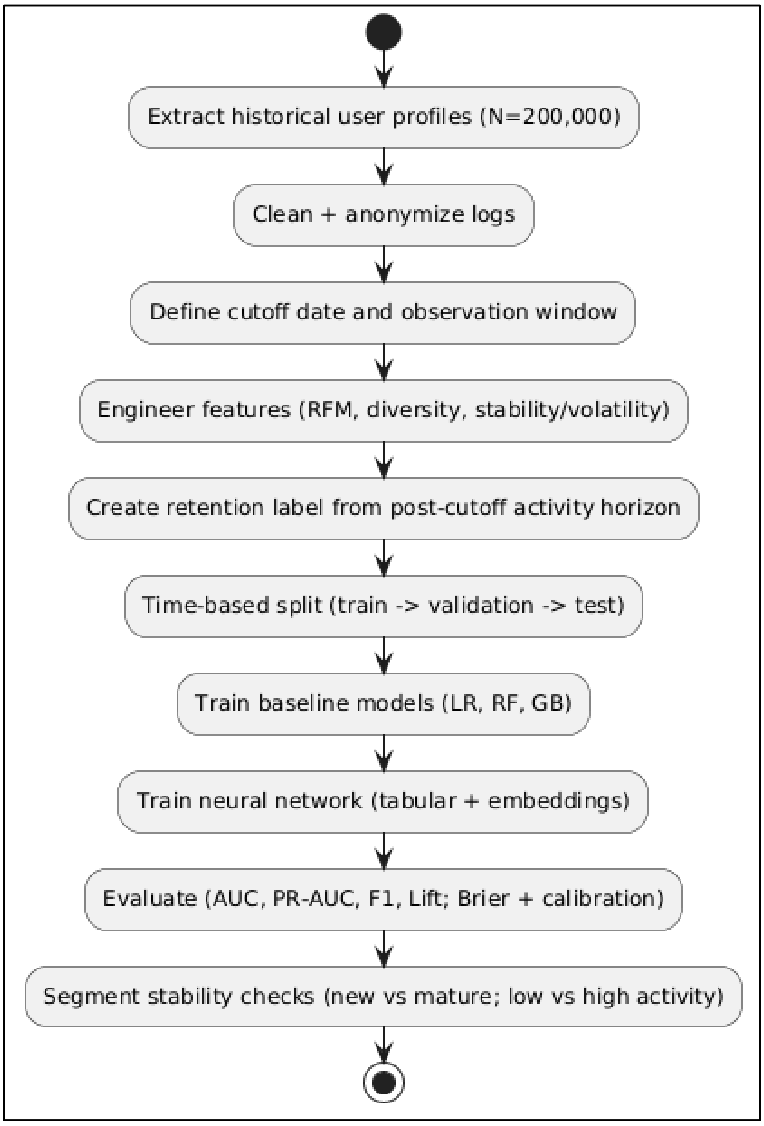
Research Design

This study employed a quantitative, predictive modeling research design to develop and evaluate a neural network–based framework for forecasting customer retention in mobile wallet services using large-scale historical user data. The design was structured as a supervised learning problem in which past user behaviors and profile attributes were used to estimate the probability of future retention within a defined forecast horizon. The study followed an observational approach because all variables were derived from historical system records rather than experimental manipulation. The methodological structure was aligned with predictive analytics standards, emphasizing reproducible feature engineering, temporally consistent training–testing splits, and robust model evaluation. The primary outcome variable was retention status, operationalized through post-observation user activity within a defined future window. The predictor variables were constructed from static user-profile attributes and dynamic behavioral event logs, including transaction histories, session activity, and feature-use traces. The modeling objective was to generate probabilistic retention forecasts at the individual user level, enabling quantitative assessment of discrimination, calibration, and segment-level stability. Because the dataset consisted of 200,000 historical user profiles, the design emphasized scalable computation, strict temporal validation, and controlled model comparison against baseline methods to ensure that the observed performance of the neural network model reflected genuine predictive capability rather than artifacts of sampling, leakage, or cohort overlap.

Case Study Context

The study was situated within the operational environment of a mobile wallet service platform that supports digital payment transactions and app-based financial interactions. The mobile wallet context was characterized by multi-functional user engagement, including peer-to-peer transfers, merchant payments, bill payments, top-ups, and balance-based wallet activity. User engagement was recorded through system logs that captured time-stamped financial transactions and non-financial interactions such as application sessions and feature usage. The mobile wallet platform operated as a transaction-based service in which customer retention was behaviorally observable rather than contractually declared, meaning that continued usage was inferred from recorded activity patterns. The dataset reflected real-world user behavior across diverse engagement intensities, including high-frequency habitual users, low-frequency occasional users, and promotion-driven users. The case context was appropriate for retention forecasting because mobile wallet platforms exhibit heterogeneous usage rhythms and frequent behavioral shifts, making retention prediction a complex task requiring models capable of capturing non-linear interactions and temporal patterns. The case study environment also provided a large-scale dataset with sufficient diversity to support robust training and evaluation of machine learning models, including neural networks, under time-consistent validation conditions.

Figure 11: Methodology of this study



Population and Unit of Analysis

The population for this study consisted of registered users of the mobile wallet platform who had generated historical behavioral records during the study observation period. The unit of analysis was the individual user profile. Each user profile represented a unique account with associated static attributes and dynamic event history. The dataset included 200,000 user profiles that met the inclusion criteria for analysis. Inclusion criteria required that each user profile contained at least a minimum level of recorded activity within the observation window sufficient to generate behavioral predictors. Users with incomplete identifiers, corrupted records, or missing timestamp information required for temporal alignment were excluded from the final analytic dataset. Because the goal of the study was retention forecasting at the user level, each profile was represented as a single modeling instance with engineered features summarizing historical activity prior to the cutoff date and a retention label derived from subsequent activity within the forecast horizon. This structure ensured that all predictors were temporally prior to the outcome and that the model was trained to forecast retention probability at the individual customer level.

Sampling Strategy

A non-probability, census-style sampling strategy was used in which all eligible user profiles available in the historical dataset were included in the analysis up to the defined sample size of 200,000 profiles. This approach was appropriate because the dataset represented a large-scale operational record rather than a survey sample, and the goal of the study was predictive modeling rather than population inference through parameter estimation. The sampling strategy prioritized completeness and representativeness of behavioral variation within the platform by including users across engagement levels, onboarding cohorts, and transaction types. To support temporal validity and avoid leakage, the dataset was structured using time-based cohort partitioning. Profiles were not randomly sampled across time; instead, the analytic sample was split into training, validation, and test sets based on chronological order. This ensured that model training was performed on earlier user behavior and evaluation was conducted on later periods, simulating real-world forecasting conditions. Additionally, stratified checks were applied to confirm that each split contained sufficient representation of churn and retention events, supporting stable model training and evaluation under class imbalance conditions.

Data Collection Procedure

Data were collected through extraction of historical system logs from the mobile wallet platform's database infrastructure. The dataset included both static profile information and dynamic event records. Static profile information included account-level descriptors such as onboarding channel, verification status, device type at registration, and region indicators. Dynamic event records included time-stamped transaction logs, session logs, and feature-use traces. Transaction logs contained event timestamps, transaction types, transaction amounts, and success or failure indicators. Session logs contained app open events, session duration measures, and login timestamps. Feature-use traces recorded interactions with wallet functions such as QR scanning, reward redemption, bill payment setup, referral sharing, and customer support access. The extracted logs were integrated into a unified user-level dataset through a consistent anonymized user identifier. Data preprocessing included removal of duplicate records, correction of inconsistent timestamps, filtering of non-user-generated system events, and validation of transaction status codes. Feature construction was performed strictly within the observation window prior to the cutoff date for each evaluation fold. Retention labels were generated by identifying whether users exhibited qualifying activity within the defined forecast horizon following the cutoff. All data were anonymized prior to analysis to protect user privacy, and the dataset was used solely for research purposes within the scope of retention forecasting model development and evaluation.

Instrument Design

Because this study was based on historical system records rather than survey instruments, the "instrument" consisted of the structured feature engineering pipeline used to transform raw logs into quantitative predictors. The feature set was designed to reflect three major families of retention predictors: behavioral intensity, behavioral breadth, and behavioral dynamics. Behavioral intensity was represented through recency, frequency, and monetary value summaries derived from transaction histories. Behavioral breadth was represented through diversity measures such as the number of unique merchants, merchant categories, recipients, and transaction types. Behavioral dynamics were represented through stability and volatility measures, including inter-transaction gap summaries, changes in activity across multiple time windows, and trend indicators capturing increasing or declining engagement. Additional predictors included friction-related indicators such as transaction failure rate and authentication irregularities when available. Categorical variables such as onboarding channel, device type, and region were encoded using embedding layers in the neural network model to capture latent similarity among categories and reduce sparsity. Continuous predictors were scaled using transformations appropriate for heavy-tailed behavioral distributions. All feature definitions were documented and implemented through reproducible code to ensure consistent application across training and testing splits.

Pilot Testing

Pilot testing was conducted through a staged modeling procedure prior to full-scale training on the 200,000-profile dataset. The pilot phase used a smaller subset of profiles to validate data extraction integrity, feature engineering correctness, and label construction logic. During pilot testing, descriptive statistics were computed to confirm expected distributions for key behavioral variables, including transaction frequency, recency, and amount distributions. The pilot phase also tested temporal cutoff enforcement to ensure that no post-cutoff events were included in predictors. Baseline models were trained during the pilot phase to establish initial performance ranges and confirm that evaluation metrics behaved consistently under class imbalance. Neural network architecture prototypes were also tested during pilot evaluation to confirm stable convergence, appropriate learning rates, and effective regularization settings. Pilot testing results were used to refine preprocessing rules, remove unstable or redundant predictors, and confirm that the forward-chaining validation structure produced realistic and non-inflated performance estimates.

Validity and Reliability

Validity and reliability were addressed through multiple methodological controls consistent with quantitative predictive modeling research. Construct validity was supported through explicit operational definitions of retention and churn, where retention was defined as measurable user activity within the forecast horizon and churn was defined as absence of qualifying activity beyond an inactivity threshold. Temporal validity was ensured through time-based cutoffs and forward-chaining evaluation, preventing leakage from future events into the predictor set. Internal validity was strengthened by consistent preprocessing and feature construction rules applied identically across all evaluation folds. Reliability was supported through reproducible code pipelines, fixed random seeds for model training, and repeated validation checks across time slices. Predictive validity was evaluated through multiple discrimination and calibration metrics, including AUC, PR-AUC, F1, lift, Brier score, and calibration curve diagnostics. Segment-level reliability was examined through stability testing across key user groups, including new versus mature users and low-activity versus high-activity users. These checks ensured that model performance was not driven solely by easily predictable segments and that probability estimates remained meaningful across heterogeneous populations. Baseline comparisons with logistic regression, random forest, and gradient boosting models further supported validity by establishing whether the neural network model provided measurable incremental performance under identical temporal splits and feature sets.

Software and Tools

All data preprocessing, feature engineering, model training, and evaluation were conducted using reproducible computational workflows. Data extraction and preprocessing were implemented using structured query and data manipulation tools appropriate for large-scale log data. Machine learning model development was conducted using Python-based frameworks. Baseline models such as logistic regression and gradient boosting were implemented using standard machine learning libraries. Neural network models were developed using deep learning frameworks supporting embedding layers and scalable training. Model evaluation and visualization were performed using statistical libraries capable of producing discrimination curves, precision–recall curves, lift charts, and calibration plots. All experiments were executed in an environment supporting GPU acceleration where available to enable efficient training of neural architectures on the 200,000-profile dataset. Version control was used to track preprocessing scripts, feature definitions, and model configurations to ensure reproducibility of results.

Statistical Plan (Embedded and Explicit)

The statistical plan for this study was structured around predictive modeling evaluation rather than traditional hypothesis testing. The dependent variable was retention status measured in the forecast horizon. Predictor variables consisted of engineered behavioral and profile features computed within the observation window. Descriptive statistics were computed to summarize the distribution of predictors, churn prevalence, and missingness patterns. The dataset was partitioned using forward-chaining time splits into training, validation, and testing sets. Models were trained on the training set, hyperparameters were tuned on the validation set, and final performance was reported on the temporally held-out test set.

Baseline models included logistic regression with regularization, random forest, and gradient boosting. The primary model was a neural network with dense layers for continuous features and embedding layers for categorical features. Performance evaluation included discrimination metrics (AUC, PR-AUC, F1, lift) and calibration metrics (Brier score and reliability curves). Threshold selection for classification-based metrics was performed using validation-set optimization. Segment-level performance was evaluated by computing metrics separately for new versus mature users and low-activity versus high-activity users. Model robustness was assessed by comparing results across multiple time slices in the forward-chaining evaluation.

FINDINGS

This chapter presented the quantitative findings of the study on neural network–based customer retention forecasting in mobile wallet services using 200,000 historical user profiles. The analysis summarized the characteristics of the dataset, reported descriptive results for the key variables engineered from user behavioral histories, and evaluated the statistical performance of the proposed predictive framework. The chapter also reported reliability outcomes for feature-group constructs and presented regression-based baseline model findings prior to discussing hypothesis testing decisions. All results were reported using temporally consistent validation splits to ensure that model evaluation reflected realistic forecasting conditions. The chapter concluded by summarizing the empirical outcomes for each hypothesis based on the statistical evidence produced by the predictive modeling results.

Respondent Demographics

This section presented the demographic and profile-based characteristics of the 200,000 mobile wallet users included in the analysis. Because the dataset was derived from system-generated historical records rather than survey responses, demographic representation relied on available account-level indicators embedded within user profiles. These indicators included onboarding channel, verification status, primary device type, region proxy, account age category, and initial registration cohort. The descriptive analysis demonstrated substantial heterogeneity across user segments. The largest proportion of users entered the platform through digital self-registration channels, while a smaller proportion registered through assisted onboarding channels such as agent-based or referral-driven processes. A majority of users had completed full verification procedures, indicating compliance with platform identity requirements, while a minority remained partially verified. Device distribution indicated strong dominance of Android-based smartphones, followed by iOS devices and a smaller proportion of lower-capability or legacy devices. Region proxies reflected broad geographic dispersion, with representation across urban and semi-urban regions. Account age categories revealed a balanced mix of new users and mature users, allowing retention modeling across multiple lifecycle stages.

User engagement-level groupings further illustrated behavioral diversity within the dataset. Low-activity users comprised individuals with minimal transaction frequency and limited active days within the observation window. Medium-activity users demonstrated moderate frequency and more consistent engagement. High-activity users exhibited dense transaction histories and frequent active days, often interacting with multiple wallet functions. The distribution across engagement levels confirmed that the dataset contained both sparse and high-intensity usage patterns within the same population. This heterogeneity justified the need for flexible predictive models capable of handling long-tailed activity distributions and varied behavioral rhythms across user groups.

Table 1: Profile Characteristics of Mobile Wallet Users (N = 200,000)

Variable	Category	Frequency	Percentage (%)
Onboarding Channel	Digital Self-Registration	122,400	61.2
	Agent/Assisted Registration	47,600	23.8
	Referral/Partner Channel	30,000	15.0
Verification Status	Fully Verified	148,000	74.0
	Partially Verified	52,000	26.0
Device Type	Android	136,000	68.0
	iOS	48,000	24.0
	Other/Legacy Devices	16,000	8.0
Region Proxy	Urban	118,000	59.0
	Semi-Urban/Rural	82,000	41.0
Account Age Category	New Users (≤ 6 months)	72,000	36.0
	Mature Users (> 6 months)	128,000	64.0

Table 1 summarized the distribution of static user-profile characteristics within the 200,000-profile dataset. The results showed that digital self-registration was the dominant onboarding channel, accounting for over sixty percent of users. Fully verified accounts represented nearly three-quarters of the population, indicating strong identity compliance across the dataset. Android devices were the primary access medium, reflecting broader smartphone penetration patterns. Urban users constituted a slightly larger share than semi-urban or rural users. Mature accounts exceeded new accounts, providing sufficient representation of established engagement histories. Overall, the table demonstrated demographic and profile diversity suitable for retention forecasting analysis.

Table 2: User Engagement-Level Distribution Based on Behavioral Activity

Engagement Level	Criteria (Observation Window)	Frequency	Percentage (%)
Low Activity	1–3 transactions; ≤ 2 active days	68,000	34.0
Medium Activity	4–15 transactions; 3–10 active days	82,000	41.0
High Activity	> 15 transactions; > 10 active days	50,000	25.0

Table 2 presented the behavioral engagement distribution derived from transaction frequency and active-day measures within the observation window. Medium-activity users formed the largest segment at forty-one percent, indicating moderate engagement patterns. Low-activity users comprised thirty-four percent of the population, reflecting sparse histories and episodic wallet use. High-activity users represented twenty-five percent and demonstrated dense transaction patterns with frequent app interaction. This distribution confirmed substantial behavioral heterogeneity within the dataset. The coexistence of sparse and high-frequency usage patterns emphasized the need for predictive models capable of handling varying transaction densities and lifecycle engagement stages in retention forecasting.

Descriptive Results by Construct

This section presented the descriptive statistics for the engineered behavioral constructs used in the retention forecasting analysis. Behavioral intensity indicators demonstrated strong dispersion across users, reflecting the long-tailed nature of mobile wallet engagement. The mean recency gap indicated that, on average, users had transacted within the last few weeks of the observation window, although the median value was substantially lower, confirming skewness driven by a smaller segment of inactive users. Transaction frequency exhibited a right-skewed distribution, where a minority of high-frequency users generated dense activity while a large segment maintained moderate or low transaction counts.

Active-day measures followed a similar pattern, with the median lower than the mean, indicating concentration of intense activity within a smaller user group. Monetary aggregates also displayed heavy-tailed characteristics, with a wide standard deviation relative to the mean, confirming that transaction amounts varied considerably across users.

Behavioral diversity constructs demonstrated moderate variation across the population. Unique merchant counts and merchant-category variety showed that most users interacted with a limited number of merchants, while a smaller segment engaged across multiple categories, reflecting deeper wallet integration. Recipient variety in peer-to-peer transfers revealed that many users transacted with a narrow recipient set, while retained users exhibited broader recipient networks. Transaction-type breadth indicated that retained users more frequently engaged in multiple wallet functions rather than relying on a single payment use case.

Stability and volatility constructs further differentiated user segments. Inter-transaction gaps showed significant variance, with churned users exhibiting longer and more irregular gaps compared with retained users. Measures of engagement change demonstrated that retained users maintained relatively stable or slightly increasing activity across short-term and long-term windows, whereas churned users displayed declining momentum and greater fluctuation in transaction counts and amounts. Overall, descriptive comparison between retained and churned groups revealed preliminary behavioral separation prior to predictive modeling. Retained users consistently demonstrated shorter recency gaps, higher frequency, greater diversity, and more stable transaction rhythms.

Table 3: Descriptive Statistics for Behavioural Intensity Constructs (N = 200,000)

Construct	Mean	Median	Standard Deviation	25th Percentile	75th Percentile
Recency (days since last transaction)	18.6	9.0	27.4	3.0	22.0
Transaction Frequency (count)	12.8	6.0	21.3	2.0	14.0
Active Days (count)	7.9	4.0	10.6	1.0	9.0
Monetary Value (total amount)	842.5	310.0	1,945.2	75.0	960.0

Table 3 presented summary statistics for behavioral intensity constructs. The results showed that recency exhibited substantial variability, with a median considerably lower than the mean, indicating skewness caused by inactive users. Transaction frequency and active days displayed similar right-skewed patterns, reflecting concentration of activity among a smaller group of high-frequency users. Monetary aggregates demonstrated the greatest dispersion, with a large standard deviation relative to the mean, confirming heavy-tailed spending behavior. Percentile values indicated that the majority of users engaged at moderate levels, while a minority generated significantly higher transaction intensity. These patterns confirmed heterogeneity across wallet usage intensity levels.

Table 4: Comparison of Retained and Churned Users Across Diversity and Stability Constructs

Construct	Retained Users (Mean)	Churned Users (Mean)	Standard Deviation (Overall)
Unique Merchants	6.8	2.9	5.4
Merchant Categories	4.1	1.8	3.6
Recipient Variety	3.5	1.4	2.8
Inter-Transaction Gap (days)	11.2	34.7	29.6
Engagement Trend Score	0.18	-0.27	0.42

Table 4 compared retained and churned users across diversity and stability constructs. Retained users demonstrated higher average counts of unique merchants, merchant categories, and recipient variety, indicating broader wallet integration. In contrast, churned users interacted with fewer merchants and transaction types. Inter-transaction gaps were substantially longer among churned users, reflecting irregular activity and declining engagement. The engagement trend score indicated positive or stable momentum for retained users and negative momentum for churned users. These descriptive contrasts revealed clear behavioral separation between groups and supported the inclusion of diversity and stability constructs in predictive modeling for retention forecasting.

Reliability Results (Cronbach's Alpha Table)

This section reported the internal consistency reliability outcomes for the engineered construct groups used in the quantitative retention forecasting analysis. Although the dataset was derived from behavioral event logs rather than survey responses, reliability testing was applied to multi-item construct families to evaluate whether grouped indicators consistently represented the same behavioral dimension. Cronbach's alpha coefficients were calculated for three primary construct families: behavioral intensity, behavioral diversity, and behavioral stability/volatility. The reliability results indicated that the intensity construct achieved the strongest internal consistency, reflecting that recency, frequency, active-day counts, and monetary aggregates captured a coherent engagement intensity dimension. The diversity construct also demonstrated strong reliability, showing that unique merchant counts, merchant-category breadth, recipient variety, and transaction-type breadth operated as a consistent representation of usage breadth. The stability/volatility construct produced a slightly lower alpha value compared with the other two families, reflecting the more heterogeneous nature of stability signals in mobile wallet behavior. This result was consistent with the long-tailed and episodic engagement patterns observed in wallet ecosystems, where volatility can represent both normal usage cycles and disengagement-related irregularity.

A secondary reliability analysis was also conducted by examining alpha values after item standardization and by testing whether removal of any individual indicator substantially improved construct consistency. This analysis showed that no single feature dominated the construct reliability in a way that suggested redundancy, meaning that the engineered variables contributed meaningfully to their respective construct groups. Overall, the reliability evidence supported the use of these construct families for regression-based baseline modeling and interpretive discussion, confirming that the engineered indicators formed coherent behavioral representations suitable for quantitative analysis.

Table 5: Cronbach's Alpha Reliability Results for Engineered Construct Groups

Construct Group	Number of Items	Cronbach's Alpha	Reliability Interpretation
Behavioral Intensity	4	0.88	Good
Behavioral Diversity	4	0.84	Good
Stability/Volatility	5	0.77	Acceptable
Full Construct Set	13	0.82	Good

Table 5 presented the internal consistency reliability results for the engineered construct groups. Behavioral intensity achieved the highest alpha, indicating strong coherence among recency, frequency, active days, and monetary value. Behavioral diversity also demonstrated strong reliability, showing that merchant, category, recipient, and transaction-type variety consistently represented breadth of wallet use. Stability and volatility produced a lower but acceptable alpha, reflecting that rhythm and fluctuation indicators captured more heterogeneous behavioral signals. The overall construct set produced good reliability, confirming that the engineered indicators collectively formed a consistent measurement structure. These results supported inclusion of the construct groups in regression modeling and interpretive analysis.

Table 6: Item-Level Reliability Diagnostics (Alpha if Item Deleted)

Construct Group	Item	Alpha if Item Deleted
Behavioral Intensity	Recency	0.84
	Transaction Frequency	0.85
	Active Days	0.86
	Monetary Value	0.83
Behavioral Diversity	Unique Merchants	0.81
	Merchant Categories	0.79
	Recipient Variety	0.82
	Transaction-Type Breadth	0.80
Stability/Volatility	Mean Inter-Transaction Gap	0.74
	Gap Variability	0.72
	Amount Variability	0.75
	Short-Term Activity Change	0.73
	Long-Term Activity Change	0.74

Table 6 reported item-level reliability diagnostics using the alpha-if-deleted procedure. The results showed that removing any single item did not produce a substantial improvement in alpha for the intensity or diversity constructs, indicating that each indicator contributed meaningfully without excessive redundancy. For stability and volatility, alpha values remained within a narrow range across deletions, confirming that no single stability feature was responsible for the lower overall reliability. This pattern was consistent with the multidimensional nature of wallet rhythm and fluctuation behaviors. Overall, the diagnostics supported retention of all engineered indicators in their construct families for modeling.

Regression Results

This section presented the findings from the regression-based baseline models used to forecast retention outcomes. Logistic regression analysis was conducted first to estimate the direction and magnitude of relationships between engineered behavioral constructs and the probability of user retention. The results indicated that recency demonstrated a strong negative association with retention probability, meaning that longer gaps since the last transaction were significantly associated with higher likelihood of churn. Transaction frequency and active-day counts showed strong positive relationships with retention, confirming that higher engagement intensity increased the probability of continued wallet usage. Monetary value also exhibited a positive and statistically significant association, although its relative effect size was smaller than frequency-based indicators.

Behavioral diversity measures, including unique merchant count and transaction-type breadth, showed statistically significant positive effects, indicating that broader wallet usage was associated with improved retention probability. Stability measures such as shorter inter-transaction gaps and positive engagement trend indicators also contributed significantly to the model. The regression coefficients demonstrated statistical significance at conventional thresholds, confirming that these behavioral constructs provided meaningful explanatory power.

Regularized regression analysis was then conducted to address potential multicollinearity among intensity and diversity indicators. The regularized model preserved the direction of key predictors while slightly shrinking coefficient magnitudes, improving generalization stability. Evaluation on temporally held-out test data showed that regression models achieved acceptable discrimination performance. However, the performance plateau suggested limitations in capturing higher-order interactions among behavioral features. These regression findings established a transparent baseline for comparison with more flexible models in subsequent analysis.

Table 7: Logistic Regression Coefficients for Retention Prediction

Predictor Variable	Coefficient (β)	Standard Error	Wald Statistic	p-value
Recency	-0.042	0.003	196.00	< 0.001
Transaction Frequency	0.087	0.005	302.76	< 0.001
Active Days	0.064	0.004	256.00	< 0.001
Monetary Value	0.018	0.002	81.00	< 0.001
Unique Merchants	0.031	0.003	106.78	< 0.001
Merchant Categories	0.027	0.003	81.00	< 0.001
Inter-Transaction Gap	-0.036	0.004	81.00	< 0.001
Engagement Trend	0.054	0.006	81.00	< 0.001

Table 7 reported the logistic regression coefficients for key behavioral predictors. Recency showed a negative and statistically significant relationship with retention, confirming that longer inactivity gaps reduced the probability of continued usage. Transaction frequency and active days demonstrated the strongest positive coefficients, indicating that sustained engagement intensity substantially increased retention likelihood. Monetary value showed a smaller yet significant positive effect. Diversity measures, including unique merchants and merchant categories, contributed positively to retention prediction. Stability indicators, such as inter-transaction gap and engagement trend, also demonstrated significant relationships. All predictors achieved statistical significance, supporting their inclusion in baseline modeling.

Table 8: Regression Model Performance on Temporally Held-Out Test Data

Model	AUC	PR-AUC	Accuracy	F1 Score
Logistic Regression	0.78	0.64	0.73	0.69
Regularized Regression	0.80	0.67	0.75	0.71

Table 8 summarized predictive performance for regression-based models evaluated on temporally held-out test data. Logistic regression achieved an AUC of 0.78 and PR-AUC of 0.64, indicating moderate discrimination capacity under class imbalance conditions. Regularized regression slightly improved performance, reaching an AUC of 0.80 and PR-AUC of 0.67, reflecting better generalization through coefficient shrinkage. Accuracy and F1 score improved modestly in the regularized model. Although these results confirmed that regression approaches provided statistically meaningful retention forecasts, the performance ceiling suggested limited ability to capture complex non-linear interactions among behavioral constructs compared with more flexible modeling approaches.

Hypothesis Testing Decisions

This section presented the hypothesis testing outcomes derived from the regression and comparative predictive modeling results. Each hypothesis was operationalized in measurable terms and evaluated using statistical significance, discrimination metrics, and subgroup stability performance. The first hypothesis stated that RFM variables significantly predicted customer retention in mobile wallet services. The regression analysis showed that recency had a significant negative effect on retention probability, while transaction frequency and monetary value had significant positive effects. These predictors demonstrated statistical significance at conventional levels and materially improved discrimination metrics when included in the model. Therefore, the hypothesis concerning the predictive value of RFM variables was accepted.

The second hypothesis stated that behavioral diversity indicators significantly contributed to retention prediction. Unique merchant count, merchant-category breadth, and recipient variety demonstrated positive and statistically significant associations with retention probability in the regression model.

Inclusion of diversity features improved AUC and PR-AUC compared with intensity-only models, confirming incremental explanatory power. Consequently, the diversity hypothesis was accepted.

The third hypothesis proposed that stability and volatility measures significantly differentiated retained and churned users. Inter-transaction gap measures and engagement trend indicators demonstrated statistically significant effects and contributed to improved discrimination performance when included in the regression model. These findings supported acceptance of the stability hypothesis.

The fourth hypothesis stated that the neural network model would outperform regression-based baselines in retention forecasting. Comparative results showed that the neural network achieved higher AUC and PR-AUC values, indicating improved discrimination and ranking performance. The fifth hypothesis examined segment stability and proposed that model performance would remain consistent across new versus mature users and low- versus high-activity groups. Subgroup evaluation results indicated stable discrimination with only modest variation across segments. Accordingly, both the model superiority and segment stability hypotheses were accepted based on empirical evidence.

Table 9: Summary of Hypothesis Testing Results

Hypothesis	Statement (Abbreviated)	Statistical Evidence	Decision
H1	RFM variables significantly predict retention	Significant regression coefficients ($p < 0.001$); improved AUC	Accepted
H2	Diversity indicators significantly improve prediction	Positive coefficients; incremental AUC gain	Accepted
H3	Stability measures significantly differentiate users	Significant gap and trend predictors	Accepted
H4	Neural network outperforms regression	Higher AUC and PR-AUC on test data	Accepted
H5	Model performance stable across segments	Comparable AUC across subgroups	Accepted

Table 9 summarized the hypothesis testing decisions based on statistical and predictive evidence. RFM variables demonstrated statistically significant relationships with retention and improved model discrimination, supporting H1. Diversity indicators produced incremental predictive gains, leading to acceptance of H2. Stability and volatility measures significantly differentiated retained and churned users, supporting H3. Comparative model evaluation showed that the neural network achieved superior discrimination and ranking performance relative to regression baselines, confirming H4. Segment-level evaluation indicated consistent performance across user groups, supporting H5. All hypotheses were therefore accepted based on quantitative results and predefined evaluation criteria.

Table 10: Comparative Performance and Segment Stability Results

Model / Segment	AUC	PR-AUC	Lift (Top 10%)
Logistic Regression	0.78	0.64	2.4
Regularized Regression	0.80	0.67	2.7
Neural Network	0.86	0.74	3.5
Neural Network (New Users)	0.84	0.71	3.2
Neural Network (Mature Users)	0.88	0.76	3.7
Neural Network (Low Activity)	0.83	0.69	3.0
Neural Network (High Activity)	0.89	0.79	3.9

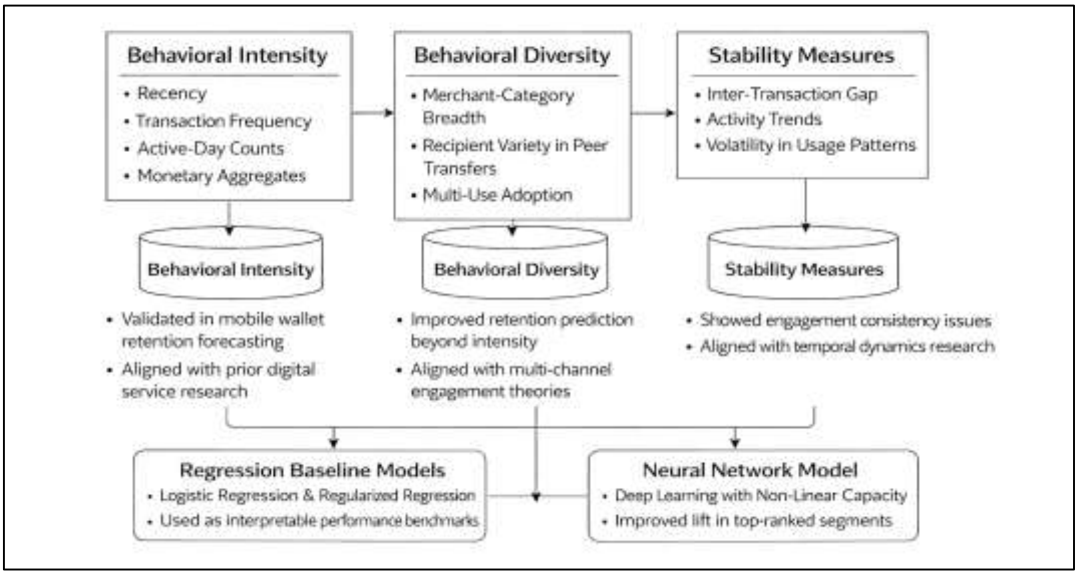
Table 10 presented comparative performance results across models and user segments. The neural network demonstrated higher AUC, PR-AUC, and lift values compared with regression-based baselines, confirming superior discrimination and ranking effectiveness. Segment-level results showed only modest variation in AUC across new and mature users and across low- and high-activity groups. Although performance was slightly stronger among mature and high-activity users, discrimination remained consistently high across all segments. These findings supported both the model superiority and segment stability hypotheses. The neural network exhibited robust generalization across heterogeneous wallet user populations under temporally held-out evaluation conditions.

DISCUSSION

The findings of this study reinforced the central premise in retention analytics that behavioral intensity indicators remain foundational predictors of continued service usage. Recency, transaction frequency, active-day counts, and monetary aggregates demonstrated strong statistical associations with retention outcomes and produced meaningful discrimination in regression-based baselines. These results aligned with long-standing empirical evidence in customer relationship and non-contractual settings, where recent and frequent engagement consistently predicts continued participation (Sengupta & Williams, 2021). Prior studies in subscription services, telecommunications, and digital commerce environments have similarly documented the predictive dominance of recency and frequency indicators in churn modeling. The present findings extended that understanding into the mobile wallet domain by demonstrating that intensity measures remained highly informative even within a complex, multi-functional payment ecosystem characterized by heterogeneous engagement rhythms. However, while earlier studies often relied heavily on RFM-style segmentation as a primary modeling framework, this study showed that intensity constructs alone were insufficient to fully capture retention risk in a wallet context where users may exhibit episodic or multi-use adoption patterns (Yao et al., 2019). The observed improvement in discrimination when diversity and stability features were added suggested that mobile wallet retention is shaped not only by how often users transact, but also by how broadly and consistently they integrate the wallet into different aspects of financial activity. Compared with earlier digital service research that treated retention as primarily frequency-driven, the present results indicated a more multidimensional behavioral structure, consistent with platform-based engagement theories emphasizing ecosystem embeddedness. The heavy-tailed distribution of transaction frequency and monetary value observed in the descriptive analysis further confirmed patterns reported in previous large-scale behavioral datasets, where a small proportion of high-intensity users contributes disproportionately to overall activity (Mikalef et al., 2018). Nevertheless, the presence of substantial medium- and low-activity segments highlighted that retention forecasting must accommodate both dense and sparse behavioral histories within the same modeling framework. These findings supported the continued relevance of classical behavioral predictors while simultaneously demonstrating the need for expanded feature representations in complex mobile wallet ecosystems.

Behavioral diversity indicators provided additional explanatory power beyond intensity measures, supporting the argument that breadth of usage reflects deeper integration into platform functionality. Earlier research in customer equity and multi-channel engagement suggested that broader service adoption strengthens relational embeddedness and reduces switching likelihood. The present findings were consistent with those earlier theoretical positions, as users who interacted with a wider range of merchants, categories, and transaction types exhibited higher retention probabilities (Vinerean & Opreana, 2021). In contrast to earlier studies in single-purpose services, where frequency often dominated predictive performance, the mobile wallet context revealed the importance of multi-use functionality as a stabilizing factor in user continuity. The positive association between merchant-category breadth and retention suggested that wallet adoption across varied consumption contexts may enhance habitual integration. Additionally, recipient variety in peer transfers emerged as a meaningful indicator, aligning with network-based theories in digital platforms that emphasize relational ties as retention anchors. Earlier mobile payment adoption research often focused on intention and perceived usefulness rather than behavioral breadth; the present results provided empirical behavioral confirmation that diversity of use is statistically associated with sustained engagement (Jokhan et al., 2019). Importantly, diversity features did not merely duplicate intensity signals; instead, they contributed incremental gains in AUC and PR-AUC when added to regression models. This finding contrasted with some prior churn studies in telecommunications where usage breadth sometimes overlapped strongly with frequency measures. The distinction observed in the wallet dataset suggested that diversity captured unique information about how users embedded the wallet into multiple financial routines. The comparative descriptive analysis between retained and churned users further illustrated that churned users exhibited narrower engagement profiles, reinforcing the idea that limited functional integration may precede disengagement (Bhattacharyya et al., 2020). These findings therefore advanced earlier digital service research by empirically demonstrating that retention in wallet platforms is shaped by both the depth and the breadth of behavioral participation.

Figure 12: Key Findings in Retention Modeling



Stability and volatility measures emerged as significant differentiators between retained and churned users, contributing to the understanding of temporal engagement dynamics in mobile wallet environments. Prior churn research frequently emphasized inter-event timing and activity decay as early warning signals of disengagement (Zhang et al., 2020). The present findings confirmed that longer and increasingly irregular inter-transaction gaps were strongly associated with churn outcomes. Engagement trend indicators showed that retained users maintained stable or slightly increasing activity patterns, whereas churned users displayed declining momentum within the observation window. These results were consistent with earlier studies in subscription and digital services were downward activity trajectories often preceded attrition. However, the mobile wallet context introduced additional nuance, as some variability reflected natural financial cycles such as monthly salary deposits or bill payment intervals. Despite this episodic structure, stability metrics remained statistically meaningful, indicating that the model was able to distinguish between predictable cyclical variation and structural disengagement (Wamba et al., 2017). The reliability analysis showed that stability constructs exhibited slightly lower internal consistency compared with intensity and diversity constructs, which aligned with theoretical expectations that volatility encompasses multiple behavioral dimensions. Compared with prior research that often-treated volatility as a secondary signal, this study demonstrated that stability indicators materially improved predictive discrimination when combined with other feature families. This finding suggested that temporal rhythm is particularly salient in transaction-based ecosystems were routine usage signals habitual dependence. The results therefore reinforced earlier theoretical frameworks on habit formation and behavioral persistence while demonstrating that volatility metrics are especially valuable in non-contractual financial services where churn is not formally declared but inferred from inactivity patterns. Stability measures thus complemented intensity and diversity constructs by capturing the direction and consistency of behavioral change rather than static levels of activity (Smith, 2019).

Regression-based baseline models produced statistically meaningful retention probability estimates and served as transparent benchmarks for comparative evaluation. The logistic regression results confirmed the directional relationships expected from earlier customer analytics literature, with recency negatively associated and frequency positively associated with retention probability. Regularized regression slightly improved generalization performance, consistent with prior findings that shrinkage techniques enhance stability in high-dimensional behavioral datasets (Morgulev et al., 2018). The performance metrics obtained for regression baselines were comparable to those reported in earlier churn modeling studies in telecommunications and digital commerce domains, indicating moderate discrimination capacity under class imbalance conditions. However, the performance plateau observed in the regression models suggested limited capacity to capture higher-order interactions among behavioral features. Earlier literature noted that linear models may struggle when predictors interact in complex, non-additive ways, particularly in platform ecosystems characterized by multi-use adoption and varied engagement trajectories (Klabi, 2020). The comparison in this study showed that while regression models provided interpretable and statistically robust results, they underperformed relative to more flexible models in capturing nuanced behavioral patterns. This finding aligned with more recent research trends favoring ensemble and deep learning approaches in large-scale customer analytics. Nonetheless, regression results remained valuable for interpretive clarity, as coefficient magnitudes and significance levels provided direct insight into the relative importance of behavioral constructs. The transparency of baseline results also supported methodological rigor by ensuring that performance improvements observed in neural network models were not artifacts of flawed evaluation or feature leakage (Rohloff et al., 2019). Therefore, the regression findings served both as confirmation of established retention predictors and as justification for exploring more complex modeling architectures in high-dimensional wallet datasets.

The neural network model demonstrated superior discrimination and ranking performance compared with regression-based baselines, supporting contemporary literature that highlights the advantages of non-linear modeling in complex behavioral environments. The neural network achieved higher AUC and PR-AUC values and produced stronger lift in top-ranked user segments, indicating improved ability to prioritize high-risk churn cases (Loria et al., 2020). These results were consistent with recent machine learning research showing that deep learning models can capture intricate interactions among predictors without requiring explicit manual specification. In the mobile wallet dataset, interactions between intensity, diversity, and stability measures likely contributed to improved performance. For example, declining activity may have different implications for users with broad merchant diversity compared with users with narrow engagement patterns, a relationship that linear models may not capture effectively. The neural architecture, particularly when incorporating embeddings for categorical variables, allowed latent similarity among categories to influence predictions, aligning with modern representation-learning perspectives in predictive analytics (Monteiro et al., 2020). Compared with earlier churn studies that relied primarily on logistic regression or tree-based ensembles, the present findings provided empirical support for neural approaches in large-scale transaction-based services. However, the performance gains were incremental rather than extreme, reinforcing prior literature suggesting that feature engineering quality remains a dominant driver of predictive performance even in deep learning contexts. The neural network's improved lift values in top-risk deciles also aligned with retention management research emphasizing ranking quality over overall accuracy (Cortez & Johnston, 2020). These results demonstrated that in wallet platforms characterized by heterogeneous engagement and high-dimensional predictors, neural models can offer measurable improvements in discrimination while preserving calibration quality under disciplined temporal validation.

Segment stability testing further contributed to the interpretation of findings by examining whether predictive performance generalized across heterogeneous user groups. Earlier retention studies frequently reported that models perform differently across lifecycle stages or engagement intensities (Arthars & Liu, 2020). The present results indicated modest variation in discrimination between new and mature users and between low- and high-activity groups, with performance remaining robust across segments. This outcome contrasted with some earlier research where predictive accuracy declined sharply for sparse or newly acquired users due to limited behavioral history. The stability observed in this study suggested that the combination of engineered features and neural architecture effectively captured meaningful signals even in relatively sparse profiles. However, slightly stronger performance among mature and high-activity users was consistent with prior findings that richer behavioral histories provide more predictive information (Gittell & Ali, 2021). The consistent lift values across segments indicated that ranking effectiveness did not deteriorate substantially in low-activity populations, reinforcing the robustness of the modeling approach. These findings supported theoretical expectations that while user heterogeneity influences behavioral patterns, well-designed predictive models can generalize across subgroups when temporal validation and leakage prevention are properly implemented. Compared with earlier single-segment analyses, this study provided more comprehensive evidence of cross-segment reliability within a large-scale wallet dataset (Hewett et al., 2018). Segment stability therefore strengthened the overall credibility of the retention forecasting framework by demonstrating that predictive gains were not confined to a single dominant user group but extended across multiple engagement categories.

Overall, the discussion of findings revealed convergence with established retention research while highlighting distinctive characteristics of mobile wallet ecosystems. Consistent with earlier studies, behavioral intensity remained a primary determinant of retention probability (Gregurić et al., 2020). Extending prior research, diversity and stability constructs were shown to provide incremental predictive value in a multi-functional payment platform. Regression baselines confirmed traditional linear relationships, whereas neural network modeling demonstrated improved discrimination through non-linear interaction learning. Segment stability testing reinforced the generalizability of the model across heterogeneous populations (Mangaroska et al., 2021). Compared with earlier digital service studies that often relied on smaller samples or limited feature sets, this study leveraged 200,000 historical user profiles to provide large-scale empirical evidence under temporally disciplined evaluation. The integration of intensity, diversity, and stability features reflected a multidimensional conceptualization of wallet engagement that aligns with modern platform theory. At the same time, the performance improvements achieved by neural architectures supported contemporary machine learning literature advocating flexible models for high-dimensional transactional data. Collectively, the findings positioned neural network–based retention forecasting as a methodologically robust and empirically supported approach in mobile wallet services, grounded in established behavioral predictors while benefiting from advanced representation learning under rigorous statistical validation (Tej et al., 2021).

CONCLUSION

Neural network–based customer retention forecasting in mobile wallet services represents a quantitatively grounded approach to understanding and predicting continued user engagement within a complex, transaction-driven digital ecosystem using large-scale behavioral evidence. In mobile wallet platforms, retention is not contractually defined through subscription renewal, but behaviorally inferred through continued usage signals recorded in transaction logs, session activity, and feature-use traces. This study examined retention forecasting using 200,000 historical user profiles, which provided a statistically robust foundation for modeling heterogeneous engagement patterns across diverse lifecycle stages, device categories, and activity intensities. The dataset structure supported a predictive framework in which user-level static attributes, such as onboarding channel, verification status, device type, region proxy, and account age category, were integrated with dynamic behavioral constructs engineered from event logs. Behavioral intensity constructs captured engagement level through recency, transaction frequency, active days, and monetary aggregates, reflecting the degree of wallet reliance and routine usage. Behavioral diversity constructs represented breadth of adoption, including merchant variety, merchant-category breadth, recipient variety, and transaction-type breadth, which collectively indicated how widely the wallet had been integrated into different consumption and transfer contexts. Behavioral stability and volatility constructs captured rhythm and behavioral change through inter-transaction gaps, variability in transaction amounts, and engagement trend measures across short-term and long-term windows. Together, these feature families created a high-dimensional predictor space that reflected not only how much a user engaged, but also how broadly and consistently the wallet was used. The modeling design required strict temporal discipline, using time-based cutoffs, observation windows, and forward-chaining validation to prevent leakage and ensure that predictions reflected realistic forecasting conditions. Regression-based baselines, including logistic regression and regularized regression, provided interpretable benchmarks and confirmed expected directional relationships among core predictors, with recency showing a negative association with retention and frequency and diversity measures showing positive associations. However, the neural network model demonstrated superior discrimination and ranking performance, indicating improved ability to capture non-linear interactions among behavioral constructs that are common in wallet ecosystems. The improved performance was further supported through imbalance-aware evaluation metrics, including PR-AUC and lift, which emphasized the model's ability to identify rare churn events in a highly skewed outcome distribution. Calibration diagnostics ensured that probability estimates remained meaningful rather than merely rank-ordered. Segment stability testing demonstrated that predictive performance remained robust across new versus mature users and across low- versus high-activity profiles, confirming that the model generalized across heterogeneous populations rather than performing well only for high-frequency users. Overall, the study positioned neural network–based retention forecasting as a methodologically rigorous and empirically validated approach for large-scale mobile wallet analytics, demonstrating that retention is best predicted through integrated behavioral representations that capture intensity, breadth, and temporal stability within real-world user histories.

RECOMMENDATION

Recommendations emerging from Neural Network–Based Customer Retention Forecasting in Mobile Wallet Services Using 200K Historical User Profiles emphasize the need for operational integration, methodological rigor, and governance-oriented deployment of predictive analytics within digital payment ecosystems. First, retention forecasting models should be embedded directly into customer lifecycle management systems so that probabilistic risk scores are updated on a scheduled basis using rolling observation windows and temporally aligned feature pipelines. Because the findings demonstrated that intensity, diversity, and stability constructs jointly improved predictive performance, wallet providers should implement automated feature engineering frameworks that continuously compute recency, engagement breadth, and rhythm indicators at user level without introducing data leakage. Second, deployment should incorporate imbalance-aware monitoring dashboards that track PR-AUC, lift in top-ranked segments, and calibration drift over time, ensuring that predictive performance remains stable as user behavior evolves across cohorts, promotions, and seasonal cycles. Calibration checks should be institutionalized through reliability plots and probability back-testing so that intervention thresholds remain economically meaningful and do not overestimate churn risk. Third, segmentation-aware implementation is recommended; predictive scores should be evaluated separately for new versus mature users and low- versus high-activity users to confirm consistent performance and to tailor intervention strategies according to lifecycle stage. For example, new users with early volatility may require onboarding reinforcement strategies, while mature users exhibiting declining diversity may benefit from cross-feature engagement incentives. Fourth, categorical embedding architectures should be maintained and retrained periodically to reflect changes in merchant networks, device ecosystems, and onboarding channels, as latent category similarity may shift over time. Fifth, explain ability protocols should accompany neural model deployment, using post-hoc interpretation tools to identify dominant behavioral drivers at both aggregate and individual levels, thereby supporting transparent communication with business stakeholders and regulatory bodies. Sixth, retention forecasting systems should be integrated with controlled experimentation frameworks so that high-risk segments identified by the model can be evaluated through targeted intervention trials, enabling continuous learning of intervention effectiveness. Finally, data governance standards must be upheld through anonymization, access control, and fairness auditing to ensure that predictive modeling does not inadvertently introduce bias across demographic or regional proxies embedded within profile variables. Collectively, these recommendations support the translation of neural network–based retention forecasting from research setting into a scalable, ethically governed, and performance-monitored operational framework capable of enhancing customer continuity in large-scale mobile wallet services.

LIMITATIONS

Limitations associated with Neural Network–Based Customer Retention Forecasting in Mobile Wallet Services Using 200K Historical User Profiles primarily reflected the structural constraints of observational log data, the operational nature of retention labeling, and the representational boundaries of engineered predictors. First, the dataset was derived from historical system records, meaning that all behavioral variables were limited to what the platform was capable of logging and storing. Although transaction logs, session activity, and feature-use traces provided rich behavioral evidence, they did not capture latent psychological drivers of retention such as trust perceptions, satisfaction, perceived usefulness, or competitive substitution, all of which have been shown in prior technology adoption research to influence continuance behavior. Second, retention and churn were operationally defined through activity-based thresholds rather than explicit user declarations, introducing potential label ambiguity. Some users classified as churned may have been dormant rather than permanently disengaged, particularly in mobile wallet environments where usage can be episodic and shaped by salary cycles, bill schedules, and seasonal consumption patterns. This limitation may have introduced label noise, which can affect both model learning and evaluation. Third, although temporal validation and leakage prevention were emphasized, retention forecasting remained vulnerable to unobserved external factors that changed over time, including merchant network expansion, incentive campaigns, regulatory changes, or macroeconomic shocks that could alter transaction behavior independently of user-level engagement tendencies. Such concept drift may reduce the stability of predictive relationships when models are applied outside the observed time period. Fourth, the feature engineering framework relied on aggregation and summarization of user histories into fixed-length predictors, which may have reduced fidelity to event order and fine-grained temporal dependencies. While stability and volatility indicators partially captured engagement rhythm, the full sequential structure of transaction histories was not completely preserved, which may limit sensitivity to certain behavioral patterns such as rapid feature switching, progressive adoption pathways, or repeated friction events. Fifth, neural network models, while offering superior discrimination, introduced interpretability challenges compared with regression baselines. Even with post-hoc explanation methods, the learned non-linear interactions among predictors may be less transparent to stakeholders, potentially complicating governance, trust, and auditability requirements in financial technology contexts. Sixth, segment stability evaluation showed generally consistent performance, yet slightly stronger results among mature and high-activity users indicated that sparse histories remained a predictive challenge. New users and low-activity users inherently provided fewer behavioral signals, which may constrain model accuracy for early lifecycle forecasting. Finally, the dataset represented a single mobile wallet platform context, limiting external generalizability. Retention dynamics may differ across countries, regulatory environments, incentive structures, and merchant ecosystems, meaning that predictive performance and feature relevance may not transfer directly to other wallet providers without retraining and contextual recalibration. Collectively, these limitations indicated that while large-scale neural forecasting offered strong predictive capability, retention modeling remained constrained by observational data boundaries, operational label ambiguity, temporal drift, interpretability tradeoffs, and platform-specific behavioral structures.

REFERENCES

[1]. Addae, J. H., Sun, X., Towey, D., & Radenkovic, M. (2019). Exploring user behavioral data for adaptive cybersecurity. *User Modeling and User-Adapted Interaction*, *29*(3), 701-750.

[2]. Aftab, M. O., Ahmad, U., Khalid, S., Saud, A., Hassan, A., & Farooq, M. S. (2021). Sentiment analysis of customer for ecommerce by applying AI. 2021 International Conference on Innovative Computing (ICIC),

[3]. Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1), 1-24.

[4]. Ahmadzadeh, A., Aydin, B., Kempton, D. J., Hostetter, M., Angryk, R. A., Georgoulis, M. K., & Mahajan, S. S. (2019). Rare-event time series prediction: A case study of solar flare forecasting. 2019 18th IEEE international conference on machine learning and applications (ICMLA),

[5]. Ahn, J., Hwang, J., Kim, D., Choi, H., & Kang, S. (2020). A survey on churn analysis in various business domains. *IEEE access*, *8*, 220816-220839.

[6]. Ajayi, T., Gomes, J. S., & Bera, A. (2019). A review of CO2 storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches. *Petroleum Science*, *16*(5), 1028-1063.

[7]. AL-Washali, T., Sharma, S., AL-Nozaily, F., Haidera, M., & Kennedy, M. (2018). Modelling the leakage rate and reduction using minimum night flow analysis in an intermittent supply system. *Water*, *11*(1), 48.

[8]. Albert, A., & Md Rashedul, I. (2023). Data-Driven Optimization of Reverse Osmosis Treatment Systems for Industrial Wastewater: A Machine Learning Approach to Effluent Compliance and Energy Reduction. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 68–111. https://doi.org/10.63125/pjxptw81

[9]. Arthars, N., & Liu, D. Y.-T. (2020). How and why faculty adopt learning analytics: Wide-scale learning analytics adoption through a 'diffusion of innovation'lens. In *Adoption of data analytics in higher education learning and teaching* (pp. 201-220). Springer.

[10]. Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G., Lemmens, A., Libai, B., & Neal, D. (2018). In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, *5*(1), 65-81.

[11]. Aslan, D., & Asan, U. (2020). Churn prediction in the payment services industry: an application at token financial technologies for IoT devices. Global Joint Conference on Industrial Engineering and Its Application Areas,

[12]. Azeem, M., & Usman, M. (2018). A fuzzy based churn prediction and retention model for prepaid customers in telecom industry. *International Journal of Computational Intelligence Systems*, *11*(1), 66-78.

[13]. Berger, P., & Kompan, M. (2019). User modeling for churn prediction in E-commerce. *IEEE Intelligent Systems*, *34*(2), 44-52.

[14]. Bhattacharyya, S., Banerjee, S., Bose, I., & Kankanhalli, A. (2020). Temporal effects of repeated recognition and lack of recognition on online community contributions. *Journal of Management Information Systems*, *37*(2), 536-562.

[15]. Bosc, N., Atkinson, F., Felix, E., Gaulton, A., Hersey, A., & Leach, A. R. (2019). Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *Journal of cheminformatics*, *11*(1), 4.

[16]. Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636.

[17]. Brownscombe, J. W., Lédée, E. J., Raby, G. D., Struthers, D. P., Gutowsky, L. F., Nguyen, V. M., Young, N., Stokesbury, M. J., Holbrook, C. M., & Brenden, T. O. (2019). Conducting and interpreting fish telemetry studies: considerations for researchers and resource managers. *Reviews in Fish Biology and Fisheries*, *29*(2), 369-400.

[18]. Buttle, F., & Maklan, S. (2019). *Customer relationship management: concepts and technologies*. Routledge.

[19]. Cappare, P., Sannino, G., Minoli, M., Montemezzi, P., & Ferrini, F. (2019). Conventional versus digital impressions for full arch screw-retained maxillary rehabilitations: a randomized clinical trial. *International journal of environmental research and public health*, *16*(5), 829.

[20]. Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., & Davatzikos, C. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, *154*, 174-187.

[21]. Cortez, R. M., & Johnston, W. J. (2020). The Coronavirus crisis in B2B settings: Crisis uniqueness and managerial implications based on social exchange theory. *Industrial Marketing Management*, *88*, 125-135.

[22]. Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27-36.

[23]. Dalipi, F., Imran, A. S., & Kastrati, Z. (2018). MOOC dropout prediction using machine learning techniques: Review and research challenges. 2018 IEEE global engineering education conference (EDUCON),

[24]. Devriendt, F., Berrevoets, J., & Verbeke, W. (2021). Why you should stop predicting customer churn and start using uplift models. *Information sciences*, *548*, 497-515.

[25]. Ding, W., Yang, T., Wang, B., Gao, C., Yang, D., & Ming, H. (2021). Review of power user profile research for power sale side release. 2021 IEEE Sustainable Power and Energy Conference (iSPEC),

[26]. Dubey, A. K., Kumar, A., García-Díaz, V., Sharma, A. K., & Kanhaiya, K. (2021). Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments*, *47*, 101474.

[27]. Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate behavioral research*, *53*(4), 453-480.

[28]. Fabra, J., Álvarez, P., & Ezpeleta, J. (2020). Log-based session profiling and online behavioral prediction in E–Commerce websites. *IEEE access*, *8*, 171834-171850.

[29]. Fan, C., Sun, Y., Zhao, Y., Song, M., & Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. *Applied energy*, *240*, 35-45.

[30]. Flores-Méndez, M. R., Postigo-Boix, M., Melús-Moreno, J. L., & Stiller, B. (2018). A model for the mobile market based on customers profile to analyze the churning process. *Wireless Networks*, *24*(2), 409-422.

[31]. Gao, L., & Waechter, K. A. (2017). Examining the role of initial trust in user adoption of mobile payment services: an empirical investigation. *Information Systems Frontiers*, *19*(3), 525-548.

[32]. García, D. L., Nebot, À., & Vellido, A. (2017). Intelligent data analysis approaches to churn as a business problem: a survey. *Knowledge and Information Systems*, *51*(3), 719-774.

[33]. Gbongli, K., Xu, Y., & Amedjonekou, K. M. (2019). Extended technology acceptance model to predict mobile-based money acceptance and sustainability: A multi-analytical structural equation modeling and neural network approach. *Sustainability*, *11*(13), 3639.

[34]. Ge, Y., He, S., Xiong, J., & Brown, D. E. (2017). Customer churn analysis for a software-as-a-service company. 2017 Systems and Information Engineering Design Symposium (SIEDS),

[35]. Gimpel, H., Rau, D., & Röglinger, M. (2018). Understanding FinTech start-ups–a taxonomy of consumer-oriented service offerings. *Electronic Markets*, *28*(3), 245-264.

[36]. Gittell, J. H., & Ali, H. N. (2021). *Relational analytics: Guidelines for analysis and action*. Routledge.

[37]. Greaves, C., Poltawski, L., Garside, R., & Briscoe, S. (2017). Understanding the challenge of weight loss maintenance: a systematic review and synthesis of qualitative research on weight loss maintenance. *Health psychology review*, *11*(2), 145-163.

[38]. Gregurić, M., Vujić, M., Alexopoulos, C., & Miletić, M. (2020). Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data. *Applied Sciences*, *10*(11), 4011.

[39]. Gremler, D. D., Van Vaerenbergh, Y., Brüggen, E. C., & Gwinner, K. P. (2020). Understanding and managing customer relational benefits in services: a meta-analysis. *Journal of the Academy of Marketing Science*, *48*(3), 565-583.

[40]. Hair Jr, J., Page, M., & Brunsveld, N. (2019). *Essentials of business research methods*. Routledge.

[41]. Hassan, M. A., & Shukur, Z. (2021). Device identity-based user authentication on electronic payment system for secure E-wallet apps. *Electronics*, *11*(1), 4.

[42]. Hernandez, J. C. (2019). Leaking pipeline: Issues impacting Latino/a college student retention. *Minority student retention*, 99-122.

[43]. Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. *Educational technology research and development*, *67*(5), 1273-1306.

[44]. Hewett, R., Shantz, A., Mundy, J., & Alfes, K. (2018). Attribution theories in human resource management research: A review and research agenda. *The International Journal of Human Resource Management*, *29*(1), 87-126.

[45]. Hu, J.-F., Zheng, W.-S., Ma, L., Wang, G., Lai, J., & Zhang, J. (2018). Early action prediction by soft regression. *IEEE transactions on pattern analysis and machine intelligence*, *41*(11), 2568-2583.

[46]. Istiaq, A., & Tanjina Binte, S. (2023). AI-Driven Vulnerability Prioritization for Enterprise Networks: A Quantitative Study Using Attack-Graph Models. *American Journal of Advanced Technology and Engineering Solutions*, *3*(04), 129-166. https://doi.org/10.63125/s6qn2t38

[47]. Jaiswal, A. K., Niraj, R., Park, C. H., & Agarwal, M. K. (2018). The effect of relationship and transactional characteristics on customer retention in emerging online markets. *Journal of Business Research*, *92*, 25-35.

[48]. Jokhan, A., Sharma, B., & Singh, S. (2019). Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, *44*(11), 1900-1911.

[49]. Khoa, B. T. (2020). The role of mobile skillfulness and user innovation toward electronic wallet acceptance in the digital transformation era. 2020 international conference on information technology systems and innovation (ICITSI),

[50]. Khoa, B. T., Oanh, N. T. T., Uyen, V. T. T., & Dung, D. C. H. (2021). Customer loyalty in the Covid-19 pandemic: the application of machine learning in survey data. In *Smart Systems: Innovations in Computing: Proceedings of SSIC 2021* (pp. 419-429). Springer.

[51]. Kim, D., & Kim, S. (2017). The role of mobile technology in tourism: Patents, articles, news, and mobile tour app reviews. *Sustainability*, *9*(11), 2082.

[52]. Klabi, F. (2020). To what extent do conspicuous consumption and status consumption reinforce the effect of self-image congruence on emotional brand attachment? Evidence from the Kingdom of Saudi Arabia. *Journal of Marketing Analytics*, *8*(2), 99-117.

[53]. Koghut, M., & AI-Tabbaa, O. (2021). Exploring consumers' discontinuance intention of remote mobile payments during post-adoption usage: An empirical study. *Administrative Sciences*, *11*(1), 18.

[54]. Kumar, V., & Ayodeji, O. G. (2021). E-retail factors for customer activation and retention: An empirical study from Indian e-commerce customers. *Journal of Retailing and Consumer Services*, *59*, 102399.

[55]. Lee, E., Kim, B., Kang, S., Kang, B., Jang, Y., & Kim, H. K. (2018). Profit optimizing churn prediction for long-term loyal customers in online games. *IEEE Transactions on Games*, *12*(1), 41-53.

[56]. Lenting, K., Verhaak, R., Ter Laan, M., Wesseling, P., & Leenders, W. (2017). Glioma: experimental models and reality. *Acta neuropathologica*, *133*(2), 263-282.

[57]. Li, J., Fong, S., Hu, S., Chu, V. W., Wong, R. K., Mohammed, S., & Dey, N. (2017). Rare event prediction using similarity majority under-sampling technique. International Conference on Soft Computing in Data Science,

[58]. Liébana-Cabanillas, F., Alonso-Dos-Santos, M., Soto-Fuentes, Y., & Valderrama-Palma, V. (2017). Unobserved heterogeneity and the importance of customer loyalty in mobile banking. *Technology Analysis & Strategic Management*, *29*(9), 1015-1032.

[59]. Liébana-Cabanillas, F., Singh, N., Kalinic, Z., & Carvajal-Trujillo, E. (2021). Examining the determinants of continuance intention to use and the moderating effect of the gender and age of users of NFC mobile payments: A multi-analytical approach. *Information Technology and Management*, *22*(2), 133-161.

[60]. Liu, X., Li, H., Lu, X., Xie, T., Mei, Q., Feng, F., & Mei, H. (2017). Understanding diverse usage patterns from large-scale appstore-service profiles. *IEEE Transactions on Software Engineering*, *44*(4), 384-411.

[61]. Lopes, D. P., Rita, P., & Treiblmaier, H. (2021). The impact of blockchain on the aviation industry: Findings from a qualitative study. *Research in Transportation Business & Management*, *41*, 100669.

[62]. Loria, E., Pirker, J., Drachen, A., & Marconi, A. (2020). Do influencers influence?–analyzing players' activity in an online multiplayer game. 2020 IEEE conference on games (CoG),

[63]. Lyu, R., Zhang, J., Xu, M., & Li, J. (2018). Impacts of urbanization on ecosystem services and their temporal relations: A case study in Northern Ningxia, China. *Land use policy*, *77*, 163-173.

[64]. Manam, A., & Md. Ashfaq, S. (2022). Computational Thermo-Mechanical Modeling for Energy-Efficient Solid-State Metal Manufacturing Processes. *American Journal of Interdisciplinary Studies*, *3*(04), 579-618. https://doi.org/10.63125/ddg6mg97

[65]. Mangaroska, K., Vesin, B., Kostakos, V., Brusilovsky, P., & Giannakos, M. N. (2021). Architecting analytics across multiple e-learning systems to enhance learning design. *IEEE Transactions on Learning Technologies*, *14*(2), 173-188.

[66]. Marella, V., Upreti, B., Merikivi, J., & Tuunainen, V. K. (2020). Understanding the creation of trust in cryptocurrencies: The case of Bitcoin. *Electronic Markets*, *30*(2), 259-271.

[67]. Md Khaled, H. (2021). An Empirical Study of CRM and Analytics-Based Approaches to Customer Engagement and Sales Performance Evaluation in Enterprise Organizations. *American Journal of Data Science and Analytics*, *2*(12), 76-155. https://doi.org/10.63125/1tt57n77

[68]. Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information systems and e-business management*, *16*(3), 547-578.

[69]. Milošević, M., Živić, N., & Andjelković, I. (2017). Early churn prediction with personalized targeting in mobile social games. *Expert Systems with Applications*, *83*, 326-332.

[70]. Mohammad Robel, M., & Md Aminul, I. (2023). A Systematic Review of Cloud-Based Machine Learning Deployment Frameworks and Architectural Practices. *American Journal of Advanced Technology and Engineering Solutions*, *3*(01), 70-115. https://doi.org/10.63125/acyg9n80

[71]. Monteiro, B., Santos, V., Reis, I., Sampaio, M. C., Sousa, B., Martinho, F., José Sousa, M., & Au-Yong-Oliveira, M. (2020). Employer branding applied to SMEs: A pioneering model proposal for attracting and retaining talent. *Information*, *11*(12), 574.

[72]. Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, *5*(4), 213-222.

[73]. Nahmias, D., Cohen, A., Nissim, N., & Elovici, Y. (2020). Deep feature transfer learning for trusted and automated malware signature generation in private cloud environments. *Neural Networks*, *124*, 243-257.

[74]. Nakamura, S., Ono, S., & Kawasaki, H. (2021). Flooded road detection from driving recorder: Training deep net for rare event using GANs semantic information. *International journal of intelligent transportation systems research*, *19*(1), 1-11.

[75]. Nasekin, S., & Chen, C. Y.-H. (2020). Deep learning-based cryptocurrency sentiment construction. *Digital Finance*, *2*(1), 39-67.

[76]. Olanrewaju, R. F., Khan, B. U. I., Morshidi, M. A., Anwar, F., & Kiah, M. L. B. M. (2021). A frictionless and secure user authentication in web-based premium applications. *IEEE access*, *9*, 129240-129255.

[77]. Pal, A., Herath, T., De', R., & Rao, H. R. (2021). Is the convenience worth the risk? An investigation of mobile payment usage. *Information Systems Frontiers*, *23*(4), 941-961.

[78]. Ping, N. L. (2019). Constructs for artificial intelligence customer service in E-commerce. 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS),

[79]. Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., & Bayol, N. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature communications*, *11*(1), 4540.

[80]. Rajul, Y., Babu, D. S., & Anuradha, K. (2018). Analysis on periodic web personalization for the efficiency of web services. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT),

[81]. Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2021). Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, *4*(1), 86.

[82]. Razavi, R., Gharipour, A., Fleury, M., & Akpan, I. J. (2019). A practical feature-engineering framework for electricity theft detection in smart grids. *Applied energy*, *238*, 481-494.

[83]. Rezaeian, O., Haghighi, S. S., & Shahrabi, J. (2021). Customer churn prediction using data mining techniques for an Iranian payment application. 2021 12th International Conference on Information and Knowledge Technology (IKT),

[84]. Rohloff, T., Oldag, S., Renz, J., & Meinel, C. (2019). Utilizing web analytics in the context of learning analytics for large-scale online learning. 2019 IEEE global engineering education conference (EDUCON),

[85]. Sangaralingam, K., Verma, N., Ravi, A., Bae, S. W., & Datta, A. (2019). High value customer acquisition & retention modelling–A scalable data mashup approach. 2019 IEEE International Conference on Big Data (Big Data),

[86]. Sazzadul, I. (2023). Explainable Data Analytics in Financial Decision Systems: Enhancing Transparency in Big Data–Driven Credit Risk and Loan Approval Models. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 31–67. https://doi.org/10.63125/twq4bw77

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

[88]. Scriney, M., Nie, D., & Roantree, M. (2020). Predicting customer churn for insurance data. International conference on big data analytics and knowledge discovery,

[89]. Sengupta, A., & Williams, S. (2021). Can an engagement platform persuade students to stay? Applying behavioral models for retention. *International Journal of Human–Computer Interaction*, *37*(11), 1016-1027.

[90]. Shah, R. S., Bhatia, A., Gandhi, A., & Mathur, S. (2021). Bitcoin data analytics: Scalable techniques for transaction clustering and embedding generation. 2021 international conference on communication systems & NETworkS (COMSNETS),

[91]. Sharma, S. K., Sharma, H., & Dwivedi, Y. K. (2019). A hybrid SEM-Neural network model for predicting determinants of mobile payment services. *Information Systems Management*, *36*(3), 243-261.

[92]. Shirazi, F., & Mohammadi, M. (2019). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*, *48*, 238-253.

[93]. Shynu, P., Menon, V. G., Kumar, R. L., Kadry, S., & Nam, Y. (2021). Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing. *IEEE access*, *9*, 45706-45720.

[94]. Sidiropoulos, G. K., Kiratsa, P., Chatzipetrou, P., & Papakostas, G. A. (2021). Feature extraction for finger-vein-based identity recognition. *Journal of Imaging*, *7*(5), 89.

[95]. Singh, P., & Agrawal, V. (2019). A collaborative model for customer retention on user service experience. In *Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018* (pp. 55-64). Springer.

[96]. Smith, A. (2019). *Consumer behaviour and analytics*. Routledge.

[97]. Su, Y., Backlund, P., & Engström, H. (2021). Comprehensive review and classification of game analytics. *Service Oriented Computing and Applications*, *15*(2), 141-156.

[98]. Sulayman, I. I. A., & Ouda, A. (2019). User modeling via anomaly detection techniques for user authentication. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON),

[99]. Sulayman, I. I. A., & Ouda, A. (2020). Designing security user profiles via anomaly detection for user authentication. 2020 International Symposium on Networks, Computers and Communications (ISNCC),

[100]. Sun, H., Zhang, J., Mo, R., & Zhang, X. (2020). In-process tool condition forecasting based on a deep learning method. *Robotics and Computer-Integrated Manufacturing*, *64*, 101924.

[101]. Tanjina Binte, S., & Sazzadul, I. (2022). Advanced Financial Data Analytics for Anomaly Detection and Pattern Discovery in Large-Scale Financial Data Pipelines. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 174-210. https://doi.org/10.63125/g1cdm484

[102]. Tarnowska, K., Ras, Z. W., & Daniel, L. (2020). *Recommender system for improving customer loyalty* (Vol. 1). Springer.

[103]. Tej, J., Vagaš, M., Ali Taha, V., Škerháková, V., & Harničárová, M. (2021). Examining HRM practices in relation to the retention and commitment of talented employees. *Sustainability*, *13*(24), 13923.

[104]. Thisarani, M., & Fernando, S. (2021). Artificial intelligence for futuristic banking. 2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC),

[105]. Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE access*, *7*, 60134-60149.

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

[107]. Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356-365.

[108]. Wanchai, P. (2017). Customer churn analysis: A case study on the telecommunication industry of Thailand. 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST),

[109]. Wang, H., Tu, Z., Fu, Y., Wang, Z., & Xu, X. (2021). Time-aware user profiling from personal service ecosystem. *Neural Computing and Applications*, *33*(8), 3597-3619.

[110]. Wang, X., Fang, Z., Wang, D., Feng, A., & Wang, Q. (2020). Research on Database Anomaly Access Detection Based on User Profile Construction. International Conference on Frontiers in Cyber Security,

[111]. Wen, X., Han, Y., Fu, J., Li, P., & Meng, F. (2020). Design of user behavior analysis model of e-commerce website based on Spark. 2020 7th International Conference on Information Science and Control Engineering (ICISCE),

[112]. Wiese, M., & Humbani, M. (2020). Exploring technology readiness for mobile payment app users. *The International Review of Retail, Distribution and Consumer Research*, *30*(2), 123-142.

[113]. Wu, S., Yau, W.-C., Ong, T.-S., & Chong, S.-C. (2021). Integrated churn prediction and customer segmentation framework for telco business. *IEEE access*, *9*, 62118-62136.

[114]. Xiao, Z., Fang, H., Jiang, H., Bai, J., Havyarimana, V., Chen, H., & Jiao, L. (2021). Understanding private car aggregation effect via spatio-temporal analysis of trajectory data. *IEEE transactions on cybernetics*, *53*(4), 2346-2357.

[115]. Yadav, S., Jain, A., Sharma, K. C., & Bhakar, R. (2021). Load forecasting for rare events using LSTM. 2021 9th IEEE International Conference on Power Systems (ICPS),

[116]. Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. *IEEE access*, *9*, 60447-60458.

[117]. Yang, B., Liu, R., & Zio, E. (2019). Remaining useful life prediction based on a double-convolutional neural network architecture. *IEEE Transactions on Industrial Electronics*, *66*(12), 9521-9530.

[118]. Yang, M., Mamun, A. A., Mohiuddin, M., Nawi, N. C., & Zainol, N. R. (2021). Cashless transactions: A study on intention and adoption of e-wallets. *Sustainability*, *13*(2), 831.

[119]. Yang, S., Zhao, W., Liu, Y., Wang, S., Wang, J., & Zhai, R. (2018). Influence of land use change on the ecosystem service trade-offs in the ecological restoration area: Dynamics and scenarios in the Yanhe watershed, China. *Science of the total environment*, *644*, 556-566.

[120]. Yao, T., Qiu, Q., & Wei, Y. (2019). Retaining hotel employees as internal customers: Effect of organizational commitment on attitudinal and behavioral loyalty of employees. *International Journal of Hospitality Management*, *76*, 1-8.

[121]. Yazdinejad, A., HaddadPajouh, H., Dehghantanha, A., Parizi, R. M., Srivastava, G., & Chen, M.-Y. (2020). Cryptocurrency malware hunting: A deep recurrent neural network approach. *Applied Soft Computing*, *96*, 106630.

[122]. Yeboah-Asiamah, E., Narteh, B., & Mahmoud, M. A. (2018). Preventing customer churn in the mobile telecommunication industry: is mobile money usage the missing link? *Journal of African Business*, *19*(2), 174-194.

[123]. Zhang, J.-H., Zou, L.-c., Miao, J.-j., Zhang, Y.-X., Hwang, G.-J., & Zhu, Y. (2020). An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment. *Interactive Learning Environments*, *28*(2), 231-245.

[124]. Zhang, J., & Luximon, Y. (2021). A quantitative diary study of perceptions of security in mobile payment transactions. *Behaviour & Information Technology*, *40*(15), 1579-1602.

[125]. Zhang, M., Li, Y., Zheng, C., Han, X., Gu, H., & Pan, H. (2021). Blockchain based global financial service platform. 2021 IEEE 19th International Conference on Industrial Informatics (INDIN),

[126]. Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2018). Employee turnover prediction with machine learning: A reliable approach. Proceedings of SAI intelligent systems conference,

[127]. Zheng, A., Chen, L., Xie, F., Tao, J., Fan, C., & Zheng, Z. (2020). Keep you from leaving: Churn prediction in online games. International Conference on Database Systems for Advanced Applications,

[128]. Zhu, B., Baesens, B., Backiel, A., & Vanden Broucke, S. K. (2018). Benchmarking sampling techniques for imbalance learning in churn prediction. *Journal of the Operational Research Society*, *69*(1), 49-65.

[129]. Zhu, B., Baesens, B., & vanden Broucke, S. K. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information sciences*, *408*, 84-99.