



Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Received	2026/09/18	تم استلام الورقة العلمية في
Accepted	2026/10/11	تم قبول الورقة العلمية في
Published	2026/10/12	تم نشر الورقة العلمية في

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case Study Using Operational Billing Logs

*Abdulsalam Bakouri¹, Mohamed Elbeshti², Husayn Mussa³,
Amrou salem Ahmed Dhunnis⁴

Faculty of Information Technology, University of Zawia, Zawia, Libya

¹abd.bakouri@zu.edu.ly*; ²mohamed.elbeshti@zu.edu.ly;

³H.mosa@zu.edu.ly, ⁴dhunnis@zu.edu.ly

1- ORCID: 0009-0005-6855-5613

2- 0009-0005-5897-542X

3- 0009-0000-8458-2121

4- 0009-0008-7783-3066

Abstract

This study examines subscriber renewal behavior at a Libyan Internet Service Provider (ISP) and frames the problem as renewal risk prediction rather than conventional churn prediction. The operational question concerns the probability that a customer whose package expires on a given day will fail to activate a new commercial package within the following 15 days. The analysis draws on raw billing and operational subscription logs collected between January 2024 and April 2026, yielding 33,308 valid subscription renewal events. Renewal failure defined operationally as the absence of any subsequent commercial package activation within 15 days after the expiry date. A rigorous temporal validation design adopted using train, validation, and test splits, supported by paired bootstrap testing to assess the reliability of model comparisons. The observational evaluation compared a rule-based baseline, Logistic Regression, Random Forest, and CatBoost, while SHAP used to support model interpretability. CatBoost performed best in operational terms after Platt scaling, which improved probability calibration while preserving ranking quality. The



Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

calibrated CatBoost model achieved a ROC AUC of 0.7091, PR AUC of 0.4268, and a Brier score of 0.1535. Operationally, targeting the highest-risk 20% of subscribers captured 41.84% of renewal failures before they occurred. These findings suggest that explainable renewal risk modeling can support more focused customer retention actions, such as targeted service follow-up, promotional renewal offers, and advertising campaigns, rather than broad messaging to all subscribers.

Keywords: Renewal risk prediction; Internet service provider (ISP); Operational billing logs; Event-level modeling; Customer retention; CatBoost; Random Forest; Probability calibration; SHAP; Libyan telecommunications.

تطبيق نماذج قابلة للتفسير لمخاطر التجديد لدى مزود خدمات إنترنت

ليبي: دراسة حالة باستخدام سجلات الفوترة التشغيلية

عبد السلام بكوري¹، محمد البشتي¹، حسين موسى¹، عمرو سالم احمد دحنس¹

¹كلية تقنية المعلومات، جامعة الزاوية، الزاوية، ليبيا

abd.bakouri@zu.edu.ly; mohamed.elbeshti@zu.edu.ly;

H.mosa@zu.edu.ly, dhunnis@zu.edu.ly

ORCID: 0009-0005-6855-5613; 0009-0005-5897-542X; 0009-0000-8458-2121

الملخص

تتناول هذه الدراسة سلوك تجديد الاشتراك لدى مزود خدمات إنترنت ليبي، إذ تعيد تأطير المشكلة بوصفها تنبؤًا بمخاطر عدم التجديد بدلا من التنبؤ التقليدي بفقدان العملاء. ويسعى السؤال التشغيلي إلى تقدير احتمال ألا يفعل المشترك، الذي تنتهي صلاحية باقته في يوم محدد، أي باقة تجارية جديدة خلال الخمسة عشر يوما التالية. وتعتمد الدراسة على سجلات الفوترة التشغيلية وسجلات الاشتراك الخام للفترة بين يناير 2024 وأبريل 2026، وقد أفرزت 33,308 حدث تجديد صالح للتحليل. وقد عُرِف فشل التجديد إجرائيا بغياب أي تفعيل لباقة تجارية لاحقة خلال 15 يوما من تاريخ انتهاء الصلاحية. واعتمد تصميم تحقق زمني صارم يقوم على تقسيم البيانات إلى مجموعات تدريب وتحقيق واختبار،

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

مدعوما باختبار إعادة المعاينة العشوائية المقترن (Paired Bootstrap) لتقييم موثوقية مقارنات النماذج. وقارن التقييم بين نموذج قاعدي قائم على القواعد، والانحدار اللوجستي، والغابة العشوائية، و CatBoost، في حين استُخدم SHAP لدعم قابلية تفسير النموذج. وقد حقق CatBoost أفضل أداء تشغيلي بعد تطبيق معايير بلات، مما أدى إلى تحسين معايرة الاحتمالات مع الحفاظ على جودة الترتيب. وبلغت قيم النموذج المعايير ROC : $AUC = 0.7091$ ، و $PR AUC = 0.4268$ ، ودرجة $Brier = 0.1535$ ومن منظور تشغيلي، أسفر استهداف المشتركين الأعلى خطرا بنسبة 20% عن النقاط 41.84% من حالات فشل التجديد قبل وقوعها. وتشير هذه النتائج إلى أن النمذجة القابلة للتفسير لمخاطر التجديد يمكن أن تدعم إجراءات أكثر تركيزا للاحتفاظ بالعملاء، من بينها المتابعة الخدمية الموجهة، وعروض التجديد الترويجية، والحملات الإعلانية، بدلا من الاعتماد على رسائل عامة موجهة إلى جميع المشتركين.

الكلمات المفتاحية: التنبؤ بمخاطر التجديد؛ مزود خدمات الإنترنت (ISP)؛ سجلات الفوترة التشغيلية؛ النمذجة على مستوى الحدث؛ الاحتفاظ بالعملاء؛ CatBoost؛ الغابة العشوائية؛ معايرة الاحتمالات؛ SHAP؛ الاتصالات الليبية.

1. Introduction

1.1 Background and Problem Context

Customer churn has long been a central concern in the telecommunications sector, and a substantial body of research has addressed it through predictive modeling. Most of this work treats churn as a single, discrete event and relies on datasets that have already aggregated into customer-level summary tables. The problem with that approach is that it discards much of the operational detail that actually explains why customers leave. Billing schedules, service states, package expiry dates, payment behavior, and administrative corrections all carry information that static tables tend to flatten (Barsotti et al., 2024).

The issue is particularly visible among Internet Service Providers (ISPs) operating in competitive private markets, where subscribers move between companies with relative ease. This study works directly with real operational billing logs rather than pre-prepared

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

datasets. Such logs record daily activity in detail, from package activations and expiry dates to financial adjustments and promotional offers, and they document every financial or technical transaction in real time. Tracking them chronologically allows a fuller reconstruction of the service lifecycle and exposes behavioral patterns that a summary table would hide.

1.2 Motivation

This data structure motivated a shift away from the traditional churn question. Instead of asking whether a customer will leave or stay, the study asks a operational question: given a subscriber's current package and prior history, how likely is it that they will fail to renew within a defined period after expiry? Framing the problem, this way sharpens the analysis and links predictions to a specific moment where the marketing department can intervene.

Recent work supports this direction. Studies emphasize that accurate predictions alone are insufficient and that the model must be explainable so that management can justify decisions and prioritize actions (Poudel et al., 2024). Research that accounts for timing goes further, arguing that customer departure is not a simple binary event but a process tied to the service lifecycle (Kimatei et al., 2025).

1.3 Research Gap

The gap addressed here is not only geographic, though studies on the Libyan market remain scarce. It is primarily a gap in how data are used. Most previous research in the telecommunications sector relies on aggregated summary tables in which payment processes and service lifecycles have already merged and condensed. In Libya and the surrounding region, published studies have tended to focus on customer satisfaction, descriptive behavior, or general churn prediction.

Alteer and Alariyibi (2024) offer an important regional reference, but their work treats churn as a single aggregated entity. The present study differs in that it reconstructs renewal risk directly from daily operational billing logs. This difference affects prediction accuracy, redefines the problem formulation, and makes the resulting model applicable within the company's marketing policies and advertising

campaigns. It allows focus on a specific segment instead of launching general campaigns that reach loyal customers as well.

1.4 Research Objective and Research Questions

This study aims to develop and evaluate an explainable renewal risk model for a Libyan Internet service provider using operational billing logs and package lifecycle features. The goal is to estimate the risk that a valid commercial package event not followed by renewal within a defined post-expiry window, while preserving the operational meaning of the billing process. The study addresses the following research questions:

- **RQ1:** Can operational billing logs from a Libyan ISP be transformed into a valid event-level renewal risk dataset without simplifying package lifecycle behavior?
- **RQ2:** Do lifecycle-aware predictors derived from expiry timing, prior commercial activity, financial behavior, promotional traces, and correction-linked events provide valid predictive signal?
- **RQ3:** Under a time-aware evaluation design, do non-linear tabular models outperform simpler baselines in this setting?
- **RQ4:** Does post-hoc calibration improve the practical usability of predicted renewal risk probabilities?
- **RQ5:** Which operational patterns contribute most strongly to renewal risk, and how can they support selective pre-expiry intervention?

1.5 Main Contributions

This paper offers five contributions to the field of telecom data analytics:

1. **New Problem Framing:** Rather than treating customer churn as a general event, the study focuses on event-level renewal risk. The model predicts risk for each specific package at its expiry, which is more practical for daily ISP operations.
2. **A Practical Data Pipeline:** The study develops a clear procedure for turning messy operational billing logs into a

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

clean dataset while preserving the lifecycle context of each subscriber.

3. Realistic Evaluation Design: Linear and non-linear models (CatBoost and Random Forest among them) tested using a chronological design, which prevents future data from leaking into the training process.
4. Focus on Probability Quality: The paper does not only rank who might leave; it also focuses on the accuracy of the predicted percentages. Post-hoc calibration ensures that probabilities such as a 70% risk are realistic and can be trusted for financial decisions.
5. Application in the Libyan Context: The final model translated into a retention targeting strategy for a Libyan ISP. Since the Libyan market underrepresented in research, this work provides a case study for similar emerging markets.

2. Related Work

2.1 Telecom Churn and Retention Modeling

Customer churn prediction in telecommunications has developed into a broad methodological field spanning linear models, tree ensembles, deep learning, and hybrid systems (Shahabikargar et al., 2025). Barsotti et al. (2024) show that while the literature is rich in modeling approaches; it is less consistent in how the prediction task itself defined. Studies differ in target construction, unit of analysis, and validation design, which limits direct performance comparison. Recent applied studies continue to report strong results using machine learning models for churn and retention tasks (Chang et al., 2024; He & Ding, 2024; Saleh & Saha, 2023; Sikri et al., 2024; Liu et al., 2024). In many cases, non-linear tabular models outperform simpler baselines when customer behavior and financial signals interact. Still, much of this literature relies on customer-level summary tables, where billing and service processes have already aggregated into static features.

2.2 Explainability, Calibration, and Time-Aware Views

A growing strand of telecom retention research emphasizes explainability. In applied settings, a ranked list of risky customers is not enough on its own; the model must also support interpretation so that intervention decisions can be justified and reviewed. Poudel et al. (2024) illustrate this direction by combining tabular machine learning models with explainability tools to identify the drivers of churn risk. Work that accounts for timing adds a further dimension, arguing that customer departure should not always reduced to a context-free binary event, especially when delay structure and service lifecycle matter (Kimiti et al., 2025). Calibration has also become increasingly relevant in applied work, since a model may rank customers well while still producing unreliable probability estimates.

2.3 Libyan and Regional Context

The Libyan and broader regional literature remains limited, though some useful directions have emerged. Omer et al. (2022) applied machine learning to customer satisfaction analysis for Libyan telecom companies using Twitter-derived signals. In the wider Maghreb, Balouli et al. (2025) analyzed churn in the Algerian telecom market through a case study of Ooredoo. Sherif and Elmabrouk (2023) explored customer behavior in the Libyan telecom market through data visualization, offering descriptive insight without directly modeling churn or renewal behavior. Awadh et al. (2025) examined service quality and customer satisfaction in Benghazi service centers, adding a service-side perspective that remains descriptive rather than predictive.

The closest published work to the present study is Alteer and Alariyibi (2024). Their paper is relevant, yet the two studies differ in formulation, data structure, and evaluation logic. Alteer and Alariyibi modeled customer churn using an aggregated customer-level dataset under severe class imbalance, whereas the present study models event-level renewal risk reconstructed directly from operational billing logs. The comparison (shown in Table I) is contextually useful, though not a strict like-for-like benchmark.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Table I. Comparison Between the Present Study and Alteer & Alariyibi (2024)

Dimension	Alteer & Alariyibi (2024)	Present study
Context	Libyan ISP	Libyan ISP
Prediction target	Customer churn prediction	Event-level renewal risk prediction
Unit of analysis	Customer-level aggregated instance	Commercial package-bearing renewal event
Data source	Aggregated customer dataset derived from ISP records	Raw operational billing logs and package lifecycle records
Observation period	January 2022 to March 2023	January 2024 to April 2026
Dataset size	3,606 instances with 11 attributes	37,301 raw records transformed into 33,308 labeled events
Class prevalence	Approximately 5% churn	25.95% renewal failure under the 15-day window
Feature structure	Aggregated behavioral and usage features	Lifecycle, billing, spending, correction, and promotion features
Imbalance handling	SMOTE and SMOTEENN	Class-weighted learning and calibrated probability modeling
Compared models	LR, DT, RF, GBM, XGBoost	Rule-based baseline, LR, RF, CatBoost, calibrated CatBoost
Validation design	80/20 split	Chronological train/validation/test split
Explainability	Not central	SHAP-based global and local interpretation
Calibration analysis	Not reported	Explicit post-hoc calibration with Brier score evaluation

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Dimension	Alteer & Alariyibi (2024)	Present study
Practical targeting analysis	Not reported	Targeting/capture analysis for pre-expiry intervention
Reported headline result	XGBoost with AUC around 94% after resampling	Calibrated CatBoost with ROC AUC 0.7091, PR AUC 0.4268, and Brier 0.1535

A smaller yet important strand of churn and retention research deals with settings where raw data not already organized as clean customer-level tables. Here the challenge extends beyond model selection. It also involves reconstructing a valid predictive dataset from transactional and operational records. Billing activity, package activations, corrections, balance movements, and service updates must all translated into meaningful observations and labels. The present study addresses this gap by combining four elements that previous work has not fully treated together. It relies on operational billing logs rather than pre-aggregated customer tables, frames the problem as event-level renewal risk prediction instead of generic customer churn, evaluates non-linear tabular models with explainability and calibration, and applies the framework in the underrepresented context of a Libyan Internet Service Provider (ISP).

3. Data

3.1 Dataset Overview

The dataset consists of operational billing records exported from a Libyan Internet service provider and stored as a semicolon-delimited transaction log. After parsing and standardizing the file, the raw dataset contained 37,301 records and 20 original fields spanning 1 January 2024 to 4 April 2026. Across this interval, the file covered 3,109 unique customers, 50 unique service descriptions, and 12 distinct operators listed under the Created by field.

Each row represents an invoice-linked operational event rather than a customer-level observation. This distinction matters because the dataset not originally structured as a supervised-learning table for

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

churn analysis. After identifying valid commercial subscription-bearing events and excluding observations without a sufficient post-expiry horizon for label generation, the final event-level sample under the primary 15-day renewal window contained 33,308 labeled observations. Of these, 25.95% labeled as renewal failures. The labeled data then divided chronologically into 22,190 training events, 5,562 validation events, and 5,556 test events. Table II below lists the core billing-log fields and their operational meaning.

Table II. Core Billing Log Fields and Their Operational Meaning

Field	Data Type	Operational Meaning	Role in This Study
Date	Date	Date of the invoice-linked operational event	Event timing and temporal ordering
Invoice number	String	Unique identifier of the transaction	Administrative identifier only; not used as predictor
Description	Categorical text	Encodes package or transaction type	Core field for package family and service regime inference
Created by	Categorical text	Operator or creator associated with the event	Used as predictor and later examined through sensitivity analysis
Customer	String	Subscriber identifier	Used for event sequencing and history construction; not used directly as predictor
Price	Numeric	Monetary value of the event	Used as financial feature
Balance	Numeric	Account balance after the transaction	Used as financial readiness indicator

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Field	Data Type	Operational Meaning	Role in This Study
Added total Bytes	Numeric	Added quota volume for volume-based packages	Used mainly for quota-based plans
Available total Bytes	Numeric	Remaining or system-recorded total quota state	Used cautiously due to heterogeneous semantics across plans
Added days	Numeric	Added service duration	Central feature for time-based plans
Expiry	Date	End of current service entitlement	Central field for renewal label construction
Comment	Text	Promotional, correctional, or operational note	Source of promotion and correction flags
Paid	Date	Payment date associated with the event	Used in temporal financial features

3.2 Field Structure and Business Meaning

The schema consists of administrative, financial, temporal, and quota fields. Administrative fields such as Date, Invoice number, created by, and Customer determine when and by whom each event occurred. The Price, Balance, and Paid fields represent the money side of the transaction. Added days and Expiry indicate entitlement to a service, while byte-related variables such as Added total Bytes and Available total Bytes indicate behavior linked to quota. The Description field is particularly important because it encodes the operational meaning of the event. When interpreted together with timing, price, balance, and byte-related variables, entries such as Infinity 03, Business 20, Swift - 30GB, Family - 1TB, and Mazaya 2TB reveal structured package behavior rather than loose text labels. The raw logs also included Arabic transaction labels, which normalized during preprocessing for analytical consistency.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

3.3 ISP Operational Logic Revealed by the Logs

Two main service regimes emerged from the billing logs. The first is time-based provisioning, represented mainly by the Infinity family and, to a lesser extent, Business and Weekly. In these records, service continuation is expressed primarily through Added days and Expiry, while Added total Bytes is usually zero or operationally uninformative. The second is volume-based provisioning, represented most clearly by Swift, Family, Mazaya, and Stream. In this regime, Added total Bytes carries meaningful information and often aligns with the package size indicated in the description. The data reflect a distinct distribution across package families and highlight the variations between time-based and volume-based plan regimes (Fig. 1).

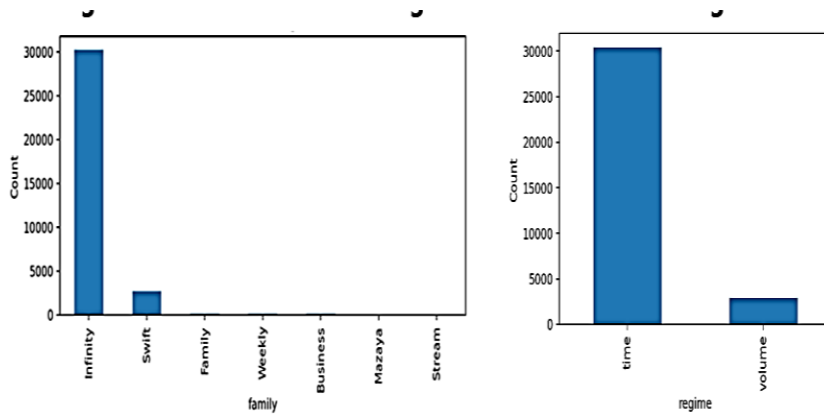


Fig. 1. Distribution of Package Families and Time-Based versus
Volume-Based Plan Regimes

The data also include a third layer related to financial activities, such as deposits and top-ups. These entries are not counted as renewals because adding money to a balance differs from actually buying a subscription. Instead, these records serve as financial signals that help understand the customer's spending habits. The logs also show that the system allows manual fixes, such as corrections and reversals, which appear through negative prices, reversed days, or specific notes in the comments. Such details matter because they indicate when a transaction was a mistake or a manual adjustment rather than a regular customer action.

3.4 Data Quality, Noise, and Constraints

The dataset is operationally rich but not uniformly clean. Some fields were clearly less useful than others. Added time, for example, contributed little because it was largely uninformative in the export, while Available online time dominated by zero values or inconsistent entries that were not reliable enough for direct modeling. Textual inconsistency also required preprocessing, particularly in operator naming, Arabic normalization, package wording, and correction-comment phrasing. The file has important limits as well. It captures billing activity and package-lifecycle traces, not direct network telemetry. It does not include indicators such as latency, packet loss, outage alarms, throughput measurements, or infrastructure-fault logs. The analysis can therefore support claims about renewal behavior as reflected in operational and commercial records, but it cannot support strong claims about network quality degradation without additional technical data sources.

4. Methodology

4.1 Problem Formulation

This study formulates the task as event-level renewal-risk prediction rather than generic customer churn classification. The billing log does not represent each customer as a single static profile. Instead, customers reappear through package activations, expiry updates, balance-related actions, promotions, migrations, and occasional corrections. Reducing that sequence to one customer-level row would simplify the task, but it would also discard much of the operational meaning embedded in the logs. The overall research workflow for this approach is illustrated in Fig. 2.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

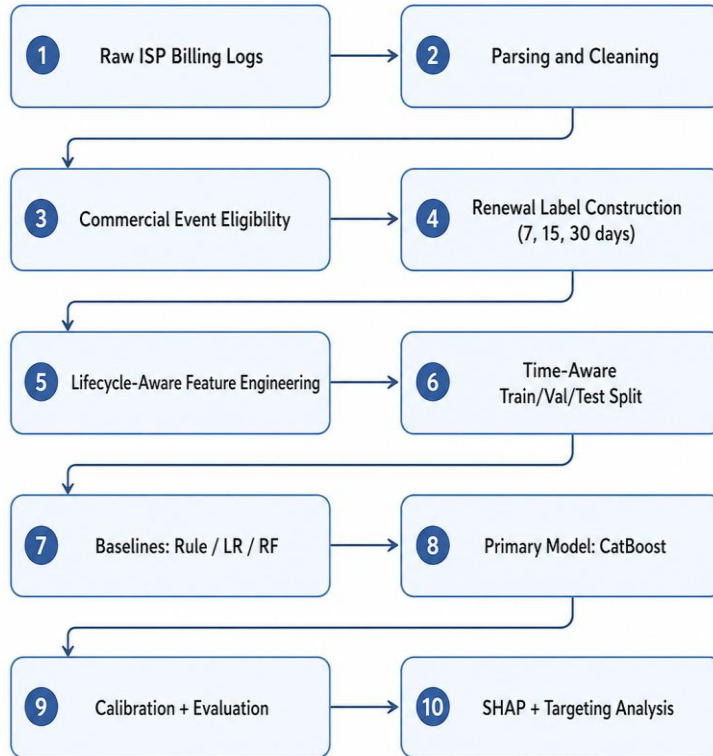


Fig. 2. Overall Research Workflow

The unit of analysis is a customer-renewal event. Each observation corresponds to a valid commercial package-bearing event that grants service entitlement through added days, added quota, or both. The objective is to estimate the probability that such an event not followed by renewal within a specified post-expiry window.

4.2 Outcome Definition and Label Construction

The target variable defined around renewal failure rather than abstract churn. For each valid subscription-bearing event, the recorded Expiry date marks the end of the current entitlement cycle. A forward search then performed within the same customer history to determine whether a later commercial subscription-bearing event occurs within a fixed window after expiry.

$y_i = 0$ if renewal occurs within W days; $y_i = 1$ otherwise

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

In this study, a subscription considered "renewed" if the customer activates a new package within 15 days after their current one expires. To ensure the reliability of our results, we performed robustness checks by testing the model again using different time limits: a shorter window of 7 days and a longer window of 30 days. These checks showed that our findings remain solid across different scenarios. It is important to note that simply adding money to an account (a Deposit) not counted as a renewal. While it changes the account balance, it does not grant the customer the actual service entitlement to use the internet. A true renewal only occurs when a service-bearing package is activated, a process illustrated in Fig. 3.

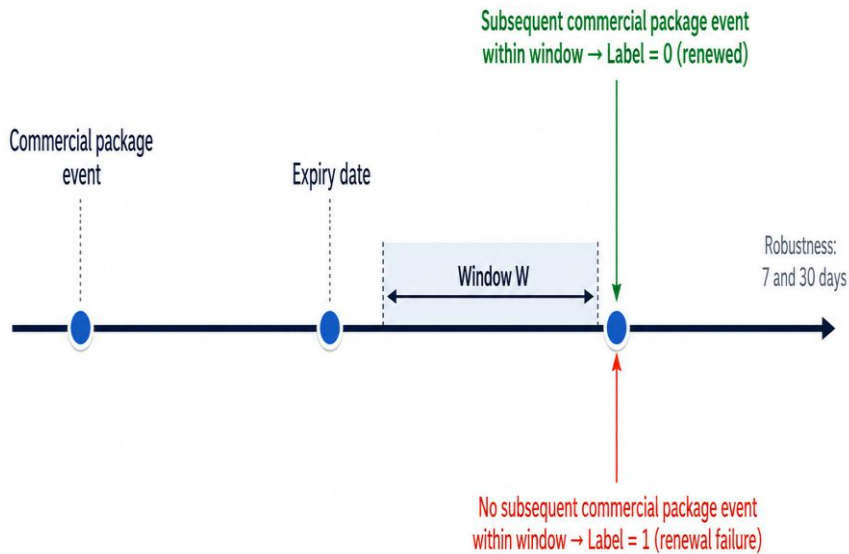


Fig. 3. Renewal Label Construction Timeline

4.3 Event Eligibility and Sample Construction

A record treated as a valid subscription-bearing anchor event if it belonged to a recognized commercial package family (primarily Infinity, Business, Weekly, Swift, Family, Mazaya, and Stream) and if it conferred service entitlement through duration, quota, or both. Top-up-only records excluded as anchors but retained in customer history as financial predictors. Entries dominated by reversals or corrections, including negative prices or negative added days, also

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

excluded as anchor events. They preserved as historical signals, however, because they reflect operational instability.

4.4 Feature Engineering from Package Lifecycle

Feature engineering treated the billing system as an operational process, not as a flat accounting table. The final feature space covered six groups: current event state, temporal lifecycle, historical commercial activity, financial behavior, promotion and operational signals, and package migration. Table III summarizes the feature groups used in modeling. Section 4.9 documents the derivation logic for selected variables.

Table III. Feature Groups Used in Modeling

Feature Group	Example Features	Analytical Purpose
Current-event features	family, regime, desc_raw, Price, Balance, Added days, Added total Bytes, Available total Bytes	Describe the immediate commercial and entitlement state of the event
Temporal lifecycle features	days_until_expiry, days_since_prev_commercial, customer_tenure_days, commercial_tenure_days, event_month, event_year	Capture lifecycle timing, continuity, and calendar context
Historical commercial activity	prior_commercial_count, commercial_events_last_90d, prior_total_commercial_spend	Measure prior engagement depth and cadence
Financial behavior	prior_topup_count, prior_total_topup_amount, days_since_prev_topup, log_balance	Capture account funding and payment-related behavior
Promotion and operational signals	has_current_promo, prior_promo_count, comment-based flags, has_current_correction, prior_correction_count, corrections_last_90d, prior_negative_count	Represent commercial intervention, offer exposure, and operational irregularity

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Feature Group	Example Features	Analytical Purpose
Package migration	prev_commercial_family, prior_family_switch_count, multimonth_flag	Represent plan- transition behavior and subscription structure

4.5 Model Development

Four predictive systems evaluated a rule-based baseline, Logistic Regression, Random Forest, and CatBoost. The rule-based model provides a minimal operational benchmark. Logistic Regression serves as the linear baseline, while Random Forest offers a strong non-linear ensemble comparator. CatBoost chosen as the primary model because the data are tabular, mixed, and rich in categorical and operational signals whose interactions are unlikely to behave linearly. CatBoost is especially suited to structured tabular data with numerical and categorical predictors (Prokhorenkova et al., 2018). Main hyperparameters and training settings reported in Table IV.

Table IV. Main Modeling Configuration

Model	Parameter / Setting	Value / Strategy
Logistic Regression	Solver	liblinear
Logistic Regression	Class handling	class_weight = balanced
Logistic Regression	Maximum iterations	1500
Logistic Regression	Regularization parameter (C)	Selected on validation set
Logistic Regression	Threshold selection	Validation-based, recall-sensitive
Random Forest	Class handling	Balanced classes

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Model	Parameter / Setting	Value / Strategy
Random Forest	Trees	100
CatBoost	Loss function	Logloss
CatBoost	Evaluation metric	PRAUC
CatBoost	Random seed	42
CatBoost	Class handling	auto_class_weights = Balanced
CatBoost	Depth	6
CatBoost	Learning rate	0.05
CatBoost	L2 regularization	3
CatBoost	Maximum iterations	120
CatBoost	Best-model selection	Validation-based (use_best_model = True)
All models	Data split strategy	Chronological train/validation/test split

For Logistic Regression, the predicted renewal-failure probability is:

$$p(y_i = 1 | x_i) = \sigma(\beta_0 + x_i^T \beta)$$

4.6 Probability Calibration

Because the initial CatBoost model showed stronger ranking performance than Logistic Regression but weaker raw probability quality, post-hoc calibration added as an explicit stage. Two alternatives evaluated on the validation set: Platt scaling and isotonic regression. The final calibrated model selected based on the trade-off between improved probability quality and preservation of discrimination. For Platt scaling, the calibrated probability is:

$$\hat{p}_i = \frac{1}{1 + \exp(A \cdot s_i + B)}$$

Where s_i is the raw model score and A and B are parameters fitted on the validation set.

4.7 Explainability and Sensitivity Checks

Explainability implemented through SHAP-based feature attribution over the final CatBoost model. SHAP provides additive feature attributions at both the global and local level and has become one of the most widely used interpretability frameworks for structured machine-learning tasks (Lundberg & Lee, 2017). Global importance assessed through mean absolute SHAP values and summary plots, while local explanations were inspected for representative test cases. Given the strong signal associated with Created by, an explicit sensitivity analysis also conducted. This check designed to determine whether the variable behaved as an informative operational proxy or as a dominant shortcut. The analysis therefore treated Created by as a potentially useful but methodologically sensitive predictor whose role required cautious interpretation.

4.8 Experimental Setup and Evaluation Protocol

The labeled 15-day sample of 33,308 events divided chronologically into 22,190 training events, 5,562 validation events, and 5,556 test events. No random shuffling used. Threshold selection performed on the validation block rather than being fixed mechanically at 0.50.

Performance evaluated at three levels. First, ranking quality assessed using ROC-AUC and PR-AUC. Second, threshold-level decision quality was measured using precision, recall, F1-score, and balanced accuracy. Third, probability quality assessed using the Brier score and calibration plots.

The equations used are:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} & \text{Recall} &= \frac{TP}{TP+FN} \\ F1 &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ \text{Brier} &= \frac{1}{N} \sum_{i=1}^N (\hat{p}_i - y_i)^2 \end{aligned}$$

Uncertainty quantified through paired bootstrap resampling, which used to estimate confidence intervals and paired performance

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

differences between competing models on the same test set. The full evaluation repeated under 7-day and 30-day renewal windows to assess robustness.

To move beyond aggregate metrics, the final calibrated model was also evaluated across operational segments, including time-based vs. volume-based plans, shorter vs. longer tenure, and lower vs. higher prior spend. A targeting analysis then performed by ranking test events according to predicted failure risk and measuring how many renewal failures would be captured when intervening only on the highest-risk segments.

$$\text{Lift}@k = \frac{\text{Failure rate within top-}k\% \text{ ranked events}}{\text{Overall failure rate}}$$

The full process implemented in Python using a workflow directly transferable to Google Colab. pandas was used for event construction and preprocessing, scikit-learn for preprocessing utilities, Logistic Regression, Random Forest, calibration, and evaluation, CatBoost for the primary model, and matplotlib for visualization. Reproducibility supported through fixed random seeds, deterministic event-generation and label-construction rules, explicit chronological splitting, and strict exclusion of direct identifiers such as customer names or invoice numbers from predictive modeling. Records without sufficient post-expiry horizon to evaluate the renewal window excluded before supervised labeling. The robustness analyses at 7, 15, and 30 days were conducted as separate end-to-end runs of the same process rather than as simple threshold reinterpretations of a single trained model.

4.9 Feature Derivation Summary

Section 4.9 presents a simplified view of the most operationally relevant feature groups. The full feature space documented in Table III.

Customer histories reconstructed chronologically using the Customer identifier and event dates. All historical features derived exclusively from information available on or before the anchor event date. The features fall into five operational groups:

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

- Commercial continuity: prior_commercial_count, commercial_events_last_90d, days_since_prev_commercial.
- Tenure: customer_tenure_days, commercial_tenure_days.
- Spending and top-up behavior: prior_total_commercial_spend, prior_topup_count, days_since_prev_topup.
- Operational corrections: prior_correction_count, corrections_last_90d.
- Package migration: prior_family_switch_count.

Event eligibility followed three rules. A record was treated as a valid commercial anchor event if the Description field mapped to a recognized package family (Infinity, Business, Weekly, Swift, Family, Mazaya, Stream) and if it conferred service entitlement through positive Added days or positive Added total Bytes. Top-up-only entries such as Deposit excluded as anchors but retained as financial predictors. Entries dominated by reversals or corrections were also excluded as anchors, though preserved as historical signals.

Text normalization covered both English and Arabic entries. Minor variation in comments, package labels, and operator wording normalized during preprocessing, and the Comment field parsed into three operational flags: promotion-related, correction-related, and special-purpose.

Leakage prevention followed four rules. Historical features used only records observed up to the anchor date. Future renewal events used only for label construction, not as predictors. Train, validation, and test sets split chronologically rather than randomly. Direct identifiers such as customer name or invoice number excluded from predictive modeling.

5. Results and Discussion

5.1 Comparative Predictive Performance

The final supervised sample contained 33,308 event-level observations derived from valid commercial subscription-bearing

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

records. Under the primary 15-day renewal definition, 25.95% of events labeled as renewal failures. It is worth noting that the chronological split produces different failure rates across the three subsets. The overall labeled sample has a failure rate of 25.95%, while the test block, which covers the most recent time, has a failure rate of approximately 22.16% (1,231 failures out of 5,556 events). The lift calculations in Section 5.5 use the test-block rate as their base, since all targeting analysis performed on held-out test events. The dataset split chronologically into 22,190 training events, 5,562 validation events, and 5,556 test events. This made the task more demanding, yet also more credible, because the test block evaluated strictly on patterns that had already appeared in the past. The billing logs preserve enough operational structure to support predictive modeling, even though they were not originally prepared as a machine-learning dataset.

The held-out test results showed a clear ordering across models. The rule-based baseline performed worst, with ROC-AUC = 0.5566, PR-AUC = 0.2459, F1 = 0.3403, and Brier = 0.3796. Logistic Regression improved substantially over that baseline, reaching ROC-AUC = 0.6397, PR-AUC = 0.3607, F1 = 0.3815, and Brier = 0.2015. The two non-linear models then moved ahead. Random Forest achieved ROC-AUC = 0.7090 and PR-AUC = 0.4258, while CatBoost produced ROC-AUC = 0.7091 and PR-AUC = 0.4268 before calibration. Visual comparisons of these metrics provided in Fig. 4.

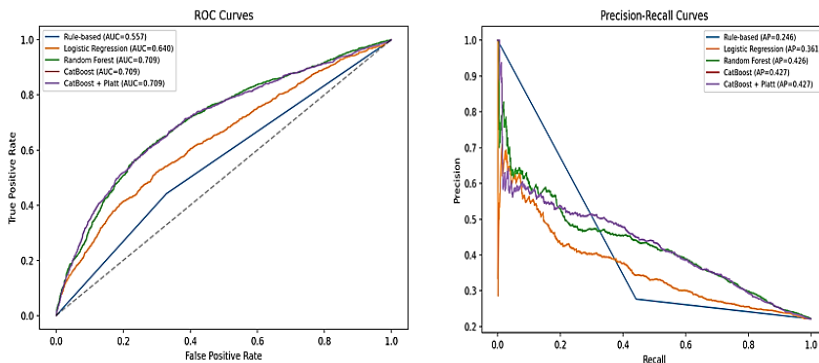


Fig. 4. ROC Curves and Precision–Recall Curves for the Main Models

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

The gap between Random Forest and CatBoost in ranking performance was very small. The more useful difference appeared after calibration. CatBoost with Platt scaling preserved ranking quality while improving threshold-level behavior and probability quality, reaching $F1 = 0.4009$, balanced accuracy = 0.5827, and Brier = 0.1535. This improvement is illustrated in Fig. 5, which compares calibration quality before and after scaling.

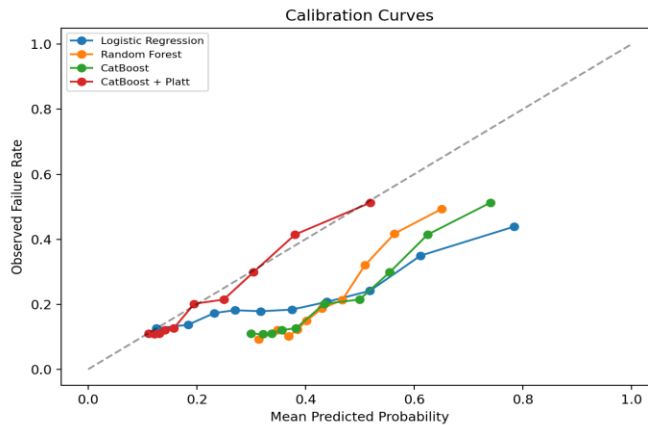


Fig. 5. Calibration Comparison Before and After Platt Scaling

These refinements matter because while the model ranked well initially, it became more operationally usable after calibration by providing reliable probability estimates. Table V summarizes the complete comparative test-set performance across all models.

Table V. Comparative Predictive Performance of Rule-Based Baseline, Logistic Regression, Random Forest, CatBoost, and CatBoost + Platt Scaling

Model	ROC-AUC	PR-AUC	Precision	Recall	F1	Balanced Accuracy	Brier	Threshold
Rule-based baseline	0.5566	0.2459	0.2767	0.4419	0.3403	0.5566	0.3796	0.5000
Logistic Regression	0.6397	0.3607	0.2433	0.8830	0.3815	0.5508	0.2015	0.2100
Random Forest	0.7090	0.4258	0.2312	0.9716	0.3736	0.5261	0.2049	0.3250
CatBoost	0.7091	0.4268	0.2490	0.9058	0.3906	0.5641	0.2084	0.3300
CatBoost + Platt	0.7091	0.4268	0.2595	0.8806	0.4009	0.5827	0.1535	0.1300
CatBoost + Isotonic	0.7090	0.4118	0.2496	0.9033	0.3911	0.5652	0.1534	0.1250

5.2 Uncertainty and Robustness

Point estimates alone can sometimes overstate a model's superiority, especially when competing systems perform similarly. To address this, we examined predictive uncertainty through paired bootstrap resampling. Relative to Logistic Regression, the calibrated CatBoost model demonstrated a significant advantage, showing a mean ROC-AUC gain of +0.0695 (95% CI: 0.0536 to 0.0870) and a mean PR-AUC advantage of +0.0659 (95% CI: 0.0478 to 0.0852). The mean Brier difference was approximately -0.0481 (95% CI: -0.0531 to -0.0431), consistently favoring the calibrated CatBoost model.

The comparison between CatBoost and Random Forest less pronounced. The differences in ROC-AUC and PR-AUC were marginal, and the bootstrap confidence intervals crossed zero, indicating that neither model clearly dominates the other in terms of ranking. The more defensible conclusion is that once calibrated, CatBoost provides similarly strong discrimination but with superior probability quality. These paired bootstrap comparisons detailed in Table VI, while the per-model confidence intervals for the main metrics reported in Table VII.

Table VI. Paired Bootstrap Performance Differences

Model A	Model B	Metric	Mean diff	CI lower	CI upper	p_boot
CatBoost + Platt	Logistic Regression	roc_auc	0.0695	0.0536	0.0870	0.0000
CatBoost + Platt	Logistic Regression	pr_auc	0.0659	0.0478	0.0852	0.0000
CatBoost + Platt	Logistic Regression	brier	-0.0481	-0.0531	-0.0431	0.0000
CatBoost + Platt	Random Forest	roc_auc	0.0001	-0.0061	0.0065	0.9840
CatBoost + Platt	Random Forest	pr_auc	0.0010	-0.0116	0.0136	0.8480
CatBoost + Platt	Random Forest	brier	-0.0513	-0.0557	-0.0466	0.0000

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Model A	Model B	Metric	Mean diff	CI lower	CI upper	p_boot
CatBoost + Platt	CatBoost	roc_auc	0.0000	0.0000	0.0000	1.0000
CatBoost + Platt	CatBoost	pr_auc	0.0000	0.0000	0.0000	1.0000
CatBoost + Platt	CatBoost	brier	-0.0549	-0.0595	-0.0502	0.0000

Table VII. Bootstrap Confidence Intervals for Selected Metrics

Model	Metric	CI Lower	CI Upper
Rule-based baseline	ROC-AUC	0.5402	0.5726
Rule-based baseline	PR-AUC	0.2313	0.2613
Rule-based baseline	Brier	0.3664	0.3920
Logistic Regression	ROC-AUC	0.6221	0.6583
Logistic Regression	PR-AUC	0.3358	0.3903
Logistic Regression	Brier	0.1967	0.2065
Random Forest	ROC-AUC	0.6931	0.7258
Random Forest	PR-AUC	0.3987	0.4554
Random Forest	Brier	0.2024	0.2073
CatBoost	ROC-AUC	0.6945	0.7261
CatBoost	PR-AUC	0.4008	0.4553
CatBoost	Brier	0.2050	0.2116
CatBoost + Platt	ROC-AUC	0.6941	0.7258
CatBoost + Platt	PR-AUC	0.4018	0.4560
CatBoost + Platt	Brier	0.1474	0.1588

The primary analysis used a 15-day post-expiry window to define renewal failures. To ensure the model's reliability not tied to a single arbitrary timeframe, we conducted robustness checks using 7-day

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

and 30-day windows through the final calibrated pipeline. The results remained consistent across these variations:

- 7-day window: ROC-AUC = 0.7076, PR-AUC = 0.4640
- 15-day window: ROC-AUC = 0.7109, PR-AUC = 0.4285
- 30-day window: ROC-AUC = 0.7240, PR-AUC = 0.3260

The stability of the ROC-AUC values across all three definitions suggests that the model has captured a fundamental predictive structure rather than a localized pattern. While the PR-AUC showed more variation, this is an expected statistical outcome as class prevalence shifts when the renewal window widens (fewer failures occur as more time is allowed). Notably, probability quality showed additional improvement under the 30-day window. Because the robustness runs were conducted as separate end-to-end executions of the same pipeline, small numerical differences relative to Table V are expected and do not indicate inconsistency. The 15-day results presented in Table VIII come from a separate end-to-end robustness rerun. While they confirm the overall trend, they intended for comparative purposes within the robustness study and should be viewed alongside the primary performance results reported earlier.

Table VIII. Robustness Across 7-Day, 15-Day, and 30-Day Renewal Windows

Window (days)	n	Target rate	ROC-AUC	PR-AUC	F1	Balanced Accuracy	Brier	Threshold
7	33,796	0.2953	0.7076	0.4640	0.3968	0.5002	0.1650	0.0900
15	33,308	0.2595	0.7109	0.4285	0.4065	0.5916	0.1532	0.1300
30	32,507	0.1721	0.7240	0.3260	0.3476	0.6491	0.1154	0.1000

5.3 Sensitivity Analysis of the Administrative Feature

The most methodologically sensitive feature in the model remained Created by. To examine whether the model was relying too heavily on operator-linked information, two sensitivity checks were conducted. First, the model was retrained after removing Created by. Performance declined from ROC-AUC = 0.7091 and PR-AUC = 0.4268 to ROC-AUC = 0.7015 and PR-AUC = 0.4141. Second, a permutation test was conducted by randomly shuffling the Created by values within the

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

test set. This produced an average drop of approximately 0.0244 in ROC-AUC and 0.0226 in PR-AUC. Taken together, these results support a cautious interpretation. Created by carries real predictive information, yet the model does not collapse when this signal removed or disrupted. The feature should read as an informative operational proxy rather than as a sufficient explanation of renewal behavior on its own. The sensitivity analysis reduces concern that the final model driven mainly by operator identity, although it does not eliminate that concern.

5.4 Stratified Performance and Error Structure

While aggregate metrics provide a general overview, a deeper examination across different customer segments reveals that model performance is not uniform. When analyzing plan regimes, the model demonstrated a stronger ability to identify potential failures within the volume-based subset. This came with a notable trade-off, since predictions in this segment were highly skewed, achieving high recall for failures at the expense of very poor specificity among renewing customers. Given that this segment is considerably smaller than the time-based subset, these results should interpret with caution.

Regarding customer tenure, the differences were modest. The model maintained consistent performance for both new and long-term customers, showing only minimal variations in ROC-AUC, PR-AUC, and F1-scores. This stability suggests the model's predictive logic is robust across different stages of the customer life cycle.

The most significant operational insight emerged from the prior-spend segments. The model performed notably better in the low-spend group, where both minority retrieval and F1-scores were stronger. From an operational standpoint, this is highly valuable; since customers in low-spend, categories often require more precise and cost-effective retention strategies. By accurately identifying at-risk customers in this group, the company can apply more selective intervention measures. The detailed stratified metrics and their associated error profiles presented in Table IX.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

**Table IX. Stratified Performance and Error Profiles of the
Final Calibrated Model**

Segment Type	Segment	ROC- AUC	PR- AUC	F1	Balanced Accuracy	n	FP	FN	FP Rate among Renewals	FN Rate among Failures
segment_regime	Time- based	0.7003	0.4023	0.3814	0.5821	5170	2863	147	0.6991	0.1367
segment_regime	Volume- based	0.6745	0.5714	0.5756	0.5000	386	230	0	1.0000	0.0000
segment_tenure	<= 1 year	0.7044	0.4273	0.4131	0.5909	1547	808	50	0.6762	0.1420
segment_tenure	> 1 year	0.7107	0.4279	0.3964	0.5798	4009	2285	97	0.7300	0.1104
segment_spend	High prior spend	0.7012	0.4023	0.3782	0.5855	4554	2502	129	0.6902	0.1389
segment_spend	Low prior spends	0.6980	0.4930	0.4826	0.5481	1002	591	18	0.8443	0.0596

5.5 Practical Value for the Company

The final calibrated model provides more than just abstract risk scores; it offers a concrete foundation for selective prioritization. By ranking test-set events based on their predicted probability of renewal failure, the company can move from blanket marketing to targeted interventions. For instance, the top 10% highest-risk segment alone captured 23.15% of all renewal failures, with an observed failure rate of 51.35%. This represents a lift of 2.32 over the baseline prevalence, meaning the model is over twice as effective as random targeting.

The efficiency of this approach increases as the target group expands: the top 20% highest-risk segment captured 41.84% of all failures (a lift of 2.09), while the top 30% segment successfully identified 55.32% of all failures. This targeting efficiency visually demonstrated in Fig. 6, which illustrates the advantage of concentrating outreach on high-risk events.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

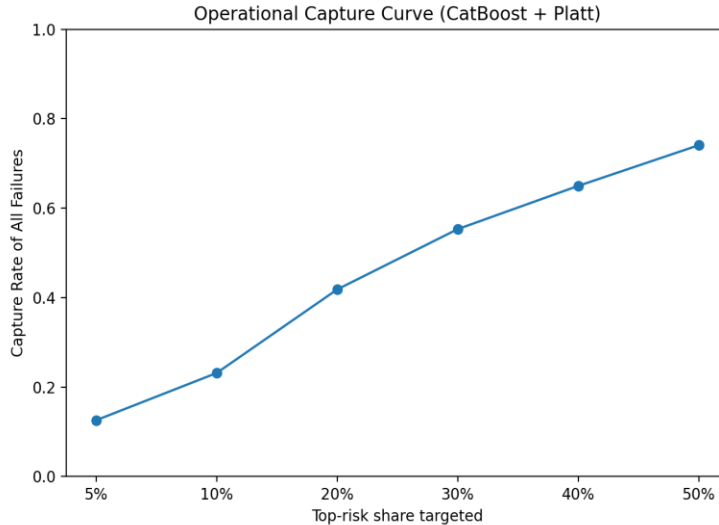


Fig. 6. Capture Curve Across Targeted High-Risk Segments

Even so, it is important to note that while these results support the model's value in prioritization, they do not serve as direct evidence of retention uplift. Since no prospective treatment experiment conducted, the findings confirm the model's ability to identify risk, not the causal effect of any specific intervention. Detailed targeting behavior summarized in Table X, while Table XI translates these ranked risk patterns into plausible operational actions for the company.

Table X. Targeting Efficiency of the Final Calibrated Model

Top share targeted	Events targeted	Failure rate in targeted set	Capture rate of all failures	Lift vs base rate
5%	277	0.5596	0.1259	2.5256
10%	555	0.5135	0.2315	2.3177
20%	1111	0.4635	0.4184	2.0922
30%	1666	0.4088	0.5532	1.8449
40%	2222	0.3600	0.6499	1.6250
50%	2778	0.3283	0.7409	1.4817

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Table XI. Suggested Operational Actions by Risk Pattern

Risk pattern identified by the model	Suggested operational action	Suggested timing	Expected operational value
Near-expiry event with weak recent commercial activity	Proactive renewal reminder with targeted follow-up	3–7 days before expiry	Improves focus on customers with weak renewal cadence
Correction-heavy history	Manual account review or workflow audit before contact	Before retention contact	Reduces waste caused by unresolved operational issues
Long gap since previous commercial event	Direct retention outreach rather than generic reminder	Close to expiry	More suitable for customers showing early disengagement
Low-spend but high-risk profile	Lightweight, low-cost retention message or digital reminder	Before expiry	Supports efficient targeting without over-investment
High-risk event within top ranked segment	Priority intervention by retention team	Earliest possible pre-expiry window	Concentrates effort on cases with highest expected yield

5.6 Feature Attribution and Error Interpretation

SHAP analysis used to clarify which variables most strongly influenced the final model. The leading contributors were Created by, days_until_expiry, commercial-events-last_90d, days_since_prev_commercial, prior_total_commercial_spend, corrections_last_90d, and tenure-related features. This ordering is consistent with the broader interpretation of the billing logs: renewal risk emerges from a mix of expiry timing, commercial continuity, spending history, and operational irregularity rather than from any single variable alone. Table XII reports the top features ranked by mean absolute SHAP contribution.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Table XII. Top SHAP Features Ranked by Mean Absolute Contribution

Feature	Mean SHAP
Created by	0.2000
days_until_expiry	0.1854
commercial_events_last_90d	0.1474
days_since_prev_commercial	0.1220
prior_total_commercial_spend	0.0561
corrections_last_90d	0.0441
commercial_tenure_days	0.0353
customer_tenure_days	0.0318
prior_commercial_count	0.0245
log_price	0.0233

The individual SHAP plots help make that structure visible at the case level. They are useful, but they should read as explanatory aids rather than as proof of causality. Figures 7 and 8 summarize the interpretation layer of the final model through global SHAP importance and local case-level explanation.

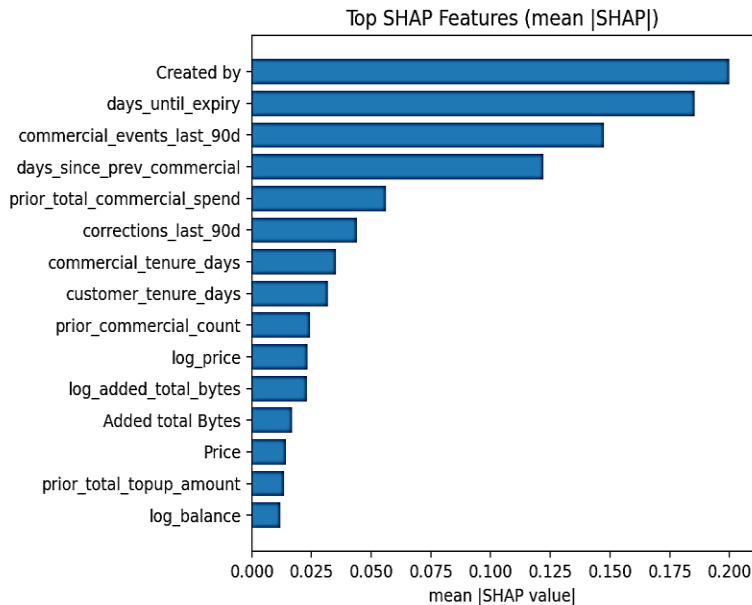


Fig. 7. SHAP Global Importance Plot

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

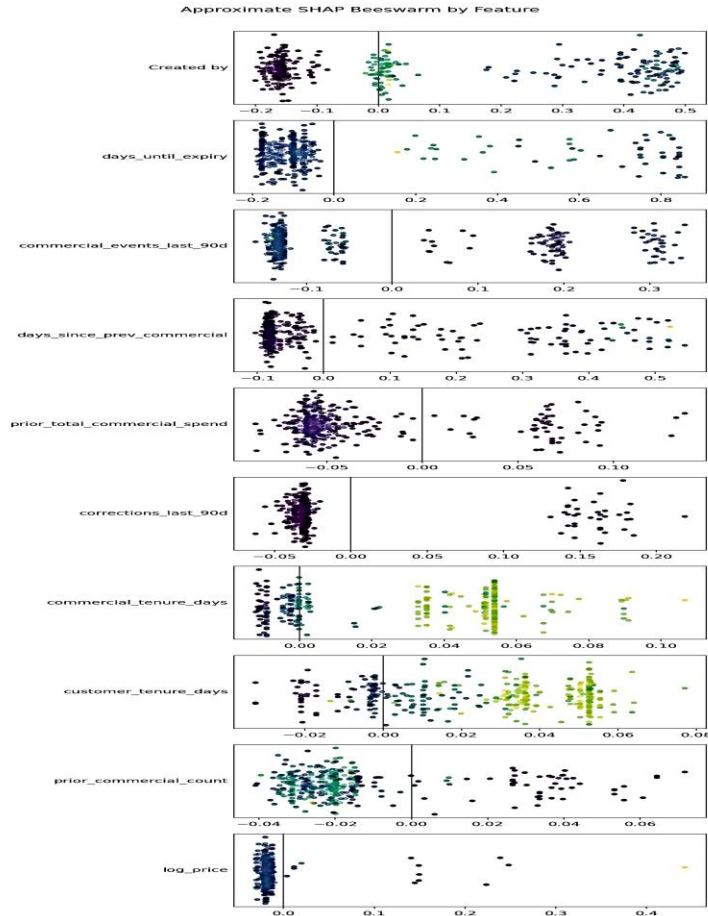


Fig. 8. SHAP Beeswarm Plot

6. Conclusion

6.1 Summary of Findings

This study examined event-level renewal-risk prediction in a Libyan ISP using operational billing logs rather than a pre-aggregated customer table. From 37,301 raw billing records, the analysis reconstructed 33,308 valid subscription-bearing events, of which 25.95% labeled as renewal failures under the primary 15-day post-expiry definition. The results show that these logs contain meaningful predictive structure. The rule-based baseline performed weakly, Logistic Regression captured part of the signal, and the non-linear models performed better overall. After Platt scaling, the final

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

CatBoost model preserved ranking quality while improving probability quality, reaching a Brier score of 0.1535.

6.2 Contributions, Limitations, and Future Work

The contribution of this work is primarily applied and contextual rather than methodological. It provides a careful case study of how a noisy operational billing system can be transformed into an event-level renewal-risk dataset, modeled with explainable tabular methods, and evaluated under a time-aware design that is closer to deployment conditions than a random split. In practical terms, the study suggests that renewal-risk modeling can support ISP retention operations when the goal is selective prioritization rather than uniform treatment of all customers.

The study remains bounded by the data that made it possible in the first place. The logs are strong in billing semantics, package lifecycle structure, and operational traces, but they do not include direct network telemetry such as latency, packet loss, outage alarms, throughput degradation, or infrastructure-fault measurements. Future work should link billing logs with complaint records, outage traces, and technical QoS indicators. The current binary renewal window could also be extended into a time-to-event framework, and the operational prioritization suggested by the model should be tested prospectively in real intervention settings. Although the billing logs were operationally messy, they proved sufficiently informative to support a credible and explainable renewal-risk modeling pipeline in an underrepresented ISP setting.

Data and Code Availability Statement

The raw operational billing logs used in this study contain commercially sensitive and customer-linked information and therefore cannot be shared publicly. The supervised dataset was derived through deterministic event-construction, eligibility-filtering, and label-generation rules described in the Methodology. Derived analytical outputs, selected aggregate results, and anonymized implementation logic may be made available from the corresponding author upon reasonable request and subject to organizational approval.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

Declaration of AI-assisted Technologies in the Writing Process

AI-assisted tools used for language editing, grammar checking, and improving the clarity of selected parts of the text. The authors reviewed and edited all outputs and takes full responsibility for the final content of the research.

References

- Alteer, S. A., & Alariyibi, A. (2024). Customer churn prediction using machine learning: A case study of Libyan internet service provider company. In 2024 IEEE 4th International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA) (pp. 605–612). IEEE.
- Awadh, D. S. A., Elburki, H. J., & Algmami, M. M. (2025). Assessment of service quality in telecommunications: A study of customer satisfaction in service centers in Benghazi, Libya. *The North African Journal of Scientific Publishing*, 3(3), 342–348.
- Balouli, H. E., Chabi, Y., & Benmoussat, M. (2025). Analysis of telecom churn in Algeria: A case study of Ooredoo. *Revue des Sciences Humaines*, 8(2).
- Barsotti, A., Gianini, G., Mio, C., Lin, J., Babbar, H., Singh, A., Taher, F., & Damiani, E. (2024). A decade of churn prediction techniques in the TelCo domain: A survey. *SN Computer Science*, 5(4), 404.
- Chang, V., Hall, K., Xu, Q. A., Amao, F. O., Ganatra, M. A., & Benson, V. (2024). Prediction of customer churn behavior in the telecommunication industry using machine learning models. *Algorithms*, 17(6), 231.
- He, C., & Ding, C. H. Q. (2024). A novel classification algorithm for customer churn prediction based on hybrid Ensemble-Fusion model. *Scientific Reports*, 14, 20179.
- Kimitei, S., Agiro, D., Ni, S., & Ni, H. (2025). Predictability & explainability of survival analysis in churn prediction. *Journal of Marketing Analytics*. Advance online publication.
- Liu, X., Xia, G., Zhang, X., Ma, W., & Yu, C. (2024). Customer churn prediction model based on hybrid neural networks. *Scientific Reports*, 14, 30707.

Applying Explainable Renewal-Risk Models to a Libyan ISP: A Case
Study Using Operational Billing Logs

<http://www.doi.org/10.62341/istj-vol39-2-as12>

- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30).
- Omer, A. A., Essgaer, M., & Ahmed, K. M. S. (2022). Using machine learning model to predict Libyan telecom company customer satisfaction. In *2022 International Conference on Engineering & MIS (ICEMIS)* (pp. 1–6). IEEE.
- Poudel, S. S., Pokharel, S., & Timilsina, M. (2024). Explaining customer churn prediction in telecom industry using tabular machine learning models. *Machine Learning with Applications*, 17, 100567.
- Prokhorenkova, L., Gusev, G., Vorobev, A. V., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems* (Vol. 31).
- Saleh, S., & Saha, S. (2023). Customer retention and churn prediction in the telecommunication industry: A case study on a Danish university. *SN Applied Sciences*, 5(7), 173.
- Shahabikargar, M., Beheshti, A., Zhang, X., & Foo, J. (2025). A comprehensive survey on customer churn analysis studies. *Journal of Information and Telecommunication*. Advance online publication.
- Sherif, A. T., & Elmabrouk, S. K. (2023). Studying customer behavior in the Libyan telecoms market using data visualization. *Journal of Pure & Applied Sciences*, 22(3), 286–290.
- Sikri, A., Jameel, R., Idrees, S. M., & Kaur, H. (2024). Enhancing customer retention in telecom industry with machine learning driven churn prediction. *Scientific Reports*, 14, 13097.