ORIGINAL ARTICLE
Electrica 2026; 26: 1-13

Explaining and Predicting Mobile Network–Driven Churn Using Machine Learning and SHapley Additive exPlanations–Based Attribution

Sümeyye Baş[1,2], Buket Bulut[1], Gizem Türkoğlu Ay[1], Ramazan Yağmur[1], Alpaslan Gökcen[1,3], Saliha Sezgin Alp[1]

[1]Private practice
[2]Department of Computer Engineering, İstanbul Technical University Faculty of Computer and Informatics, İstanbul, Türkiye
[3]Department of Computer Engineering, İstanbul Gelişim University Faculty of Engineering and Architecture, İstanbul, Türkiye

Cite this article as: S. Baş, B. Bulut, G. Türkoğlu Ay, R. Yağmur, A. Gökcen and S. Sezgin Alp, "Explaining and predicting mobile network–driven churn using machine learning and SHapley Additive exPlanations–based attribution," *Electrica*, 26, 0240, 2026. doi:10.5152/electrica.2026.25240.

WHAT IS ALREADY KNOWN ON THIS TOPIC?

- *Customer churn prediction in the telecommunications industry has been widely studied using machine learning methods, with tree-based ensemble models such as random forest and gradient boosting algorithms demonstrating strong predictive performance on demographic and billing data.*
- *Explainable AI techniques, particularly SHAP (SHapley Additive exPlanations), have been applied in churn prediction to provide global feature importance rankings and improve model transparency.*
- *Network quality indicators such as signal strength and throughput are*

Corresponding Author:
Alpaslan Gökcen

E-mail:
alpaslan.gokcen@turkcell.com.tr

Received: September 1, 2025
Revision Requested: December 29, 2025
Last Revision Received: May 4, 2026
Accepted: May 18, 2026
Publication Date: September 8, 2026

DOI: 10.5152/electrica.2026.25240

Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ABSTRACT

Customer churn remains a critical concern in the mobile telecommunications industry, where retaining subscribers is essential for long-term profitability. This study presents an explainable machine learning framework to detect and interpret network-driven churn by integrating demographic data with key network performance indicators, including signal quality metrics e.g., reference signla received power, signal to interference plus noise ratio, download/upload throughput, and latency-related measurements derived from the radio access network. Using the CatBoost classifier on a proprietary dataset of 115 264 subscribers constructed with a balanced 50/50 class distribution to prevent majority-class bias, the model achieves an F1 score of 81.0% and an area under the receiver operating characteristic curve of 0.884, demonstrating strong discriminative performance. A central contribution is the introduction of a systematic SHapley Additive exPlanations (SHAP)-based attribution method that distinguishes network-driven churn from other churn types. For each churn prediction, positive SHAP values are extracted and categorized as network-related or non-network-related; instances where network features contribute the dominant share of the total positive SHAP attribution are labeled as network-driven churn. The framework is validated on both the proprietary Turkcell dataset and three publicly available telecom churn benchmarks, demonstrating generalizability. The proposed approach bridges predictive modeling and real-world network optimization, offering telecom operators a scalable and interpretable tool for targeted customer retention.

***Index Terms*—**Customer churn prediction, customer retention strategies, explainable AI, key performance indicators, machine learning, SHAP analysis

I. INTRODUCTION

In the highly competitive mobile telecommunications industry, customer retention is a key strategic priority, driven by high churn rates and low switching costs [1]. In Türkiye, Turkcell is one of the major operators, holding 41.2% of the mobile market as of Q2 2024 [2]. Maintaining this position requires not only attracting new customers but also minimizing churn by understanding and addressing its root causes.

Customer churn, defined as the discontinuation of service, may occur voluntarily when customers seek better offers or involuntarily, due to external constraints such as financial hardship [3]. Moreover, churn behavior is heterogeneous: contractual churn is typically observable, while non-contractual churn often lacks explicit indicators [4]. This diversity increases the challenge of identifying and mitigating churn across the customer base.

Industry reports suggest that churn rates can exceed 30%, motivating a strategic shift from customer acquisition to customer retention [5]. Beyond financial losses, churn has social effects;

recognized as contributors to customer dissatisfaction, yet most existing churn models treat all churn types uniformly without distinguishing network-driven cases from those caused by pricing, competition, or customer service failures.

WHAT DOES THIS STUDY ADD ON THIS TOPIC?

- *This study introduces a systematic SHAP-based attribution method that identifies and demarcates network-driven churn as a distinct, actionable sub-class, enabling targeted engineering interventions rather than generic commercial retention strategies.*
- *A large-scale, real-world dataset of 115 264 mobile subscribers integrating radio access network key performance indicators (reference signla received power, signal to interference plus noise ratio, throughput, latency) with demographic data is used to train and validate the CatBoost classifier, achieving an F1 score of 81.0% and an area under the receiver operating characteristic curve of 0.884.*
- *The proposed framework is validated across both a proprietary Turkcell dataset and three publicly available telecom churn benchmarks, demonstrating generalizability and offering telecom operators a scalable, interpretable tool for customer retention.*

dissatisfied users can influence others in their network to leave, amplifying attrition through indirect peer effects [6].

Machine learning (ML) has emerged as a dominant approach for churn prediction due to its ability to uncover complex, nonlinear relationships in high-dimensional customer data [1]. However, many ML models act as "black boxes," offering high accuracy at the expense of interpretability, limiting their usefulness in operational decision-making [7].

Despite progress in churn modeling, a critical dimension remains underexplored: the direct influence of network-layer technical performance on customer attrition. Poor signal quality, degraded throughput, and frequent service disruptions are among the leading causes of involuntary and technically motivated churn, yet most existing frameworks do not distinguish these cases from churn driven by pricing, competitor offers, or customer service failures. This distinction is operationally vital; network-driven churn demands engineering interventions, not commercial incentives, and its absence limits the actionability of conventional churn prediction systems. Furthermore, while explainability techniques such as SHapley Additive exPlanations (SHAP) have been applied in churn contexts, no prior work has systematically leveraged SHAP attribution to demarcate network-driven churn as a distinct, actionable sub-class. The dataset used in this study was constructed to be representative of the predominant individual subscriber segment in Türkiye, comprising approximately 83.4% of all mobile users, with a balanced 50/50 class distribution to prevent majority-class bias and validated by domain experts to ensure semantic consistency across customer segments.

To address this gap, this study proposes an explainable ML framework to predict and interpret network-driven churn cases primarily caused by poor technical service quality. The framework integrates demographic information with key performance indicators (KPIs) and applies the CatBoost classifier to achieve high prediction performance. To ensure transparency and trust, SHAP-based attribution is used to explain model decisions at both global and individual levels, enabling targeted engineering interventions and personalized retention strategies.

The majority of existing churn prediction studies rely exclusively on demographic or billing data, overlooking the direct impact of network-layer technical performance on customer attrition. While SHAP-based explainability has been applied in churn contexts, no prior work has systematically employed SHAP attribution to demarcate network-driven churn as a distinct sub-class requiring specialized engineering responses. Moreover, most models provide only global feature importance rankings, lacking the instance-level granularity necessary for personalized retention actions. The present study directly addresses these gaps.

The primary contributions of this study are as follows:

1. A systematic SHAP-based attribution method is introduced to distinguish network-driven churn from other churn types, enabling targeted engineering interventions at the individual level.
2. A robust preprocessing pipeline is developed for large-scale, real-world telecom data, addressing class imbalance, multicollinearity, and missing values at scale.
3. The CatBoost classifier is benchmarked against five competing algorithms: decision tree (DT), random forest (RF), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) across four datasets, including a proprietary operator dataset and three public benchmarks, demonstrating consistent superiority in both F1 score and area under the receiver operating characteristic curve (AUC-ROC).
4. Per-user churn explanations are generated via SHAP values, supporting individualized retention strategies and network performance optimization.
5. The framework is validated on both proprietary and publicly available datasets, confirming its generalizability across different telecom environments.

The remainder of this article is organized as follows. Section II reviews related work on churn prediction and explainable AI. Section III describes the proposed methodology. Section IV details the dataset construction, model development, and evaluation process. Section V presents and discusses the experimental results. Section VI concludes the paper and outlines directions for future work. Finally, Section VII presents limitations.

II. RELATED WORK

Several publicly available datasets are commonly used in churn prediction research, including Churn-Data-UCI (5000 samples) [8], IBM Telco (7043 samples) [9], and Churn-in-Telecom (3333 samples) [10]. These datasets present various common challenges that must be carefully addressed to build robust and generalizable solutions. First, null values in the data can distort analysis and lead to unreliable predictions if not properly handled. Second, strong correlations among features lead to redundancy and make models harder to interpret, which is particularly important when explainability matters. Third, a significant class imbalance, often due to a much higher number of non-churned customers compared to churned ones, makes it difficult for models to accurately detect the minority class. Finally, many real-world churn phenomena involve nonlinear relationships between input features and churn outcomes, which limits the effectiveness of simple linear models and calls for the use of more sophisticated algorithms such as tree-based models or neural networks. Addressing these challenges is crucial for improving model performance and ensuring reliable deployment in real-world scenarios.

Recent studies have demonstrated the effectiveness of unsupervised and hybrid deep learning approaches for churn prediction, particularly in complex customer behavior settings. Hasumoto and Goto proposed a variational autoencoder–based framework that leverages unsupervised deep learning to extract latent features from customer purchase histories in platform businesses, where consumer interactions are inherently complex and non-linear [11]. Their approach emphasizes unsupervised representation learning to capture underlying purchasing behaviors without relying on extensive labeled data, and incorporates these latent representations into churn prediction models, enabling improved predictive capability while providing insights into customer behavior patterns.

Similarly, Bilal et al. introduced an ensemble-based methodology that integrates unsupervised clustering with supervised classification algorithms to enhance churn prediction in the telecom industry [12]. In this hybrid approach, customer data is first segmented into meaningful groups using clustering techniques such as K-means and K-medoids, and subsequently, multiple classification methods are applied within an ensemble framework. The study highlights that combining unsupervised and supervised learning paradigms enables more effective modeling of customer heterogeneity and behavioral diversity, emphasizing the value of hybrid strategies in large-scale churn prediction tasks.

Ahmad et al. used nine months of data to predict the two-month period. They addressed missing data by removing records with over 90% missing features and features with more than 70% missing values. Then, missing categorical values were replaced with "other," numerical values with the mean. Features with numeric values and limited duplicates were excluded due to poor model performance. Trying both oversampling and undersampling, they revealed that some algorithms yield better results with oversampling, while others perform more effectively with undersampling [13]. Poudel et al. worked with data having 7043 samples and 30 features with a 26% churn rate. They dropped null values, duplicated columns, and unique value columns, and employed one-hot encoding to categorical features. These preprocessing steps ended up with 1183 features. They utilized SHAP to understand both feature importances and the reason individual customer churns [14]. Cenggoro et al. used a publicly available telecom customer churn dataset with 3333 customers and 19 attributes. The authors proposed deep learning models with vector embedding to detect customer churn in an interpretable way [15]. Ljubičić et al. focused on spatio-temporal patterns and entropy of customers' financial transactions. Their method offers insights into how behavioral features influence churn across different demographic groups, incorporating dynamic behavioral traits derived from individual transaction records [16].

Several key outcomes that may be useful for handling datasets have been identified in previous works. Manzoor et al. revealed that older individuals are less likely to churn, and interactions with churners increase churn probability [6]. According to Ahmad et al., most churners stay longer than non-churners who do not make any purchases, and the majority of churners also have low balances. They also demonstrated that younger customers (under age 32) are more likely to churn [17]. As can be seen in these examples, integrating social network analysis (SNA) with statistical features is crucial for churn prediction.

While the importance of behavioral and social context in churn prediction has been emphasized in past studies, few have provided instance-level explanations that align with model logic and are accessible to stakeholders. Despite the diverse real-world reasons behind customer churn, such as high pricing, poor network coverage, frequent service disruptions, attractive competitor offers, unresponsive customer support, billing issues, and lack of loyalty incentives, it is often modeled in a uniform manner. These gaps are addressed through a hybrid framework that combines model-level insights with instance-level rules to generate interpretable, customer-specific churn reasons. Additionally, network features and loyalty indicators are incorporated to uncover insights that may be overlooked by purely statistical approaches.

III. METHODOLOGY

The proposed methodology is designed to integrate diverse customer data sources, process and cleanse the information, predict churn using advanced ML models, and explain each prediction through explainable AI techniques to distinguish network-driven churners. The entire process is structured in two main stages: extract, transform, load (ETL) Pipeline and Model Pipeline, as shown in Fig. 1.

A. Extract, Transform, Load Pipeline: Data Collection and Preparation

The first phase focuses on aggregating heterogeneous data sources relevant to churn behavior:

Before any processing or modeling, all customer data was fully anonymized to prevent the identification of individuals. This ensured compliance with institutional data privacy guidelines and regulatory standards of Turkish personal data protection law.

Customer experience management tools provide subjective quality indicators such as customer satisfaction surveys and feedback.

Telco KPI Data from sites include objective, technical measurements (e.g., reference signla received power (RSRP), signal to interference plus noise ratio (SINR), download/upload speeds) that capture network performance from the radio access layer.

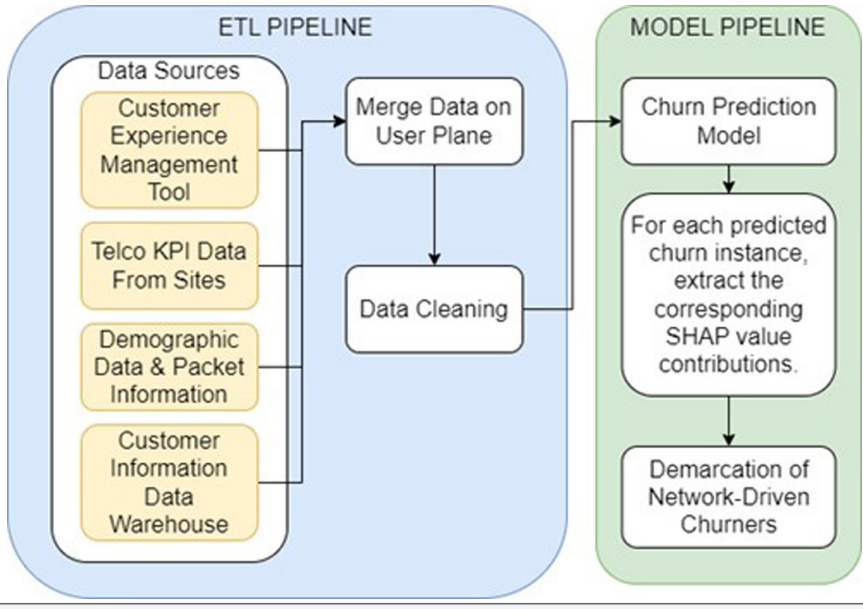

Fig. 1. Architecture of the churn prediction and explainable AI workflow.

Demographic and packet information (e.g., age, gender, location, plan type) is crucial for understanding customer segments and personalized behaviors.

Customer information data warehouse offers historical and behavioral insights, including past complaints, tenure, and usage patterns.

All datasets are merged at the user level through a component titled Merge Data on User Plane, ensuring a unified record per customer. Following the integration, a data cleaning process is applied, addressing missing values, outliers, and feature formatting to optimize model readiness.

B. Model Pipeline: Prediction and Explanation

A supervised classification model is developed to predict churn. As discussed in Section IV, tree-based gradient boosting methods such as CatBoost [18] are particularly suitable due to their superior performance on tabular and categorical datasets.

Once the model generates churn predictions, SHAP (SHapley Additive exPlanations) values are employed to explain the output of each individual prediction [19] . For every customer predicted to churn, feature-level contribution scores are extracted.

This step enables the identification of feature importance per individual, moving beyond global feature rankings to case-specific interpretations [20].

C. Demarcation of Network-Driven Churners

The final step involves the demarcation of network-driven churners, a sub-class of churners whose model explanations are dominated by network-related features (e.g., low signal strength, high drop rates). A thresholding mechanism on SHAP values is used to flag churners primarily affected by technical service quality rather than pricing, customer service, or other factors. This allows for targeted action plans by engineering teams.

IV. MODEL DEVELOPMENT AND EVALUATION PROCESS

A. Turkcell Dataset Construction for Network-Centric Churn Prediction

This study concentrates on predicting intentional churn among individual mobile customers, a segment that comprises approximately 83.4% of all mobile subscribers in Türkiye [2]. Focusing on this predominant customer group allows for more precise modeling outcomes that align with the most common churn patterns observed in practice. In line with institutional privacy standards and ethical research practices, all datasets were fully anonymized prior to analysis, ensuring the protection of personally identifiable information.

The dataset utilized in this study integrates multiple heterogeneous data sources to offer a multidimensional view of customer behavior and network experience. Specifically, it includes:

- Key performance indicators: KPIs are technical metrics derived from probes that monitor interactions between mobile devices and network infrastructure. At the user plane level, KPIs are captured across multiple service categories including instant messaging, streaming, web browsing, voice, user datagram protocol (UDP), and quick UDP internet connections and encompass a broad set of performance dimensions such as packet loss, delay, jitter, uplink/downlink throughput, and round-trip time. Each of these metrics is further broken down across multiple temporal and spatial granularities, enabling a fine-grained characterization of the end-user network experience. In addition, radio access layer measurements such as signal quality indicators (e.g., RSRP, SINR) and downlink/uplink speed metrics are incorporated to capture the underlying network conditions driving user experience degradation.
- Demographic and location information: It includes customer attributes such as age, gender, region, and device usage patterns, as well as the geographical location of service delivery.
- Network traffic categories: Packets transmitted over the network were segmented into classes based on protocol and content type,

including UDP, TCP, video, instant messaging, and voice over internet protocol. These categories were used as predictive features to capture user behavior patterns and application-level network usage.
- Operations support system data: It includes Infrastructure-related metrics, including base station, cell, and antenna-level performance information, providing technical granularity on the quality of service delivered to individual users.

This rich dataset was initially collected for a 30-day window corresponding to the month of December, a period characterized by end-of-year behavioral shifts in mobile usage. To generalize behavior over time and reduce noise, hourly KPI data was aggregated using median values over three distinct sliding windows of 7, 15, and 30 days. This temporal aggregation strategy enabled the generation of three separate features per KPI, helping to uncover both short-term fluctuations and long-term patterns in customer experience.

1) Data Quality Challenges and Preprocessing Solutions

The original dataset consisted of over 650 features and required extensive preprocessing to ensure analytical validity and modeling robustness. The following procedures were implemented:

Handling missing values: Null entries were prevalent across KPI fields, primarily due to varied customer behavior and inconsistent network exposure (e.g., customers using only 3G services lacked 4G data entirely). Furthermore, some columns included corrupted or incomplete values. Based on expert domain input, missing numeric values were imputed using median values, while missing categorical features were filled with either the mode or a placeholder (e.g., "unknown"). Features with over 80% missing data were excluded entirely to preserve modeling reliability.

Addressing multicollinearity and sparse categorical classes: High correlations among certain KPI metrics posed challenges, especially for linear or less robust models. Multicollinearity was mitigated by either removing redundant features or combining highly similar variables. In cases where categorical variables had sparse class distributions, minor categories were merged to form broader, analytically meaningful groups.

Class imbalance correction: As is common in churn prediction tasks, the original dataset exhibited a significant imbalance between churn and non-churn instances, with non-churners constituting the majority class. To address this, random undersampling was applied by randomly reducing the non-churn class until a balanced distribution was achieved. The resulting dataset comprises 57 632 churn and 57 632 non-churn instances (total: 115 264 samples), representing a 50/50 class distribution. This approach improved the model's sensitivity to churn signals while avoiding bias toward the majority class and is consistent with methods commonly recommended in the churn prediction literature.

Feature Engineering: After cleaning, the dataset was reduced to 625 features across 115 264 samples. Of these, 29 features related to demographics and packet usage, while the remaining 596 captured technical aspects of network performance. In addition to the temporal aggregation discussed above, further feature selection was guided by variance, mutual information scores, and expert feedback to retain the most informative predictors.

This rigorous data preparation strategy ensured the reliability and integrity of the modeling process while addressing real-world constraints such as class imbalance, incomplete data, and correlated metrics. More importantly, by incorporating network-layer performance data, protocol-level traffic categorization, and customer-centric attributes, the dataset laid a strong foundation for distinguishing network-driven churn from other types of customer attrition. This distinction is pivotal for telecom operators seeking actionable insights that link technical performance to customer retention.

B. Model Training

The churn prediction task was formulated as a supervised binary classification problem, where the goal is to classify customers as churners or non-churners.

- Algorithm choice: The CatBoost gradient boosting classifier was selected due to its robust handling of categorical variables, resistance to overfitting, and superior performance on tabular data.
- Training-testing split: All the datasets were divided into training and testing (70% train and 30% test) subsets using a stratified split to preserve the original class distribution.
- Hyperparameter optimization: A grid search procedure was employed to fine-tune key hyperparameters such as learning rate, depth, and number of iterations, optimizing for the F1 score to balance precision and recall.
- Cross-validation: To ensure generalizability and prevent overfitting, k-fold cross-validation (with $k=5$) was performed on the training data, with the final model selected based on average validation performance. Table I summarizes the final hyperparameter configuration of the CatBoost classifier, determined through a grid search procedure optimized for the F1 score using five-fold stratified cross-validation.

C. Model Evaluation

Selecting an appropriate algorithm is critical for both the effectiveness of churn prediction and the interpretability of the model's decisions, especially when dealing with complex and highly correlated customer data. Prior research frequently favors tree-based algorithms due to their superior ability to manage such feature characteristics. Among these, boosting methods have demonstrated particular promise by balancing high predictive accuracy with a degree of interpretability that surpasses more opaque models like deep neural networks [7].

Decision trees (DTs) serve as the foundational models, creating predictions by learning simple decision rules derived from data features [17]. However, single DTs tend to suffer from overfitting and instability. To mitigate this, ensemble methods such as RF aggregate multiple DTs trained on bootstrap samples with added feature randomness at each split, enhancing accuracy and reducing variance [21].

Boosting algorithms further improve performance by sequentially focusing on difficult-to-classify instances. Adaptive Boosting adjusts the weights of misclassified samples to build subsequent classifiers that emphasize these challenging cases, thereby improving overall accuracy [22]. Extreme Gradient Boosting refines this concept by incorporating regularization, second-order gradient optimization, and parallelized tree construction, offering scalable and efficient performance improvements [23].

TABLE I. CATBOOST HYPERPARAMETER SEARCH SPACE AND FINAL SELECTED VALUES, DETERMINED VIA GRID SEARCH WITH FIVE-FOLD STRATIFIED CROSS-VALIDATION. THE F1 SCORE WAS USED AS THE OPTIMIZATION CRITERION TO BALANCE PRECISION AND RECALL IN THE CONTEXT OF CHURN PREDICTION

Hyperparameter	Search Range	Selected Value	Description
Iterations	100, 300, 500, 1000	500	Number of boosting trees to build
learning_rate	0.01, 0.05, 0.1	0.05	Step size for gradient update per iteration
depth	4, 6, 8, 10	6	Maximum depth of each decision tree
l2_leaf_reg	1, 3, 5, 10	3	L2 regularization coefficient to prevent overfitting
loss_function	—	Logloss	Objective function for binary classification
eval_metric	—	F1	Metric used for model selection during optimization
early_stopping_rounds	—	50	Training stops if no improvement after 50 rounds
random_seed	—	42	Fixed seed for reproducibility
fold_count (CV)	—	5	Number of folds in stratified cross-validation
train/test split	—	70%/30%	Stratified split preserving class distribution

Light Gradient Boosting Machine introduces additional scalability and efficiency, using a leaf-wise tree growth strategy that contrasts with level-wise approaches like XGBoost. Its native handling of categorical features further improves performance on tabular data [24].

CatBoost stands out as a gradient boosting framework explicitly designed for categorical data. Utilizing ordered boosting and feature selection mechanisms to combat overfitting, CatBoost achieves strong predictive performance while maintaining interpretability [18]. Its effectiveness on datasets with a large proportion of categorical variables made it the preferred model in this study.

All these models were trained and evaluated using consistent preprocessing, including handling missing values, multicollinearity, and addressing class imbalance through random undersampling. Hyperparameter tuning was conducted via grid search with stratified k-fold cross-validation to ensure robust and fair comparison across models.

Model evaluation metrics included accuracy, precision, recall, and F1 score, providing a comprehensive view of each model's ability to correctly classify churners while balancing false positives and negatives.

In total, six models were compared in this study: DT, RF, AdaBoost, XGBoost, LightGBM, and CatBoost. These models were evaluated on multiple datasets, including the proprietary Turkcell dataset as well as three publicly available telecom churn datasets: the Churn-Data-UCI dataset [8], the IBM Telco dataset [9], and the Churn-in-Telecom dataset [10]. This comprehensive evaluation across diverse datasets allowed us to identify the most effective and interpretable model for churn prediction in different telecom environments.

This comprehensive comparison allowed for identifying the best-performing and most interpretable model for churn prediction within the given dataset.

D. Model Evaluation Metrics

The metrics are formally defined as follows:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Recall = TP / (TP + FN) \quad (2)$$

$$Precision = TP / (TP + FP) \quad (3)$$

$$F1Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

Where (1-4) use the following definitions:

- TP: True Positives - number of correctly predicted churned customers
- TN: True Negatives - number of correctly predicted non-churned customers
- FP: False Positives - number of non-churned customers incorrectly predicted as churned
- FN: False Negatives - number of churned customers incorrectly predicted as non-churned

Accuracy measures the overall correctness of the model, while recall focuses on capturing as many actual churners as possible, minimizing missed cases. Precision reflects the trustworthiness of churn predictions, i.e., how many predicted churns are true churns. The F1 score harmonizes precision and recall to provide a balanced single metric accounting for both false positives and false negatives.

E. Identification of Network-Driven Churn

To distinguish churn cases primarily influenced by network-related factors, a systematic SHAP-based attribution method was developed. This approach focuses on the positive SHAP values generated by the churn prediction model, isolating features that contribute most strongly to increasing the likelihood of churn.

1) SHapley Additive exPlanations Values: Definition and Usage in this Study

To enhance model interpretability and provide actionable insights, SHAP values were computed for each prediction. SHapley Additive exPlanations values quantify the contribution of individual features to the prediction outcome, allowing both global and local explanation of model behavior [20]. This transparency supports the identification of key churn drivers and underpins the subsequent classification of network-driven churn cases.

SHapley Additive exPlanations values quantify the contribution of each feature to a specific model prediction in a fair and consistent manner.

Given a model prediction $f(x)$ for an instance $x=(x1,x2,...,x_m)$, where m is the number of features, the SHAP value for a feature is defined as shown in (5):

$$\varnothing_i = \sum_{s\subseteq F\setminus\{i\}} \frac{|s|!(M-|s|-1)!}{M!}\left[f_{s\cup\{i\}}\left(X_{s\subseteq F\setminus\{i\}}\right)-f_s\left(x_s\right)\right] \quad (5)$$

Where:

- M is the total number of input features.
- F is the set of all features $\{1,2,3,\ldots,M\}$.
- S is a subset of features not containing feature$i$.
- $f_S(x_S)$ the model's prediction using only features in subset S.
- The term $|S|!(M-|S|-1)!/M!$ is the Shapley weight, assigning importance based on subset size.

This formula in (5) calculates the average marginal contribution of feature x_i across all possible feature subsets of features, capturing how much the prediction changes when the feature is added [19, 25].

For each churn-predicted instance, only positive SHAP values (those that increase the predicted probability of churn) are considered in (6):

$$\varnothing_i^+ = \max(\varnothing_i, 0) \quad (6)$$

The following is then calculated:

Total positive SHAP attribution as shown in (7):

$$\Phi_{total}^+ = \sum_{i=1}^{M} \varnothing_i^+ \quad (7)$$

Network-related positive SHAP attribution as shown in (8):

$$\Phi_{network}^+ = \sum_{i\in N} \varnothing_i^+ \quad (8)$$

Non-network positive SHAP attribution as shown in (9):

$$\Phi_{non-network}^+ = \sum_{i\in F\setminus N} \varnothing_i^+ \quad (9)$$

Where:

- $N \subseteq F$ is the set of all network-related features.
- Φ_i^+ is the positive SHAP value of feature i defined as $max(\Phi_i, 0)$.

Churn classification rule:

$$If\ \Phi_{network}^+ > \Phi_{non-network}^+ \Rightarrow label\ as\ Network-Driven\ Churn \quad (10)$$

This rule in (10) identifies churn instances where technical/network issues are the primary contributors to the model's decision.

For each customer predicted to churn, the positive SHAP values indicating feature contributions toward the churn class were extracted and ranked in descending order. From this ranking, a fixed number of the top contributing features were selected to represent the most influential factors driving churn for that specific instance.

Each selected feature was then categorized as network-related or non-network-related based on its semantic and technical characteristics. Subsequently, the sum of the SHAP values of all network-related features within this subset was calculated and compared to the sum of the remaining non-network features.

If the cumulative contribution of network-related features exceeded that of non-network features, the churn instance was classified as network-driven. This thresholding strategy enables a transparent and interpretable separation of churn cases dominated by network performance issues from those influenced by other factors such as customer profile or service plans.

To identify whether churn was primarily driven by network-related conditions, SHAP values of the top-K features with the highest positive contribution were analyzed for each churned instance. These features were categorized as network-related or non-network-related based on domain semantics. Instances where the cumulative contribution of network-related features exceeded the alternative group were classified as network-driven churn.

The empirical rationale behind the selection of K and the performance of this attribution scheme are detailed in Section V.

V. RESULTS AND DISCUSSION

To preserve the confidentiality of internal network architecture and service design, individual feature names and their corresponding user-level distributions cannot be disclosed. This is due to commercial sensitivity and data protection policies defined by the service provider. Nevertheless, all features were semantically grouped based on their technical origin into network-related and non-network categories by subject-matter experts. These groupings were applied consistently during SHAP attribution analysis, allowing the classification of network-driven churn without exposing proprietary details.

While exact numbers and feature identities are withheld, the selection and labeling process is thoroughly documented in methodology and validated on public datasets, where full transparency is possible.

All experiments were conducted on a high-performance Linux-based server to ensure repeatability and consistent computational performance. The hardware and software configuration of the system is detailed below in Table II.

All model training, hyperparameter tuning, SHAP computations, and evaluation metrics were performed within this environment. All analyses were executed using Jupyter Notebooks running on the same environment, ensuring interactive experimentation and transparency in preprocessing, modeling, and evaluation steps.

Across all datasets, CatBoost consistently achieved either the best or near-best performance, especially excelling in the Churn-in-Telecom dataset with a 93.5% accuracy and in the IBM Telco dataset with an

TABLE II. SYSTEM CONFIGURATION OVERVIEW

Component	Description
Operating system	CentOS Linux 8 (Kernel 4.18.0 - 553.50.1.el8_10.x86_64)
Processor	Dual Intel® Xeon® CPU E5-2695 v3 @ 2.30GHz
Logical cores	56 logical cores (2 sockets × 14 cores × 2 threads)
Max clock speed	3.3 GHz
Memory (RAM)	128 GB RAM (approximately 109 GB available during execution)
GPU	Matrox MGA G200EH (used only for display, no CUDA acceleration)
Python version	Python 3.11.13
Virtual environment	Conda/venv-managed, with key packages including CatBoost, XGBoost, LightGBM, SHAP, and Scikit-learn

Conda, Anaconda environment manager; CPU, central processing unit; CUDA, Compute Unified Device Architecture; GHz, gigahertz; GPU, graphics processing unit; OS, operating system; RAM, random access memory; venv, Python virtual environment tool.

F1 score of 87.2%, as shown in Table III. Notably, even on the proprietary Turkcell dataset which features high dimensionality, multicollinearity, and real-world noise, CatBoost outperformed all other models, achieving an F1 score of 81.0%, confirming its robustness and scalability. These results validate the selection of CatBoost as the final model for both prediction and interpretability.

To further assess the discriminative ability of the CatBoost model, the ROC curve was computed on the Turkcell test set, as illustrated in Fig. 2. The AUC was measured at 0.884, indicating strong separability between churners and non-churners. An AUC value approaching 1.0 reflects the model's high capability to correctly rank a randomly chosen churner above a randomly chosen non-churner. This result complements the F1 score of 81.0% reported in Table III and confirms that the CatBoost classifier is not only accurate but also well-calibrated in distinguishing the two classes, even in the presence of real-world noise and high dimensionality inherent to the Turkcell dataset.

To support the explainability of churn predictions and enable transparent categorization of network-driven churn, an empirical analysis was conducted to determine an optimal number of features (Top-K)

TABLE III. PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS ACROSS DIFFERENT TELECOM CHURN DATASETS

Dataset	Model	Accuracy	Precision	Recall	F1 Score
Churn Data - UCI	Decision Tree	0.821	0.823	0.821	0.822
	Random Forest	0.913	0.916	0.913	0.911
	Adaboost	0.812	0.810	0.812	0.808
	XGBoost	0.901	0.901	0.901	0.900
	LightGBM	0.904	0.905	0.904	0.902
	CatBoost	0.913	0.913	0.913	0.912
IBM Telco	Decision Tree	0.763	0.764	0.763	0.763
	Random Forest	0.866	0.875	0.866	0.865
	Adaboost	0.809	0.820	0.809	0.808
	XGBoost	0.852	0.859	0.852	0.851
	LightGBM	0.859	0.862	0.859	0.858
	CatBoost	0.873	0.879	0.873	0.872
Churn-in-Telecom	Decision Tree	0.859	0.862	0.859	0.860
	Random Forest	0.924	0.924	0.824	0.923
	Adaboost	0.818	0.809	0.814	0.811
	XGBoost	0.921	0.920	0.921	0.919
	LightGBM	0.921	0.920	0.921	0.920
	CatBoost	0.935	0.934	0.935	0.933
The Dataset	Decision Tree	0.732	0.730	0.731	0.731
	Random Forest	0.793	0.789	0.786	0.792
	Adaboost	0.792	0.789	0.796	0.792
	XGBoost	0.807	0.805	0.806	0.806
	LightGBM	0.809	0.809	0.807	0.808
	CatBoost	0.812	0.809	0.811	0.810

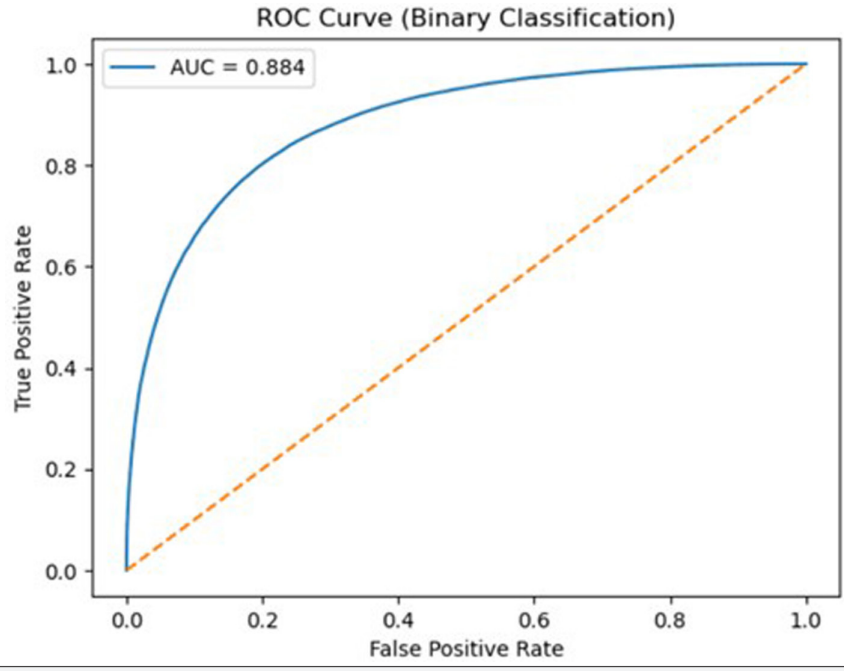

Fig. 2. Receiver operating characteristic curve of the CatBoost classifier evaluated on the Turkcell test set. The area under the curve score of 0.884 demonstrates strong discriminative performance in distinguishing churners from non-churners.

that would capture a meaningful proportion of the positive SHAP attribution at the individual user level.

For each user predicted as churned, SHAP values were sorted in descending order based on their positive contribution to the churn prediction. The cumulative impact was then computed across increasing values of K, ranging from 1 to 100. This process aimed to quantify how much of the predictive explanation could be accounted for by the most influential features alone.

The results of this analysis, visualized in Fig. 3, reveal a distinct trend: as more features are included, the total explanatory power increases, but the marginal gain from adding additional features declines sharply beyond the 10th feature. On average, the top 10 features were found to account for approximately 50% of the total positive SHAP attribution across the user base. This finding identifies a natural inflection point at K = 10, where a substantial portion of the predictive reasoning is captured without introducing unnecessary complexity.

The choice of this threshold was supported not only by empirical evidence but also through validation by domain experts. Network engineers and data scientists at Turkcell reviewed the top contributing features and confirmed that the most technically and semantically meaningful variables were consistently captured within the top 10. From an operational standpoint, limiting interpretability to 10 features provides a pragmatic trade-off between analytical depth and human readability, ensuring that downstream teams can act upon the model's explanations in a scalable and comprehensible way.

It is important to note that this threshold is not intended as a universal rule. The optimal value of K may vary depending on the feature set, customer behavior, and infrastructure specifics of different telecom operators. In this context, K = 10 represents an operator-specific and empirically determined configuration that balances model transparency with interpretability. Other practitioners may recalibrate this value based on their own technical and business constraints.

Following the identification of the top 10 features, each was categorized as either network-related or non-network-related based on its semantic origin and technical function. For each churned instance, the sum of SHAP values within these two groups was compared. If the aggregated contribution from network-related features exceeded that of non-network features, the case was labeled as network-driven churn. This method enabled the detection of churn cases primarily influenced by degraded service quality and technical performance issues, thereby supporting targeted operational interventions.

By combining SHAP attribution intensity with expert-validated feature selection, the proposed approach enables a principled and scalable strategy for churn explainability in real-world telecom settings. The threshold selection process, while grounded in operator-specific constraints, is methodologically sound and repeatable in other deployment scenarios.

VI. CONCLUSION AND FUTURE WORK

This study introduced a comprehensive and explainable ML framework to predict and interpret customer churn in the mobile telecommunications sector, with a special focus on identifying network-driven churners. By integrating high-resolution technical KPIs with demographic and behavioral data and leveraging the CatBoost classifier alongside SHAP values, the framework offers both high predictive performance and transparent, actionable insights.

From a practical standpoint, the proposed framework has direct real-world applicability. Network engineers can use the network-driven churn labels to prioritize cell sites and coverage areas requiring

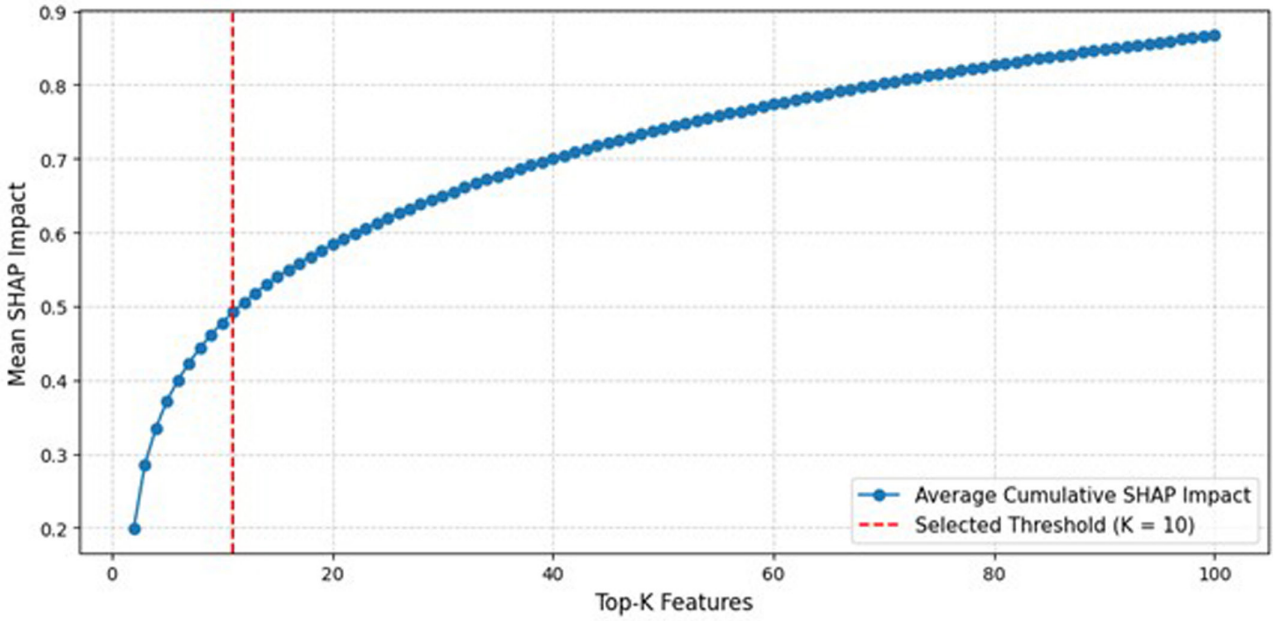

Fig. 3. Mean cumulative SHapley Additive exPlanations impact across Top-K most influential features. The red dashed line indicates the selected interpretation threshold at K = 10, after which additional features contribute marginal explanatory power.

immediate technical attention, while customer relationship management (CRM) teams can deploy per-user SHAP explanations to design personalized retention offers. This dual utility technical and commercial positions the framework as an actionable decision-support tool within live telecom operations, rather than a purely academic exercise.

Among the key contributions of this research is the development of a SHAP-based attribution method that distinguishes network-induced churn from other types of churn. The study also introduces a robust preprocessing pipeline specifically designed for large-scale, real-world telecommunications data, effectively addressing common challenges such as class imbalance, multicollinearity, and missing values. The proposed approach is validated using both proprietary Turkcell data and publicly available datasets, demonstrating its generalizability across different data sources. Additionally, the method enables per-user churn explanations, which support targeted customer retention strategies and network performance optimization.

Results showed that CatBoost consistently outperformed other models in both accuracy and F1 score. Moreover, the SHAP-based attribution analysis demonstrated that the framework is capable of isolating churn cases likely influenced by network-related factors, offering valuable insights into potential technical drivers of churn.

While this study lays a strong foundation for explainable churn prediction in the telecom domain using the CatBoost algorithm, several avenues remain open for future enhancement. One direction is the incorporation of temporal modeling techniques such as long short term memory or temporal attention networks in future iterations, which could complement the current tree-based approach by capturing sequential patterns and early warning signs of churn. Another promising area involves dynamic SHAP thresholding; instead of using a fixed cutoff to identify network-driven churners, future work could explore adaptive or probabilistic methods that account for user-specific context and behavioral history.

Building upon the SHAP outputs, integrating rule-based systems or reinforcement learning could help construct root-cause recommendation engines that automatically suggest actionable interventions for churn mitigation. Incorporating social influence and peer effects through SNA could further enhance the model's ability to detect group-driven churn behavior. The addition of geospatial visualization tools to map churn hotspots and correlate them with network KPIs would allow engineering teams to localize and prioritize interventions more effectively. Finally, integrating the framework into real-time customer care platforms or CRM systems would enable personalized service actions and retention strategies based on explainable churn drivers.

Finally, this study bridges the gap between ML accuracy and business interpretability in churn prediction. By making model decisions explainable and actionable, the framework offers a practical path forward for telecom operators seeking to reduce churn, improve network quality, and build stronger customer relationships.

VII. LIMITATIONS

This study presents significant findings on predicting mobile network-driven churn and explaining its underlying causes using ML and SHAP-based attribution. However, several limitations remain that should be addressed in future research.

First, the dataset primarily focuses on a specific mobile network operator's data and may not fully capture the diversity of customer behaviors and network conditions across different regions or operators. Extending the analysis to multiple operators and integrating additional data sources, such as social media or customer service logs, could enhance the model's generalizability and robustness.

Second, although CatBoost combined with SHAP provides interpretable predictions, the explanation methods used are mostly post-hoc and may not fully reflect causal relationships. Future work could explore causality-aware models or integrate domain knowledge to improve the reliability of churn explanations.

Third, the current model considers static snapshots of network KPIs and customer data without fully incorporating temporal dynamics or seasonality effects. Incorporating time-series models or recurrent architectures could better capture customer behavior trends and improve prediction accuracy.

Moreover, the impact of network infrastructure changes and marketing campaigns on churn has not been explicitly modeled. Future studies could include such external factors to better understand their influence and enable proactive network management.

While the study employs SHAP for explainability, emerging explainable AI techniques could be evaluated to further enhance transparency and user trust in churn prediction systems.

Another limitation of this study is that the dataset used is private and specific to a single mobile network operator. Due to confidentiality agreements and privacy concerns, the data cannot be publicly shared, which limits the reproducibility and external validation of the results by other researchers.

Addressing these limitations will contribute to the development of more accurate, interpretable, and actionable churn prediction frameworks that can better support network operators in maintaining customer satisfaction and optimizing resources.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Artificial Intelligence Usage Statement: This study used ChatGPT for language editing and grammar improvement during the preparation of the manuscript.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – A.G., S.A.; Design – S.B., B.B., G.T., R.Y., A.G.; Supervision – S.A.; Resources – R.Y., A.G.; Materials – S.B., G.T.; Data Collection and/or Processing – S.B., G.T.; Analysis and/or Interpretation – R.Y., A.G., S.A; Literature Search – S.B., B.B.; Writing – S.B., A.G.; Critical Review – A.G., S.A.

Acknowledgements: The authors thank the DAWN6G consortium and TUBITAK TEYDEB for their support.

Declaration of Interests: The authors have no conflicts of interest to declare.

Funding: This work is conducted within the EU Project "DAWN6G: Data Integration of Networks and Applications for Enhanced Management towards 6G" under the EUREKA Celtic-Next Programme and funded by the Scientific and Technological Research Council of Turkey (TUBITAK) under the TEYDEB 1509 Programme - Project No: 9240049.

REFERENCES

1. A. Sikri, R. Jameel, S. M. Idrees, and H. Kaur, "Enhancing customer retention in telecom industry with machine learning driven churn prediction," *Sci. Rep.*, vol. 14, no. 1, 13097, 2024. [CrossRef]
2. Information and communication technologies authority (BTK), "*Turkish Electronic Communications Market Quarterly Report: 2024 Q2*," BTK, Ankara, Turkey, Rep. No. 2024-Q2, 2024 [Accessed: Jan. 23, 2025] Available: https://www.btk.tr/elektronik-haberlesme-pazar-verileri.
3. Kriti, *Customer Churn: A Study of Factors Affecting Customer Churn Using Machine Learning* M.S. thesis. Ames, IA, USA: Iowa State University, 2019.
4. W. Park, and H. Ahn, "Not all churn customers are the same: investigating the effect of customer churn heterogeneity on customer value in the financial sector," *Sustainability*, vol. 14, no. 19, p. 12328, 2022. [CrossRef]
5. V. Chang, K. Hall, Q. Xu, F. Amao, M. Ganatra, and V. Benson, "Prediction of customer churn behavior in the telecommunication industry using machine learning models," *Algorithms*, vol. 17, no. 6, p. 231, 2024. [CrossRef]
6. A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "A review on machine learning methods for customer churn prediction and recommendations for business practitioners," *IEEE Access*, vol. 12, pp. 70434–70463, 2024. [CrossRef]
7. A. Duval, "*Explainable Artificial Intelligence (XAI)*," MA4K9 Scholarly Report, Mathematics Institute, University of Warwick, 2019, doi: [CrossRef].
8. Earino, "Churn dataset", data.world, [Online]. Available: https://data.world/earino/churn. [Accessed: Dec. 21, 2024].
9. BlastChar, "Telco customer churn." [Online], *Kaggle*. Available: https://www.kaggle.com/datasets/blastchar/telco-customer-churn. [Accessed: Dec. 21, 2024].
10. david_becks, "Churn in telecom's dataset." [Online], *Kaggle*. Available: https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset. [Accessed: Dec. 21, 2024].
11. K. Hasumoto, and M. Goto, "Predicting customer churn for platform businesses: using latent variables of variational autoencoder as consumers' purchasing behavior," *Neural Comput. Appl.*, vol. 34, no. 21, pp. 18525–18541, 2022. [CrossRef]
12. S. Fakhar Bilal, A. Ali Almazroi, S. Bashir, F. Hassan Khan, and A. Ali Almazroi, "An ensemble based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in telecom industry," *PeerJ Comput. Sci.*, vol. 8, e854, 2022. [CrossRef]
13. A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *J. Big Data*, vol. 6, no. 1, 2019. [CrossRef]
14. K. Ljubičić, A. Merčep, and Z. Kostanjčar, "Churn prediction methods based on mutual customer interdependence," *J. Comput. Sci.*, vol. 67, 101940, 2023. [CrossRef]
15. W.-Y. Loh, "Classification and regression trees," *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, vol. 1, no. 1, pp. 14–23, Jan. 2011. [CrossRef]
16. S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Mach. Learn. Appl.*, vol. 17, p. 100567, 2024. [CrossRef]
17. T. W. Cenggoro, R. A. Wirastari, E. Rudianto, M. I. Mohadi, D. Ratj, and B. Pardamean, "Deep learning as a vector embedding model for customer churn," *Procedia Comput. Sci.*, vol. 179, pp. 624–631, 2021. [CrossRef]
18. L. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, and A. Gulin, *"CatBoost: Unbiased Boosting with Categorical Features," arXiv Preprint ArXiv:1706.09516*, 2018.
19. L. Ren, G. Sun, and J. Wu, "RoNGBa: A Robustly Optimized Natural Gradient Boosting Training Approach with Leaf Number Clipping," *arXiv preprint arXiv:1912.02338*, 2019.
20. A. Terra, R. Inam, S. Baskaran, P. Batista, I. Burdick, and E. Fersman, "Explainability methods for identifying root-cause of SLA violation prediction in 5G network," in *Proc. IEEE GLOBECOM*. New York: IEEE, 2020, pp. 1–7. [CrossRef]
21. L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [CrossRef]
22. Y. Freund, and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–137, Aug. 1997. [CrossRef]
23. T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
24. C. Molnar, *"SHAP (Shapley Additive Explanations)," Interpretable Machine Learning* Book, 2022. [Accessed: Jan. 30, 2025].
25. L. S. Shapley, "A value for n-person games," in *Contributions to the Theory of Games*, Vol. II, H. W. Kuhn, and A. W. Tucker, Ed. Princeton, NJ, USA: Princeton University Press, 1953, pp. 307–318. [CrossRef]

Sümeyye Baş graduated as the top student from İstanbul Technical University's (ITU) Artificial Intelligence and Data Engineering program and is currently pursuing her master's degree in Computer Engineering at ITU. Sümeyye also works at Turkcell as a Next Generation Researcher. Her main areas of focus include Time Series Prediction, Anomaly Detection, Explainable AI, and Graph Neural Networks, with applications primarily in telecommunications and network analysis.

Buket Bulut graduated with a degree in Mathematical Engineering from Yıldız Technical University in 2021. Since then, she has been working as a Data Scientist at Turkcell.

Gizem Türkoğlu received her M.Sc. degree in Statistics from Marmara University. She works as a Data Scientist at Turkcell, focusing on machine learning, big data, and network analytics. Her interests include predictive modeling and anomaly detection, and she is currently improving her Python skills for large-scale data analysis.

Ramazan Yağmur graduated with a degree in Electrical and Electronics Engineering from Mersin University in 2007. Since then, he has worked in the telecommunications industry, gaining extensive experience in engineering. For the past three years, he has been working as a Data Scientist at Turkcell.

Alpaslan Gökcen graduated with a Ph.D. degree in Computer Engineering from Istanbul Commerce University in 2025. His research interests are machine learning, parallel processing, and real-time image processing. Currently, he works as a Lead Data Scientist at Turkcell.

Saliha Sezgin Alp currently works at Turkcell as Data Intelligence & Analytics Manager. She holds a degree in Electrical Engineering from Istanbul Technical University and an MBA from Istanbul University. She has held various roles in the telecommunications sector, including Planning & Optimization and Access Network Architecture.