Journal of Applied Automation Technologies

Research Paper

Self-Adaptive Feature Engineering Driven Deep Learning Model for Telecom Churn Prediction

Sławomir Bronisław Herdzik[1], Seweryn Grzelak[1] and Michał Dąbrowski[1, *]

[1] Faculty of Electrical Engineering, Automatics and Computer Science, Kielce University of Technology, Kielce, 25-314, Poland
*Corresponding author: michal.da@tu.kielce.pl

Abstract. Telecom customer churn prediction must accurately identify clients who are likely to quit or cut back on their service in the future. In order to increase the accuracy of the telecom customer churn prediction model while maintaining the interpretability of business choices, this study proposes a novel adaptive feature engineering technique. The new approach consists of four parts: machine learning classification, adaptive feature selection, behavior-oriented feature generation, and data cleansing. Predictive features can be extracted from customer profile attributes, service subscription records, usage patterns, billing fluctuations, past complaints, and contract status. Adaptively choose a method to evaluate the features' significance, redundancy, and marginal contribution for the churn classifier. After preprocessing, a telecom customer dataset with 7043 user records and 42 engineered attributes was employed for the experiment. The feature dimension has been lowered from 42 to 24 while the accuracy is 89.42%, the F1-score is 86.17%, and the AUC is 91.36%. When compared to static feature engineering, the adaptive approach decreased feature redundancy by 31.6% and raised the AUC by 3.28 percentage points. Therefore, by using compact and behavior-sensitive customer features, adaptive feature engineering has improved churn forecast accuracy. The suggested framework for telecom customer retention analysis offers a workable method to enhance the feature selection.

Keywords: *Telecom customer churn prediction; Adaptive feature engineering; Feature selection; Machine learning; Customer behavior analysis; Classification model*

Received on 22 September 2023, Accepted on 15 February 2024, Published on 23 February 2024

Introduction

The rate of client attrition for telecom businesses has increased due to market saturation, intense pricing rivalry, and growing customer acquisition expenses. Broadband and integrated-service scenarios in mobile communications will result in lower direct subscription costs and less prospects for future cross-selling to former clients. With accurate churn prediction, the operator may anticipate which consumers are most likely to depart and take prompt action to retain them, like tailored promotions or service enhancements [1]. A decrease in service utilization, billing discontent, an increase in complaints, contract expiration, and decreased use of value-added services are some behavioral indications of customer churn [2]. Customer relationship management platforms, billing systems, network usage logs, service records, and marketing response databases are just a few of the corporate systems that often receive these signals [3]. The fragmented records can be used to create a prediction model that produces useful features for customer lifecycle management [4]. In real-world telecom operations, churn prediction is both a binary classification problem and a data-driven decision support challenge; that is, the selected attributes must help explain why customers are likely to depart and what retention strategies might work [5]. Feature engineering has been utilized to develop numerous high-reliability churn risk indicators for the raw customer data due to the ongoing growth in the size and complexity of telecoms data [6].

To better identify high-risk consumers, previous churn prediction research has employed statistical models, machine learning classifiers, and ensemble learning techniques [7]. The non-linear link between usage behavior, service package modifications, and customer complaints cannot be explained by logistic regression, despite its easy-to-understand coefficients [8]. When it comes to modeling non-linear feature interactions and evaluating feature relevance, Random Forest, Gradient Boosting Decision Trees, XGBoost, and LightGBM are typically more successful [9]. These models are still heavily reliant on the caliber of the input features, though. Fixed attributes, such as length of employment, monthly costs, subscription models, number of support calls and complaint records, etc., are typically used in static manual feature engineering; nevertheless, these indications might not reflect changes in user behavior before attrition [10]. Missing values, category service features, unbalanced churn labels, correlated variables, and redundant behavior indicators are common in telecom data [11]. A high-complexity model will have poor generalization if the features are unstable or irrelevant [12]. Weak churn indications may be missed by aggressive feature reduction that disregards business semantics [13]. The dimensionality of data can be decreased using feature selection techniques including correlation filtering, recursive elimination, embedded tree significance, and regularized regression; however, many of these techniques are static and do not change feature contributions in response to model feedback [14]. In order to prevent multicollinearity and determine whether a customer's characteristics are likely to predict the target variable, adaptive feature engineering can be used [15].

The steps of feature development, feature transformation, feature selection, and model input should all be integrated into a single workflow for adaptive feature engineering for telecom churn prediction. Keep information on user behavior, billing risk, service dependency, and interaction history; remove variables that are recurring or unstable. In order to create behavior-sensitive features from telecom data and dynamically choose a high-value subset of features for churn classification, this study builds an adaptive feature engineering system. The framework's three challenges are how to transform diverse customer data into predictive feature groups, assess the impact of features under redundancy limitations, and maintain prediction accuracy while using fewer variables. In a customer retention scenario, the churn prediction model based on the design will be smaller, more accurate, and simpler to comprehend.

Related Work

Customer Churn Prediction in Telecom Services

Due to the cheap cost of retaining current consumers, telecoms service management has long predicted client desertion. In order to assess churn risks, early research merely gathered basic information on the reasons for staying or leaving and billed products. The customer profiles don't always reflect behavioral shifts before service termination. Changes in usage, payment risk, service interactions, complaint behavior, and customer involvement have recently been included in numerous research on telecom churn prediction as markers of a customer's discontent and inclination to depart [16].

In general, telecom churn statistics are irregular. Age group, account type, area, tenure, and contract form are examples of characteristics of the customer profile. Voice, broadband, streaming, roaming, value-added, and bundle subscriptions are examples of service qualities. The length of the call, data usage, frequency of recharges, payment delays, service outages, and customer service interactions are examples of behavioral characteristics [17]. The overall prediction accuracy will also be determined by the prediction strength of a single factor and its interactions. For instance, a relatively high monthly fee might not be a sign of a significant churn risk for a long-term enterprise user; instead, it might be the result of a recent complaint from a short-term user and a decline in data usage. As a result, feature construction is likewise somewhat complicated.

Class disparity is the other. There is not a significant percentage of churned subscribers in the majority of telecom statistics. A model that is solely focused on overall accuracy may ignore minority churn samples and still produce positive outcomes. Thus, in addition to accuracy, churn prediction studies frequently employ recall, F1-score, AUC, and cost-sensitive evaluation [18]. By developing behavior-change indicators like usage-decline ratios, complaint density, payment instabilities, and contract-expiry-proximity, feature engineering can enhance the identification of the minority class. The characteristics lessen reliance on raw information and are more clear indications of churn.

Feature Engineering and Feature Selection Methods

Raw business data is transformed into a feature set that is suitable for models through feature engineering. Missing value imputation, categorical encoding, numerical normalization, time-window aggregation, feature discretization, interaction creation, and behavioral trend extraction are common processes in telecom churn prediction [19]. There are other preparation steps as well. They assess the classifier's ability to identify a valuable pattern in the customer data. An absolute monthly charge, for instance, might be appropriate, but a relative shift in monthly charges over the previous few billing cycles is frequently more suggestive of a customer's intention to depart. The frequency of complaints each service month may more accurately reflect the level of unhappiness than the overall number of complaints.

Feature selection techniques fall into three broad categories: filters, wrappers, and embedding approaches. The chi-squared score, variance threshold, correlation, and mutual information are statistical filters for features. Although they may ignore nonlinear feature interactions, they are computationally straightforward. Although wrapper approaches are computationally costly, they continually train models with various feature subsets and assess their predictive performance to identify a favorable combination [20]. During model training, embedded techniques—such as L1 regularization for linear models or feature significance in tree-based models—select features. The chosen features might not be appropriate if the classifier or data distribution is changed because the techniques are somewhat straightforward and not heavily reliant on the model.

A generic extension of feature selection, adaptive feature engineering takes into account input from prediction performance as well as the contribution, redundancy, and stability of features. Instead of choosing the variables all at once and training the model, adaptively update the feature weights based on the model's response and validation outcomes [21]. The client groups will have distinct churn indicators, and it is appropriate for churn prediction. A recharge failure will indicate a prepaid user's risk, whereas an increase in complaints or contract expiration will indicate a postpaid contract user's danger. Adaptive selection can minimize the number of variables while retaining those that contribute to the explanation of challenging churn cases.

Machine Learning Models for Churn Classification

Because customer behavior and service parameters frequently have non-linear correlations, many machine learning models are utilized to categorize customer attrition. Although logistic regression is appropriate for interpretability, it may underfit due to its inability to capture complicated patterns. Although Support Vector Machines can use kernel functions to manage non-linear boundaries, their scalability is not very good when dealing with a lot of features and customers [22]. Although decision trees are simple to comprehend, they become unstable when noise is present. Through ensemble learning, Random Forests increase the model's stability and offer feature relevance indications for feature selection.

Gradient Boosting Models are now widely used as baselines for predicting churn. For effective prediction performance in structured data, XGBoost and LightGBM can effectively handle mixed-type features and represent non-linear feature interactions [23]. These models are appropriate for adaptive feature engineering since they also facilitate feature importance analysis. However, the issue of inadequate feature design cannot be resolved by a powerful classifier. The model will learn unstable rules and perform badly under data drift if the consumer behavior is expressed as raw or redundant variables.

Neural networks, including multi-layer perceptrons, embedding-based models, and hybrid feature representation networks, have also been used to structured churn data in recent studies [24]. Although these methods can discover latent correlations between consumer qualities, they are frequently less interpretable and require more training data. To prevent non-churners from dominating the minority churn samples, class imbalance learning has been used [25]. A feature-sensitive churn model has also been constructed by combining Bayesian churn prediction, large data analysis, sequential feature selection, logistic regression, and information-gain filtering [26]. In order to increase the accuracy of telecom churn prediction by feature processing of customer behavior, scale-level analysis has also been carried out on the data [27]. Before building a model, unnecessary attributes can be removed via sequential feature selection [28]. After the feature space has been optimized, logistic regression can clearly explain the customer's risk [29]. A few potential predictor variables for the churn model can be chosen via information-gain filtering [30]. According to the findings, telecom operations demand business explanations and prediction accuracy. In order to identify customers who are likely to churn and maintain the meaning of the chosen characteristics for the creation of a retention strategy, a practical churn prediction system should use adaptive feature engineering and a trustworthy classifier.

Adaptive Feature Engineering Framework

Overall Framework of Adaptive Feature Engineering

Using organized customer data, a telecom business has developed a novel adaptive feature engineering method to forecast client attrition. A binary churn indicator, 19 raw variables, and 7043 customer records make up the initial set of data. 42 candidate variables, including customer profile attributes, service subscription data, billing information, usage behavior metrics, complaint records, and contract details, have been added to the feature pool as a result of preprocessing and feature creation. The system consists of four modules: churn model generation, adaptive feature selection, behavior-oriented feature extraction, and data preprocessing. Sort the different telecom data into a limited number of features that are not redundant and contain the predictive information.

First, a heterogeneous feature vector is created from the input customer record. Monthly charges, tenure, total charges, call durations and data usage are numerical variables that have been standardised; contract types, payment methods, service packages and customer groups are categorical variables that have been converted into numerical codes. Impute missing values according to the missing pattern and feature type. The following is the cleaned feature vector:

$$x_c = \Psi_{num}(x_n) + \Psi_{cat}(x_g) + \Psi_{mis}(x_r) \qquad \text{Eq.(1)}$$

Here, x_n denotes numerical attributes, x_g denotes categorical service attributes, and x_r denotes variables containing missing or irregular records. The transformation functions describe normalization, encoding, and missing-value correction. This representation prevents raw data inconsistency from directly affecting downstream churn modeling.

Use a framework to extract behavior-sensitive features from customer usage and service-interaction data. Even with a high monthly fee, a user with consistent service use and no history of complaints will have a lower churn risk; on the other hand, a user who has lately cut back on usage and frequently complained will be more likely to quit. Profile, billing, use, complaint, and contract features make up the candidate feature pool:

$$F_0 = \Phi_p(x_c) \cup \Phi_b(x_c) \cup \Phi_u(x_c) \cup \Phi_s(x_c) \qquad \text{Eq.(2)}$$

In the experiment, 42 attributes make up the final feature pool. Tenure ratio, frequency of payment delays, monthly charge volatility, service downgrade flag, complaint density, usage decrease ratio, contract remaining time, and service bundle reliance are the indications mentioned above. The general framework of the suggested adaptive feature engineering procedure is shown in Figure 1.

The produced features are subjected to Adaptive Feature Selection. Use the framework to dynamically weight the features according to their predictive strength, redundancy relationships, and stability during the validation cycle. Train a churn classifier using the chosen subset, then return the validation feedback to update the feature score. The following are the Framework's objectives:

$$F^* = \arg\max_F A(F) - \alpha R(F) + \beta S(F) \qquad \text{Eq.(3)}$$

In this expression, $A(F)$ denotes prediction contribution, $R(F)$ denotes redundancy penalty, and $S(F)$ denotes feature stability. The coefficients α and β control the balance between compactness and robustness. In the experiment, the adaptive process reduces the feature dimension from 42 to 24 while maintaining strong predictive performance.

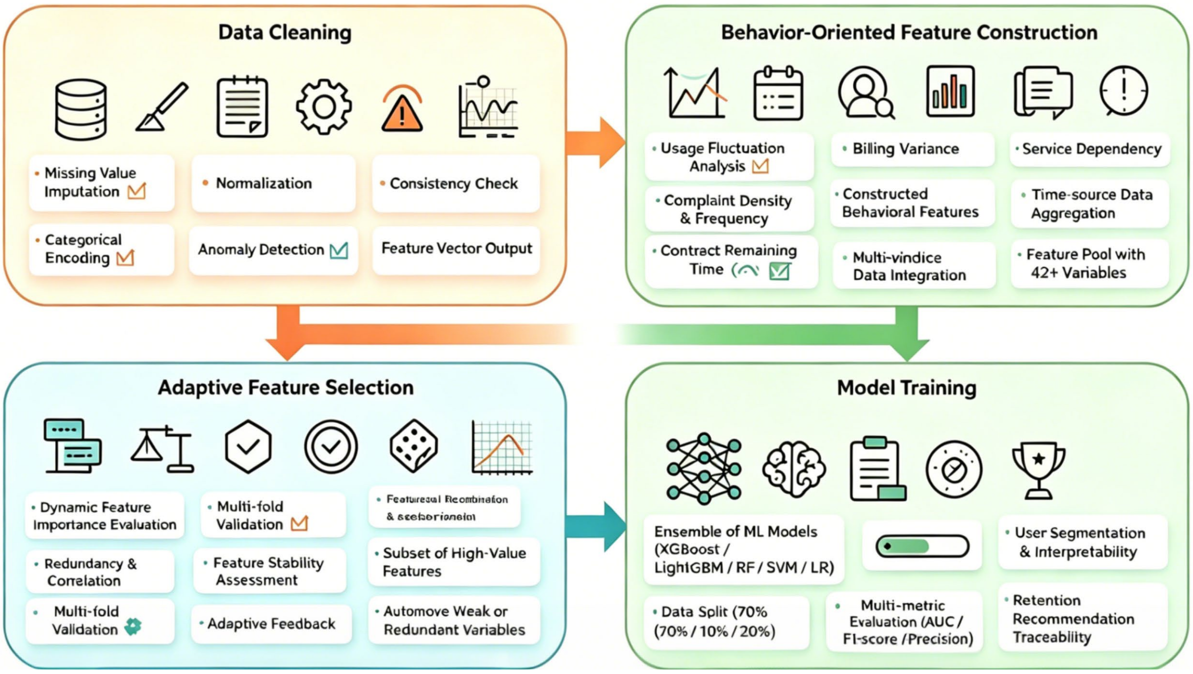


Figure 1. Overall framework of adaptive feature engineering for telecom churn prediction

Behavioral Feature Construction and Transformation

Behavior extraction is the initial stage of feature engineering for churn prediction using telecom data. The customer's existing circumstances are often the raw factors, and any changes to these circumstances could indicate a higher chance of attrition. More information can probably be obtained from a decline in data use over the previous few months than from a single total quantity. The overall amount billed may not be as sensitive as the frequency of payment delays. As a result, the above-mentioned technique uses telecom data to create trend-, ratio-, and interaction-based features.

A decrease in current service use relative to the historical average is known as a use decline. The experiment's historical window is set to the preceding ninety days, while the recent window is set to the most recent thirty days. The design is not overly sensitive to daily variations, but it will detect a brief decline in behavior.

$$d_u = 1 - \frac{U_{30}}{U_{90} + \varepsilon} \qquad \text{Eq.(4)}$$

The feature d_u increases when recent usage drops below the historical level. In churn samples, the average usage decline score reaches 0.286, while in non-churn samples it remains around 0.117. This difference indicates that usage reduction is a meaningful behavior-level churn signal.

Monthly charge changes and late payments are the causes of billing instability. Users are more likely to quit if their expenses have recently increased or if they have repeatedly missed payments. Relative charge variation and the frequency of delays in the most recent billing cycle combine to create billing risk.

$$r_b = \frac{|B_c - B_h|}{B_h + \varepsilon} + \eta D_b \qquad \text{Eq.(5)}$$

Here, B_c denotes current monthly charge, B_h denotes historical average charge, and D_b denotes payment delay frequency. The coefficient η is set to 0.35 in the experiment to prevent payment delay from overwhelming billing variation. This feature is especially useful for users with short tenure and flexible contracts.

The purpose of service reliance is to illustrate the degree to which a user is dependent on bundled telecom services. Users will stay longer if they utilize all of the company's services, including streaming, mobile data, high-speed internet, and technical assistance. The quantity of active services and contract obligations determine the service dependency score.

$$s_d = \log(1 + N_s)(1 + \lambda C_t) \qquad \text{Eq.(6)}$$

In this expression, N_s denotes the number of active services, and C_t denotes contract commitment strength. The coefficient λ is initialized to 0.4. A higher service dependency score usually indicates stronger retention potential, but it may also interact with billing pressure when monthly charges become excessive.

Adaptive Feature Selection Strategy

Prediction strength, redundancy with previously chosen attributes, and stability under various validation sets are the three categories of criteria used to choose candidate attributes. Only when a feature improves churn discrimination and doesn't include redundant data that has already been included into other variables is it retained. In this manner, a large number of the linked variables in the 42-dimensional feature pool are incorporated, including monthly fee, total charge, billing variation, service count, and contract type.

The degree to which the validation loss rises when a feature is altered determines predictive importance. A feature is deemed to be related to churn prediction if substituting its mean value causes a significant drop in validation performance.

$$v_i = L(x_i \leftarrow \bar{x}_i) - L(x_i) \qquad \text{Eq.(7)}$$

The feature importance score v_i is normalized within each iteration. In the experiment, usage decline ratio, contract type, complaint density, monthly charge volatility, and payment delay frequency consistently rank among the top features.

Correlation and contribution overlap are used to illustrate redundancy. Although correlated features are not intrinsically harmful, the model will become more complex and less interpretable if several variables contain almost identical information. The set that has been chosen thus far determines a candidate feature's redundancy penalty.

$$r_i = \max_{j \in F_{sei}} \left|\rho\left(f_i, f_j\right)\right| \qquad \text{Eq.(8)}$$

A candidate feature with redundancy above 0.85 is retained only if its prediction contribution is higher than the correlated selected feature. This prevents the framework from removing useful behavioral indicators solely because they are statistically related to billing or service variables.

Feature stability is evaluated across repeated validation splits. A feature that appears important only in one split may reflect sampling noise rather than stable churn behavior. The stability score is calculated according to the selection frequency over K validation iterations.

$$s_i = \frac{1}{K} \sum_{k=1}^{K} z_i^k \qquad \text{Eq.(9)}$$

Here, z_i indicates whether the feature is selected in the k-th validation iteration. In the experiment, K is set to 5, and features with stability lower than 0.40 are penalized unless they provide a strong contribution to minority churn samples.

Ultimately, the adaptive score is created by combining the three elements into a single feature rank. Iteratively add features until the validation set's AUC does not rise for three rounds in a row. The following is the feedback optimization process: Figure 2.

$$q_i = \sigma(v_i - \alpha r_i + \beta s_i) \qquad \text{Eq.(10)}$$

The preserved feature's upper bound is set at 0.55. From the initial 42 designed variables, choose 24 features in an adaptive manner. Usage drop, billing fluctuation, service dependency, complaint density, contract type, tenure, and payment behavior are among the features that were kept, providing a reasonably succinct but behavior-rich representation for churn modeling.

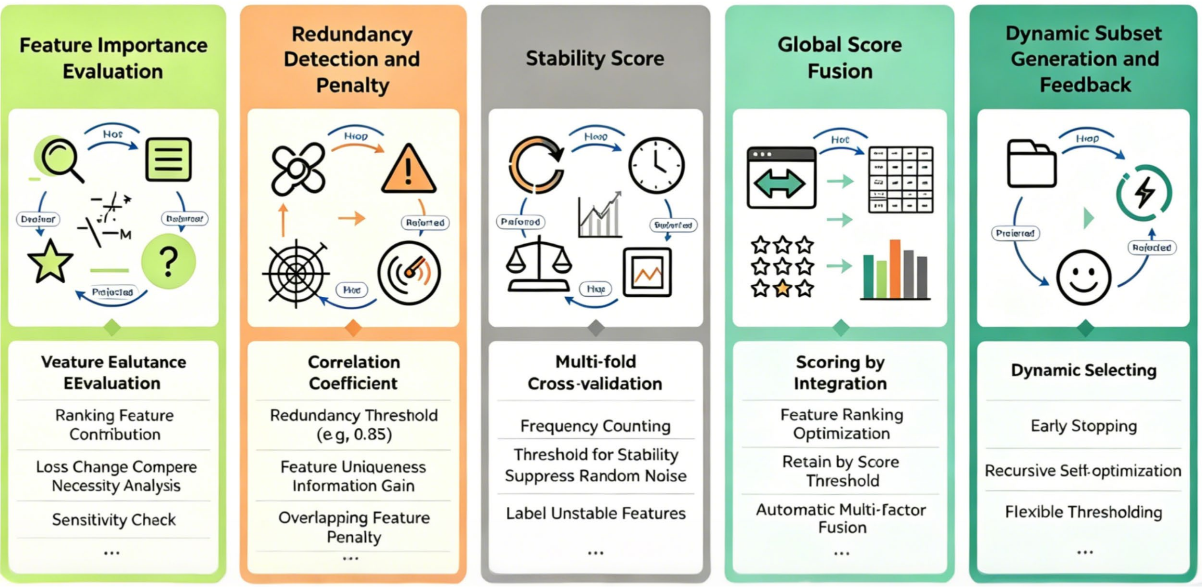


Figure 2. Adaptive feature selection and feedback optimization process

Experimental Scheme and Evaluation Metrics

Dataset Description and Preprocessing Settings

The experiment uses a structured telecom customer churn dataset containing 7043 customer records. Each record includes customer profile information, service subscription status, contract type, payment method, monthly charge, total charge, tenure, service usage indicators, and churn label. The churn label is binary, where churn users account for 26.5% of all samples. After preprocessing and feature construction, the candidate feature pool contains 42 variables. The dataset is divided into training, validation, and testing subsets according to a 7: 1: 2 ratio, corresponding to 4930 training records, 704 validation records, and 1409 testing records.

The median is used to replace missing values in numerical variables, while a distinct "unknown" category is used to fill in missing categorical items. Standardize the continuous variables and minimize the dimension of the categorical variables by using target-aware frequency encoding. Give churn samples a higher training weight to address the issue of class imbalance.

$$w_i = \frac{N}{2N_{c_i}} \qquad \text{Eq.(11)}$$

Here, N denotes the total number of training samples, and N_{c_i} denotes the number of samples in the class of the i-th instance. This weighting strategy reduces the tendency of the classifier to favor non-churn users. XGBoost, random forest, LightGBM, logistic regression, and support vector machine are used as comparison models under the same train-test split.

Prediction and Feature Evaluation Indicators

Prediction accuracy, minority class recognition, ranking ability, feature compression, and redundancy reduction are the evaluation criteria. Because the percentage of non-churners is rather high, accuracy—which gauges the overall correctness of classification—is not appropriate for churn prediction. In order to balance the trade-off between recall and precision in churn detection, the F1-score is used.

$$F_1 = \frac{2PR}{P + R + \varepsilon} \qquad \text{Eq.(12)}$$

Recall provides the percentage of actual churners that have been successfully detected, whereas precision displays the percentage of predicted churners that are truly churners. A comparatively high F1-score shows that the model has successfully balanced the risk of missing real churners with the requirements of preventing false alarms.

AUC shows how well churn risk may be ranked. Only high-risk users are being targeted at this moment due to the operator's limited funding for retention measures. More real-churning users can be assigned to the high-risk group by a model with a higher AUC.

$$AUC = \Pr(s_{\text{churn}} > s_{\text{nonchurn}}) \qquad \text{Eq.(13)}$$

Feature efficiency is evaluated by considering feature compression and redundancy reduction. After adaptive selection, the number of features decreases from 42 to 24, corresponding to a compression ratio of 42.9%. The redundancy score is calculated from the average maximum correlation among selected features.

$$C_r = 1 - \frac{|F^*|}{|F_0|} \qquad \text{Eq.(14)}$$

If the number of features drops and the prediction performance rises, the chosen feature subset is deemed superior. In addition to being simpler to comprehend, a small collection of features can be utilized to minimize computation and streamline data integration for telecom implementation.

Comparative Experimental Design

To ascertain whether the outcomes of churn prediction following adaptive feature engineering are superior to those attained by conventional preprocessing and static feature selection, a comparison experiment will be conducted. The raw cleaned features, devoid of any additional behavior building, are the basis of the baseline setting. The second setting does not use adaptive selection; instead, it makes use of manually constructed behavior features. The third is the conventional filter-based feature selection technique, which relies on mutual information and correlation. Behavior-oriented feature construction and adaptive feature selection based on model feedback are the suggested settings.

The group's classifiers include Support Vector Machine, XGBoost, Random Forest, Logistic Regression, and LightGBM. Random forest and gradient boosting models are employed as nonlinear classifiers, whereas logistic regression is chosen as an interpretable linear baseline. Both XGBoost and LightGBM can handle non-linear correlations between contract, billing, and behavior variables, and they will probably do well on the structured telecom data. Every model is trained using the same evaluation methodology, class weighting approach, and data split.

The contribution of each feature engineering component is determined by ablation tests. The complete framework is contrasted with versions that do not have adaptive feedback, redundancy penalty, stability score, or behavior modification. It can be concluded that adaptive feature engineering has provided pertinent predictive information rather than merely increasing feature complexity if the entire framework can increase the AUC and F1-score by lowering the number of features.

Result Interpretation and Comparative Analysis

Prediction Performance Comparison

The prediction findings show that all models used adaptive feature engineering to characterize customer churn more accurately. Using XGBoost as the classifier, the suggested method's accuracy, F1-score, and AUC on the test set are 89.42%, 86.17%, and 91.36%, respectively. The baseline raw feature accuracy is 84.76%, the F1-score is 79.84%, and the AUC is 86.91%. Although the AUC is improved to 88.08% by manual feature generation without adaptive selection, there are redundant variables and the feature dimension is 42. Due to the exclusion of some weak but complimentary churn indicators, the traditional filter-based selection's accuracy of 88.21% and AUC of 89.04% are lower than those of the adaptive technique [31].

The comparison of the predicted outcomes is displayed in Figure 3. The accuracy with various feature engineering parameters is displayed in Figure 3(a). When compared to the baseline using raw features, the adaptive technique raises the overall accurate rate by 4.66%. The F1-score is shown in Figure 3(b), and the suggested approach raises it from 79.84% to 86.17%. This indicates that it has enhanced the identification of churn consumers compared to only raising the percentage of the majority class. Adaptive feature engineering increases the churn risk ranking and places more real churn users in the high-risk probability range [32], as shown by the AUC values in Figure 3(c).

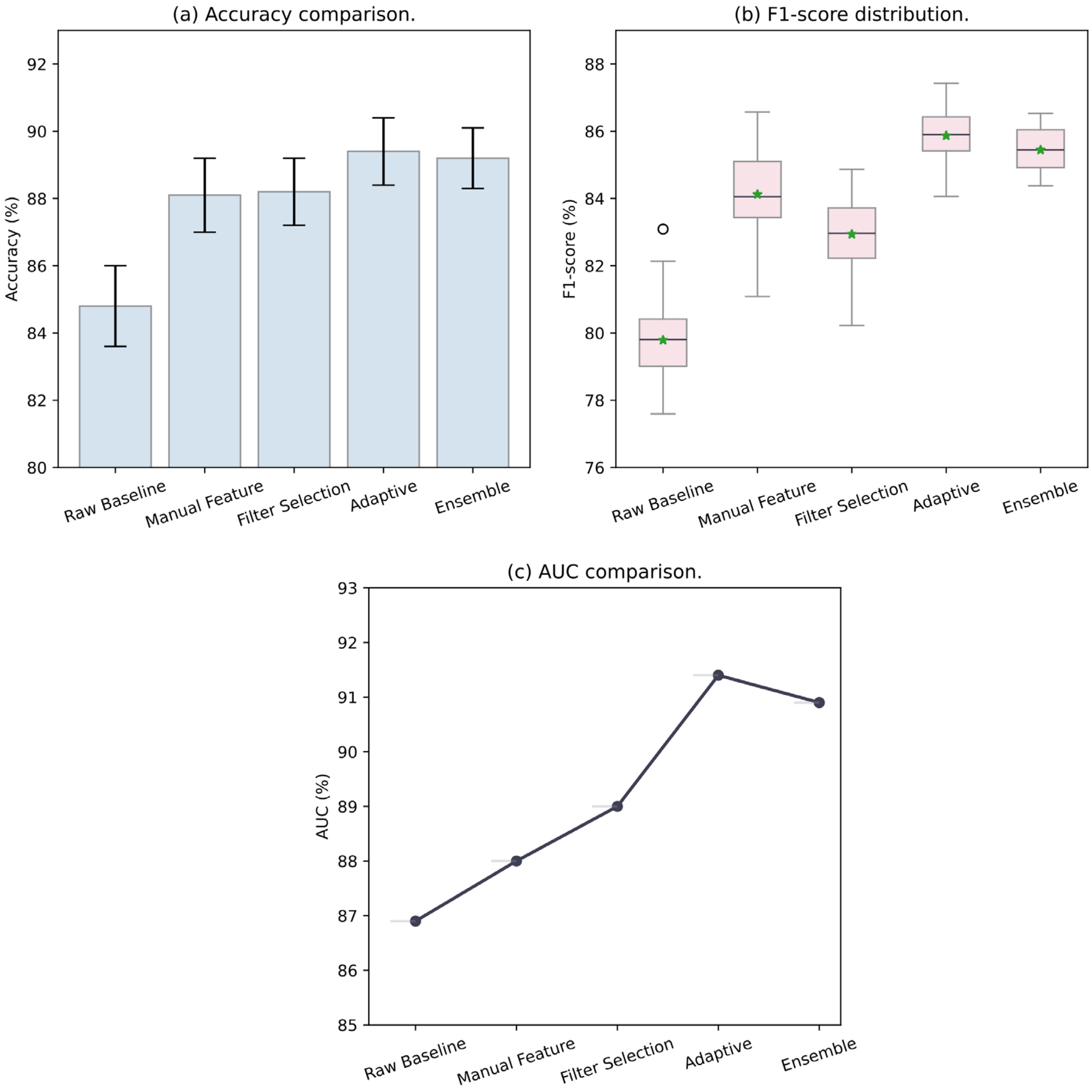
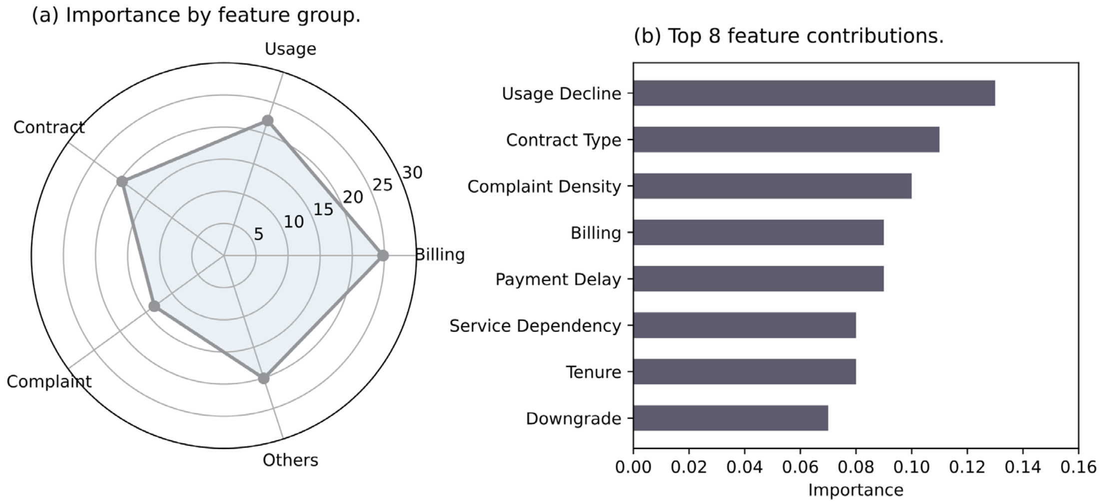


Figure 3. Prediction performance comparison among churn models. (a) Accuracy comparison. (b) F1-score comparison. (c) AUC comparison.

Feature Contribution and Redundancy Analysis

Both created behavior-change indicators and direct business indicators have an impact on churn prediction, according to feature contribution analysis. Contract type, duration, usage drop ratio, monthly fee variation, frequency of complaints, occurrence of payment delays, service reliance score, and recent downgrading flag are the top-ranked attributes. Among them, contract type and duration mostly enhance probability calibration, and usage decrease rate has a significant impact on churn recall. Complaint density can be utilized to detect high-risk consumers with a short service history, while having a relatively low global relevance value [33].

The feature contributions are shown in greater detail in Figure 4. Adaptive feature construction has contributed information beyond that included in the original customer attributes, as seen in Figure 4(a), which displays the top 15 selected features, seven of which are built variables. The distribution of the behavioral feature groups is as follows, as seen in Figure 4(b): Of the total importance, billing-related features account for 24.8%, usage-change features for 22.1%, contract features for 19.6%, and complaint-related features for 13.4%. The stability of features during five validation iterations is depicted in Figure 4(c); 18 of the 24 chosen features were present in at least four iterations, indicating that they were not the result of random sampling variations [34].


(a) Importance by feature group.

(b) Top 8 feature contributions.

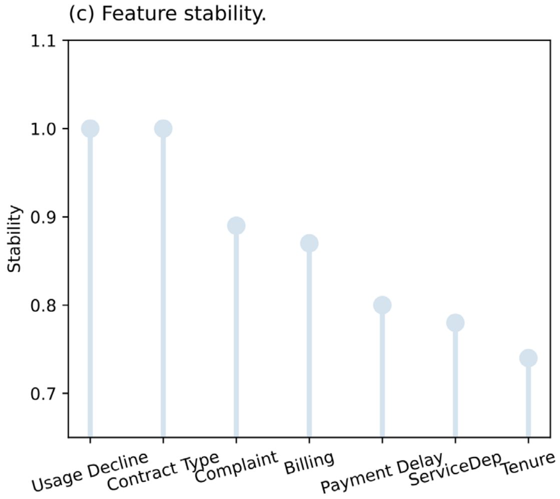
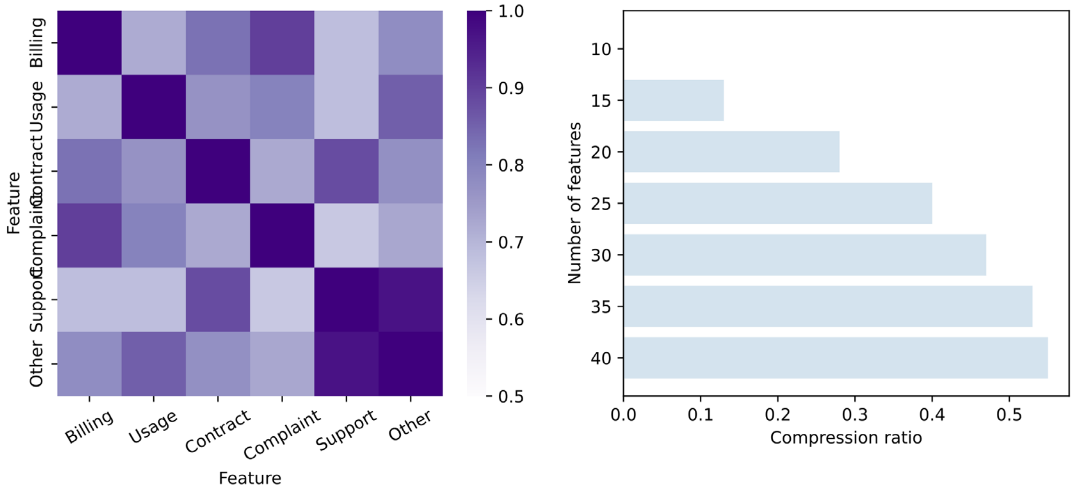

(c) Feature stability.

Figure 4. Feature importance and contribution analysis. (a) Top-ranked feature importance. (b) Contribution of behavioral feature groups. (c) Feature stability across model iterations.

Adaptive selection is necessary after feature construction, according to redundancy analysis. The average maximum feature correlation after retaining all 42 designed variables is 0.71, and certain billing-related variables include information that overlaps. The average redundancy score is now 0.49 thanks to adaptive selection. Because the eliminated variables are primarily duplicates of data previously included in charge volatility, payment delay frequency, or service dependency score, the decrease does not lower prediction accuracy [35].

The results of compression and redundancy are shown in Figure 5. The billing, service count, and tenure-related variables in the initial engineered feature pool have dense correlation blocks, as seen in Figure 5(a). Adaptive selection maintains representative variables from each feature group while breaking many of these correlation blocks, as seen in Figure 5(b). The prediction variations with different feature dimensions are depicted in Figure 5(c); the AUC climbs quickly as the number of selected features increases from 10 to 22, reaches 91.36% at 24 features, and then stays mostly stable after adding more variables [36].


(a) Feature redundancy matrix.

(b) Compression ratio across feature sets.

(c) AUC vs number of features.

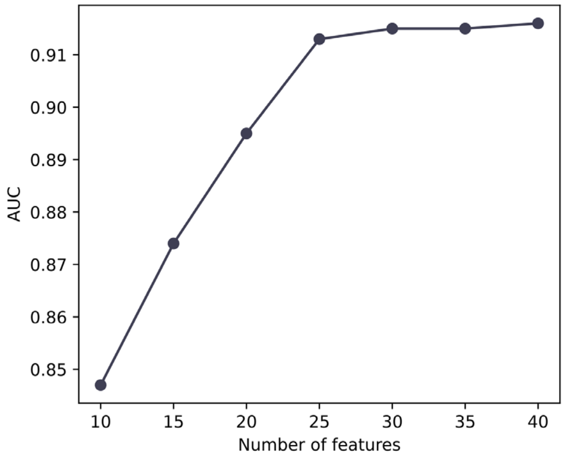
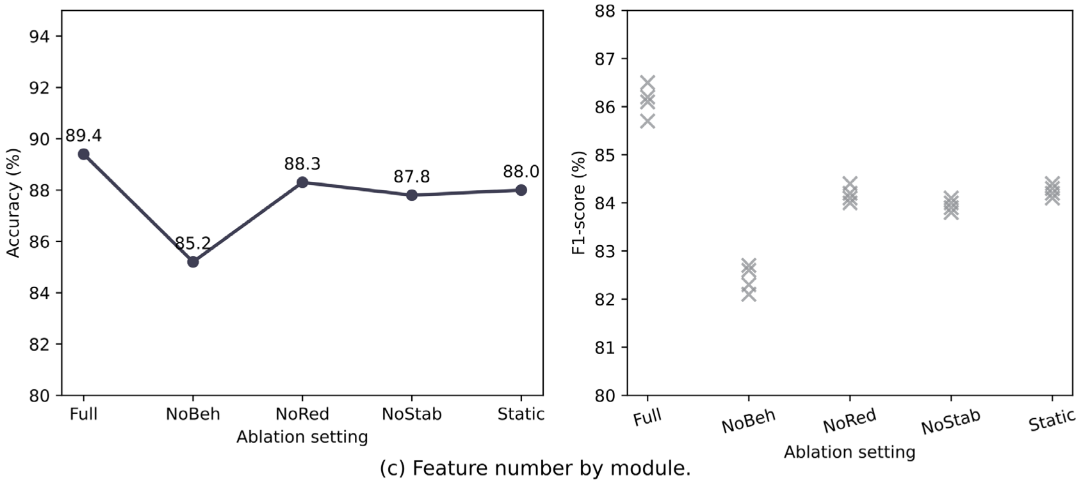


Figure 5. Feature redundancy and compression analysis. (a) Redundancy score before selection. (b) Redundancy score after selection. (c) Prediction changes under different feature dimensions.

Ablation and Robustness Discussion

Every element of the suggested framework is required for churn prediction based on the ablation experiment. The F1-score drops to 82.43% when the behavior modification is removed, indicating that the raw variables are insufficient to accurately indicate the intention to depart. The majority of the generated features are retained when the redundancy penalty is removed, but the AUC drops from 91.36% to 89.72%. This suggests that frequently used variables may be interfering with the model's ranking. For short-tenure clients, eliminating feature stability lessens the performance decline following a validation split change. The static subset has 27 characteristics and an AUC of 0.9018 in the absence of adaptive feedback [37].

The prediction and feature-dimension ablation results are displayed in Figure 6. The accuracy under various module settings is displayed in Figure 6(a); the version without behavior transformation exhibits the biggest decline, while the whole framework obtains the maximum value. The F1-score under the same settings is displayed in Figure 6(b), and it is evident that overall accuracy is less sensitive to feature development than churn-user recognition. Adaptive selection is rather small, as seen in Figure 6(c), where the full technique has 24 characteristics, the no-redundancy version has 31 features, and the static selection version has 27 features [38].



(a) Accuracy by ablation module.

(b) F1-score for ablation modules.

(c) Feature number by module.

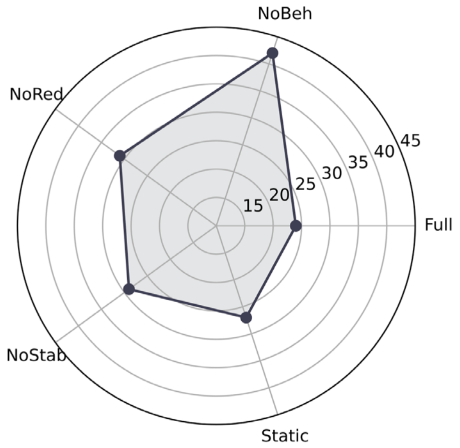
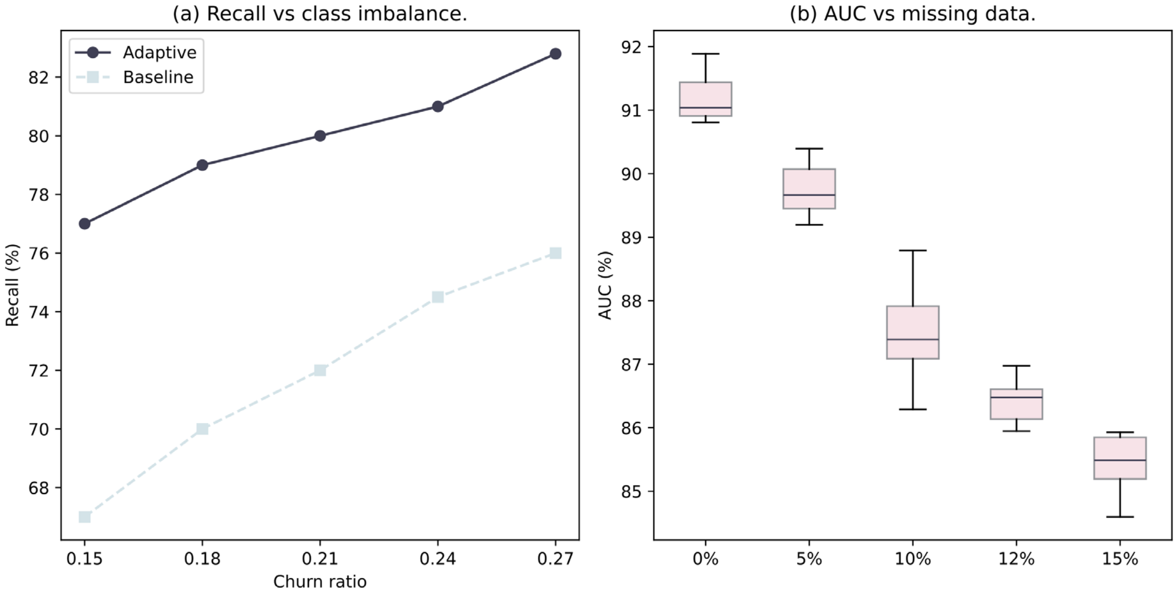


Figure 6. Ablation analysis of adaptive feature engineering modules. (a) Accuracy under different module settings. (b) F1-score under different module settings. (c) Feature number changes under different module settings.

Reduced training samples, missing data, and class imbalance all affect robustness analysis. The suggested method's F1-score remains 82.76% when the churn ratio is lowered from 26.5% to 18.0%, although the raw feature baseline falls to 75.41%. Because reliable information from contract and service dependence groups can be employed for prediction, the suggested method still achieves 88.03% accuracy when 10% of the billing and usage records are randomly masked. The AUC is still 0.8874 and greater than the raw baseline of the entire dataset when the training sample size is lowered to 60% [39].

Three forecasts' robustness results are displayed in Figure 7. The evaluation of class imbalance in Figure 7(a) demonstrates that adaptive features have higher churn recall when there are less churn samples. The results of the missing data rate at 5%, 10%, and 15% are displayed in Figure 7(b), and the corresponding AUC values are 91.36%, 89.82%, and 87.63%, respectively. The suggested approach yields an AUC of 88.74% with 60% of the training data and 90.27% with 80% of the training data [40]. Figure 7(c) compares various training sample sizes.



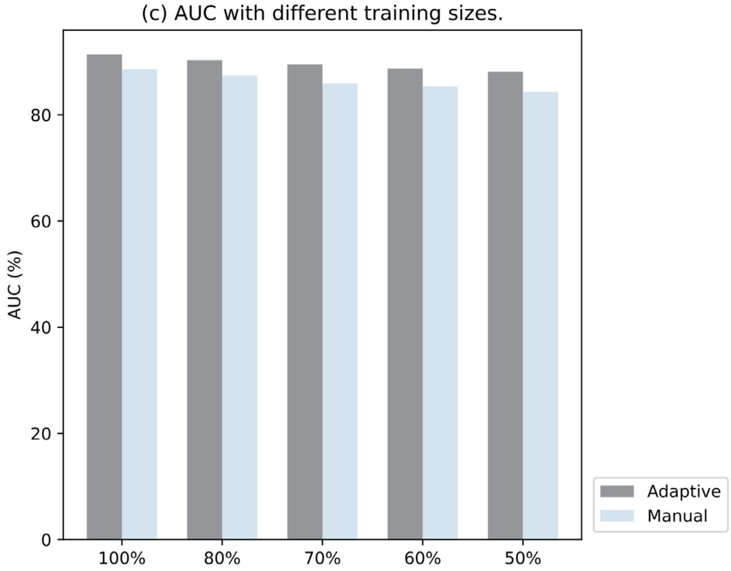


Figure 7. Robustness analysis under different churn prediction scenarios. (a) Performance under class imbalance. (b) Performance under missing data rates. (c) Performance under different training sample sizes.

Conclusion

The impact of feature construction, feature selection, prediction accuracy, and business interpretability are all examined in this study on adaptive feature engineering for telecom customer churn prediction. The methodology dynamically chooses compact variables based on their predictive power, redundancy, and stability after converting the various forms of telecom customer data into behavior-sensitive characteristics. As demonstrated by the tests, adaptive feature engineering can improve churn prediction performance while lowering the number of features. The suggested approach obtained an accuracy of 89.42%, an F1-score of 86.17%, and an AUC of 91.36% with 24 chosen features; consequently, the quality of the features directly affects the churn classification performance.

Telecom churn prediction should be done utilizing a combination of raw customer data and other factors, according to the technical results mentioned above. If a consumer stops utilizing the service, their bills are inconsistent, they have a lot of complaints, and they no longer require the service, they are likely to quit. The issues have been resolved by the new behavioral feature building, and an adaptive selection technique has kept the stable predictors while eliminating the unnecessary variables. Under class imbalance and missing data, the two can enhance the model's compactness and churn-user recognition skills.

Telecom businesses can employ a method to determine the reasons behind some customers' departures. Because they match observed consumer behavior and service conditions, the selected features can be utilized to create machine-learning prediction models and are also quite simple to comprehend for business decision-making. Future research will investigate time-series consumer behavior mining, build automated marketing intervention tactics, and integrate cost-sensitive retention models and adaptive feature engineering. This method can also be used in multi-service telecommunications settings, allowing for the joint analysis of mobile, broadband, cloud, and value-added service records.

Author Contributions

Sławomir Bronisław Herdzik contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Seweryn Grzelak and Michał Dąbrowski contribute to methodology, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0191-6

[2] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE Access, 7, 60134-60149. https://doi.org/10.1109/ACCESS.2019.2914999

[3] Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. Journal of Business Research, 94, 290-301. https://doi.org/10.1016/j.jbusres.2018.03.003

[4] Sai, B. N. K., & Sasikala, T. (2019). Predictive analysis and modeling of customer churn in telecom using machine learning technique. In 2019 3rd International Conference on Trends in Electronics and Informatics (pp. 6-11). https://doi.org/10.1109/ICOEI.2019.8862625

[5] Varun, E., Ravikumar, P., Chandana, S., & Spandana, M. K. (2019). An efficient technique for feature selection to predict customer churn in telecom industry. In 2019 1st International Conference on Advances in Information Technology (pp. 174-179). https://doi.org/10.1109/ICAIT47043.2019.8987317

[6] Maw, M., Haw, S. C., & Ho, C. K. (2019). Customer churn prediction in telecommunication: An analysis on issues, techniques and future trends. In 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (pp. 148-153). https://doi.org/10.1109/CSUDET47057.2019.9214592

[7] Sniegula, A., Poniszewska-Maranda, A., & Popovic, M. (2019). Study of machine learning methods for customer churn prediction in telecommunication company. In Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services (pp. 640-644). https://doi.org/10.1145/3366030.3366109

[8] Sandhya, N., Samuel, P., & Chacko, M. (2019). Feature intersection for agent-based customer churn prediction. Data Technologies and Applications, 53(3), 318-332. https://doi.org/10.1108/DTA-03-2019-0043

[9] Bhuse, P., Gandhi, A., Meswani, P., Muni, R., & Katre, N. (2020). Machine learning based telecom-customer churn prediction. In 2020 3rd International Conference on Intelligent Sustainable Systems (pp. 1297-1301). https://doi.org/10.1109/ICISS49785.2020.9315951

[10] Kavitha, V., Hemanth Kumar, G., Mohan Kumar, S. V., & Harish, M. (2020). Churn prediction of customer in telecom industry using machine learning algorithms. International Journal of Engineering Research and, V9(05). https://doi.org/10.17577/IJERTV9IS050022

[11] Tang, P. (2020). Telecom customer churn prediction model combining K-means and XGBoost algorithm. In 2020 5th International Conference on Mechanical, Control and Computer Engineering (pp. 1128-1131). https://doi.org/10.1109/ICMCCE51767.2020.00248

[12] Verhelst, T., Caelen, O., Dewitte, J. C., Lebichot, B., & Bontempi, G. (2020). Understanding telecom customer churn with machine learning: From prediction to causal inference. Communications in Computer and Information Science, 182-200. https://doi.org/10.1007/978-3-030-65154-1_11

[13] Tang, Q., Xia, G., Zhang, X., & Long, F. (2020). A customer churn prediction model based on XGBoost and MLP. In 2020 International Conference on Computer Engineering and Application (pp. 608-612). https://doi.org/10.1109/ICCEA50009.2020.00133

[14] Geetha, V., Punitha, A., Nandhini, A., Nandhini, T., Shakila, S., & Sushmitha, R. (2020). Customer churn prediction in telecommunication industry using random forest classifier. In 2020 International Conference on System, Computation, Automation and Networking (pp. 1-5). https://doi.org/10.1109/ICSCAN49426.2020.9262288

[15] Shumaly, S., Neysaryan, P., & Guo, Y. (2020). Handling class imbalance in customer churn prediction in telecom sector using sampling techniques, bagging and boosting trees. In 2020 10th International Conference on Computer and Knowledge Engineering (pp. 082-087). https://doi.org/10.1109/ICCKE50421.2020.9303698

[16] Bhattacharya, P., Ladha, A., Kumar, A., Verma, A., & Bodkhe, U. (2020). BaY cP: A novel Bayesian customer churn prediction scheme for telecom sector. In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (pp. 260-265). https://doi.org/10.1109/PDGC50313.2020.9315766

[17] Li, W., & Zhou, C. (2020). Customer churn prediction in telecom using big data analytics. IOP Conference Series: Materials Science and Engineering, 768(5), 052070. https://doi.org/10.1088/1757-899X/768/5/052070

[18] Yulianti, Y., & Saifudin, A. (2020). Sequential feature selection in customer churn prediction based on Naive Bayes. IOP Conference Series: Materials Science and Engineering, 879(1), 012090. https://doi.org/10.1088/1757-899X/879/1/012090

[19] Rani, K. S., Thaslima, S., Prasanna, N. G. L., Vindhya, R., & Srilakshmi, P. (2021). Analysis of customer churn prediction in telecom industry using logistic regression. International Journal of Innovative Research in Computer Science & Technology, 9(4). https://doi.org/10.21276/IJIRCST.2021.9.4.6

[20] Saheed, Y. K., & Hambali, M. A. (2021). Customer churn prediction in telecom sector with machine learning and information gain filter feature selection algorithms. In 2021 International Conference on Data Analytics for Business and Industry (pp. 208-213). https://doi.org/10.1109/ICDABI53623.2021.9655792

[21] RajaGopal Kesiraju, V. L. N., & Deeplakshmi, P. (2021). Dynamic churn prediction using machine learning algorithms: Predict your customer through customer behaviour. In 2021 International Conference on Computer Communication and Informatics (pp. 1-6). https://doi.org/10.1109/ICCCI50826.2021.9402369

[22] Zhu, M., & Liu, J. (2021). Telecom customer churn prediction based on classification algorithm. In 2021 International Conference on Aviation Safety and Information Technology (pp. 268-273). https://doi.org/10.1145/3510858.3510945

[23] Xu, T., Ma, Y., & Kim, K. (2021). Telecom churn prediction system based on ensemble learning using feature grouping. Applied Sciences, 11(11), 4742. https://doi.org/10.3390/app11114742

[24] Goyal, K., Kanishka, K., Vasisth, K., Kansal, S., & Srivastava, R. (2021). Telecom customer churn prediction: A survey. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (pp. 276-280). https://doi.org/10.1109/ICAC3N53548.2021.9725621

[25] Karimi, N., Dash, A., Rautaray, S. S., & Pandey, M. (2021). Customer profiling and retention using recommendation system and factor identification to predict customer churn in telecom industry. Studies in Big Data, 155-172. https://doi.org/10.1007/978-981-33-6518-6_9

[26] Praseeda, C. K., & Shivakumar, B. L. (2021). Fuzzy particle swarm optimization-based feature selection and hybrid kernel distance based possibilistic fuzzy local information C-means clustering for churn prediction in telecom industry. SN Applied Sciences, 3(6). https://doi.org/10.1007/s42452-021-04576-7

[27] Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: A machine learning approach. Computing, 104(2), 271-294. https://doi.org/10.1007/s00607-021-00908-y

[28] Xia, Y., Cui, B., & Duan, Y. (2022). Analysis and prediction of telecom customer churn based on machine learning. Highlights in Science, Engineering and Technology, 16, 131-145. https://doi.org/10.54097/HSET.V16I.2495

[29] Sudharsan, R., & Ganesh, E. N. (2022). A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. Connection Science, 34(1), 1855-1876. https://doi.org/10.1080/09540091.2022.2083584

[30] Naing, Y. T., Raheem, M., & Batcha, N. K. (2022). Feature selection for customer churn prediction: A review on the methods and techniques applied in the telecom industry. In 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (pp. 1-5). https://doi.org/10.1109/ICDCECE53908.2022.9793315

[31] Zhang, T., Moro, S., & Ramos, R. F. (2022). A data-driven approach to improve customer churn prediction based on telecom customer segmentation. Future Internet, 14(3), 94. https://doi.org/10.3390/fi14030094

[32] Sana, J. K., Abedin, M. Z., Rahman, M. S., & Rahman, M. S. (2022). A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection. PLOS ONE, 17(12), e0278095. https://doi.org/10.1371/journal.pone.0278095

[33] Shrestha, S. M., & Shakya, A. (2022). A customer churn prediction model using XGBoost for the telecommunication industry in Nepal. Procedia Computer Science, 215, 652-661. https://doi.org/10.1016/j.procs.2022.12.067

[34] Nguyen, N. Y., Ly, T. V., & Son, D. V. T. (2022). Churn prediction in telecommunication industry using kernel support vector machines. PLOS ONE, 17(5), e0267935. https://doi.org/10.1371/journal.pone.0267935

[35] Liu, Y., Zuo, X., & Xie, Y. (2022). Customer churn prediction by classification models in machine learning. In 2022 9th International Conference on Electrical and Electronics Engineering (pp. 399-407). https://doi.org/10.1109/ICEEE55327.2022.9772553

[36] Alhaqui, F., Elkhechafi, M., & Elkhadimi, A. (2022). Machine learning for telecoms: From churn prediction to customer relationship management. In 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (pp. 1-5). https://doi.org/10.1109/ICMLANT56191.2022.9996496

[37] Phadke, M., Bhattacharya, A., Shethia, M., & Shah, S. (2022). Feedback based telecom churn prediction using machine learning. In 2022 5th International Conference on Advances in Science and Technology (pp. 481-485). https://doi.org/10.1109/ICAST55766.2022.10039530

[38] Dalli, A. (2022). Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector. Mathematical Problems in Engineering, 2022, 1-11. https://doi.org/10.1155/2022/4720539

[39] Geiler, L., Affeldt, S., & Nadif, M. (2022). An effective strategy for churn prediction and customer profiling. Data & Knowledge Engineering, 142, 102100. https://doi.org/10.1016/j.datak.2022.102100

[40] De, S., & Prabu, P. (2022). A sampling-based stack framework for imbalanced learning in churn prediction. IEEE Access, 10, 68017-68028. https://doi.org/10.1109/ACCESS.2022.3185227