Enhancing Telecom Churn Prediction Using Emotion-Driven and Behavioral Engagement Features

Huthaifa Aljawazneh
Faculty of Business-Department of Business Intelligence, Al-Zaytoonah University of Jordan, Amman, Jordan

***Abstract*—Accurate churn prediction enables service providers to develop effective retention strategies and promotes revenue stability in the telecommunication industry. This study enhances churn prediction performance by extracting five emotion-driven and behavioral engagement features from a telecom churn dataset. The new features represent derived, experience-oriented indicators constructed from operational usage data rather than direct psychological or survey-based measurements. To assess the effect of these engineered features on predictions, three powerful classifiers (i.e., CatBoost, Random Forest, and XGBoost) were trained and tested in a structured three-stage experimental design. In the first stage, the classifiers were trained and tested using the original dataset (original features only). In the second stage, the original dataset was enriched with five newly derived features (i.e., frustration index, trust score, satisfaction index, service usage score, and international experience index). Finally, in the third stage, only the engineered features were used in the classification process to evaluate their standalone predictive capability. Because the dataset is imbalanced, SMOTE and SMOTE-Tomek were applied to address this issue. The results demonstrate that incorporating these engineered features improves churn prediction performance across the reported evaluation metrics (accuracy, precision, recall, and specificity) for the classifiers and balancing techniques combinations presented. The enriched dataset (original + engineered features) achieves the strongest overall performance compared to using either original features only or engineered features only. Compared to the original features only, the enriched dataset achieved improvements of up to 3.6% in accuracy and 5.8% in recall. These findings indicate that emotion-driven and behavioral engagement features provide meaningful complementary information that enhances churn prediction effectiveness.**

Keywords*—Customer churn prediction; emotion-driven features; feature engineering; imbalanced data; classification

I. Introduction

In the telecom sector, where acquiring new customers is more expensive than retaining existing ones, customer attrition remains a major challenge [1]. Subscriber churn can be viewed as an outcome of the overall service experience that the subscriber has with the service provider, for instance, in Jordan's telecom environment, switching to another service provider is affected by the market competitiveness, costs, and service delivery [2]. Accordingly, the main purpose of telecom churn prediction is to support the development of effective retention strategies. In addition, it allows telecom companies to address customer concerns and meet expectations. Therefore, by predicting churn, service providers can identify and pay more attention to customers who are likely to churn to a competing provider, and develop targeted solutions and personalized offers to improve retention [3]. From a practical perspective, churn prediction models are most useful when they not only predict churn accurately, but also help explain the underlying reasons for customer churn.

However, most existing studies have focused on improving churn prediction using sampling techniques, feature selection, or model optimization, while relying mainly on original telecom attributes [3], [4], [5], [6], [7], [8]. These attributes provide descriptive information about service usage, but do not fully reflect customers' experiences or level of engagement with the service. In particular, traditional telecom features typically capture what customers did, while offering limited insight into how customers perceived the service, such as whether they were frustrated or satisfied. Accordingly, many churn models achieve strong predictive performance, but provide limited explainability from a customer experience perspective. Therefore, a clear research gap exists in the literature regarding the development and evaluation of experience-oriented features that represent customers' emotional states and engagement levels using operational telecom data.

To address this limitation, this study proposes a set of emotional and behavioral indicators derived from telecom data to reflect customer experience and engagement. In other words, five engineered features are grouped into two categories: emotion-driven features (i.e., frustration index, satisfaction index, and trust score) and behavioral engagement features (i.e., service usage score and international experience index) are derived from existing telecom data and aim to better represent customer experience, which supports a more behavior-oriented churn prediction.

In addition, to evaluate the effects of the proposed features on the classification performance, three ensemble classifiers (i.e., Random Forest [9], CatBoost [10], and XGBoost [11]) were trained and tested in three sequential experimental stages. In the first stage, classifiers were trained using the original telecom dataset to provide an initial reference for churn prediction performance. In the second stage, the original dataset was enriched with the proposed emotion-driven and behavioral engagement features. In the third stage, only the newly engineered features were used to train and test the classifiers to assess their standalone predictive capability. Moreover, since the dataset used in this study is imbalanced, SMOTE [12] and SMOTE-Tomek [13] balancing techniques were applied to address the data balancing problem. Finally, as accuracy alone is insufficient for evaluating churn prediction models on imbalanced data, additional performance metrics including precision, recall, and specificity were employed to ensure a reliable evaluation.

Accordingly, the main contributions of this study are summarized as follows: 1) Extracting emotion-driven and behavioral engagement features that reflect customer experience using telecom usage data, 2) demonstrating the effects of combining the newly generated features with the original telecom features on the classifiers' performance. To achieve this, a structured experimental design is conducted, which compares original telecom data, enhanced telecom data (original + emotion-driven and behavioral engagement features), and the newly generated features only in telecom churn classification tasks.

The remainder of this study is organized as follows: Section II reviews related work on telecom customer churn prediction, with particular focus on machine learning approaches and feature engineering techniques. Section III describes the dataset used in this study and outlines its main characteristics. Section IV presents the proposed methodology, including the feature engineering process, the classifiers employed, the data balancing techniques, and the evaluation metrics. Section V reports and analyzes the experimental results obtained from the three-stage evaluation framework. Finally, Section VI concludes the study by summarizing the main findings and discussing limitations and directions for future research.

II. Related Works

Several researchers have addressed customer churn prediction in the telecommunications sector using machine learning and deep learning models. In this context, Wagh et al. [3] developed a machine learning-based system to predict customer churn in the telecom sector using Random Forest, K-Nearest Neighbors, and Decision Tree, relying on the original dataset to train and test classifiers. Furthermore, Gaur and Dubey [4] employed four machine learning algorithms, including Logistic Regression, SVM, Random Forest, and Gradient Boosted Trees, to build and compare effective churn prediction systems. Similarly, Poudel et al. [5] demonstrated that telecom customer churn can be effectively predicted using a Gradient Boosting Machine (GBM) enhanced by explainer models and visualization tools for deeper insights. Moreover, Lalwani et al. [6] proposed a six-phase methodology for telecom churn prediction using machine learning models (i.e., Logistic Regression, Support Vector Machine, Random Forest) and boosting techniques.

However, a limited number of studies have focused on feature selection and feature transformation, which can reveal deeper patterns related to dissatisfaction or reduced service quality. For instance, Huang et al. [14] proposed a feature-selection approach that prioritizes churn-related attributes in an imbalanced telecom dataset. Their method outperformed conventional selection techniques such as Chi-Square and Information Gain, showing that carefully identifying discriminative features can significantly enhance the prediction performance. Moreover, Sana et al. [15] demonstrated that combining univariate feature selection with several transformation methods produces more informative feature representations. Their approach improved the prediction accuracy of telecom churn.

More recent studies have focused on hybrid feature-selection strategies. For example, Mozaffari et al. [16] combined several feature selection methods and aggregated their outputs to identify the most relevant and reliable features. The results show that this hybrid feature selection significantly improves churn prediction performance.

Therefore, these studies highlight the effectiveness of feature engineering in improving churn prediction, but they mainly transformed existing dataset attributes or removed weak ones. None explicitly attempted to derive advanced emotion-related features that capture customer frustration, satisfaction, trust, usage sentiment, or service engagement. Therefore, the literature lacks structured attempts to quantify psychological or experience-based patterns that could affect churn decision-making. Addressing this limitation, the present study introduces newly engineered emotion-driven and behavioral engagement features and evaluates their effect on telecom churn prediction performance.

III. Dataset Description

The dataset used in this study was obtained from the Kaggle community. It consists of 3,333 customer records and 20 attributes, 16 of which are numerical and 4 are categorical. The dataset is imbalanced, with 2,850 non-churner customers and 483 churners. In addition, three of the categorical attributes are binary (Yes/No), which were converted into a binary indicator (0/1), while the remaining categorical attribute (i.e., *State*) was excluded from the analysis due to its high cardinality and limited contribution to predictive performance. Furthermore, the data were divided into stratified training and testing sets, and the class balancing problem was addressed using SMOTE and SMOTE-Tomek, applied only to the training data. All engineered features were derived exclusively from the original input variables to prevent data leakage. However, more details about the dataset used in this study are available in: https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets, and no missing values were observed.

IV. Methodology

In this work, we extracted five engineered features from the original dataset: three emotion-driven features (frustration index, trust score, and satisfaction index) and two behavioral engagement features (service usage score and international experience index). The extracted features were then inserted into the original dataset, and were used to train and test three powerful classifiers (i.e., CatBoost, RF, and XGBoost). The dataset is highly imbalanced; thus, two advanced balancing techniques were used to handle this issue (i.e., SMOTE, and SMOTE-Tomek). Moreover, four metrics were used to evaluate the performance of classifiers in predicting churn (i.e., accuracy, precision, recall, and specificity). On the other hand, to observe the impact of adding the emotion-driven and behavioral engagement features to the original dataset on the classifiers' performance, the original dataset (with no emotion-driven and behavioral engagement features) was used to train and test classifiers. Furthermore, to obtain a broader view on the impact of the emotion-driven features on the classification performance, the emotion-driven and behavioral engagement features were used exclusively to train classifiers.

Fig. 1 illustrates the overall research framework adopted in this study.

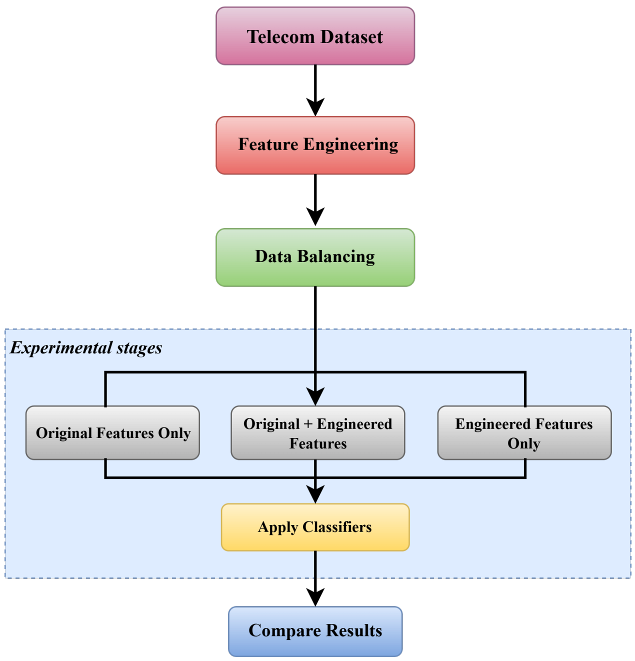

Fig. 1. Overall research framework of the proposed telecom churn prediction methodology.

A. Emotion-Driven and Behavioral Engagement Feature Engineering

To enrich the telecom churn dataset, a set of emotion-driven and behavioral engagement features was extracted from existing attributes and added to the original dataset. These features are grouped into two categories:

- Emotion-driven features: *frustration index*, *satisfaction index*, and *trust score*.
- Behavioral engagement features: *service usage score*, and *international experience index*.

The following list explains the extraction process for each newly derived feature.

1) Frustration index: Customer frustration is extracted from the Customer Service Calls feature, which measures how many times a subscriber contacted support. Therefore, frequent calls indicate unresolved problems or dissatisfaction with service quality.

$$Frustration_c = \frac{\min(CustomerServiceCalls_c, 10)}{10} \tag{1}$$

where, $CustomerServiceCalls_c$ is the number of service calls made by customer c. The $min(x, 10)$ function limits the number of calls to 10 to prevent excessive values from affecting the results. The division by 10 rescales the values; 0 corresponds to no calls, whereas 1 corresponds to 10 or more calls.

2) Trust score: Trust is derived from the Account length feature, which represents the number of subscription days. A longer subscription period indicates stronger loyalty and confidence in the provider. To measure Trust score, a min/max normalization was applied.

$$Trust_c = \frac{AccountLength_c - \min(AccountLength)}{\max(AccountLength) - \min(AccountLength)} \tag{2}$$

In this formula, $AccountLength_c$ presents the subscription duration of customer c. The minimum and maximum values of $AccountLength$ represent the shortest and longest subscription periods found in the dataset.

3) Satisfaction index: Satisfaction is extracted from the efficiency of daytime service usage, which is expressed as the ratio of *Total day minutes* to *Total day charge* features. Customers who receive more minutes per unit cost are assumed to have higher satisfaction. The satisfaction index is then scaled between 1 and 5.

$$Satisfaction_c = 1 + 4 \times \frac{\dfrac{TotalDayMinutes_c}{TotalDayCharge_c + \varepsilon}}{\max\left(\dfrac{TotalDayMinutes}{TotalDayCharge + \varepsilon}\right)} \tag{3}$$

The formula includes $TotalDayMinutes_c$ which represents daytime minutes used by customer c, and $TotalDayCharge_c$ which represents the corresponding cost. Also, the formula includes a small constant ε to avoid division by zero. The denominator represents the highest efficiency ratio found in the dataset, which ensures that the customer with the highest efficiency gets a satisfaction score of 5, while all remaining customers are proportionally scaled between 1 and 5.

4) Service usage score: The calculation of off-peak hours billing fairness through balance ratio requires the integration of evening and night usage data with their respective charges according to the following formula:

$$Balance_c = \frac{TotalEveMinutes_c + TotalNightMinutes_c}{TotalEveCharge_c + TotalNightCharge_c + \varepsilon} \tag{4}$$

For customer c, $TotalEveMinutes_c$ and $TotalNightMinutes_c$ show the total evening and night call times, while $TotalEveCharge_c$ and $TotalNightCharge_c$ represent the corresponding charges. A tiny value of ε was added to avoid division by zero. This balance ratio was used to calculate the $ServiceUsageScore_c$, as shown in the following formula:

$$ServiceUsageScore_c = \frac{Balance_c - \min(Balance)}{\max(Balance) - \min(Balance)} \tag{5}$$

where, $Balance_c$ represents the ratio of off-peak usage to charge for customer c, while $min(Balance)$ and

$max(Balance)$ represent the minimum and maximum balance ratios found in the dataset. The scores increase with better off-peak efficiency, indicating improved billing performance.

5) International experience index: Satisfaction with international services was represented by the average duration of international calls. This was calculated for each customer and then normalized across the dataset, as described in the following formulas:

$$IntlExp_c = \frac{TotalIntlMinutes_c}{TotalIntlCalls_c + \varepsilon} \tag{6}$$

$$IntlExperienceIndex_c = \frac{IntlExp_c - \min(IntlExp)}{\max(IntlExp) - \min(IntlExp)} \tag{7}$$

Here, $IntlExp_c$ represents the average international call duration for the customer c. The total international call duration for customer c is represented by $TotalIntlMinutes_c$. Whereas the total number of international calls made by customer c is denoted as $TotalIntlCalls_c$. The fraction computes the average international call length, with ε included to avoid division by zero. The minimum and maximum observed averages, $min(IntlExp)$ and $max(IntlExp)$, are then used to rescale the measure into the [0,1] interval. Higher values reflect better international call experiences.

B. Classifiers Employed

1) Categorical boosting (CatBoost): It is a powerful classification algorithm that builds ensembles of binary decision trees to perform gradient boosting operations. It introduces a technique called ordered boosting, where the model is trained on specific permutations of the dataset. For each data point, CatBoost computes predictions and target statistics using only the preceding examples within the permutation. This prevents target leakage, leading to better generalization and reduced bias [10].

2) eXtreme gradient boosting (XGBoost): It is an efficient and scalable machine learning algorithm based on the gradient boosting framework. Its procedure is based on building an ensemble of decision trees sequentially, where each new tree attempts to correct the errors made by the previous ones. Furthermore, the addition of regularization techniques in XGBoost enhances its performance compared to conventional boosting algorithms [11]. Additionally, it provides parallel processing capabilities, internal handling of missing data and optimized data structures which enhance speed and accuracy in different classification tasks [17].

3) Random forest (RF): It is a supervised machine learning algorithm that generates decision trees and combines their outputs to improve the accuracy and reliability of the model. Each tree is trained on a randomly selected subset of the dataset, and at each decision point within the tree, only a random subset of features is considered. The random selection process creates different trees which work together to prevent overfitting and improve predictions for new data points [9]. Additionally, in classification, it selects the most common prediction among all trees, while in regression, it calculates the average of their outputs [18].

However, hyperparameter tuning was not performed in this study. All classifiers were implemented using their default parameter settings as provided in their respective libraries to ensure a fair comparison across the three experimental stages.

C. Balancing Techniques Used

1) Synthetic Minority Over-sampling Technique (SMOTE): It is an advanced oversampling technique that generates new synthetic samples through interpolation between minority class instances and their *k*-nearest neighbors instead of replicating the existing minority class samples. This is done by selecting a random point on the line segment connecting the two feature vectors. Consequently, this technique effectively expands the decision space for the minority class and reduces the risk of overfitting [12].

2) Synthetic Minority Over-sampling Technique (SMOTE) with Tomek Links (SMOTE-Tomek): SMOTE-Tomek is a hybrid resampling technique that combines SMOTE with Tomek Links to better handle class balancing issue and enhance classification performance. As stated before, SMOTE generates synthetic samples by interpolating between minority class instances and their nearest neighbors, thereby increasing the representation of the minority class. On the other hand, Tomek Links identify pairs of instances belonging to different classes that are each other's closest neighbors; these pairs often occur near class boundaries and are likely to be noisy. Removing such pairs produces a cleaner dataset with clearer class separation, which helps classifiers achieve more robust performance [13].

D. The Evaluation Metrics Considered

The performance of the classifiers was evaluated using four metrics: accuracy, precision, recall, and specificity [19]. These metrics provide a comprehensive evaluation of model performance, particularly for imbalanced classification problems.

- Accuracy: measures the overall proportion of correctly classified instances in both classes; churn and non-churn.
- Recall: measures the proportion of actual minority class instances (churn) that are correctly identified by the classifier.
- Specificity: measures the proportion of majority class instances (non-churn) that are correctly identified.
- Precision: identifies the proportion of predicted minority class instances (churn) that are truly positive, indicating the reliability of positive predictions.

V. Results

This section presents the experimental findings obtained in three stages of analysis. The first experiment evaluates classifier performance using the original telecom dataset with traditional features only. The second experiment introduces emotion-driven and behavioral engagement features alongside the original attributes to assess their added value. Finally, the third experiment examines the predictive power of emotion-driven and behavioral engagement features in isolation. The results are compared across multiple classifiers and balancing

TABLE I. Evaluation Metrics Obtained from the Combinations of Classifiers and Balancing Techniques on the Original Dataset Only.

Model	Accuracy	Precision	Recall	Specificity
CatBoost + SMOTE	0.9272	0.9620	0.8895	0.9649
CatBoost + SMOTE-Tomek	0.9488	0.9686	0.9276	0.9700
RF + SMOTE	0.9395	0.9700	0.9070	0.9719
RF + SMOTE-Tomek	0.9559	0.9657	0.9452	0.9665
XGBoost + SMOTE	0.9421	0.9755	0.9070	0.9772
XGBoost + SMOTE-Tomek	0.9559	0.9674	0.9435	0.9683

techniques to highlight the role of the newly generated features in enhancing churn prediction.

A. Results of the First Experiment: Original Dataset Only

In the first part of the experiments, customer churn prediction was performed using CatBoost, RF, and XGBoost on the original telecom dataset without the inclusion of the emotion-driven and behavioral engagement features. To address class imbalance, two resampling methods, SMOTE and SMOTE-Tomek, were applied. The performance of the classifiers was evaluated using accuracy, precision, recall, and specificity metrics. As is common in similar works [20], we compare the performance of classifiers to determine the best performer.

As shown in Table I, all classifiers yielded strong performance in predicting customer churn, and SMOTE-Tomek consistently outperformed SMOTE. The performance of the combinations of CatBoost , RF, and XGBoost with SMOTE-Tomek performs better than the combinations of the same classifiers with SMOTE in predicting the majority class and the minority class; the combinations with SMOTE-Tomek obtained higher recall and specificity. With respect to precision, the combinations of RF and XGBoost with SMOTE obtained slightly higher values compared to SMOTE-Tomek. Accordingly, these results indicate that all three classifiers are effective in predicting customer churn in telecom. However, SMOTE-Tomek obtained more balanced outcomes as stated in recall and specificity. This confirms the importance of advanced balancing techniques in improving classifier performance, particularly in predicting the minority class (i.e., churn).

B. Results of the Second Experiment: Incorporating Emotion-Driven and Behavioral Engagement Features

In the second part of experiments, three emotion-driven features (i.e., *frustration index*, *trust score*, *satisfaction index*), and two behavioral engagement features (i.e., *service usage score*, and *international experience index*) were derived from the original dataset and added as new features. As in the first part of experiments, SMOTE and SMOTE-Tomek were employed to solve class balancing problem, and the classifiers' performance was evaluated using accuracy, precision, recall, and specificity metrics.

The results presented in Table II, show that the integration of emotion-driven and behavioral engagement features with the original dataset improved the performance of all classifiers compared to the previous experiment. However, in this part of the experiments all of the combinations of classifiers with

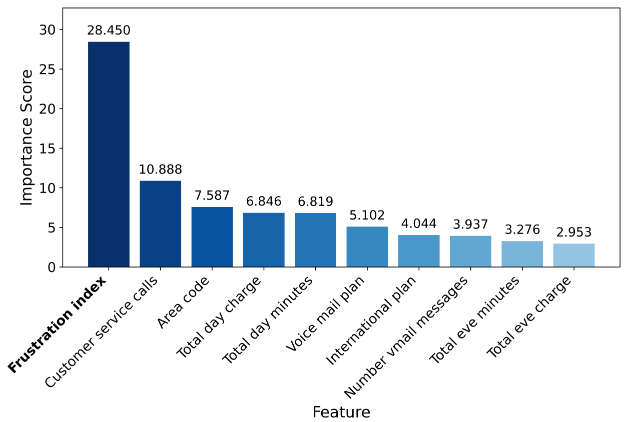

Fig. 2. Ten most important features in the classification process using CatBoost combined with SMOTE.

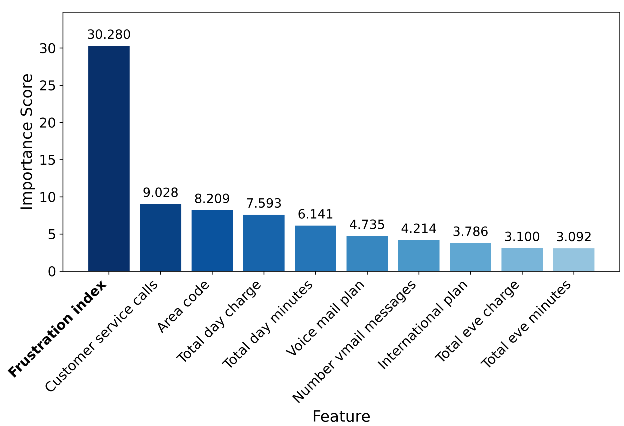

Fig. 3. Ten most important features in the classification process using CatBoost combined with SMOTE-Tomek.

SMOTE and SMOTE-Tomek balancing techniques obtained a significant accuracy value; all of them obtained more than 0.96, except the combination of RF with SMOTE, it obtained slightly lower value. In addition, the combination of CatBoost with SMOTE-Tomek showed superior performance in predicting the minority class (i.e., churn); it obtained the highest recall value. On the other hand, the combination of XGBoost with SMOTE showed the best performance in predicting the majority class (i.e., non-churn), as stated in the specificity metric values.

TABLE II. Evaluation Metrics Obtained from the Combinations of Classifiers and Balancing Techniques Applied on the Original Dataset + Emotion-Driven and Behavioral Engagement Features

Model	Accuracy	Precision	Recall	Specificity
CatBoost + SMOTE	0.9632	0.9783	0.9474	0.9789
CatBoost + SMOTE-Tomek	0.9673	0.9783	0.9558	0.9788
RF + SMOTE	0.9570	0.9745	0.9386	0.9754
RF + SMOTE-Tomek	0.9629	0.9712	0.9541	0.9718
XGBoost + SMOTE	0.9675	0.9819	0.9526	0.9825
XGBoost + SMOTE-Tomek	0.9656	0.9765	0.9541	0.9771

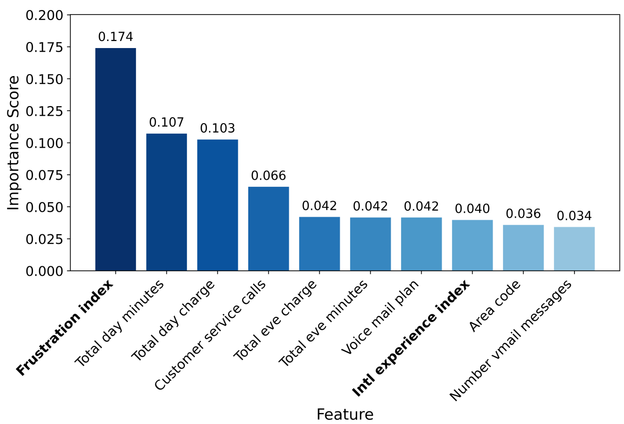

Fig. 4. Ten most important features in the classification process using RF combined with SMOTE.

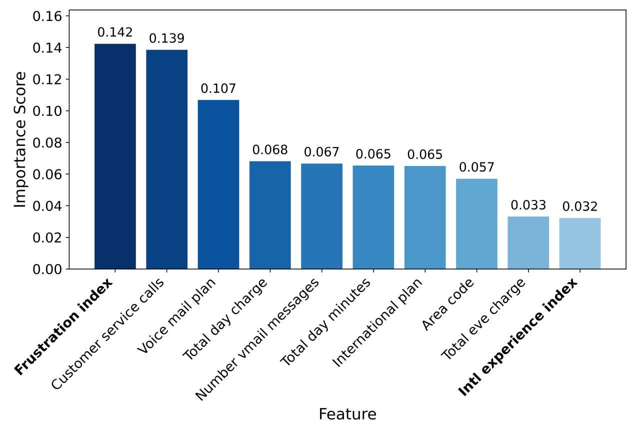

Fig. 6. Ten most important features in the classification process using XGBoost combined with SMOTE.

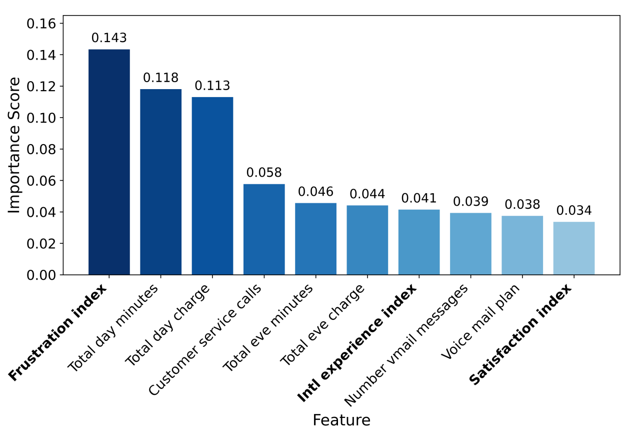

Fig. 5. Ten most important features in the classification process using RF combined with SMOTE-Tomek.

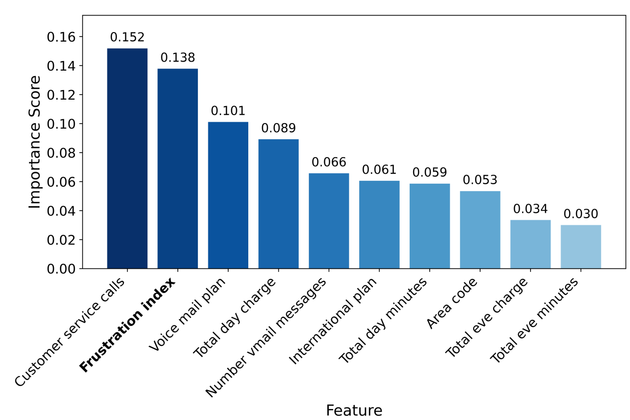

Fig. 7. Ten most important features in the classification process using XGBoost combined with SMOTE-Tomek.

Furthermore, feature importance analysis demonstrates how new emotion-driven and behavioral engagement features affect classifiers' performance. As shown in Fig. 2 to Fig. 7, the frustration index was the most important feature for all of the classifiers during the classification process, surpassing all of the original dataset's features, except with the combination of XGBoost with SMOTE, it was the second most important. The Satisfaction index and International experience index appeared in the top feature rankings obtained from RF + SMOTE, RF + SMOTE-Tomek, and XGBoost + SMOTE. Consequently, the repeated appearance of multiple emotional features in the top-ranked features for classifiers demonstrates their ability to improve the performance of classifiers in predicting customer churn.

A firm summary constructed from the second part of the experiments is that adding emotion-driven and behavioral engagement features consistently improves model performance in predicting both classes (i.e., churn and non-churn). This result is consistent with prior research indicating that customer experience and perceived service interactions play a critical role in shaping customer retention behavior and future decisions [21], [22].

TABLE III. Evaluation Metrics Obtained from the Combinations of Classifiers and Balancing Techniques Applied on the Emotion-Driven and Behavioral Engagement Features Only

Model	Accuracy	Precision	Recall	Specificity
CatBoost + SMOTE	0.8202	0.8163	0.8263	0.8140
CatBoost + SMOTE-Tomek	0.8444	0.8450	0.8435	0.8453
RF + SMOTE	0.8728	0.8571	0.8947	0.8509
RF + SMOTE-Tomek	0.8803	0.8707	0.8932	0.8674
XGBoost + SMOTE	0.8263	0.8142	0.8456	0.8070
XGBoost + SMOTE-Tomek	0.8508	0.8547	0.8453	0.8564

C. Results of the Third Experiment: Emotion-Driven and Behavioral Engagement Features Only

In the third part of the experiments, only five newly generated features were used to train and test classifiers. In addition, SMOTE and SMOTE-Tomek were adopted to solve the data balancing problem. Table III shows the results obtained by the combinations of classifiers and balancing techniques on the emotion-driven and behavioral engagement features.

Regarding the evaluation metrics values, the combinations of RF with SMOTE and SMOTE-Tomek outperformed all other combinations; they obtained the highest values for all evaluation metrics. On the other hand, the other combinations of classifiers with balancing techniques obtained considerable results with respect to predicting customer churn using the emotion-driven and behavioral engagement features only. Accordingly, the results obtained in this part of experiments proved that using the new features only to train and test classifiers to predict churn is not sufficient to achieve the same high levels of accuracy, recall, and specificity as when combined with original telecom dataset features. But it can still obtain relatively high values for the evaluation metrics.

D. Comparative Summary of the Three Experimental Stages

The three parts of the experiments gradually demonstrated the impact of inserting the emotion-driven and behavioral engagement features to the original dataset on the performance of classifiers in predicting churn. In the first part of experiments, only the original dataset's features were used to train and test the classifiers to predict churn. Thus, all classifiers showed strong performance, particularly when combined with SMOTE-Tomek. The second experiment introduced the emotion-driven and behavioral engagement features into the dataset, and yielded the highest performance across all models compared to the first part of experiments. In other words, there is an improvement in the performance of all classifiers and balancing techniques combinations regarding predicting both classes. All of the combinations trained and tested on the original dataset + emotion-driven and behavioral engagement features obtained better evaluation metrics values compared to the original dataset only. Moreover, feature importance analysis revealed that emotion-driven features, particularly, frustration index and satisfaction index, consistently ranked among the top ten important features in the classification process. This indicates the effectiveness of the emotion-driven features in predicting customer churn. Furthermore, the third part of experiments relied exclusively on the engineered features, and resulted in lower accuracy, recall, and specificity compared to the first and second parts of experiments. This suggests that while emotion-driven features provide significant improvement in the classifiers' performance, they are most effective when integrated with original telecom dataset features rather than used in isolation. Accordingly, the three parts of experiments confirm that the integration of emotion-driven features offers consistent added value for churn prediction.

VI. Conclusion

This study demonstrates that integrating standard (original) telecom features with five emotion-driven and behavioral engagement features, namely: frustration index, trust score, satisfaction index, service usage score, and international experience index, consistently improves churn prediction performance. Moreover, in three stages of evaluation, the results show that combining the telecom churn data with the proposed features yields better predictive outcomes than relying on the original telecom attributes only.

The experimental findings further indicate that emotion-driven and behavioral engagement features provide meaningful information that complements customer usage features (original data features), particularly in improving the prediction of customers at risk of churn. The consistent performance enhancement observed across Random Forest, CatBoost, and XGBoost indicates that the proposed features are effective across different classifiers, rather than being customized to a single classifier. In addition, SMOTE and SMOTE-Tomek proved their efficiency in addressing the data balancing problem, ensuring a reliable evaluation of churn prediction performance using multiple metrics (i.e., accuracy, precision, recall, and specificity).

From a practical perspective, the proposed features provide clear indicators of customer experience and engagement using existing operational data. Those indicators can be integrated into current churn prediction systems to improve performance and support retention strategies, without replacing existing prediction models.

Future research may extend this work by extracting similar features from richer data sources, such as customer surveys, textual interactions, or real-time service logs, and by evaluating their impact on deep learning–based churn prediction models. In addition, future research may analyze model explanations at both individual customer and customer segment levels to support more targeted and actionable retention strategies.

Declaration on the Use of Generative AI

The author declares that generative artificial intelligence (AI) tools were used solely for language refinement and grammatical improvement during the preparation of this manuscript. All scientific ideas, methodological design, feature engineering, experimental implementation, data analysis, interpretation of results, and conclusions were entirely conceived and developed by the author. The author has critically reviewed and edited all AI-assisted content and takes full responsibility for the accuracy, originality, and integrity of the work.

References

[1] A. Omari, O. Al-Omari, T. Al-Omari, and S. M. Fati, "A predictive analytics approach to improve telecom's customer retention," *Frontiers in Artificial Intelligence*, vol. 8, p. 1600357, 2025.

[2] L. R. N. Deiranieh, E. A. Alahmad, M. M. Qadri, A. Abdallah, F. Omeish, N. N. Ali, N. S. Al-Akash, and A. Ahmad, "Factors influencing switching behavior of customers in the telecommunication industry of jordan: The mediating role of csr," *Tec Empresarial*, vol. 19, no. 1, pp. 45–74, 2024.

[3] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," *Results in Control and Optimization*, vol. 14, p. 100342, 2024.

[4] A. Gaur and R. Dubey, "Predicting customer churn prediction in telecom sector using various machine learning techniques," in *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)*. IEEE, 2018, pp. 1–5.

[5] S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Machine Learning with Applications*, vol. 17, p. 100567, 2024.

[6] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, 2022.

[7] H. Aljawazneh and F. A. Al-Ragheb, *The Effect of Neural Networks Depth on Forecasting Performance: A Study on Customer Churn Prediction in Telecom Industry*. Cham: Springer Nature Switzerland, 2025, pp. 111–123.

[8] C. Wang, C. Rao, F. Hu, X. Xiao, and M. Goh, "Risk assessment of customer churn in telco using fclcnn-lstm model," *Expert Systems with Applications*, vol. 248, p. 123352, 2024.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, October 2001.

[10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.

[11] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 785–794. [Online]. Available: https://doi.org/10.1145/2939672.2939785

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[13] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

[14] Y. Huang, B. Huang, and M. Kechadi, "A new filter feature selection approach for customer churn prediction in telecommunications," in *2010 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, 2010, pp. 338–342.

[15] J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman, "A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection," *Plos one*, vol. 17, no. 12, p. e0278095, 2022.

[16] F. Mozaffari, I. R. Vanani, P. Mahmoudian, B. Sohrabi *et al.*, "Application of machine learning in the telecommunications industry: Partial churn prediction by using a hybrid feature selection approach," *Journal of Information Systems and Telecommunication*, 2023.

[17] S. Ramraj, N. Uzir, R. Sunil, and S. Banerjee, "Experimenting xgboost algorithm for prediction and classification of different datasets," *International Journal of Control Theory and Applications*, vol. 9, no. 40, pp. 651–662, 2016.

[18] H. A. Salman, A. Kalakech, and A. Steiti, "Random forest algorithm overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, 2024.

[19] B. J. Erickson and F. Kitamura, "Magician's corner: 9. performance metrics for machine learning models," p. e200126, 2021.

[20] A. Althunibat, R. Amro, B. Hawashin, H. AlNuhait, S. Almanasra, and H. A. Al-Khawaja, "Automated classification of user requirements written in arabic using machine learning algorithms," *Applied Mathematics & Information Sciences*, vol. 17, no. 6, pp. 1155–1170, 2023.

[21] P. C. Verhoef, K. N. Lemon, A. Parasuraman, A. Roggeveen, M. Tsiros, and L. A. Schlesinger, "Customer experience creation: Determinants, dynamics and management strategies," *Journal of Retailing*, vol. 85, no. 1, pp. 31–41, 2009.

[22] K. N. Lemon and P. C. Verhoef, "Understanding customer experience throughout the customer journey," *Journal of Marketing*, vol. 80, no. 6, pp. 69–96, 2016.