Regular paper

Journal of the KIECS. pp. 1-8, vol. 12, no. 1, Feb. 28 2017, t. 81, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2017.12.1.1

고객 이탈 예측을 위한 효율적인 스타킹 앙상블 학습 방법

홍길동[*] · 홍동순[**](휴먼명조 10.5포인트, 장평 100%, 자간 0)

An Efficient Stacking Ensemble Learning Method for Customer Churn Prediction

Gil-dong Hong[*] · Dong-sun Hong[**](휴먼명조 9.2포인트, 장평 95%, 자간 0)

요 약

최근 몇 년 동안 고객 이탈은 통신 회사의 중요한 문제이자 가장 중요한 관심사 중 하나였습니다. 신규 고객을 확보하기보다는 기존 고객을 유지하는 데 집중하는 것은 통신 산업에서 비용을 절감하고 수익을 늘리는 데 중요한 전략입니다. 이 연구에서는 불균형 데이터, 누적 모델 및 소프트 투표를 위한 SMOTE(합성 소수 과잉 샘플링) 기술로 구성된 앙상블 학습 기술을 사용하는 효율적인 고객 이탈 예측 모델을 제안합니다. 랜덤 포레스트, 익스트림 그래디언트 부스팅(XGBoost), Catboost 및 MLP(MultiLayer Perceptrons) 머신 러닝 알고리즘을 선택하여 스태킹 앙상블 모델을 구축하고 네 가지 알고리즘의 결과를 소프트 투표에 사용합니다.다른 예측 모델과 비교하여 제안된 모델은 원래 불균형 데이터 세트와 새로운 균형 데이터 세트에 대해 각각 78.79% 및 88.20%의 최고의 정확도를 보였습니다. 우리가 제안한 모델은 통신 산업에서 고객 이탈을 조기에 감지할 수 있습니다.

ABSTRACT

In recent years, customer churn has been a significant issue and one of the top concerns for telecommunications companies. Focusing on retaining existing customers rather than acquiring new customers is an important strategy for reducing costs and increasing revenues in the telecommunications industry. This study proposes an effective customer churn prediction model that uses an ensemble learning technique consisting of unbalanced data, stacking models, and synthetic minority oversampling (SMOTE) techniques for soft voting. Random Forest, Extreme Gradient Boosting (XGBoost), Catboost, and MultiLayer Perceptrons (MLPs) machine learning algorithms are selected to create the stacking ensemble model, and the results of the four algorithms are used for soft voting. Compared with other prediction models, the proposed model showed the best accuracy of 78.79% and 88.20% for the original imbalanced dataset and the newly balanced dataset, respectively. Our proposed model enables early detection of customer failures in the telecommunications industry.

키워드

Churn Prediction, Ensemble learning, Stacking, Soft voting, Telecommunication company

이탈 예측, 앙상블 학습, 스타킹, 소프트 보팅, 통신사

Ⅰ. Introduction

Most telecom services view customers as their most valuable asset. As a result, one of the most challenging problems that telco companies face nowadays is when clients switch to another

* 한국□□□□ 연구원(Email)
** 교신저자 : □□대학교 □□□□학과
・접 수 일 : 0000. 00. 00
・수정완료일 : 0000. 00. 00
・게재확정일 : 0000. 00. 00
・Received : May 00, 0000, Revised : July 00, 0000, Accepted : July 00, 0000
・Corresponding Author : Dong-Sun Hong
Dept. □□□□□□□□□, □□ University,
Email : □□□@□□□.ac.kr
(중고딕 8포인트, 장평 90%, 자간 -5, 들여쓰기 0, 줄간격 130%)

service provider for whatever reason. Because consumers can easily switch services, in many cases, churn can significantly impact company profitability [1]. Customer churn forecasting enables decision-makers to identify the causes of relationship breakdowns and develop strategies to reduce churn while improving profitability. Therefore, it is crucial and seen as a competitive advantage for telecom to anticipate customers' desire to cancel their subscriptions. Various approaches have been proposed for customer churn prediction cases. Linear Regression and Support Vector Machine (SVM) models were adopted and showed promising results for churn prediction problems in the early millennium. Recently, ensemble learning methods have become popular in customer churn prediction [2]. Ensemble methods are meta-algorithms that combine multiple machine learning models to make predictions better. There are two types of ensemble method: Bagging and Boosting. Random Forest is the most popular Bagging method, with XGBoost, LightGBM and CatBoost being to Boosting methods that have attracted a lot of attention in recent studies on binary classification. Previous studies of customer churn prediction have shown superior performance of Random Forest compared to traditional categorization methods. For example, showed that Random Forest achieved higher scores than Naïve Bayes, SVM, Decision Tree, Bagging and Boosting [3]. Author of showed higher rates of Random Forest compared to SVM, KNN and Logistics Regression [4]. Gradient Boosted Decision Trees (GBDT's) achieved the highest performance, followed by a Decision Tree and a Random Forest. KNN (0,575) and Neve Bayes (0,646) achieved low results [5]. Boosting-based algorithms as ensemble methods have shown outstanding classification effectiveness in various research areas, but have yet to show a broad advantage in research on customer churn prediction. XGBoost (Extreme Gradient Boosting) is a recently proposed ensemble method. This is an advanced way to gradient boosting [6]. In this paper, we first compared the most popular supervised machine learning algorithms with this problematic situation in the telecom industry. Then, an effective stacked ensemble model is proposed to overcome the problems associated with churn prediction. The rest of the paper is structured as follows: we review some churn prediction-related works in Section II. Dataset is described in Section III. Section IV analyzes different classification algorithms and ensemble methods with some essential approaches and explains the chosen algorithms. The section V explains the results followed by section VI that concludes the paper.

II. Related works

In order to understand and estimate the churn rate of telecommunication services, the industry has used various methods. Data mining, machine learning (ML) and deep learning (DL) algorithms have been widely used in the field. While most relevant studies have focused on using only one machine learning method for data extraction, some have compared multiple methods for classification disorders. Compared linear regression, decision trees, and MLPs to predict customer churn based on features associated with customer complaints. They split the data set evenly between non-churners and churners with a ratio of 50%. The authors demonstrated that MLPs are capable of accurately predicting churners [7]. Authors of presented an efficient data mining model for predicting customer turnover, utilizing a dataset of 3333 call details and 21 characteristics with two values of churn labels: Yes or No. They adopted the Principal Component Analysis (PCA) method to reduce the feature dimension before implementing Bayes Networks, SVM, and MLPs for churn prediction modeling. There were no missing values in the dataset utilized in this study since it was tiny [8]. In telco companies, focused on evaluating and analyzing the performance of a set of tree-based machine learning methods and algorithms for predicting churn based on the big data platform. After implementing data preprocessing, feature engineering, and feature selection techniques, the authors have experimented with several algorithms, namely Random Forest, Decision Tree, XGBoost, and Gradient Boosting to build the predictive model of customer churn. Working on a massive dataset obtained by processing enormous raw data given by the Syriatel telecom firm, the model was constructed and validated using the Spark environment. The dataset was used to train, test, and evaluate Syriatel's system, which includes all customer information for the entire nine months [9]. Data imbalance is also a key issue in churn prediction tasks, similar to other machine learning problems. In an imbalanced dataset, the number of churned customer labels is lower than the number of current customer labels. Several studies have looked into the issue of data imbalance. More details on how class imbalance affects different classification algorithms can be found in [10]. Many other solutions have been proposed to address this problem; they can be broadly classified into three categories: data-level, algorithm-level, and ensemble solutions. Intelligent sampling methods have been proposed to solve the problem of data imbalance, including synthetic minority oversampling methods (SMOTE) is probably the most popular one. SMOTE is commonly used as a benchmark for oversampling algorithm [11. 12]. Therefore, evaluated how well Weighted Random Forests, Gradient Boosting Model, Advanced Under Sampling, and Random Sampling performed in churn prediction models using unbalanced datasets. The under-sampling strategy paid off other strategies considered according to the results [13]. To solve the problem of telco outage prediction, the authors in considered six different sampling strategies. According to the results, Mega-Trend-Diffusion Function (MTDF) and rules-generation based on genetic algorithms outperformed the other oversampling methods [14]. Recently, designed factor analysis to analyze telco business characteristics building a discriminant model and a logistic regression model for forecasting customers and telecommunications Customer attrition using customer segmentation data from three major Chinese telecom companies [15]. The authors implemented a logistic regression model to predict the telecommunications customer churn in a new way. CatBoost is a new boosting algorithm compared to XGBoost and LightGBM. This algorithm is based on Gradient Boosted Decision Trees (GBDT's) and used for regression and classification tasks [16]. One of its strengths is its superiority in handling categorical features [17]. Recent studies have shown the algorithm's superior performance in binary format classification problems. Moreover, used CatBoost to predict corporate failure and found that CatBoost achieved higher accuracy and Area Under Curve Receiver Operator Characteristic (ROC-AUC) scores than discriminant analysis, Logistic Regression, SVM, Artificial Neural Network, Random Forest, Gradient Boosting, Deep Neural Network (DNN) and XGBoost [18].

Ⅲ. Datasets explanation

We used the Telco customer churn dataset [19], which contains information on a telecom company that provided home phone and Internet services to 7,043 customers in California in the third quarter. There are 21 features in the dataset

"Churn" is the target/dependent variable, and rest 30 are independent variables which we need to explore further. Only 3 features are numeric: "SeniorCitizen" (categorical), and "tenure" and "MonthlyCharges" (continuous). Datatype of the rest of the features is object, looking at the sample data they look like to be of type string. Some of these features are categorical, which we will map into numerical values. We split the data into the training set, validation set and test set in the ratio of 6:2:2. It indicates which customers have left, stayed, or signed up for their service. Multiple important demographics are included for each customer, as well as a Satisfaction Score, Churn Score, and Customer Lifetime Value (CLTV) index. The features of the dataset are the following:

Customers who left within the last month - the column is called churn;

Services that each customer has signed up for - phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies;

Customer account information - how long they have been a customer, contract, payment method, paperless billing, monthly charges, and total charges;

Demographic info about customers - gender, age range, and if they have partners and dependents.

Ⅳ. Applied methods

First, we process our dataset and then go for candidates of model selection to create an ensemble model classifier.

4.1 Data processing

In the real world, churn datasets typically contain irrelevant information, which includes irrelevant features, imbalanced datasets, disparate feature scales, non-numeric features, and missing values. We present five procedures in this research to process data before submitting it to training.

4.1.1 Missing values

We implemented several missing value solutions depending on the amount of missing data in the feature. The dataset excludes attributes with more than 90% missing values. For the empty data, column means and modes for categorical and numeric features, accordingly, are substituted in place of incomplete data in the dataset.

4.1.2 Categorical variables

Although many machine-learning algorithms work only with numerical values, many important real-world features are categorical rather than numerical. As categorical properties, they take degrees or values. Since these categorical features can not be directly used in most machine learning algorithms, categorical features must be converted to numerical features. Although there are many techniques for changing these properties, the most common method is one hot encoding.

4.1.3 Class imbalance

The classes of churn data in the real world are typically unbalanced (Fig. 1); specifically, the ratio of churners to non-churners is frequently substantially smaller. In order to have the same amount of data for different labels, we applied Synthetic Minority Oversampling Technique (SMOTE) to generate the synthetic data [20], which boosted the size of the minority class by random sampling (Fig. 2).

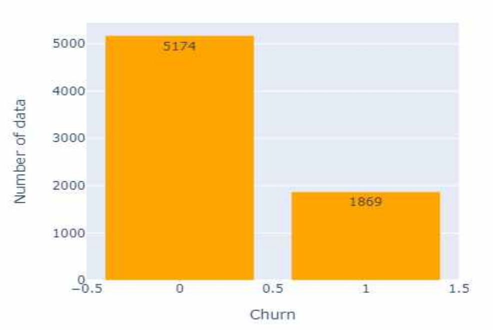

Figure 1. Distribution between churn and non-churn

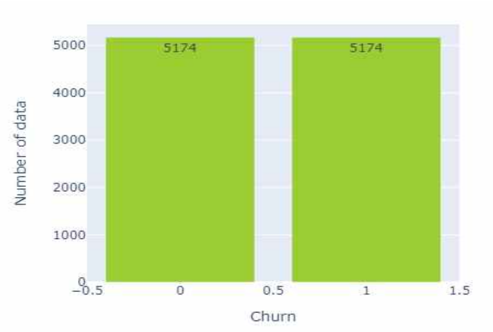

Figure 2. Distribution between churn and non-churn after smote algorithm

4.1.4 Feature selection

We select the most favorable features as churn indicators. First, new features are created based on the feature selection algorithm. Secondly, highly correlated features are eliminated since they often increase computing costs without improving model prediction capability. If we had 7043 data with 21 features at the beginning, now we possess 31 features after the feature selection process accordingly.

4.1.5 Normalization

Lastly, each feature is scaled through standardization, involving rescaling the features with the properties of a standard normal distribution with a mean of zero and a standard deviation of one. It helps the network training to converge better and faster, accelerating the model process speed. A Min-Max normalization is the process of taking data measured in its engineering units and transforming it to a value between 0.0 and 1.0. Where by the lowest (min) value is set to 0.0 and the highest (max) value is set to 1.0. This provides an easy way to compare values that are measured using different scales or different units of measure. The normalized value is defined as: Equation 1.

$$\mathrm{min}-\mathrm{max}-norm = \frac{x - x_{\mathrm{min}}}{x_{\mathrm{max}} - x_{\mathrm{min}}} \qquad (1)$$

4.2 Hyperparameter optimizationn

We used the default hyper-parameters which have been preset by the package in the previous step. However, these hyper-parameters are not guaranteed to be optimal. Therefore, we need a technique to tune the machine learning models and the optimal hyper-parameters for it, which is called hyper-parameter tuning. In this work, we used "GridSearch" optimization to automate the tuning of hyperparameters for the top models.

Gridsearch is essentially an optimization algorithm which lets you select the best parameters for your optimization problem from a list of parameter options that you provide, hence automating the 'trial-and-error' method.

Although it can be applied to many optimization problems, but it is most popularly known for its use in machine learning to obtain the parameters at which the model gives the best accuracy.

4.3 Evaluation metrics

There is a few measures that are commonly used within machine learning. Some steps have more than one name, which can lead to confusion, and the user names often depend on the technical area in which they are used. The essential measures used in this study are precision, recall, and F1-score, which in turn rely on the concepts of true positive rate and false positive rate [21]. Equation 2 calculates the accuracy metric. It identifies number of instances that were correctly classified.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

Here:

★ True Positives (TP): predicted positive, true value positive

★ False Positives (FP): predicted positive, true value negative

★ False Negatives (FN): predicted negative, true value positive

★ True Negatives (TN): predicted negative, true value negative

These tells us what portion of the data is correctly classified as positive. For any classifier, the TP rate must be high. TP rate is calculated by using Equation 3.

$$TP\,rate = \frac{True\,Positives}{Actual\,Positives} \qquad (3)$$

FP Rate tells us which part of the data are incorrectly classified as positive. The result of the FP rate must be low for any classifier. It is calculated by using Equation 4.

$$FP\,rate = \frac{False\,Positives}{Actual\,Potives} \qquad (4)$$

Precision, also known as Positive Predictive Value (PPV), indicates which part of the prediction data is positive. It is calculated by using Equation 5.

$$Precision = \frac{TP}{TP + FP} \qquad (5)$$

Recall in Machine Learning is defined as the ratio of Positive samples that were properly categorized as Positive to the total number of Positive samples. It is the probability that all the relevant instances are selected by the system. The low value of recall means many false negatives. It is calculated by using Equation 6.

$$Recall = \frac{TP}{TP + FN} \qquad (6)$$

The F1-score is a trade-off between correctly classifying all the data points and ensuring that each class contains points of only one class. It is calculated by using Equation 7.

$$F1 - score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \qquad (7)$$

A confusion matrix (CM) shown in Table 1 is a popular evaluation metric for classification problems. To illustrate the idea, we can think of the classification problem as a binary problem where the instance is classified correctly or not.

Table 1. Confusion matrix for customer churn prediction

	Predicted	
	Not-churn	Churn
Not-churn	**True Positive**	**False Positive**
Churn	**False Negative**	**True Negative**
	Not-churn	Churn

4.4 Receiver operating characteristic curve

A standard method for evaluating performance in churn prediction is to use Receiver Operating Characteristic (ROC) curves. To extract a measure from ROC curves, it is common to use the area under the curve (AUC). This metric can, for instance, be used to compare different types of classifiers to each other or the same kind of classifier but with varying parameter values [22].

Unlike accuracy, AUC is applicable when there is a class imbalance and evaluates the ability of a model to distinguish between classes based on the class membership probabilities. Previous research has also found that AUC is generally a better evaluation metric than accuracy regarding statistical consistency and discrimination. This makes it a suitable evaluation metric also when the data is balanced. The ROC AUC metric was used in this study to evaluate the final performance of the model.

Area Under the Receiver Operating Characteristics is an experiment analysis for the classification problem for given varying thresholds. AUC (Area Under The Curve) indicates the measurement or degree of distinction, whereas ROC (Receiver Operating Characteristics) is a likelihood curve. It shows how well the model can discriminate between classes. The higher AUC score indicates how well the model at predicting between classes. In our case, The higher the curve, the better the model distinguishes between churn and non-churn cases.

Ⅴ. Experimental Results

5.1 Introduce to proposed model

The procedure used for churn predection in this study is shown in (Fig. 3). The figure illustrates the techniques and algorithms used in this work. The main goal of our study is to compare two approaches to ensemble learning such as stacking and voting classifier, to investigate how they improve the performance of models to assist with customer churn prediction. Several models can be used as base learners for ensemble learning purpose. To investigate the performance of ensemble methods in customer churn prediction context the following steps were performed. First, individual models are trained with the training data set and then evaluated against the test set. Then, we select the best models with high accuracy and low log loss to create the ensemble model selection. In addition, we give a voting classifier technique as ensemble learning. A soft voting classifier is chosen to search for optimal parameters. When soft voting is used, the final prediction is made based on the average of the class probabilities predicted by each model. This allows assigning weights to each classifier, meaning that a more robust classifier gets more votes in predicting the results. This study proposes an efficient customer-churn prediction model consisting of model selection, a stacking model, and a soft voting approach.

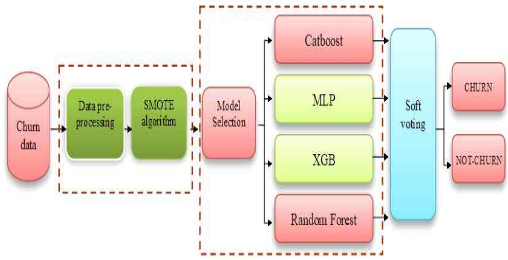

Figure 3. Proposed an efficient customer-churn prediction model

5.2. Results with imbalanced data

Accuracy and AUC scores for all individual models showed in Table 2. The research study indicates that Random Forest, XGBoost, Multilayer Perceptrons (MLPs), and CatBoost are the top four classifiers in terms of performance with accuracy scores of 0.7288, 0.7397, 0, 8286, and 0.8452, respectively. In summary, preliminary model exploration findings revealed that both Boosting classifiers and Multilayer Perceptrons (MLPs) perform well on this dataset. Compared to these four classifiers, other models showed less performance and efficiency.

Table 2. Results of individual models

Model	Accuracy	ROC-AUC
Catboost	0.8452	0.7135
MLP	0.8286	0.7014
XGBoost	0.7397	0.7162
Random Forest	0.7288	0.6994
SVM	0.7161	0.7090
KNN	0.7024	0.6895
Naïve Bayes	0.6753	0.7257

5.3. Results with balanced data

In the previous section, we mentioned that the churn prediction dataset differs significantly between churn and non-churn classes, with one class having a higher value than the other one. It showed in Table 3 that before and after results of smote algorithm, minority class "churn" is increased and balanced with "non-churn" classes. The new points added here are synthetically generated points, not exact replications of existing minority class instances. Thus, the overfitting problem caused by random oversampling was solved by SMOTE algorithm. Therefore, we compared model performance before applying SMOTE technique to the training data.

Table 3. Results of smote algorithm before and after balancing the dataset

Model	Accuracy	
	Imbalance dataset	Balanced dataset
Catboost	0.7936	0.8452
MLP	0.7813	0.8286
XGB	0.7993	0.7397
RF	0.7893	0.7288

5.4. Model selection

Furthermore, using those high-performed models, we defined a soft voting ensemble model.

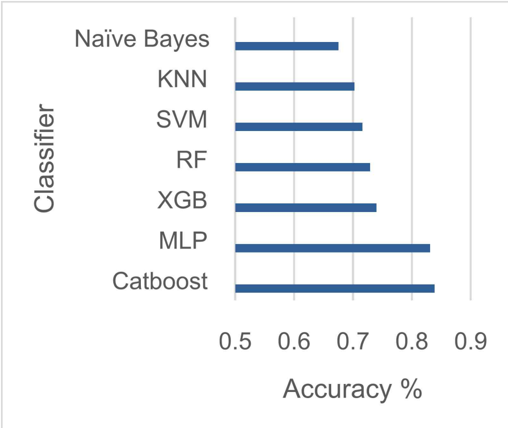

Figure 4. Classifier models in terms of accuracy

To choose which models for the ensemble, we compared seven machine learning models with accuracy and log loss, then got the four top models that showed the highest accuracy and lowest log loss. As we see from (Fig. 4 and 5), Catboost, MLP, XGB and Random Forest achieved the best results in terms of accuracy and loss.

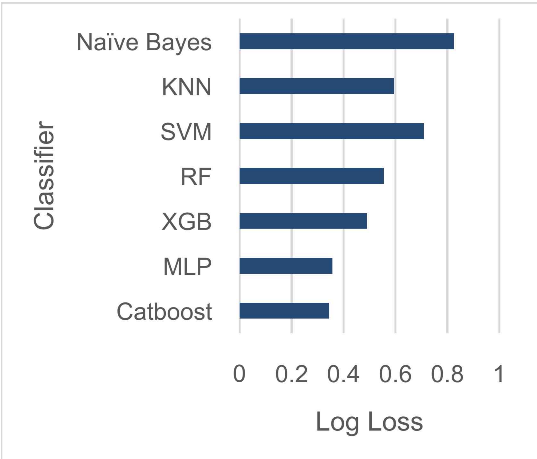

Figure 5. Classifier models in terms of Log Loss

5.4.1 Soft voting results for combination of the models

Another way of constructing an ensemble classifier is by voting among classifiers. Each independent classifier assigns a class label for each instance. Then using a voting scheme the class label of each instance is determined in this work.

Additionally, models with high accuracy, precision, recall, and f1-score are chosen in the soft voting process to achieve better results. Seven individual model combinations are calculated to implement the soft voting method. In Table 4, we can see that the particular model results are compared. Our proposed ensemble model achieved better results than other individual models.

Table 4. Soft voting results for combination of the models

Model selection	Models	Accuracy %	Precision	Recall	F1_score
1	1, 2, 3, 4	0.7804	0.5514	0.8940	0.7926
2	2, 3, 4, 5	0.8457	0.6471	0.9120	0.8527
3	3, 4, 5, 6	0.8622	0.6852	0.8833	0.8672
Proposed Model	**4, 5, 6, 7**	**0.8820**	**0.7088**	**0.8958**	**0.8797**
1. Naïve Bayes 2. KNN 3. SVM 4. Random Forest 5. XGB 6. MLP 7. CatBoost					

5.4.2 Comparison with other works

Table 6. Compares the proposed an efficient stacking ensemble method with other works recently. The proposed model shows the best accuracy.

Table 5. Comparison with other works

Works	Model	Accuracy %
Afifah Ratna Safitri [23]	Using smote and genetic algorithms	78.46%
Takuma Kimura [24]	Hybrid resampling and ensemble learning	77.10%
M. Imron [25]	Z-score normalization and particle swarm optimization	82.50%
Our work	An efficient stacking ensemble method	88.20%

5.5 Further analysis

It is only sometimes helpful to calculate the accuracy of the metric score, especially for unbalanced data. Therefore, to compare the different algorithms more clearly, we present a graphical type of confusion matrix and ROC-AUC curve analysis in (Fig. 6 and 7), respectively. As can be seen from these two graphs, our proposed ensemble model showed the best performance and balance in false positive (FP) and false negative (FN) rates, respectively.

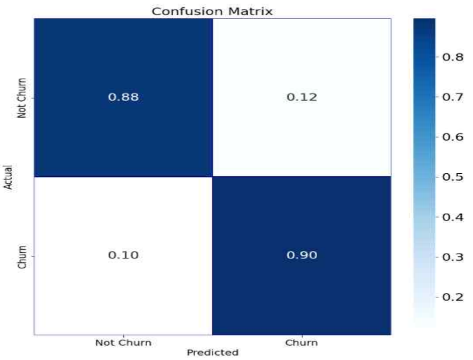

Figure 6. Confusion matrix for proposed ensemble model

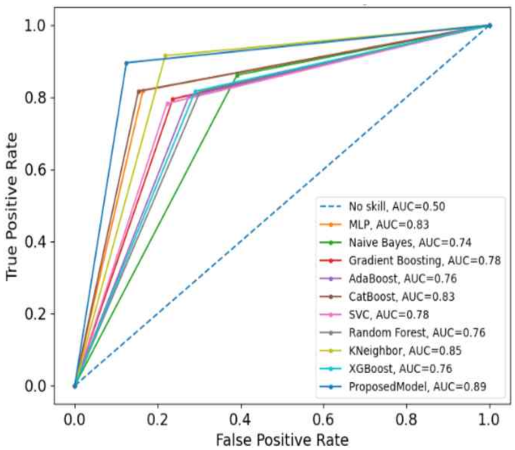

Figure 7. Comparison of roc auc curves for our work and other ml models.

VI. Conclusions

Nowadays, various machine-learning techniques have been used in telecommunications for customer churn. Our research comprehensively measures the most popular state-of-the-art machine learning methods. Quality measures of all candidate models were evaluated on a public dataset in the telecom industry. We have chosen the top four models—MLP, Random Forest, CatBoost, and XGBoost—based on an efficient stacking churn prediction model. We eventually created an ensemble model utilizing the soft voting approach. In addition to these four separate models with the best hyper-parameters, which had the best AUC score. The result of the proposed model showed the best accuracy of 78.79% and 88.20% for the original imbalance dataset and the balanced dataset, respectively, compared to other prediction models. This proposed model can provide early detection of customer churn in the telecom industry.

References

[1] O. Adwan, H. Faris, K. Jaradat, O. Harfoushi and N. Ghatasheh, "Predicting customer churn in telecom industry using multilayer perceptron neural networks: modeling and analysis", Igarss. 2014.

[2] X. Liang, S. Chen, C. Chen, & T. Zhang, "Research on Telecom Customer Churn Prediction Method Based on Data Mining". CCF Conference on Computer Supported Cooperative Work and Social Computing, 2019.

[3] A. Mishra, and U. Reddy, "A comparative study of customer churn prediction in telecom industry using ensemble based classifiers". 2017 International Conference on Inventive Computing and Informatics (ICICI). IEEE, 2017.

[4] M. Singh, S. Singh, N. Seen, S. Kaushal and H. Kumar, "Comparison of learning techniques for

prediction of customer churn in telecommunication", IEEE. 2018.

[5] S. Raeisi, & H. Sajedi, "E-Commerce Customer Churn Prediction By Gradient Boosted Trees", IEEE, 2020.

[6] T. Chen & C. Guestrin, Xgboost: "A scalable tree boosting system", Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016.

[7] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, “Churn prediction using complaints data”, in Proceedings Of World Academy Of Science, Engineering and Technology, 2006.

[8] I. Brandusoiu, G. Toderean, B. Ha "Methods for churn prediction in the prepaid mobile telecommunications industry”, In International conference on communication, 2016.

[9] A. K. Ahmad, A. Jafar, and Kadan Aljoumaa,” Customer churn prediction in telecom using machine learning in big data platform”, J Big Data, 2019.

[10] Y. Sun, A.K.Wong, M.S. Kamel, Classification of imbalanced data: A review, International Journal of Pattern Recognition and Artificial Intelligence 23(4), 2009.

[11] J. Wang, M. Xu, H. Wang and J. Zhang, "Classification of imbalanced data by using the smote algorithm and locally linear embedding", Proc. 8th Int. Conf. Signal Process., vol. 3, pp. 1-4, 2006..

[12] C. S. Ertekin, "Adaptive oversampling for imbalanced data classification", Proc. 28th Int. Symp. Comput. Inf. Sci.,vol. 264, pp. 261-269, Sep. 2013.

[13] D. Burez and V. den Poel, "Handling class imbalance in customer churn prediction”, Expert Syst Appl, 2009.

[14] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, A. Hussain, "Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study”, IEEE Access, 2016

[15] T. Zhang, S. Moro, R. F. Ramos, "A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation”, Future Internet, 2022.

[16] V. Dorogush, V. Ershov “CatBoost: gradient boosting with support for categorical features”, 2018.

[17] J. Hancock and T. Khoshgofar, “CatBoost for Big Data: An Interdisciplinary Review”. Journal of Big Data, 2020.

[18] Sami Ben Jabeura, Cheima Gharib, Salma Mefteh-Wali, Vissal Ben Arfi “CatBoost Model and Artificial Intelligence Techniques for Corporate Failure Prediction”, Technological Forecasting and Social Change, 2021.

[19] IBM, "Predict behavior to retain customers. You can analyze all real-world customer data and develop focused customer retention programs”, 2019.

[20] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. “SMOTE: synthetic minority over-sampling technique”. Journal of Artificial Intelligence Research, 2020.

[21] Max Bramer, “Principles of Data Mining," Springer London, 2 edition, 2013.

[22] Koen W, De Bock, and Dirk Van den Poel, “Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models”, Expert Systems with Applications, 2012.

[23] Afifah Ratna Safitri and M. A. Muslim, “Improved accuracy of Naive Bayes classifier for determination of customer churn uses SMOTE and genetic algorithms”, Journal of Soft Computing Exploration, 2020.

[24] T. Kimura, “Customer churn prediction with hybrid resampling and ensemble learning”. Journal of Management Information and Decision Sciences, 2022.

[25] M. Imron and Budi Prasetyo, “Improving algorithm accuracy k-nearest neighbor using Z-score normalization and particle swarm

optimization to predict customer Churn", Journal of Soft Computing Exploration, 2020.

저자 소개

홍길동(Gil-dong Hong)

1995년 □□대학교 □□학과 졸업(공학사)
0000년 □□대학교 대학원 □□학과 졸업(공학석사)
0000년 □□대학교 대학원 □□학과 졸업(공학박사)
0000년 □□대학교 컴퓨터공학과 교수
0000년 ~현재 □□시 □□□ 위원
※ 관심분야 : □□통신시스템, □□통신