Interdisciplinary Journal of Information, Knowledge, and Management

An Official Publication of the Informing Science Institute
InformingScience.org

IJIKM.org

Volume 18, 2023

A NOVEL TELECOM CUSTOMER CHURN ANALYSIS SYSTEM BASED ON RFM MODEL AND FEATURE IMPORTANCE RANKING

Tianpei Xu	Hulunbuir University, HulunBuir, China	xutianpei@hlbec.edu.cn
Ying Ma	Nanchang Hangkong University, Nanchang, China	marian0225@163.com
Changyu Ao	Chonnam National University, Yeosu, Korea	aochangyu@jnu.ac.kr
Min Qu	Jiangsu Vocational Institute of Commerce, Nanjing, China	220018@jvic.edu.cn
XiangHong Meng*	Inner Mongolia Minzu University, Tongliao, China	mxh-hlr@163.com

* Corresponding author

ABSTRACT

Aim/Purpose	In this paper, we present an RFM model-based telecom customer churn system for better predicting and analyzing customer churn.
Background	In the highly competitive telecom industry, customer churn is an important research topic in customer relationship management (CRM) for telecom companies that want to improve customer retention. Many researchers focus on a telecom customer churn analysis system to find out the customer churn factors for improving prediction accuracy.
Methodology	The telecom customer churn analysis system consists of three main parts: customer segmentation, churn prediction, and churn factor identification. To segment the original dataset, we use the RFM model and K-means algorithm with an elbow method. We then use RFM-based feature construction for customer churn prediction, and the XGBoost algorithm with SHAP method to obtain a feature importance ranking. We chose an open-source customer churn dataset that contains 7,043 instances and 21 features.

Accepting Editor Zaenal Akbar | Received: May 26, 2023 | Revised: July 29, September 2, September 20, 2023 | Accepted: September 21, 2023.
Cite as: Xu, T., Ma, Y., Ao, C., Qu, M. & Meng, X. (2023). A novel telecom customer churn analysis system based on RFM model and feature importance ranking. *Interdisciplinary Journal of Information, Knowledge, and Management*, *18,* 719-737. https://doi.org/10.28945/5192

Contribution	We present a novel system for churn analysis in telecom companies, which encompasses customer churn prediction, customer segmentation, and churn factor analysis to enhance business strategies and services. In this system, we leverage customer segmentation techniques for feature construction, which enables the new features to improve the model performance significantly. Our experiments demonstrate that the proposed system outperforms current advanced customer churn prediction methods in the same dataset, with a higher prediction accuracy. The results further demonstrate that this churn analysis system can help telecom companies mine customer value from the features in a dataset, identify the primary factors contributing to customer churn, and propose suitable solution strategies.
Findings	Simulation results show that the K-means algorithm gets better results when the original dataset is divided into four groups, so the K value is selected as 4. The XGBoost algorithm achieves 79.3% and 81.05% accuracy on the original dataset and new data with RFM, respectively. Additionally, each cluster has a unique feature importance ranking, allowing for specialized strategies to be provided to each cluster. Overall, our system can help telecom companies implement effective CRM and marketing strategies to reduce customer churn.
Recommendations for Practitioners	More accurate churn prediction reduces misjudgment of customer churn. The acquisition of customer churn factors makes the company more convenient to analyze the reasons for churn and formulate relevant conservation strategies.
Recommendations for Researchers	The research achieves 81.05% accuracy for customer churn prediction with the Xgboost and RFM algorithms. We believe that more enhancements algorithms can be attempted for data preprocessing for better prediction.
Impact on Society	This study proposes a more accurate and competitive customer churn system to help telecom companies conserve the local markets and reduce capital outflows.
Future Research	The research is also applicable to other fields, such as education, banking, and so forth. We will make more new attempts based on this system.
Keywords	CRM, churn prediction, feature construction, RFM, K-means, XGBoost, SHAP method

INTRODUCTION

Due to the intensification of competition in the telecommunication industry leading to a serious problem of customer churn in the companies concerned, for better business development it is necessary to provide not only high-quality technical services but also high-quality personalized services to prevent churn (Sudharsan & Ganesh, 2022). It is estimated that the cost of acquiring a new customer is 5-6 times more than the cost of retaining an existing customer (Santharam & Krishnan, 2018). A study shows that the average annual customer churn rate in the telecom industry is about 20% (Kuznietsova et al., 2021). Therefore, understanding the needs of customers to reduce churn became one of the main concerns of telecommunication companies (Fujo, 2022). Traditional churn systems (Bhattacharyya & Dash, 2021) cannot effectively tap into the needs of each individual and it is difficult to assist companies in accurately predicting possible churn. With the emergence of emerging technologies such as artificial intelligence and machine learning, telecommunication companies can utilize customer data to provide diversified services and derive greater value from their customer base, which also brings new opportunities for telecom companies (Sharaf Addin et al., 2022). In 2021, a customer churn system was proposed to be applied in the telecommunication industry (Sağlam & El-Montaser, 2021). The system reduced the company's customer churn rate by 5% and

increased profitability from 25% to 85%. Therefore, a novel system for customer churn analysis can help companies grow better and become more competitive.

Given that customers represent a telecommunications company's most valuable asset, it is imperative to comprehend and manage them effectively. A framework designed to oversee the interactions between a business and its present or prospective customers is commonly referred to as customer relationship management (CRM) (Guerola-Navarro et al., 2021).

The Recency, Frequency, and Monetary (RFM) model is an important tool and instrument for measuring customer value and customer profitability. Among the many analytical models for Customer Relationship Management (CRM), the RFM model is widely mentioned (Wan et al., 2022). The model describes the value status of a customer through three indicators: the customer's recent purchasing behavior, the overall frequency of purchases, and the amount of money spent. Through the analysis of customer data, organizations employing CRM strategies can enhance profitability, reduce customer churn, and identify new market opportunities (Abdullah, 2021). The importance of CRM is heightened for businesses when the market transitions from a commodity-centric model to a customer-centric one (Bonacchi & Perego, 2012). To optimize the value of each customer, customer segmentation involves dividing consumers into distinct clusters based on their characteristics, behaviors, preferences, and other relevant factors. This technique can be utilized to inform budget allocation, product development, marketing strategies, and promotional tactics (T. Chen et al., 2012; Jovanov & Disoska, 2021; Tauni et al., 2014).

The methods employed for customer segmentation are not standardized, and businesses often choose the most appropriate technique based on the unique characteristics of their respective industries. In advanced churn prediction techniques, K-means and RFM models are applied to generate new features. The addition of features makes the model easier to understand the data and thus learn better, with the aim of making more accurate predictions (Barus & Nathasya, 2022). The application of the SHAP method (Wu & Zhou, 2022) for ranking the importance of features to assist the machine learning method for final prediction has proven its superior performance in the e-commerce field. The ordering of feature importance allows the model to prioritize the learning of more important features thus improving the training efficiency and accuracy of the model.

However, telecom customer data is often severely incomplete, and the information content of each feature is relatively low, making many methods difficult to predict accurately. This fact is a common issue for telecommunications companies worldwide. In addition, this research usually neglects to assist companies in finding the cause of customer churn. Therefore, this paper aims to propose a more accurate and novel system for a customer churn analysis system based on feature importance ranking and data extension techniques for the company to gain stronger competitiveness.

The contributions of this paper are described as follows:

1. A comprehensive system for evaluating customer churn that encompasses feature importance ranking, customer churn prediction, and customer segmentation is proposed.
2. The K-means algorithm and RFM model are combined to segment customers and generate new features for data extension.
3. The SHAP method was applied to determine the importance rankings of individual features, and the XGBoost algorithm is applied to facilitate more accurate churn prediction.

The structure of this paper is structured as follows. The next section provides a literature review of the main concepts of clustering, RFM models, and feature importance. The paper then describes the specific dataset and research methodology. Next, we provide some discussion of the simulation results. Finally, the conclusion is presented.

LITERATURE REVIEW

In telecom companies, customer segmentation and customer management are important methods to prevent customer churn. Customer segmentation means that a company divides its customers into several categories according to certain rules based on their attributes, characteristics, and so forth. Customer management refers to a means of improving the competitiveness of an enterprise by increasing customer satisfaction through in-depth analysis of customer details (Guerola-Navarro et al., 2021). In recent years, telecom companies are facing a serious problem of subscriber churn. Some research (Mohamed & Al-Khalifa, 2023) has been proposed to predict customer churn. However, these models usually focus on more accurate prediction and neglect to assist companies in customer segmentation and management. This problem makes some current research difficult to apply in real-world telecom scenarios. Therefore, balancing both churn prediction accuracy and churn cause analysis is an urgent need for telecom companies. This paper not only improves the prediction accuracy but also analyzes the reasons for customer churn to better assist the company's decision-making.

Some researchers think about improving the accuracy of model prediction by clustering method.

CLUSTERING

Clustering refers to the process of grouping data points in a way that maximizes the similarity between data points within the same group and minimizes the similarity between data points in different groups. Several clustering techniques have been proposed in the literature, and many studies have compared and discussed their relative merits (Fu et al., 2020; Guyeux et al., 2019).

The clustering approach has yielded good results in other customer churn-related areas. The study conducted by Mulyawan et al. (2019). showed that the K-means algorithm effectively supported the accuracy of their recommendations. The clusters were defined using the K-means method, which minimized the total within-cluster sum of squares (WSS) (Claypo & Jaiyen, 2015). The WSS is calculated as the sum of the squares of the distance between each data point and its cluster center. The value of WSS varies depending on the number of clusters K, and there are several methods for determining K, including the elbow technique, silhouette score, and gap statistic approach (Nandapala & Jayasena, 2020; Ogbuabor & Ugwoke, 2018; Thorndike, 1953). Additionally, there are clustering algorithms based on spatial density, such as DBSCAN and OPTICS. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) employs a minimal set of parameters to identify clusters of arbitrary shapes. Several researchers have utilized DBSCAN, a density-based spatial clustering algorithm, to delineate regions that are susceptible to wildfires (Anwar et al., 2019). Various data mining techniques, such as PAM, CLARA, DBSCAN, and multiple linear regression, were employed to analyze agricultural data, to identify optimal parameters for enhancing crop productivity (Majumdar et al., 2017). In large-scale agricultural data clustering tasks, DBSCAN was found to outperform PAM and CLARA clustering algorithms, as demonstrated in the results. Additionally, a density-based cluster identification approach in geographic data, known as OPTICS (Ordering Points to Identify the Clustering Structure), was employed to enhance the ability to anticipate the quality of assembled chips by extracting features from cell-level data from the probing test process. This method led to a reduction in manufacturing costs as a side effect (Kim & Baek, 2014). The K-means technique is a widely used method in data mining, data analysis, and customer segmentation. It is a simple yet effective approach for clustering datasets quickly and efficiently (Cen et al., 2022).

In the telecom industry, Rachmahwati et al. (2022) utilized the K-means algorithm to segment customers in telecom companies into five groups. Their study found that customers with high income levels and high spending scores were prime targets for marketing campaigns. Xian et al. (2022) applied the K-means algorithm to cluster the transaction behavior of telecom companies to help companies gain a better understanding of their customer's demands.

These methods demonstrate the potential of clustering methods to improve the accuracy of telecom customer churn prediction. In this paper, we utilize clustering methods to construct more new features based on the original features, thus expanding the original dataset so that the model learns comprehensively from data to improve prediction accuracy.

RFM Model

The RFM company model has been successfully applied in numerous areas related to customer churn. The RFM company model prioritizes analyzing and quantifying the behavior of existing customers over acquiring new ones (Hallishma, 2023). The RFM model utilizes three key metrics to assess customer behavior: recency, frequency, and monetary value. Recency refers to the time elapsed since the customer's last purchase, while frequency relates to the number of purchases made within a particular time period, and monetary is the consumption in a period. The RFM score, which varies by customer cluster, is used to quantify customer value (Ha & Park, 1998). Parikh and Abdelfattah (2020) applied RFM models to the retail sector and demonstrated the significant cost-reduction benefits of identifying valuable customer groups for marketing efforts. Maryani et al. (2018) employed the RFM approach to conducting data mining and successfully divided 102 customers from 82,648 transaction data into two groups based on distinct characteristics, enabling companies to make tailored marketing decisions for each customer group. However, when the factors that influence consumer behavior are complex, it can be difficult to segment customers using the conventional RFM approach. To improve the RFM model's ability to segment customers, researchers have proposed various updated RFM models. For instance, Hu et al. (2020) developed the RFMT model as a demonstration. Huang et al. (2020) proposed an improved RFM model, called the RFMC model, which incorporates the community relationship value C as an indication of customer loyalty and the potential for ongoing consumption. In the case of video-on-demand services, Guney et al. (2020) employed a data mining technique using the LRFMP model, K-means algorithm, and a-priori algorithm to segment customers based on their consumption patterns. The length (L) and periodicity (P) factors were introduced to the RFM model to better understand the characteristics of consumers with different consumption behaviors.

The key advantage of the RFM company model is that it can identify the most valuable customer groups, and in any industry, high-quality customers are the core assets. Therefore, in the telecom industry, the application of the RFM company model can assist companies in customer segmentation, find the most valuable customer groups, and prioritize the prevention of the loss of this part of the customers can be in the same labor cost for the telecom company to ensure the highest efficiency. In this paper, the use of the RFM company model for telecom customer segmentation aims to provide customer prioritization for better decision-making assistance. In addition, the most valuable customer group is identified and tagged to enable the model to learn customer distinctions to further improve churn prediction accuracy.

Feature importance ranking

Feature importance ranking is widely used to enhance the interpretability of predictive models to find the causes of customer churn (Wojtas & Chen, 2020). Various techniques such as Random Forest (Fei et al., 2022), Regularized Linear models (Fonti & Belitser, 2017), LightGBM (Machado et al., 2019), and XGBoost (extreme gradient boosting) algorithm (M. Chen et al., 2019) are available for feature importance ranking. Among these, the XGBoost algorithm has demonstrated superior performance on a range of issues. Lundberg and Lee's (2017) SHapley Additive exPlanations (SHAP) is a method that utilizes cooperative game theory to determine the importance of input features in a prediction model. SHAP calculates the marginal contribution of each input feature by altering its value while keeping other conditions constant, and then recombining the contributions (Guo et al., 2021). With the SHAP approach, it is possible to better understand how the input characteristics affect the prediction results thus assisting organizations in finding the causes of customer churn.

In summary, in the field related to customer churn, clustering algorithms are used to improve prediction accuracy through feature expansion. The RFM model assists the model to learn better by identifying the most valuable group of customers and prioritizing customer acquisition for the telecom companies. The feature importance ranking approach allows organizations to identify the reasons for customer churn. Thus, a comprehensive system for better evaluating and predicting telecom customer churn is possible.

MATERIALS AND METHODS

DATASET

Research on customer churn analysis systems utilizes an open-source customer churn dataset from the Kaggle (n.d.) website, which contains information about a telecom company that provided home phone and internet service to 7,043 customers in California in the third quarter of 2019. While the dataset may not fully represent all aspects of the global telecommunications industry, specific regions or countries may have their unique characteristics or issues. However, it may still encompass many of the common problems and key features faced by the global telecommunications industry, such as customer churn and consumer behavior. It indicates which customers have churned, stayed, or signed up for their service, among 21 other items of customer information. Each customer contains multiple important demographics. The target variable "churn" contains two categories: churn (yes) and non-churn (no). Table 1 lists the features, feature descriptions, and data types.

Table 1. Dataset description

Feature	Description	Data type
customerID	Customer ID	Nominal
Gender	Whether the customer is a male or a female	Nominal
SeniorCitizen	Whether the customer is a senior citizen or not	Nominal
Partner	Whether the customer has a partner or not	Nominal
Dependents	Whether the customer has dependents or not	Nominal
Tenure	Number of months the customer has stayed with the company	Numeric
PhoneService	Whether the customer has a phone service or not	Nominal
MultipleLines	Whether the customer has multiple lines or not	Nominal
InternetService	Customer's Internet service provider	Nominal
OnlineSecurity	Whether the customer has online security or not	Nominal
OnlineBackup	Whether the customer has an online backup or not	Nominal
DeviceProtection	Whether the customer has device protection or not	Nominal
TechSupport	Whether the customer has tech support or not	Nominal
StreamingTV	Whether the customer has streaming TV or not	Nominal
StreamingMovies	Whether the customer has streaming movies or not	Nominal
Contract	The contract term of the customer	Nominal
PaperlessBilling	Whether the customer has paperless billing or not	Nominal
PaymentMethod	The customer's payment method electronic check, mailed check, bank transfer	Numeric
MonthlyCharges	The amount charged to the customer monthly	Numeric
TotalCharges	The total amount charged to the customer	Nominal
Churn	Whether the customer churned or not	Numeric

Proposed Customer Churn Analysis System

The proposed customer churn analysis system includes customer segmentation, feature construction, customer churn prediction, and customer churn factor identification, as shown in Figure 1. The system is aimed at discovering churned customers and factors of customer churn. K-means algorithm combined with the RFM model is used to segment customers and new features of customers are constructed based on the segmentation results. The XGBoost algorithm is used to predict churned customers and the XGBoost algorithm and SHAP algorithm are used to identify churn factors.

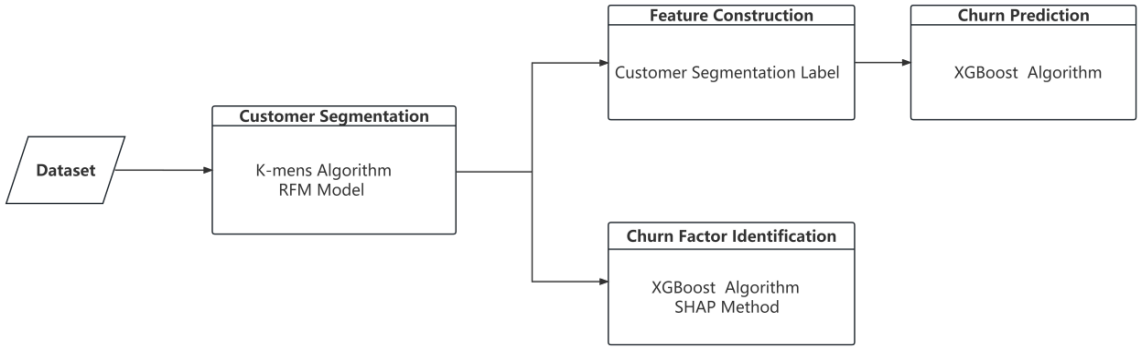


Figure 1. Churn analysis system structure

Customer Segmentation

The customer segmentation is performed by using the K-means algorithm combined with the RFM model. The K-means algorithm is a cluster analysis algorithm for iterative solutions. The principle of the algorithm is to divide the data into K groups, randomly select K objects as the initial clustering centers, then calculate the distance between each object and each seed clustering center and assign each object to the clustering center closest to it. The cluster centers and the objects assigned to them represent a cluster. The elbow approach, a heuristic for estimating the number of clusters in a data set, is employed in the K-means algorithm. The calculation of the explained variation as a function of the number of clusters (K), the procedure entails choosing the elbow of the curve as the appropriate number of clusters (Sari et al., 2022).

The RFM model is a method used to objectively rank and organize clients based on the recency, frequency, and monetary value of their most recent transactions to determine the best customers and execute focused marketing campaigns. Recency means the time of the customer's most recent purchase. Frequency means that the customer buys more often in a certain period. Monetary value means the amount of money the customer spent in a given period. When RFM scores are used in this paper, the up and down symbols used previously in the literature are also used here. The technique shows that the R, F, and M values of the cluster are above the aggregate average; therefore, the up symbol (↑); otherwise, the down (↓) symbol (Ha & Park, 1998). The customer segmentation process is shown in Figure 2.

Step 1: RFM variables (recency, frequency, and monetary) correspond to dataset features. The three metrics are used to represent a customer's value profile. In the telecom industry, most customers opt for monthly payments for their communication expenses, and thus the variable 'Recency' is defined as the monthly payment, with a fixed value of 1 for all customers in this RFM model. As the feature 'Tenure' describes how long a customer has been with the company, it is related to the variable 'Frequency.' Regular payments made by telecom customers can impact their customer loyalty and purchasing frequency, with higher frequency resulting in increased loyalty and purchases. Furthermore, the 'MonthlyCharges' attribute aligns with the monetary variable, as it reflects the customer's value

over a given period. As such, the RFM model parameters are defined as follows: Recency corresponds to 'Monthly payments,' Frequency corresponds to 'Tenure,' and Monetary corresponds to 'Monthly charges.'

Step 2: Determine the number of clusters (K). Since the K-means method is an unsupervised clustering technique, K must be predetermined to divide the dataset into K clusters according to the K-means guidelines. Instances in the same cluster are highly similar, while instances in different clusters have low similarity. To identify the clusters, the K-means method minimizes the total within-cluster sum of squares (WSS), which is the sum of squares of the distance between each location and the cluster center. The WSS value varies depending on the number of clusters K. The ideal K value is obtained using the elbow approach.

Step 3: Clustering by K-means. All customers into K clusters based on the results of Steps 1 and 2, using the RFM variables and the number of clusters as input data.

Step 4: Segmentation by RFM score. When RFM scores are used in this paper, the up and down symbols used previously in the literature (Ha & Park, 1998) are also used here. The technique shows that the R, F, and M values of the cluster are above the aggregate average; therefore, the up symbol (↑); otherwise, the down (↓) symbol. Based on RFM scores we can segment customers and discover the customer value of each customer cluster.

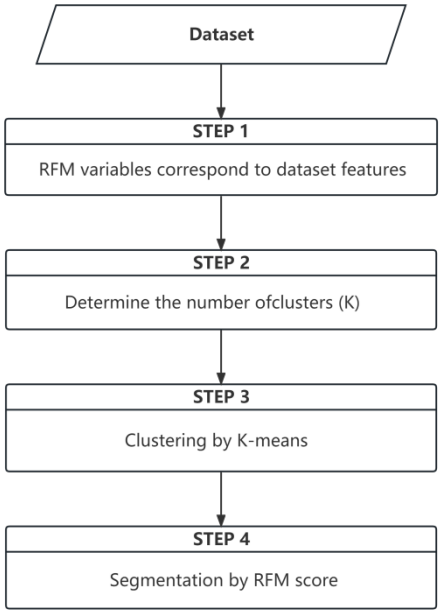


Figure 2. Customer segmentation process

Feature Construction

Feature construction is a technique used to bridge gaps in knowledge about the relationships between features by inferring or creating new features based on existing ones. This technique aims to improve data mining objectives, such as increased accuracy, better understanding, identification of true clusters, and discovery of hidden patterns. By creating new features, the original feature space is transformed into a new one, thereby enabling better analysis and interpretation of the data (Liu & Motoda, 1998).

In this paper, we construct the customer segmentation result as a new customer feature used to increase the feature space of the dataset.

CHURN PREDICTION

Telecom customers often switch providers for various reasons, leading to high churn rates. Companies can take measures to reduce churn after identifying potential churners. To achieve this, advanced algorithms must be employed. Therefore, the XGBoost algorithm is used in this paper.

The XGBoost algorithm, an improvement on gradient boosting decision trees, is known for its effective boosting decision tree construction and parallel processing capabilities (T. Chen & Guestrin, 2016). This algorithm employs the first-order and second-order Taylor expansions of the loss function and features a powerful supervised learning system consisting of several decision trees. Additionally, the XGBoost technique includes a regularization component in the objective function that simplifies the model and prevents overfitting, resulting in remarkable prediction accuracy.

When the model completes the prediction task, the model performance can be evaluated using some metrics such as accuracy, precision, recall, and f1-score (Hossin & Sulaiman, 2015).

CHURN FACTOR IDENTIFICATION

When a telecom company identifies potential churn customers, using advanced methods to fully understand the underlying factors of churn is critical to developing effective strategies and services to retain productive employees. However, some advanced machine learning models are often referred to as 'black box' models because their internal computations are not fully understood. While cross-validation can demonstrate the high accuracy of modern machine learning models, their reliability is questioned due to their poor interpretability (Shams Amiri et al., 2021). Researchers may have more confidence in a model if they can precisely quantify the importance of each input feature. Therefore, this paper uses the XGBoost algorithm combined with the SHAP approach for calculating feature importance ranking (churn factor importance ranking). This method means that it can accurately calculate the feature importance of each feature of the customer to the target feature and also explain the feature importance ranking. Based on this approach, telecom companies can develop more targeted marketing strategies.

The SHAP method uses the SHAP value (Rozemberczki et al., 2022) to calculate the average of the marginal contributions of all features, clarifying the results of any machine learning method. The SHAP value is calculated as shown below:

$$\Phi_i = \sum_{S \subseteq F, \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S\left(x_S\right) \right] \qquad (1)$$

Where Φ_i is i^{th} feature's SHAP value. F is the set of all features. S is the subset of all features obtained from F after removing the i^{th} feature. Two models $f_{S\cup\{i\}}$ and f_S, are trained, and predictions of these two models are compared to the current input $f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S\left(x_S\right)$, where x_S means the values of the input features in the set S.

RESULTS

CUSTOMER SEGMENTATION

The elbow method is applied to determine the optimal number of clusters based on the curve of WSS values. As shown in Figure 3, the WSS value generated by the K-means algorithm ranges from 2 to 10, and a substantial decrease occurs in the value as K increases from 2 to 4. However, there is little change in the WSS value beyond K=4. Therefore, the optimal number of clusters is determined to be 4.

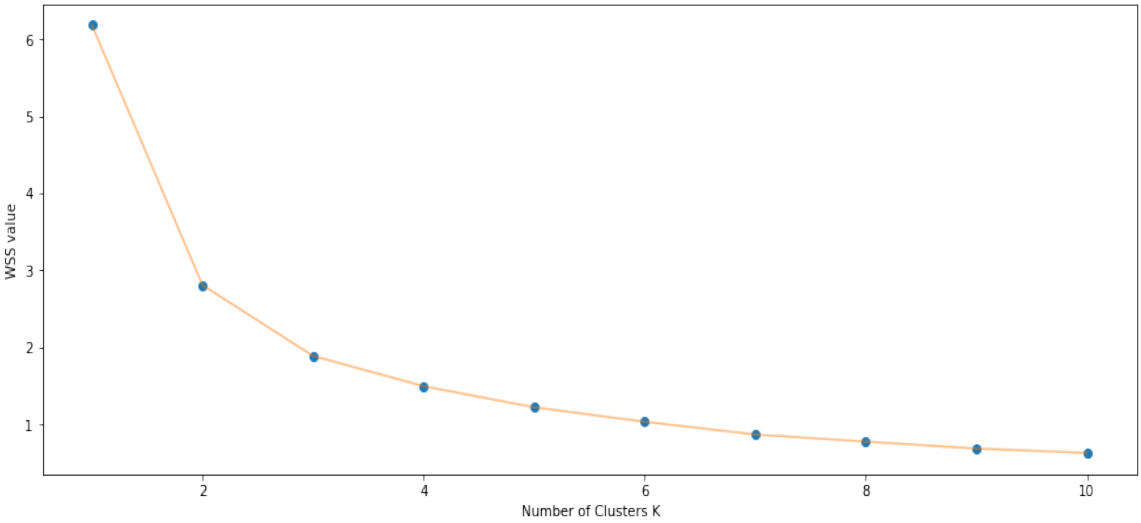


Figure 3. WSS value curve

The dataset is partitioned into 4 clusters by utilizing the K-means method with K=4 and the RFM model, with 'MonthlyPayment,' 'Tenure,' and 'MonthlyCharges' as R, F, and M, respectively. Figure 4 exhibits a scatter plot of the four clusters, which are subsequently subdivided into four groups based on the 'Tenure' and 'MonthlyCharges' features.

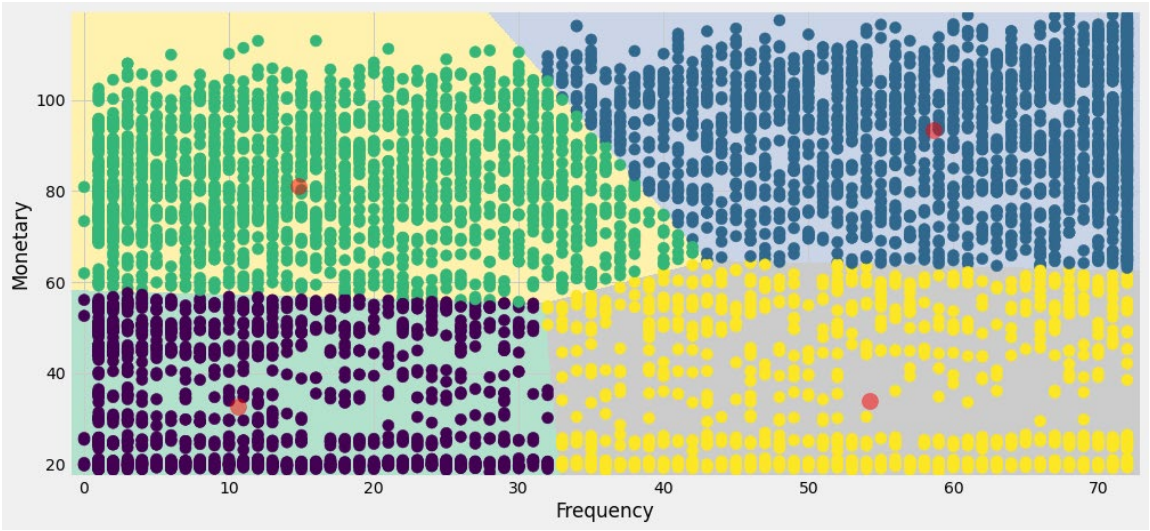


Figure 4. Scattered plot for 4 clusters

Table 2 shows the mean values of the RFM variable, RFM score, name, and percentage for the four clusters in the dataset. Customers with F values above the mean value exhibit a strong commitment to the company and tend to remain members for a longer period. Conversely, customers with F values below the average have a more transient relationship with the business. In terms of M, customers who spend more than the average generate significant profits for the company, while those who spend less typically purchase discounted services.

Table 2. Customer segmentation results

Data	R(mean)	F(mean)	M(mean)	RFM score	Name of Cluster	Amount (%)
Cluster1	1	10.59	32.64	R-F↓M↓	LCSTC	24.8%
Cluster2	1	58.60	93.27	R-F↑M↑	HCLTC	27.8%
Cluster3	1	14.78	81.09	R-F↓M↑	HCSTC	31.1%
Cluster4	1	54.16	33.96	R-F↑M↓	LCLTC	16.4%
Dataset	1	32.37	64.76			

In cluster 1, F and M are lower than the average value and the customers in this cluster are named 'Less Consuming Short-Term Customers (LCSTC).' In cluster 2, F and M are higher than the average value and the customers in this cluster are named 'High Consuming Long-Term Customers (HCLTC).' In cluster 3, F is lower than the average value of the dataset, and M is higher than the average value. The customers in this cluster are named 'High Consuming Short-Term Customers (HCSTC).' In cluster 4, F is higher than the average value, and M is lower than the average value, so the customers in this cluster are named 'Less Consuming Long-Term Customers (LCLTC).'

The HCLTC cluster comprises 27.8% of all customers and represents the telecom provider's high-quality customers who exhibit long-term engagement. To enhance customer loyalty, the company should maintain regular communication and offer a range of promotions. The HCSTC cluster, which constitutes 31.1% of all customers, comprises individuals with a preference for high-quality goods and services. In order to create and promote premium items, the firm should evaluate their preferences. The LCLTC cluster comprises 16.4% of all customers and consists of individuals who have a high frequency of consumption but spend less overall. To encourage good consumption habits and long-term high spending, the company should provide appropriate suggestions for high-quality products and services. The LCSTC cluster, which accounts for 24.8% of all consumers, comprises individuals who are less consuming and short-term in nature. To increase their engagement and reduce churn, the company should analyze and evaluate the characteristics of this cluster.

Feature Construction

According to the results of customer segmentation, it is shown that the dataset of this paper is categorized into four customer clusters with different customer value attributes. Customers in the same cluster have the same value label value and customers in different clusters have different value label values. Therefore, different value attributes among customers can be used to construct a new customer feature for distinguishing customer value differences among customers. The process of constructing new data from the original dataset after feature construction is shown in Figure 5.

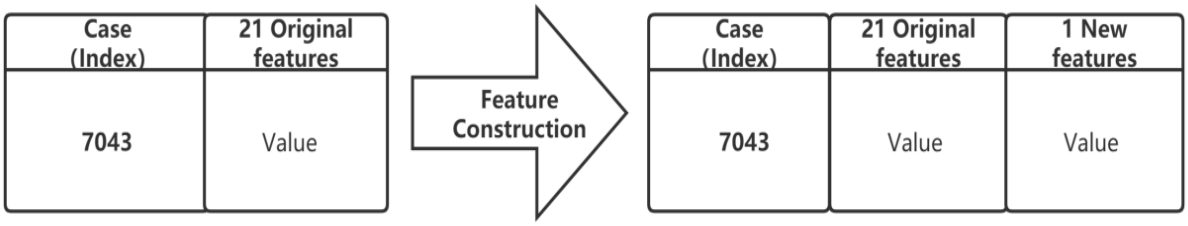


Figure 5. The process of dataset feature construction

Churn Prediction

As shown in Table 3, we applied the XGBoost algorithm to both the original and new datasets and used four kinds of metrics to evaluate the impact of the new features on the model performance. The results show that the new features lead to a significant improvement in the performance of the model. The accuracy is improved by approximately 2%.

Table 3. Comparison of model performance after feature construction

Dataset	Model	Accuracy	Recall	Precision	F1-score
Original dataset	XGBoost	0.793	0.79	0.80	0.79
New dataset	XGBoost	0.8105	0.81	0.81	0.80

Table 4 shows that the accuracy of the proposed customer churn analysis system in this paper is compared with other advanced machine learning methods on the same data set. The results show that our proposed churn prediction method based on customer segmentation obtains the highest accuracy.

Table 4. Comparison with others' work

Work	Model	Accuracy %
(Halibas et al., 2019)	Naive Bayes	73.0
	Generalized linear model	75.7
	Logistic regression	75.7
	Deep learning	74.3
	Decision tree	70.7
	Random forest	75.2
	Gradient boosted tree	79.1
(Rahmaty et al., 2022)	The Grey Wolf Optimizer and Ensemble Neural Networks	80.84
Our method	xgboost	79.3
	Customer segmentation + xgboost	81.05

Churn Factor Identification

Table 5 shows the top 5 feature importance rankings (churn factor importance rankings) based on the target feature 'churn' for the original dataset and the 4 new clusters. Each cluster has a different feature importance ranking, which means that the main churn factors are significantly different for each cluster. When XGBoost combined with the SHAP method predicts customer churn in the original dataset, the feature 'Contract' is the most important feature, but the most important features in LCSTC, HCLTC, HCSTC, and LCLTC are 'Tenure,' 'MonthlyCharges,' 'Tenure,' and 'TotalCharges' respectively.

Table 5. Top 5 important features for the original dataset and 4 clusters

Ranking	Original Dataset	LCSTC	HCLTC	HCSTC	LCLTC
1	Contract	Tenure	MonthlyCharges	Tenure	TotalCharges
2	Tenure	TotalCharges	Contract	TotalCharges	MonthlyCharges
3	OnlineSecurity	MonthlyCharges	Tenure	MonthlyCharges	Contract
4	MonthlyCharges	TechSupport	TotalCharges	InternetService	Tenure
5	TechSupport	Contract	TechSupport	MultipleLines	TechSupport

In order to reduce customer churn, it is commonly believed that telecom companies should offer uniform services and strategies to all customers, based on the ranking of features such as 'Contract,', 'Tenure,' and 'TechSupport' as shown in Table 5. However, our research suggests that the original dataset should be divided into four groups, each with a different ranking of important features. By doing so, telecom companies can provide appropriate and customized services and strategies to each group.

For example, in the LCSTC group, the most important feature is 'Tenure.' Customers who have been with the company for a specific period of time are at a higher risk of churning, and therefore, the company should provide customized strategies to prevent churn for this specific group of customers. By understanding the importance ranking of different features, companies can identify the main factors behind customer churn, and develop tailored strategies to address these weaknesses.

Overall, our research suggests that, instead of offering uniform services to all customers, telecom companies should divide their customer base into distinct groups, and apply customized strategies based on the unique characteristics of each group. This approach is more effective in reducing customer churn and improving overall customer satisfaction.

DISCUSSION

This paper proposes a novel customer churn analysis system for the telecommunications industry, designed to segment customers, identify those at risk of churn, and ascertain the key factors contributing to customer attrition. High-value customers constitute the core assets of telecommunications companies. However, previous research has predominantly focused on isolated areas, such as telecom customer churn prediction (Fujo, 2022), telecom customer segmentation (Vieri et al., 2023), and churn customer factor identification (Ramesh et al., 2022).

In the study by Fujo (2022), two feature selection methods were employed, along with early stopping techniques, random oversampling, and activity regularization strategies, which enhanced the model's comprehensiveness and robustness. However, despite the exemplary results achieved, the study overlooked the computational cost and methodological complexity in actual commercial activities. In contrast, this paper improves model accuracy solely using feature construction methods for customer segmentation, offering a more straightforward approach with lower computational complexity compared to the study by Fujo (2022).

Vieri et al. (2023) conducted an empirical examination of consumer behavior, which facilitated a more precise understanding of actual consumer behaviors. Additionally, the study applied machine learning techniques, specifically unsupervised learning, aiding in the identification of hidden structures or patterns within the data. Ramesh et al. (2022) utilized datasets from the telecom industry and

applied artificial neural networks (ANN) and random forests (RF) to ascertain the factors influencing consumer attrition. The study compared the performance of ANN and RF, ultimately identifying the optimal model structure. The proposed model exhibited high accuracy, demonstrating the method's efficacy. However, despite the notable achievements of the aforementioned studies, their focus on singular domains meant that telecom companies could not directly discern the reasons for different categories of customer churn in practical applications, thereby hindering the implementation of customized strategies for specific customer groups.

The customer churn prediction analysis system proposed in this paper integrates three components: telecom churn prediction, telecom customer segmentation, and telecom churn customer factor identification, thereby providing a more comprehensive analysis of telecom customer churn. In this process, we enhanced the accuracy of customer churn prediction to 81.05% solely through feature engineering methods rather than hybrid models, which helps reduce methodological complexity. Additionally, we segmented customers into four categories and identified the main churn factors for different telecom customer groups. This approach will enable telecom companies to comprehend actual consumer behaviors more accurately, ultimately enhancing the reliability and accuracy of business decisions. Our dataset may partially represent the characteristics and issues of the global telecommunications industry, but it also has limitations, such as the inability to cover all region-specific factors. While our research may provide some reference value for the global telecommunications industry, it requires further studies from other regions or countries to supplement and perfect it.

CONCLUSION

In this paper, we integrate the three research areas to study telecom customer churn and propose a new churn prediction and analysis system that analyzes churn in a more comprehensive way than previous studies. It not only has a high prediction accuracy but also identifies the causes of churn for different categories of customers. The system comprises three key components: customer segmentation, churn prediction, and feature importance ranking. To achieve these objectives, the paper utilizes the K-means algorithm for data extension. The RFM model is used to identify high-value customer groups and labeling improves accuracy by nearly 2%. The XGBoost algorithm is employed for customer churn prediction. In addition, the XGBoost algorithm combined with the SHAP method is used to derive a feature importance ranking. In this way, the reasons for customer churn can be obtained to assist companies in developing targeted churn prevention strategies.

Overall, this system helps telecom companies to implement effective CRM and marketing strategies aimed at reducing churn customers and enhancing business benefits. Due to the privacy of the telecom dataset in slightly fewer data, in the future, we will try to obtain a larger and more comprehensive dataset and try to retrain and apply the proposed system to achieve better system performance.

ACKNOWLEDGMENT

We acknowledge the support we received from Hulunbuir University under the grant reference number 2023BSJJ16.

REFERENCES

Abdullah, F. A. (2021). Improving Customer Relationship Management (CRM) system development for Zain Telecom Company. *Форум молодых ученых, 6*(58), 8-12. https://cyberleninka.ru/article/n/improving-customer-relationship-management-crm-system-development-for-zain-telecom-company

Anwar, M. T., Hadikurniawati, W., Winarno, E., & Supriyanto, A. (2019). Wildfire risk map based on DBSCAN clustering and cluster density evaluation. *Advance Sustainable Science Engineering and Technology, 1*(1), 0190102. https://doi.org/10.26877/asset.v1i1.4876

Barus, O., & Nathasya, C. (2022). The implementation of RFM analysis to customer profiling using K-means clustering. *Mathematical Modelling of Engineering Problems*, *10*(1), 298-303. https://doi.org/10.18280/mmep.100135

Bhattacharyya, J., & Dash, M. K. (2021). What do we know about customer churn behaviour in the telecommunication industry? A bibliometric analysis of research trends, 1985–2019. *FIIB Business Review, 11*(3), 280-302. https://doi.org/10.1177/23197145211062687

Bonacchi, M., & Perego, P. (2012). Improving profitability with customer-centric strategies: The case of a mobile content provider. *Strategic Change*, *20*(7-8), 253-267. https://ssrn.com/abstract=2091221

Cen, S., Yoo, J. H., & Lim, C. G. (2022). Electricity pattern analysis by clustering domestic load profiles using discrete wavelet transform. *Energies, 15*(4), 1350. https://doi.org/10.3390/en15041350

Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C.-H., & Liu, R. (2019). XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. *IEEE Access, 7,* 13149–13158. https://doi.org/10.1109/ACCESS.2019.2893448

Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA,* 785-794. https://doi.org/10.1145/2939672.2939785

Chen, T., Lu, A., & Hu, S.-M. (2012). Visual storylines: Semantic visualization of movie sequence. *Computers & Graphics, 36*(4), 241–249. https://doi.org/10.1016/j.cag.2012.02.010

Claypo, N., & Jaiyen, S. (2015, January). Opinion mining for Thai restaurant reviews using K-Means clustering and MRF feature selection. *Proceedings of the 7th International Conference on Knowledge and Smart Technology, Chonburi, Thailand,* 105-108. https://doi.org/10.1109/KST.2015.7051469

Fei, H., Fan, Z., Wang, C., Zhang, N., Wang, T., Chen, R., & Bai, T. (2022). Cotton classification method at the county scale based on multi-features and random forest feature selection algorithm and classifier. *Remote Sensing, 14*(4), 829. https://doi.org/10.3390/rs14040829

Fonti, V., & Belitser, E. (2017). Feature selection using LASSO. *VU Amsterdam Research Paper in Business Analytics, 30, 1–25*https://scholar.google.com.hk/scholar?cluster=8592243949114933511&hl=zh-CN&as_sdt=0,5

Fu, L., Lin, P., Vasilakos, A. V., & Wang, S. (2020). An overview of recent multi-view clustering. *Neurocomputing, 402,* 148–161. https://doi.org/10.1016/j.neucom.2020.02.104

Fujo, S. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters, 11*(1), 185–198. https://doi.org/10.18576/isl/110120

Guerola-Navarro, V., Gil-Gomez, H., Oltra-Badenes, R., & Sendra-García, J. (2021). Customer relationship management and its impact on innovation: A literature review. *Journal of Business Research, 129*, 83–87. https://doi.org/10.1016/j.jbusres.2021.02.050

Guney, S., Peker, S., & Turhan, C. (2020). A combined approach for customer profiling in video on demand services using clustering and association rule mining. *IEEE Access, 8,* 84326–84335. https://doi.org/10.1109/ACCESS.2020.2992064

Guo, M., Yuan, Z., Janson, B., Peng, Y., Yang, Y., & Wang, W. (2021). Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost. *Sustainability, 13*(2), 926. https://doi.org/10.3390/su13020926

Guyeux, C., Chrétien, S., Bou Tayeh, G., Demerjian, J., & Bahi, J. (2019). Introducing and comparing recent clustering methods for massive data management in the Internet of Things. *Journal of Sensor and Actuator Networks, 8*(4), 56. https://doi.org/10.3390/jsan8040056

Ha, S. H., & Park, S. C. (1998). Application of data mining tools to hotel data mart on the Intranet for database marketing. *Expert Systems with Applications, 15*(1), 1–31. https://doi.org/10.1016/s0957-4174(98)00008-6

Halibas, A. S., Cherian Matthew, A., Pillai, I. G., Reazol, J. H., Delvo, E. G., & Bonachita Reazol, L. (2019, January). Determining the intervening effects of exploratory data analysis and feature engineering in Telecoms customer churn modelling. *Proceedings of the 4th MEC International Conference on Big Data and Smart City, Muscat, Oman*. https://doi.org/10.1109/ICBDSC.2019.8645578

Hallishma, L. (2023). Customer segmentation based on RFM analysis and unsupervised machine learning technique. In I. Woungang, S. K. Dhurandher, K. K. Pattanaik, A. Verma, & P. Verma (Eds.), *Advanced network technologies and intelligent computing* (pp. 46–55). Springer. https://doi.org/10.1007/978-3-031-28183-9_4

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process, 5*(2), 1-11. https://doi.org/10.5121/ijdkp.2015.5201

Hu, X., Shi, Z., Yang, Y., & Chen, L. (2020, April). Classification method of internet catering customer based on improved RFM model and cluster analysis. *Proceedings of the IEEE 5th International Conference on Cloud Computing and Big Data Analytics, Chengdu, China,* 28-31. https://doi.org/10.1109/ICCCBDA49378.2020.9095607

Huang, Y., Zhang, M., & He, Y. (2020, June). Research on improved RFM customer segmentation model based on K-Means algorithm. *Proceedings of the 5th International Conference on Computational Intelligence and Applications, Beijing, China,* 24-27. https://doi.org/10.1109/ICCIA49625.2020.00012

Jovanov, T., & Disoska, V. (2021). Implementation of e-commerce webshop, CRM and marketing planning. *Eprints.ugd.edu.mk*. https://eprints.ugd.edu.mk/27784/

Kaggle. (n.d.). *Telco customer churn*. https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Kim, K. H., & Baek, J. G. (2014). A prediction of chip quality using OPTICS (Ordering Points to Identify the Clustering Structure)-based feature extraction at the cell level. *Journal of Korean Institute of Industrial Engineers, 40*(3), 257-266. https://doi.org/10.7232/JKIIE.2014.40.3.257

Kuznietsova, N., Bidyuk, P., & Kuznietsova, M. (2021). Data mining methods, models and solutions for big data cases in telecommunication industry. In S. Babichev, & V. Lytvynenko (Eds.), *Lecture notes in computational intelligence and decision making* (pp. 107-127). Springer. https://doi.org/10.1007/978-3-030-82014-5_8

Liu, H., & Motoda, H. (Eds.). (1998). *Feature extraction, construction and selection: A data mining perspective*. Springer. https://doi.org/10.1007/978-1-4615-5725-8

Lundberg, S., & Lee, S.-I. (2017, December). A unified approach to interpreting model predictions. *Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA,* 4768-4777. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Machado, M. R., Karray, S., & de Sousa, I. T. (2019, August). LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. *Proceedings of the 14th International Conference on Computer Science & Education, Toronto, ON, Canada,* 1111-1116. https://doi.org/10.1109/ICCSE.2019.8845529

Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: Application of big data. *Journal of Big Data, 4*, Article 20. https://doi.org/10.1186/s40537-017-0077-4

Maryani, I., Riana, D., Astuti, R. D., Ishaq, A., Sutrisno, & Pratama, E. A. (2018, October). Customer segmentation based on RFM model and clustering techniques with K-Means algorithm. *Proceedings of the Third International Conference on Informatics and Computing, Palembang, Indonesia.* https://doi.org/10.1109/IAC.2018.8780570

Mohamed, F. A., & Al-Khalifa, A. K. (2023, January). A review of machine learning methods for predicting churn in the telecom sector. *Proceedings of the International Conference on Cyber Management and Engineering, Bangkok, Thailand,* 164-170. https://doi.org/10.1109/CyMaEn57228.2023.10051108

Mulyawan, B., Viny Christanti, M., & Wenas, R. (2019). Recommendation product based on customer categorization with K-Means clustering method. *Proceedings of the IOP Conference Series: Materials Science and Engineering, 508,* 12123. https://doi.org/10.1088/1757-899X/508/1/012123

Nandapala, E. Y. L., & Jayasena, K. P. N. (2020, November). The practical approach in Customers segmentation by using the K-Means algorithm. *Proceedings of the IEEE 15th International Conference on Industrial and Information Systems, Rupnagar, India,* 344-349. https://doi.org/10.1109/ICIIS51140.2020.9342639

Ogbuabor, G., & Ugwoke, F. N. (2018). Clustering algorithm for a healthcare dataset using silhouette score value. *International Journal of Computer Science and Information Technology, 10*(2), 27–37. https://doi.org/10.5121/ijcsit.2018.10203

Parikh, Y., & Abdelfattah, E. (2020, October). Clustering algorithms and RFM analysis performed on retail transactions. *Proceedings of the 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference, New York, NY, USA,* 506-511. https://doi.org/10.1109/UEMCON51285.2020.9298123

Rachmahwati, D. S., Andreswari, R., & Hamami, F. (2022, November). Customer segmentation of cellular telecommunication company using K-Means algorithm (case study of PT Indosat). *Proceedings of the International Conference of Science and Information Technology in Smart Administration, Denpasar, Bali, Indonesia,* 45-50. https://doi.org/10.1109/ICSINTESA56431.2022.10041532

Rahmaty, M., Daneshvar, A., Salahi, F., Ebrahimi, M., & Pourghader Chobar, A. (2022). Customer churn modeling via the grey wolf optimizer and ensemble neural networks. *Discrete Dynamics in Nature and Society, 2022,* Article 9390768. https://doi.org/10.1155/2022/9390768

Ramesh, P., Jeba Emilyn, J., & Vijayakumar, V. (2022). Hybrid artificial neural networks using customer churn prediction. *Wireless Personal Communications*, *124*(2), 1695-1709. https://doi.org/10.1007/s11277-021-09427-7

Rozemberczki, B., Watson, L., Bayer, P., Yang, H.-T., Kiss, O., Nilsson, S., & Sarkar, R. (2022). The Shapley value in machine learning. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence,* 5572-5579. https://doi.org/10.24963/ijcai.2022/778

Sağlam, M., & El-Montaser, S. (2021). The effect of customer relationship marketing in customer retention and customer acquisition. *International Journal of Commerce and Finance, 7*(1), 191–201. http://ijcf.ticaret.edu.tr/index.php/ijcf/article/view/259/pdf_157

Santharam, A., & Krishnan, S. B. (2018). Survey on customer churn prediction techniques. *International Research Journal of Engineering and Technology, 5*(11), 131-137. https://scholar.google.com.hk/scholar?hl=zh-CN&as_sdt=0%2C5&q=+Survey+on+customer+churn+prediction+techniques.+International+Research+Journal+of+Engineering+and+Technology&btnG=

Sari, R. Y., Oktavianto, H., & Sulistyo, H. W. (2022). Algoritma K-Means dengan metode elbow untuk mengelompokkan kabupaten/kota di jawa tengah berdasarkan komponen pembentuk indeks pembangunan manusia (K-means algorithm with the elbow method to cluster districts/cities in central java based on the components forming the human development index). *Jurnal Smart Teknologi, 3*(2), 104–108. http://jurnal.unmuhjember.ac.id/index.php/JST/article/view/6928

Shams Amiri, S., Mottahedi, S., Lee, E. R., & Hoque, S. (2021). Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption. *Computers, Environment and Urban Systems, 88,* 101647. https://doi.org/10.1016/j.compenvurbsys.2021.101647

Sharaf Addin, E. H., Admodisastro, N., Mohd Ashri, S. N. S., Kamaruddin, A., & Chong, Y. C. (2022). Customer mobile behavioral segmentation and analysis in telecom using machine learning. *Applied Artificial Intelligence, 36*(1), 1-21. https://doi.org/10.1080/08839514.2021.2009223

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

Tauni, S., Khan, R., Khan, M., & Aslam, S. (2014). Impact of customer relationship management on customer retention in the telecom industry of Pakistan. I*ndustrial Engineering Letters, 4*(10), 54-59. https://core.ac.uk/download/pdf/234685274.pdf

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika, 18*(4), 267–276. https://doi.org/10.1007/BF02289263

Vieri, J. K., Munandar, T. A., & Srisulistiowati, D. B. (2023). Exclusive clustering technique for customer segmentation in national telecommunications companies. *International Journal of Information Technology and Computer Science Applications, 1*(1), 51-57. https://doi.org/10.58776/ijitcsa.v1i1.19

Wan, S., Chen, J., Qi, Z., Gan, W., & Tang, L. (2022). Fast RFM model for customer segmentation. *Companion Proceedings of the Web Conference,* 965-972. https://doi.org/10.1145/3487553.3524707

Wojtas, M., & Chen, K. (2020, December). Feature importance ranking for deep learning. *Proceedings of the 34th Conference on Neural Information Processing Systems, Vancouver, Canada.* https://proceedings.neurips.cc/paper/2020/hash/36ac8e558ac7690b6f44e2cb5ef93322-Abstract.html

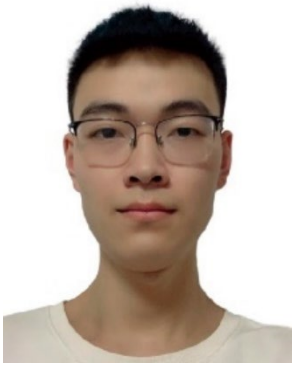
Wu, Y., & Zhou, Y. (2022). Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. *Construction and Building Materials, 330*, 127298. https://doi.org/10.1016/j.conbuildmat.2022.127298

Xian, Z., Keikhosrokiani, P., XinYing, C., & Li, Z. (2022). An RFM model using K-Means clustering to improve customer segmentation and product recommendation. In P. Keikhosrokiani (Ed.), *Handbook of research on consumer behavior change and data analytics in the socio-digital era* (pp. 124-145). IGI Global. https://doi.org/10.4018/978-1-6684-4168-8.ch006

AUTHORS

Tianpei Xu was born in Inner Mongolia, China in 1994. He received his B.S. degree in computer application software from Jilin University in 2016. He received his Ph.D. degree in computer engineering from Chonnam National University, Korea, in 2022. He has been a lecturer at Hulunbuir University since 2022. He has authored a book and more than 3 articles since 2020. His research interests include machine learning, deep learning, data analysis, and feature engineering.

Ying Ma was born in Beijing, China, in 1993. She received her PhD degree in computer engineering from Chonnam National University, Korea, in 2023. Since 2023, she has been working as a lecturer at Nanchang Hongkong University. She has published many peer-reviewed journal and international conference papers. Her research interests include machine vision, machine learning, deep learning, data analysis and artificial intelligence.



Changyu Ao is currently a Ph.D. at Chonnam National University in Korea. He obtained his Master's degree in computer science from the same institution, with a research focus on the application of blockchain technology. He is currently pursuing research in the field of deep learning.

Min Qu is a lecturer in the department of Digital Commerce at Jiangsu Vocational Institute of Commerce, China. She worked as a post-doc at Brain Korea 21 Team, and she used to be assistant professor at Chonnam National University. She received her doctoral degree in electronic commerce from Chonnam National University. Her papers have been published in the international journal *Asia Pacific Journal of Information Systems*, and many Korean journals, such as *APJIS*, *JIECR*, *JIS*, *JITS* and *JIEB*. She participated in 18th and 21st PACIS international conferences. Her current research interests include social networking service, electronic commerce, social media, live-streaming commerce, and new media marketing.

Xianghong Meng, Professor, was born in 1971 in Chifeng City, Inner Mongolia Autonomous Region. He obtained a doctoral degree in Management from Renmin University of China in 2009 and is currently working at the Inner Mongolia University for Nationalities. His main research areas include e-government and information technology. He has authored several monographs and has led numerous national research projects.