

Article

Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping

Tianpei Xu, Ying Ma and Kangchul Kim *

Department of Computer Engineering, Chonnam National University, Yeosu 59626, Korea;
197525@jnu.ac.kr (T.X.); 207939@jnu.ac.kr (Y.M.)

* Correspondence: kkc@jnu.ac.kr

Abstract: In recent years, the telecom market has been very competitive. The cost of retaining existing telecom customers is lower than attracting new customers. It is necessary for a telecom company to understand customer churn through customer relationship management (CRM). Therefore, CRM analyzers are required to predict which customers will churn. This study proposes a customer-churn prediction system that uses an ensemble-learning technique consisting of stacking models and soft voting. Xgboost, Logistic regression, Decision tree, and Naïve Bayes machine-learning algorithms are selected to build a stacking model with two levels, and the three outputs of the second level are used for soft voting. Feature construction of the churn dataset includes equidistant grouping of customer behavior features to expand the space of features and discover latent information from the churn dataset. The original and new churn datasets are analyzed in the stacking ensemble model with four evaluation metrics. The experimental results show that the proposed customer churn predictions have accuracies of 96.12% and 98.09% for the original and new churn datasets, respectively. These results are better than state-of-the-art churn recognition systems.



Citation: Xu, T.; Ma, Y.; Kim, K. Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping. *Appl. Sci.* **2021**, *11*, 4742. <https://doi.org/10.3390/app11114742>

Academic Editor: João Carlos de Oliveira Matias

Received: 20 April 2021

Accepted: 19 May 2021

Published: 21 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: customer churn; CRM; machine learning; ensemble learning; feature grouping

1. Introduction

Owing to fierce competition among telecom companies, customer churn is inevitable. Customer churn is the act of a customer ending a subscription to a service provider and choosing the services of another company.

Companies must reduce customer churn because it weakens the company. A survey showed that the annual churn rate in the telecom industry ranges from 20% to 40%, and the cost of retaining existing customers is 5–10 times lower than the cost of obtaining new customers [1]. The cost of predicting churn customers is 16 times lower than that for obtaining new customers [2]. Decreasing the churn rate by 5% increases the profit from 25% to 85% [3]. This shows that customer-churn prediction is important for the telecom sector. Telecom companies consider customer relationship management (CRM) an important factor in retaining existing customers and preventing customer churn.

To retain existing customers, CRM analyzers must predict which customers will churn and analyze the reasons for customer churn. Once the at-risk customers are identified, the company must perform marketing campaigns for churn customers to maximize the churn-customer retention. Therefore, customer-churn prediction is an important part of CRM [4].

The accuracy of the prediction systems used by CRM analyzers is important. If analyzers are inaccurate in predicting customer churn, no campaigns can be performed. Owing to recent advancements in data science, data mining and machine learning technologies provide solutions to customer churn. However, there are several limitations in existing models. For example, logistic regression, a common churn-prediction model based on older data-mining methods, is relatively inaccurate. Furthermore, feature construction [5]

is neglected during modeling development. Therefore, a good churn prediction system is necessary.

This study proposes a new customer-churn prediction system and feature construction to improve accuracy, and the contributions of this study can be summarized as follows:

- (1) A new prediction system based on ensemble learning with relatively high accuracy is proposed.
- (2) New features derived from equidistant grouping of customer behavior features are used to improve the system performance.

The rest of this paper is arranged as follows: Section 2 presents literature review, Section 3 proposes the ensemble model and equidistance feature grouping, and Section 4 describes the prediction system and experimental results. Finally, Section 5 concludes the paper.

2. Literature Review

Many methods such as machine learning and data mining are used for churn prediction. The decision-tree algorithm is a reliable method for churn prediction [6]. In addition, a neural network method [7], data certainty [8], and particle swarm optimization [9] are used for churn prediction.

Moreover, an artificial neural network (ANN) and decision trees are compared for customer-churn predictions [10], and the literature review shows that the decision-tree algorithm is better than ANNs for customer churn prediction.

A.T. Jahromi [11] studied the effect of customer loyalty on customer churn in prepaid mobile phone companies. In this study, features were segmented, and multiple algorithms such as decision trees and neural networks were used to predict the processed data. Results showed that a hybrid approach is better than a single algorithm. The KNN-LR is a hybrid approach using logistic regression and the K-nearest neighbor (KNN) algorithm [12]. Researchers surveyed KNN-LR, logistic regression and the radial basis function (RBF) network and found that KNN-LR showed the best performance. Y. Zhang [13] proposed a distributed framework for data mining techniques to predict customer churn. The framework improves the CRM quality of service.

Ruba Obeidat proposed a hybrid genetic programming approach [14]. This study used the K-means algorithm and genetic programming to predict customer churn. Sahar F. Sabbeh used AdaBoost based on the boosting algorithm [15], which summarized existing machine-learning techniques and proposed that AdaBoost showed the best results. Hossam Faris proposed a hybrid swarm intelligent neural network model [16], which proposed an intelligent hybrid model based on particle swarm optimization and feedforward neural network for churn prediction. The results show that the model can improve the coverage of churn customers.

3. Materials and Methods

3.1. Dataset Preparation

The customer churn dataset is an open-source dataset [17] that contains 21 features and 3333 observations. The feature 'Churn' shows customer churn or non-churn based on existing conditions. Approximately 14.5% of the 'Churn' is 'T' label, and 84.5% of 'non-churn' is 'F' label. Table 1 shows the data features. In this experiment, 80% (2666 instances) and 20% (667 instances) of the dataset are used for training and test datasets, respectively.

Table 1. Data-set description.

Feature Name	Description	Object
State	Customer State	Object
Account length	Account used days	Int64
Area code	Phone area code	Int64
phone number	Customer phone number	Object
International plan	Whether the customer starts international business	Object
Voice mail plan	Whether the customer starts the voice mail service	Object
Number vmail messages	Number of customer vmail messages	Int64
Total day minutes	Total minutes of talk during the day	Float64
Total day calls	Number of calls in the day	Int64
Total day charge	Call charges during the day	Float64
Total eve minutes	Total minutes of talk last night	Float64
Total eve calls	Number of calls last night	Int64
Total eve charge	Charges for calls last night	Float64
Total night minutes	Night total call minutes	Float64
Total night calls	Total number of calls in the evening	Int64
Total night charge	Total charge for calls at night	Float64
Total intl minutes	Total minutes of international business calls	Float64
Total intl calls	Total number of international business calls	Int64
Total intl charge	Total charges for international business calls	Float64
Customer service calls	The number of calls for customer service	Int64
Churn	Is the customer churn	Bool

3.2. Proposed Method

This study proposes a new customer-churn prediction system consisting of feature construction, stacking model and soft voting as shown in Figure 1.

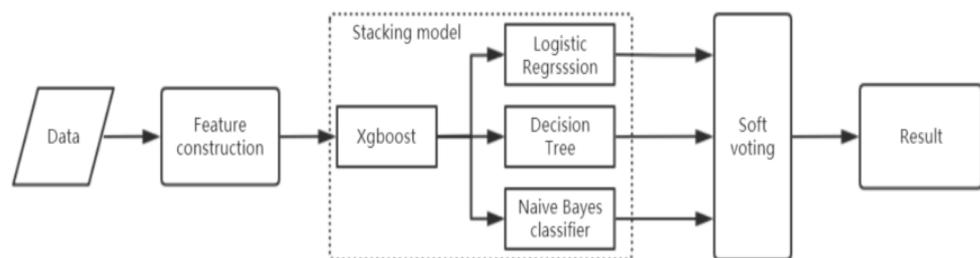


Figure 1. The proposed customer-churn prediction system.

The original features with customers’ consumption and behavior are equidistantly grouped to construct new features. The stacking model consists of two levels with four algorithms: Xgboost (XGB), Logistic regression (LR), Decision tree (DT) and Naive Bayes classifier (NBC) to achieve better prediction accuracy. The third step consists of a soft voting. The results of the stacking model are input to the soft voting.

3.2.1. New Feature Construction with Equidistant Grouping

Feature engineering is important for data processing. Good feature selection and construction are essential for achieving high performance in machine-learning jobs [18]. Feature construction is the process of inferring or constructing additional features from original features, and it discovers missing information about the relationships between features. Feature construction transforms the original representation space to a new one to help better achieve data-mining objectives: improved accuracy, easy comprehensibility, truthful clusters, revealing hidden patterns, etc. [5].

Feature grouping correlates relevant features. Features from the same group are more related, compared to features from a different group. Therefore, it is possible to generate groups of correlated features that are resistant to sample-size variations [19].

Some features in the churn dataset have small–large integer numbers, for example, from 0 to 365. The customers with similar values have similar churn trends. The churn accuracy can be improved when divided by equidistance groups.

When customers have similar consumption–expense behaviors, they may have similar churn. In this study, an equidistant grouping method for features of consumption–expense is used to construct new features. The original observations in a feature are equidistantly grouped to new groups according to the range of the feature value. Customers with similar consumption–expense patterns can be characterized into a group, resulting in identical values of the new feature. The process for feature grouping is shown in Figure 2. The instances of the original feature are grouped equidistantly based on the number of corresponding groups. Instances in the same group get the same value in the new feature.

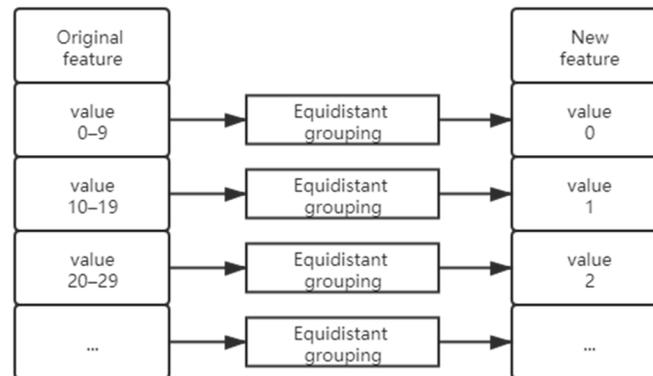


Figure 2. Example of constructing a new feature by equidistance.

Sturges’ formula formulates a method of choosing the optimum number of bins in a histogram [20] for a normally distributed dataset.

As shown in Figure 3, the histograms of some features in the dataset have a shape of normal distribution.

This study uses Sturges’ formula to determine the optimal number of groups, which is given by

$$K = 1 + \log(n)/\log(2) = 1 + 3.322 \log(n) \quad (1)$$

where K represents the optimal number of groups and n is the largest feature value.

3.2.2. Stacking Model

(1) Classifiers

XGB is a decision-tree-based ensemble machine-learning algorithm that uses a gradient boosting framework. It accurately predicts a target class by combining simple and weak models [21].

LR is a machine-learning method that solves binary classification problems and predicts classification possibility. LR has advantages such as simple implementation and strong explanatory power, and it is widely used in the industry [22].

Currently, DT is a mainstream prediction and classification technology. It is similar to human thinking. It adopts a recursive top-to-bottom method, carrying out internal node attribute comparisons and judging to split down from the node depending on attribute values, and concludes on the DT leaf [23].

In probability and statistics, the Bayesian rule is based on prior knowledge of events probability. In machine learning, NBC is a probability statistical classifier based on Bayes theorem. The classifier uses conditional independence assumptions and chooses a very likely category as the sample final category. The algorithm is simple, easy to implement, and less sensitive to missing data, showing small error and stable performance [24].

(2) Stacking

The stacking model is a general method of using a higher-accuracy algorithm to combine lower-accuracy algorithms to achieve greater predictive accuracy. The best results are obtained when the higher-accuracy algorithm is in a first level and lower-accuracy

algorithms are in a second level [25]. In this study, the stacking model consists of two layers as shown in Figure 4, level 1 and level 2. A higher accuracy model is used in the first layer (level 1), while lower-accuracy models are used in the second layer (level 2).

XGB with a kind of ensemble function got the best accuracy, which is chosen as a first-level classifier. And LR, DT and NBC have primary and different characteristics, and they can complement XGB for the mismatched samples, which are chosen in a second level. The results of XGB and data are combined to generate data 1 for the second level classifiers.

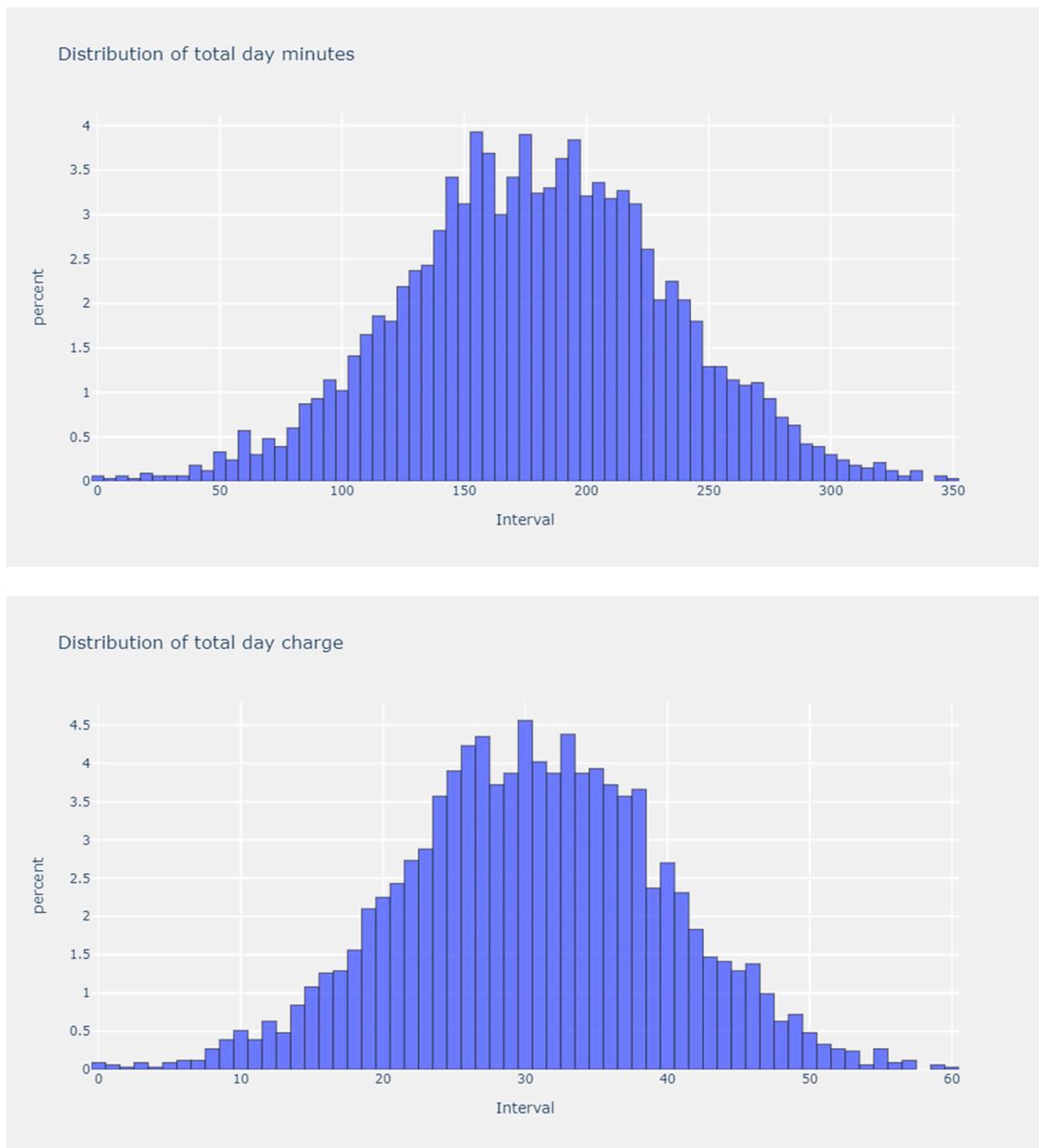


Figure 3. Cont.

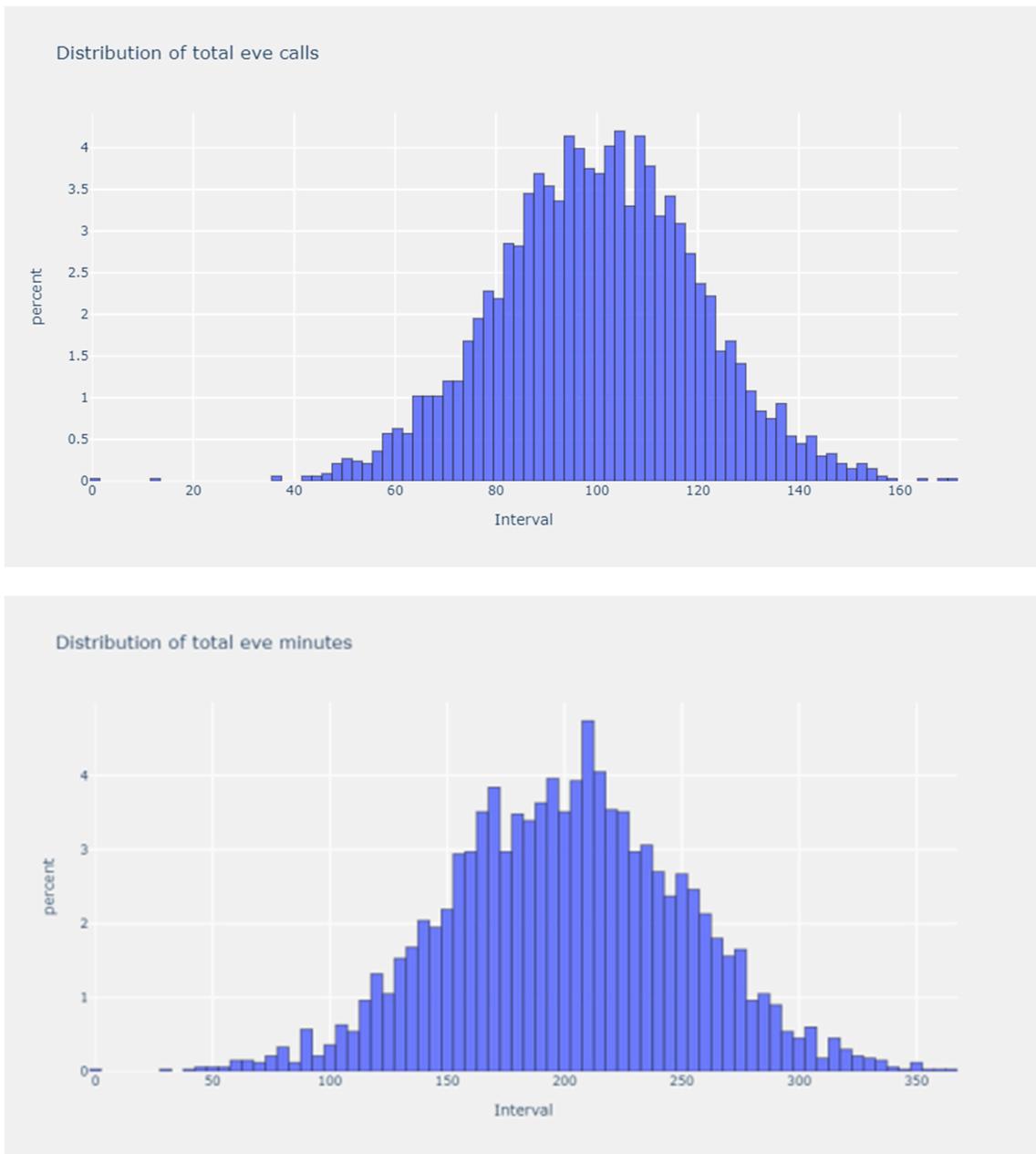


Figure 3. Histograms for data distributions of 4 features.

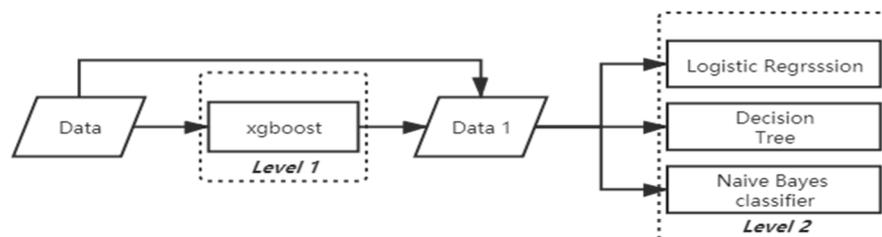


Figure 4. The proposed stacking model.

3.2.3. Soft Voting

Soft voting estimates the class probability with different algorithms having contrasting approaches to improve the accuracy of prediction. It assigns a larger weight to the important classifier, and the highest category is selected by summing the probabilities

predicted by the model. The soft-voting classifier prediction is mathematically represented as follows:

$$\hat{y} = \operatorname{argmax}_i \sum_{j=1}^m w_j p_{ij} \quad (2)$$

where argmax function is the function that outputs a maximum, and w_j represents the weight associated with the classifier prediction; p_{ij} is the probability of the classifier in predicting a certain class. In the process of assigning weights for soft voting, high-confidence models are given more weight based on the importance and accuracy of the classifier [26].

The prediction ensemble system in this study uses soft voting to obtain the final prediction results. Soft voting usually requires different classifiers to compensate for their drawbacks. Based on the accuracy and algorithmic differences, this study chose LR, DT, NBC algorithms as level 2 classifiers, and the results of level 2 classifiers with different weights are used for soft voting.

3.3. Evaluation Measures

In this study, the proposed ensemble system for predicting customer churn is evaluated using accuracy, precision, recall and F1 score.

Equation (3) calculates the accuracy metric. It is defined as the ratio of the number of samples correctly classified by the classifier to the total number of samples for a given test dataset.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

In Equation (3) 'TN' is True Negative, 'TP' is True Positive, 'FN' is False Negative and 'FP' is False Positive. Equation (4) is the formula for Precision. It identifies that the part of the prediction data is positive.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (4)$$

The recall is another measure for completeness, i.e., the true hit of the algorithm. It is calculated by using Equation (5).

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (5)$$

The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contributions of precision and recall to the F1 score are equal. It is calculated by using Equation (6).

$$\text{F1score} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (6)$$

4. Results

4.1. Feature Construction

As shown in the Table 1, 16 features with numerical value in the dataset can be applied to equidistant grouping. The features 'account length' and 'area code' are not the customer's daily behavior and consumption features. The features 'total intl charge' and 'customer service calls' cannot be effectively grouped equidistantly because the value range is too small. However, the value range of the selected 12 features is relatively large and can be used to mine the hidden information between the data. Equation (1) is used to determine the number (K) of groups for 12 original features. The results are shown in Table 2.

For example, the feature 'Total day calls' has numerical value, ranging from 0 to 165. The feature construction results using Equation (1) are shown in Table 3. The feature 'Total

day calls' is divided into 8 groups equidistantly through Equation (1). There are instances in the same group, where the same value is assigned in the new feature.

Table 2. Twelve features and values for grouping.

Feature	Value	K
Number vmail messages	0–51	6
Total day minutes	0–350.8	10
Total day calls	0–165	8
Total day charge	0–59.64	6
Total night minutes	0–363.7	10
Total night calls	0–170	8
Total night charge	0–30.91	7
Total eve minutes	0–395	10
Total eve calls	0–175	8
Total eve charge	0–17.77	5
Total intl minutes	0–20	5
Total intl calls	0–20	5

Table 3. The grouping results of the feature 'Total day calls'.

Instance (Index)	Original Feature	New Feature
1	110	6
2	123	7
3	114	6
4	71	4
5	113	6
6	98	5
7	88	5
8	79	4
...		
3333	113	6

The histogram of the new features also approximates the normal distribution of the new features as shown in Figure 5.

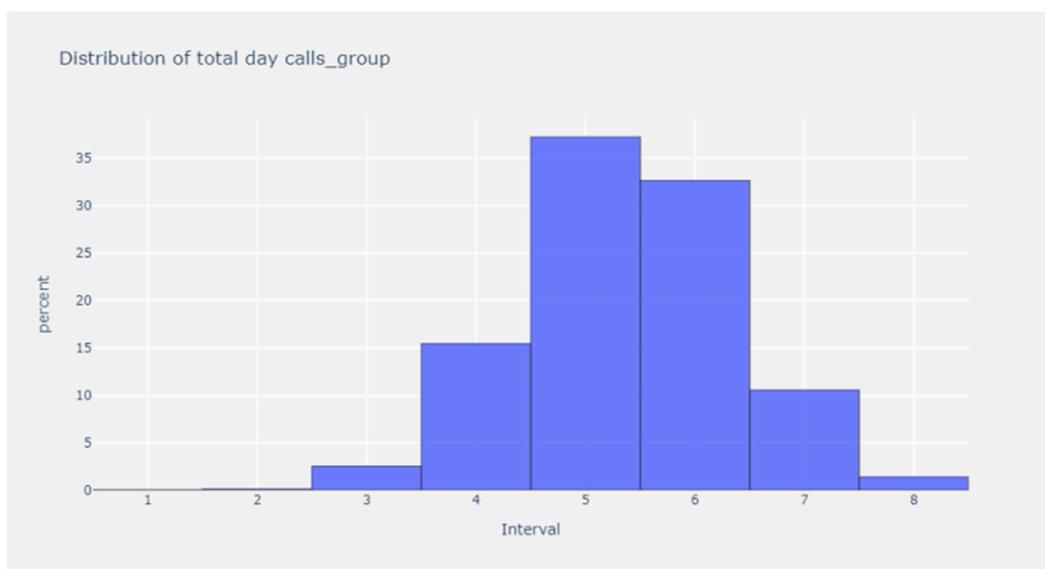


Figure 5. Histogram for the new feature 'total day calls group'.

Feature construction is to improve accuracy by expanding feature space. If the original dataset and the new dataset are not combined, it cannot increase feature space and improve accuracy. As shown in Figure 6, 12 new features are derived from the original features through feature construction. The new dataset is generated based on 21 original features combined with 12 new features.

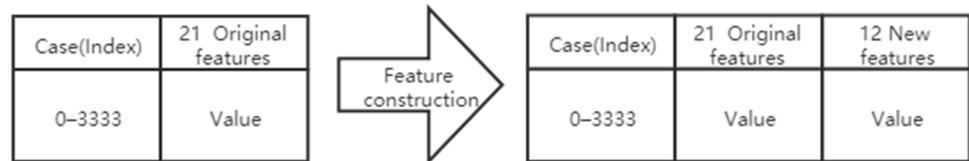


Figure 6. The new dataset after feature construction.

We use four machine learning algorithms with two datasets, namely, the original dataset and the new dataset to compare its performance. The results in Table 4 show that the new dataset shows better accuracy for all classifiers and that the proposed feature construction method improves the stacking model performance. The new feature grouping discovers missing latent factors from the churn dataset, regarding the relationship between features, and expands the feature space by creating additional features.

Table 4. Model performance comparison on different datasets.

Model	Accuracy		Precision		Recall		F1-Score		Time(S)	
	ODS	NDS	ODS	NDS	ODS	NDS	ODS	NDS	ODS	NDS
LR	0.8482	0.8586	0.7978	0.8412	0.8576	0.8745	0.8054	0.8412	0.0327	0.0606
DT	0.8513	0.8605	0.8176	0.8487	0.8029	0.8078	0.8173	0.8287	0.0423	0.0463
NBC	0.8514	0.8575	0.8445	0.8773	0.8573	0.8631	0.8545	0.8664	0.0048	0.0053
XGB	0.9543	0.9554	0.9467	0.9547	0.9528	0.9548	0.9523	0.9554	0.5853	0.7207

(ODS: Original dataset, NDS: New dataset).

4.2. Stacking Model

The LR, DT, and NBC algorithms have relatively low accuracy compared to the XGB algorithm. However, by using high-accuracy XGB to stack the three relatively low-accuracy algorithms of LR, DT, and NBC, the model performance of these three machine-learning algorithms is improved. Table 5 shows that the accuracy of the stacked model for these classifiers with the new dataset increases by approximately 10%.

Table 5. Accuracy after stacking for three classifiers.

Model	Accuracy	
	Before Stacking	After Stacking
LR	0.8586	0.9585
DT	0.8605	0.9560
NBC	0.8575	0.9535

Table 6 shows that the stacking model accuracy of the new dataset increases by around 1% compared to the original dataset.

Table 6. Accuracy of stacking model with ODS and NDS.

Model	Accuracy	
	ODS	NDS
LR	0.9463	0.9585
DT	0.9493	0.9560
NBC	0.9471	0.9535

(ODS: Original dataset, NDS: New dataset).

4.3. Soft Voting and Final Results

The stacked model outputs three different types of model results. These three results are input into soft voting. For better results, the models with high confidence and accuracy are given more weight in the soft-voting process [26]. Among the three models, LR has the best accuracy as show in Table 6 and is used as a main classifier and given more weight value assigned to 0.4. The accuracy of NBC and DT are slightly lower than that of LR, and their weight values are assigned to 0.3. Each weight for NBC and DT is lower than the main classifier LR, but the sum of their weights is bigger than the weight of LR. This can make up for the shortcomings of LR and improve the accuracy of soft voting. Soft voting outputs the label of highest probability. The results in Table 7 show that the accuracy of the proposed stacking ensemble system is 98.09% for the new dataset and 96.12% for the original dataset.

Table 7. Soft voting results for different datasets.

Model	Accuracy	
	ODS	NDS
Proposed model	0.9612	0.9809

4.4. Comparison with Other Works

Table 8 compares the proposed ensemble system with other works. The proposed customer-churn prediction ensemble system shows the best accuracy.

Table 8. Comparison with other works.

Work	Model	Accuracy
Ruba Obiedat [14]	Hybrid genetic programming approach	0.9140
Sahar F. Sabbeh [15]	AdaBoost	0.9639
Hossam Faris [16]	Hybrid swarm intelligent neural network model	0.9630
This study	Ensemble system	0.9809

5. Conclusions

Various machine-learning techniques have been used for customer churn in CRM. This study proposes a customer-churn prediction system based on a stacking ensemble of machine learning, which consists of XGB in level 1; LR, DT, and NBC in level 2; and soft voting. Feature construction is used to expand the feature space and discover latent information from implicit features. The proposed feature construction through feature grouping improves the prediction accuracy compared to the original customer-churn dataset. The result of proposed system showed the best accuracies of 96.12% and 98.09% for the old and new datasets respectively compared to other prediction systems. This proposed system can determine important factors affecting customer purchasing behavior in the telecommunications industry.

Author Contributions: Conceptualization, T.X.; Data curation, T.X.; Formal analysis, T.X.; Funding acquisition, K.K.; Methodology, T.X.; Project administration, K.K.; Resources, K.K.; Software, T.X.;

Supervision, K.K.; Validation, Y.M.; Visualization, T.X. and Y.M.; Writing—original draft, T.X.; Writing—review & editing, Y.M. and K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository. The data presented in this study are openly available in (<https://www.kaggle.com/becksdff/churn-in-telecoms-dataset>, accessed on 21 May 2021) website download. This is a dataset in the book “Discovering Knowledge in Data” by Daniel T. Larose [17].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. García, D.L.; Nebot, À.; Vellido, A. Intelligent data analysis approaches to churn as a business problem: A survey. *Knowl. Inf. Syst.* **2017**, *51*, 719–774. [CrossRef]
2. Borja, B.; Bernardino, C.; Alex, C.; Ricard, G.; David, M.-M. The Architecture of a Churn Prediction System Based on Stream Mining. *Front. Artif. Intell. Appl.* **2013**, *256*, 157–166. [CrossRef]
3. Kotler, P.T. *Marketing Management: Analysis, Planning, Implementation and Control*; Prentice-Hall: London, UK, 1994.
4. Ngai, E.; Xiu, L.; Chau, D. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Syst. Appl.* **2009**, *36*, 2592–2602. [CrossRef]
5. Motoda, H.; Liu, H. *Feature Selection, Extraction and Construction*; Communication of IICM (Institute of Information and Computing Machinery): Taipei, Taiwan, 2001; Volume 5, pp. 67–72.
6. Edwards, R.A.H.; Sári, M.; Huld, T.A.; Dallemand, J.F. GIS-Based Assessment of Cereal Straw Energy Resource in the European Union. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?> (accessed on 10 February 2020).
7. Sharma, A.; Panigrahi, D.P.K. A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. *Int. J. Comput. Appl.* **2013**, *27*, 26–31. [CrossRef]
8. Amin, A.; Al-Obeidat, F.; Shah, B.; Adnan, A.; Loo, J.; Anwar, S. Customer churn prediction in telecommunication industry using data certainty. *J. Bus. Res.* **2019**, *94*, 290–301. [CrossRef]
9. Vijaya, J.; Sivasankar, E. An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. *Clust. Comput.* **2019**, *22*, 10757–10768. [CrossRef]
10. Umayaparvathi, V.; Iyakutti, K. Applications of Data Mining Techniques in Telecom Churn Prediction. *Int. J. Comput. Appl.* **2012**, *42*, 5–9. [CrossRef]
11. Jahromi, A.T.; Moeini, M.; Akbari, I.; Akbarzadeh, A. A Dual-Step Multi-Algorithm Approach for Churn Prediction in Pre-Paid Telecommunications Service Providers. *J. Innov. Sustain. RISUS* **2010**, *1*, 2179–3565. [CrossRef]
12. Zhang, Y.; Qi, J.; Shu, H.; Cao, J. A hybrid KNN-LR classifier and its application in customer churn prediction. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montréal, QC, Canada, 7–10 October 2007; pp. 3265–3269. [CrossRef]
13. Reichheld, F.F.; Sasser, W.E. Zero defections: Quality comes to services. *Harv. Bus. Rev.* **1990**, *68*, 105–111. [PubMed]
14. Obiedat, R.; Al-kasassbeh, M.; Faris, H.; Harfoushi, O. Customer churn prediction using a hybrid genetic programming approach. *Sci. Res. Essays* **2013**, *8*, 1289–1295. [CrossRef]
15. Sabbeh, S.F. Machine-Learning Techniques for Customer Retention: A Comparative Study. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*. [CrossRef]
16. Faris, H. A Hybrid Swarm Intelligent Neural Network Model for Customer Churn Prediction and Identifying the Influencing Factors. *Information* **2018**, *9*, 288. [CrossRef]
17. Larose, D.T.; Larose, C.D. *Discovering Knowledge in Data: An Introduction to Data Mining*; John Wiley & Sons: New York, NY, USA, 2014.
18. Domingos, P. A few useful things to know about machine learning. *Commun. ACM* **2012**, *55*, 78–87. [CrossRef]
19. García-Torres, M.; Gómez-Vela, F.; Becerra-Alonso, D.; Melián-Batista, B.; Moreno-Vega, J.M. Feature grouping and selection on high-dimensional microarray data. In Proceedings of the 2015 International Workshop on Data Mining with Industrial Applications (DMIA), San Lorenzo, Paraguay, 14–16 September 2015.
20. Sturges, H.A. The choice of a class interval. *J. Am. Stat. Assoc.* **1926**, *21*, 65–66. [CrossRef]
21. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016. [CrossRef]
22. Chatterjee, S.; Simonoff, J. *Handbook of Regression Analysis with Applications in R. Logistic Regression*; Wiley: Hoboken, NJ, USA, 2020; pp. 143–171. [CrossRef]
23. Suzuki, J. Decision Trees. In *Statistical Learning with Math and R*; Springer: Singapore, 2020; pp. 147–170. [CrossRef]

-
24. Larose, C.D.; Larose, D.T. NAÏVE BAYES CLASSIFICATION. In *Data Science Using Python and R*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2019. [[CrossRef](#)]
 25. Ting, K.M.; Witten, I.H. Issues in Stacked Generalization. *J. Artif. Intell. Res.* **1999**, *10*, 271–289. [[CrossRef](#)]
 26. Zhou, Z. *Ensemble Methods: Foundation Sand Algorithms*; CRC Press: Boca Raton, FL, USA, 2012; ISBN 978-1-439-830031.