




Article

A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation

Tianyuan Zhang ¹, Sérgio Moro ^{1,*} and Ricardo F. Ramos ^{1,2,3}

¹ Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura (ISTA), Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisbon, Portugal; zhangty0612@gmail.com (T.Z.); ricardo_ramos@iscte.pt (R.F.R.)

² Instituto Politécnico de Coimbra, ESTGOH, Rua General Santos Costa, 3400-124 Oliveira do Hospital, Portugal

³ CICEE—Centro de Investigação em Ciências Económicas e Empresariais, Universidade Autónoma de Lisboa, Rua de Santa Marta, Palácio dos Condes do Redondo, 56, 1169-023 Lisboa, Portugal

* Correspondence: sergio.moro@iscte-iul.pt

Abstract: Numerous valuable clients can be lost to competitors in the telecommunication industry, leading to profit loss. Thus, understanding the reasons for client churn is vital for telecommunication companies. This study aimed to develop a churn prediction model to predict telecom client churn through customer segmentation. Data were collected from three major Chinese telecom companies, and Fisher discriminant equations and logistic regression analysis were used to build a telecom customer churn prediction model. According to the results, it can be concluded that the telecom customer churn model constructed by regression analysis had higher prediction accuracy (93.94%) and better results. This study will help telecom companies efficiently predict the possibility of and take targeted measures to avoid customer churn, thereby increasing their profits.

Keywords: telecommunications; customer segmentation; data mining; targeted marketing



Citation: Zhang, T.; Moro, S.; Ramos, R.F. A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation. *Future Internet* **2022**, *14*, 94. <https://doi.org/10.3390/fi14030094>

Academic Editor: Paolo Bellavista

Received: 14 February 2022

Accepted: 14 March 2022

Published: 16 March 2022

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



Copyright: © 2022 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/>).

1. Introduction

Client churn is a significant problem for telecommunication companies as it results in decreased profit [1]. Moreover, this is particularly relevant since telecommunication companies operate in a saturated global market, meaning it is increasingly challenging to retain customers. Although such companies make considerable marketing investments to acquire new users, retaining a customer is usually less expensive than acquiring a new one [2]. For these reasons, avoiding customer churn has become a significant concern for telecommunications companies.

Customer churn refers to the loss of a customer in favor of a competitor [3], reflecting the end of the relationship. Customer churn prediction allows one to identify the reasons for the end of the relationship and assemble a strategy that will minimize the churn rate, increasing profits. Thus, anticipating a customer's intention to end a relationship is instrumental for telecommunication companies and is considered a competitive advantage.

Previous studies have attempted to understand customer churn. For instance, Bach et al. [1] suggested a clustering and classification framework for churn management. Fathian et al. [4] proposed a new combined model based on ensemble and clustering classifiers. Holtrop et al. [5] aimed to anticipate customer churn using the principles of data anonymization. Although multiple studies have aimed to explain and predict customer churn, no study has tried to predict telecom client churn through discriminant analysis and logistic regression.

Facing this identified gap in the literature, this study aims to use factor analysis to investigate the business characteristics of telecom clients and to build a discriminant model and a logistic regression model to predict telecom client churn using customer

segmentation data from three major Chinese telecommunication companies. Data are collected from China Mobile, China Unicom, and China Telecom and analyzed using a data mining approach to understand the factors that influence and allow one to predict telecom customer churn. Our study extends the previous work by Zhang [6] by innovatively showing how logistic regression analysis can be applied to build a telecom customer churn prediction model. Thus, we propose the following research questions: (1) Which factors will lead to customer loss? (2) How can one predict customer loss using the approach of data mining? (3) How can one develop a model to predict customer churn? It is expected that the results of this study will help telecommunications managers to identify the customer churn profile and create strategies to retain customers.

2. Literature Review

Technological progress is crucial in determining who will be the market leader and to achieve better market performance [7]. Meanwhile, technological progress has already changed the competition and the game's rules in the telecom industry. In the past, telecom operators generally won customers through price competition. However, today's consumers pay more attention to differentiated and value-added services, which has increased switching costs while making consumers more loyal [8]. In the telecom section, technological progress could help companies identify customers with a high risk of churn and to establish a business strategy with customer retention as the core goal, which will make the companies healthier and allow for long-term operation [9]. Finally, the development of telecommunication technologies has also brought about more market competition and higher customer churn rates. The customer churn rate for telecom customers in the European market has reached 30%, while in Asia it has reached 60% [10].

Through a Bayesian belief network analysis, it was concluded that the average tariff amount will affect customer churn. The two other factors are the average call time and tariff type [11]. The tariff structure will affect customers' perceptions of value, affecting customer churn [12]. Through a multilayer perceptron (MLP) analysis of a sample of five thousand Jordanian telecom customers, it was concluded that the monthly tariff is the most significant factor affecting customer churn [13]. Tariffs for domestic calls are essential in predicting customer loss [14]. There are two types of pricing in the telecom section: two-part tariff and pay-per-use pricing. Compared with two-part tariffs, pay-per-use pricing can reduce the customer churn rate by 10.5% [12]. Through a discriminant analysis and t-test of one thousand Indian telecom customers, it was concluded that the tariff rates for calls and customer satisfaction with the telecom service offered are the two key factors determining customer churn [13].

As competition in the telecommunications market intensifies, providing tariff price promotions and differentiated services for the key customers will be an efficient method to avoid customer churn [15]. In the Korean market, the tariff rate is one of the critical factors determining customer churn. Tariffs and customer care services are the two main factors influencing customer satisfaction and churn, as shown using discriminant and regression analyses [16]. Service quality in the telecom industry refers to Internet signal quality. Good service quality will improve customer satisfaction and loyalty, lowering the risk of customer loss [17]. Additionally, it will also help to attract new customers. Through a factor analysis and regression analysis, it was concluded that tariffs and service quality are key factors in prepaid customer churn. Hence, companies need to monitor and improve their service quality [13].

Customer retention and loss are influenced by the customers' sociodemographic characteristics and satisfaction [13]. The customers' sociodemographic data, for example regarding gender, could be used to predict whether customers will be lost or not [18]. Age and gender will influence telecom customers' preferences and behavior. People aged less than thirty years value customer service quality, value-added services, and mobile service fees. The tariff is not a key factor in determining churn for this segment. However, those

aged older than thirty years pay more attention to tariff pricing, which will largely influence their retention or loss [19].

Predicting customer churn is not an easy task, since customer behaviors are heterogeneous [20]. In the past, companies have tended to investigate customer churn using traditional methods such as surveys. However, the data mining approach has been proven to be an efficient and better solution [21]. Specifically, a customer churn prediction model could be established to understand the factors that lead to customer churn and to predict customer loss. The model could be optimized through data mining to improve its prediction accuracy [18]. Moreover, customer segmentation is often combined with customer churn prediction for greater management effectiveness [22].

By comparing the accuracy of telecom customer churn prediction models constructed using different data mining methods, we can measure which data mining method is best [23]. In addition to accuracy, there are other metrics for measuring the performance of customer churn prediction models, such as the understandability and intuitiveness of the model [24]. Idris et al. [25] established a telecom customer churn prediction model with good understandability and intuitiveness using the GP-AdaBoost method.

There are two well-known data mining methods with outstanding prediction accuracy and understandability. One of them is the decision trees (DT) method and the other one is logistic regression. However, both methods have shortcomings: it is difficult for DT to deal with the linear relations of variables, and it is hard for logistic regression to handle the interaction impacts of variables. Thus, the logit leaf model (LLM) method performs better in classifying data. Compared with DT or logistic regression, LLM has shown better performance and understandability [26].

Vo et al. [27] stated that the current churn prediction methods mainly use structured rather than unstructured data to conduct analyses. Moreover, unstructured data and telephone communication voice content are innovatively used to build customer churn prediction models.

Machine learning (ML) and deep learning (DL) are suitable for customer loss prediction. An optimized synthetic minority oversampling method named the ISMOTE-OWELM model was used to improve accuracy in customer churn prediction [28].

3. Hypotheses and Proposed Model

Customer consumption tags distinguish and characterize customers by expense-related information, such as by monthly fee, package type, or mobile terminal price [29]. Precision marketing can be performed using telecom data to classify and identify customers. Using such information will allow telecom operators to concentrate on the target customers and convert them into potential customers. This could significantly optimize marketing expenses and avoid customer churn [29].

Expense-related data could be applied to understand the reasons for customer loss. Customers with similar consumption–expense behaviors have similar reasons for churn. Users with similar expense-related characteristics could be segmented into groups to conduct an analysis [30]. Thus, we propose the following hypotheses:

Hypothesis 1 (H1). *The total fee receivable for the month positively impacts customer loss;*

Hypothesis 2 (H2). *The fixed monthly cost has a positive impact on customer loss;*

Hypothesis 3 (H3). *The local fee has a positive impact on customer loss;*

Hypothesis 4 (H4). *The roaming fee has a positive impact on customer loss;*

Hypothesis 5 (H5). *China Unicom's network fee has a positive impact on customer loss;*

Hypothesis 6 (H6). *The fee with China Mobile positively impacts customer loss;*

Hypothesis 7 (H7). *The fixed-line fee positively impacts customer loss.*

Taiwan's telecommunication industry has experienced fierce competition since it removed the restriction of wireless telecom services, and customer churn management has become the operators' focus in order to retain telecom customers by satisfying their needs. One of the main challenges is to predict customer churn [31].

Using empirical analysis, different data mining methods that can be used to allocate 'propensity-to-churn' scores were evaluated from customer and operator perspectives. The results showed that call data along with neural network and DT methods could be applied for accurate customer churn prediction models. Furthermore, the customers' recent six-month transactions can be applied to predict customer churn for the coming month. The call data can also be included in the transaction data. Thus, we proposed the following hypotheses:

Hypothesis 8 (H8). *The total monthly caller MOU positively impacts customer loss;*

Hypothesis 9 (H9). *The total monthly called MOU has a positive impact on customer loss;*

Hypothesis 10 (H10). *The total local called MOU positively impacts customer loss.*

The Data Warehouse system, which accumulates telecom data, such as for SMS, was used to increase the customer retention rate for SyriaTel. Generally, all SMS and MMS data that indicate customer behavior should be used, as it is unknown which features will be valuable in predicting churn.

The SMS and MMS data for daily, weekly, and monthly users in the past nine months were aggregated for the research to identify related variables and see how they relate to each other. Three charts were built using three kinds of weights: (1) the standardized SMS and MMS quantities; (2) the standardized customer calling times; (3) the mean of the first two standardized weights. Two features for each chart were produced by applying the SenderRank and PageRank algorithms according to the directed charts [23].

The Indian liberalization and globalization process has influenced the telecom industry. The marked leader Airtel was selected to conduct a case study through its value proposition approach by concentrating on new value-added services such as the new SMS Pack plan. Consequently, the following hypotheses were also assessed:

Hypothesis 11 (H11). *China Unicom's SMS quantity positively impacts customer loss;*

Hypothesis 12 (H12). *China Mobile's SMS quantity positively impacts customer loss;*

Hypothesis 13 (H13). *China Telecom's SMS quantity positively impacts customer loss.*

Our hypotheses are listed in Table 1.

Table 1. The hypotheses of the study.

Hypotheses	Description
H1	The total fee receivable for the month positively impacts customer loss.
H2	The fixed monthly cost has a positive impact on customer loss.
H3	The local fee has a positive impact on customer loss.
H4	The roaming fee has a positive impact on customer loss.
H5	China Unicom's network fee has a positive impact on customer loss.
H6	The fee with China Mobile has a positive impact on customer loss.
H7	The fixed-line fee has a positive impact on customer loss.
H8	The total monthly caller MOU has a positive impact on customer loss.
H9	Total monthly called MOU has a positive impact on customer loss.
H10	The total local caller MOU has a positive impact on customer loss.
H11	China Unicom's SMS quantity has a positive impact on customer loss.
H12	China Mobile's SMS quantity has a positive impact on customer loss.
H13	China Telecom's SMS quantity has a positive impact on customer loss.

4. Methodology

4.1. Data Collection

Client data were provided by three major Chinese telecommunication operators: China Mobile, China Unicom, and China Telecom. These data included the information for 4126 clients from 2007 to 2018, as well as anonymous demographic information, business information, and basic metadata information regarding the clients' fees, calls, and SMS and MMS activity.

The information from the dataset is shown in Table 2.

Table 2. Dataset information.

Information	Characterization
Demographic and business information	<ul style="list-style-type: none"> • Sex • Age • Career • Non-fixed monthly fee • Fixed monthly fee
Phone calls (mobile and fixed line)	<ul style="list-style-type: none"> • Monthly minutes in local calls • Monthly minutes in long distance calls using mobile phone roaming service • Monthly minutes in calls using fixed line • Minutes of usage (MOU)
SMS	<ul style="list-style-type: none"> • Number of SMS
MMS	<ul style="list-style-type: none"> • Number of MMS

4.2. Data Analysis

For data analysis, we used SPSS. Factor analysis, Pearson correlation, chi-square, and discriminant and logistic regression analysis methods were used to predict customer churn [32].

The meanings of the independent variables from F1 to F6 are shown in Table 3.

Table 3. The meanings of independent variables—adapted from Zhang [6].

Independent Variable	Meaning
F1	Common factor of non-monthly fixed cost
F2	Common factor of monthly fixed cost
F3	Common factor of the calls MOU
F4	Common factor of long-distance and roaming call
F5	Common factor of SMS
F6	Common factor of China Unicom's MMS

4.3. Dataset Description

The samples' sex characteristics are shown in Table 4 and Figure 1. Of the 4126 customers, 1184 were females (28.7%) and 2942 were males (71.3%).

Table 4. Sex characteristics of the sample.

	Number	Proportion	Valid Proportion	Accumulative Proportion
F	1184	28.7	28.7	28.7
M	2942	71.3	71.3	100.0
Total	4126	100.0	100.0	

Sex: M, male; F, female.

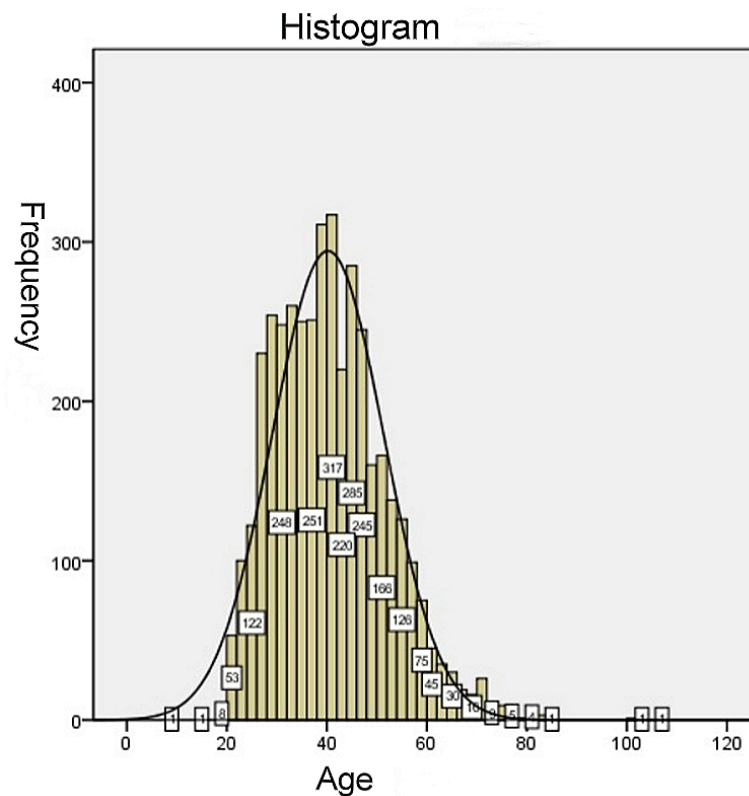


Figure 1. Age distribution of the sample.

Among the 4126 customers, the ages ranged from 9 to 107. However, the most common ages ranged from 20 to 60 years old, representing 95% of the total. Customers aged 40 years were most represented, with 165 cases (4%).

5. Factor Analysis to Characterize Expense, Call, and SMS Attributes

5.1. Expense Factor Analysis

5.1.1. Variable Selection

Factor analysis refers to the concept that significant and measured variables can be decreased to less latent variables with common variance [33]. Some factors are unobservable and unmeasurable, but variables can be reduced into the same group based on similar characteristics to test the relationships [34]. Expense data, such as monthly fee, package type, or mobile terminal price data, can be used to distinguish and characterize customers into different customer consumption tags [29]. Cost and expense management is critical to the operation of companies, and the factor analysis approach could be used to study the expense and cost data and to understand the relationships between the variables [35]. Telecom customer cost data, such as wireless data fees, are suitable for use in factor analysis and could be used to understand customer behavior [36]. Thus, the telecom customers' expense data were selected to conduct the following factor analysis. All expense-related factors, including the (1) total fee receivable for the month, (2) fixed monthly costs, (3) local fee, (4) roaming fee, (5) Unicom's network fee, (6) China Mobile's fee, and (7) fixed-line fee, were used to conduct the factor analysis and analyze the characteristics of the cost factors. Later, Kaiser–Meyer–Olkin (KMO) and Bartlett tests were applied to identify whether these factors are suitable for factor analysis.

5.1.2. Research Hypothesis Testing: KMO and Bartlett Sphericity Tests

The KMO and Bartlett tests were carried out to identify whether the data could be used to conduct a factor analysis with good effect. If the KMO measures of sampling adequacy are >0.5 or the value of Sig is <0.05 , the data can be used to conduct a factor

analysis with good effect. The KMO and Bartlett test results for expense data are shown in Table 5. The KMO measures of sampling adequacy were $0.599 > 0.5$, and the value of Sig was $0.000 < 0.05$. Therefore, it was concluded that the data were suitable for factor analysis.

Table 5. KMO and Bartlett tests.

Test of KMO and Bartlett		
KMO measures of sampling adequacy		0.599
	Value of Chi-square	22,244.842
Bartlett testing of sphericity	Value of df	21
	Value of Sig.	0.000

5.1.3. Common Factor Variance of Expenses

Factor analysis needs to extract overlapping information for variables in order to reduce them. This requires that the original variables must have strong correlations with each other. If there is no overlapping information between the variables, they cannot be integrated and concentrated, and there is no need to perform the factor analysis.

We applied the common factor variance to judge the degree of information condensing via factor analysis (Table 6). The common extracted factor values reached a maximum value of 87.8% and a minimum of 57.8%, with most being greater than 60%. The effect was good, and the information loss was low for each variable. It can be concluded that the results were representative and reliable.

Table 6. Common factor variance results—adapted from Zhang [6].

	Initial Value	Extraction Value
Total fee receivable for the month	1	0.857
Fixed monthly cost	1	0.878
Local fee	1	0.682
Roaming fee	1	0.578
Unicom network fee	1	0.590
Fee with China Mobile	1	0.838
Fixed line fee	1	0.672

Principal component analysis applied to extract values.

5.1.4. Total Interpretation Variance

The cumulative variance of the first two factors was 72.798%, suggesting most of the observed variables were fully represented (Table 7). Therefore, the common factors F1 and F2 were selected.

Table 7. Total interpretation variance.

Component	Eigenvalues Starting Value			Squared Sum Extraction Loading			Square Sum Rotation Loading		
	Sum	Variance %	Accumulative %	Sum	Variance %	Accumulative %	Sum	Variance %	Accumulative %
1	4.086	58.369	58.369	4.086	58.369	58.369	4.054	57.910	57.910
2	1.010	14.429	72.798	1.010	14.429	72.798	1.042	14.888	72.798
3	0.912	13.024	85.822						
4	0.473	6.756	92.578						
5	0.298	4.263	96.841						
6	0.171	2.447	99.288						
7	0.050	0.712	100.000						

Figure 2 shows a screen plot. The horizontal axis shows the component numbers, while the vertical axis shows the eigenvalues. The eigenvalues for the first two common factors 1 and 2 were greater than 1, which meant they were suitable for analysis.

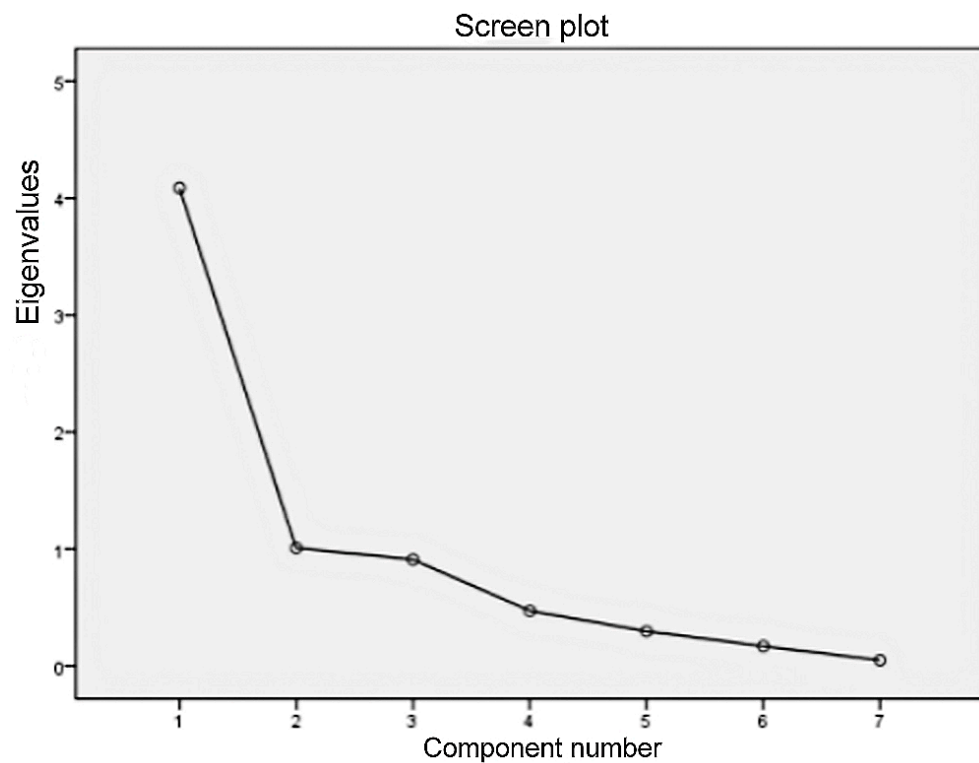


Figure 2. Scree plot.

5.1.5. Component Matrix

A component score coefficient matrix is shown in Table 8. F2 had a more significant load for the number of fixed monthly costs. Additionally, except for the small load on the fixed monthly cost, the first factor has the same load on the other cost factors. Therefore, the first factor F1 can explain the non-monthly fixed cost factor.

Table 8. Component score coefficient matrix.

	Ingredient	
	1	2
Total fee receivable for the month	0.217	0.100
Fixed monthly cost	−0.063	0.918
Local fee	0.217	−0.241
Roaming fee	0.157	0.226
Unicom network fee	0.198	−0.123
Fee with China Mobile	0.229	−0.043
Fixed line fee	0.195	0.062

Principal component analysis applied to extract values.

Therefore, we confidently concluded that F1 (common factor of non-monthly fixed costs) and F2 (common factor of monthly fixed costs) could characterize the expense attributes. The formulas used are shown below, which were adapted from Zhang [6]:

$$F1 = 0.217 \times \text{Total fee receivable for the month} - 0.063 \times \text{Fixed monthly cost} + 0.217 \times \text{Local fee} + 0.157 \times \text{Roaming fee} + 0.198 \times \text{Unicom network fee} + 0.229 \times \text{Fee with China Mobile} + 0.195 \times \text{Fee with fixed line}.$$

$$F2 = 0.100 \times \text{Total fee receivable for the month} + 0.918 \times \text{Fixed monthly cost} - 0.241 \times \text{Local fee} + 0.226 \times \text{Roaming fee} - 0.123 \times \text{Unicom network fee} - 0.043 \times \text{Fee with China Mobile} + 0.062 \times \text{Fee with fixed-line}.$$

5.2. Factor Analysis of Telecom Customer Calls

5.2.1. Variable Selection

Customer call data, such as data for total monthly calls, long-distance calls, and roaming calls, are suitable for use in a factor analysis to investigate the main factors influencing customer preference for the service provider [37]. Factor analysis was conducted on several variables, including customer call data, to identify the main factors determining customer loyalty. It was concluded that better call quality and service will positively influence customer loyalty [38]. Thus, the telecom customers' call data were selected to conduct the following factor analysis. The following call-related factors were used to conduct the factor analysis and analyze the characteristics of cost factors: (1) total monthly traffic MOU; (2) total monthly caller MOU; (3) total monthly called MOU; (4) total local MOU; (5) total local called MOU; (6) total long-distance MOU; (7) total roaming MOU. Later KMO and Bartlett tests of sphericity were applied to identify whether these factors were suitable for factor analysis.

5.2.2. Research Hypotheses Testing: KMO and Bartlett Sphericity Tests

The KMO and Bartlett test results for call data are shown in Table 9. The KMO measures of sampling adequacy were $0.555 > 0.5$, and the value of Sig was $0.000 < 0.05$. It was concluded that the data were suitable for factor analysis.

Table 9. KMO and Bartlett test results.

KMO and Bartlett's Test		
KMO measures of sampling adequacy		0.555
	Value of Chi-square	102,964.374
Bartlett testing of sphericity	Value of df	21
	Value of Sig.	0.000

5.2.3. Common Factor Variance

The common factor variance results are shown in Table 10. The common extracted factor values ranged between 47.5% and 99.6%. Most of these extraction values were greater than 80%, revealing an ideal overall effect. The results were considered scientific and representative, as each variable's loss rate was low.

Table 10. Common factor variance.

	Initial Value	Extraction Value
Total monthly traffic MOU	1	0.996
Total monthly caller MOU	1	0.882
Total monthly called MOU	1	0.931
Total local MOU	1	0.978
Total local called MOU	1	0.961
Total long-distance MOU	1	0.620
Total Roaming MOU	1	0.475

Principal component analysis applied to extract values.

5.2.4. Total Interpretation Variance

The cumulative variance reached 83.463% (Table 11), suggesting most of the observed variables were represented. Therefore, most of the original information was replaced by factors F1 and F2.

Table 11. Total interpretation variance.

Component	Eigenvalue Starting Value			Squared Sum Extraction Loading			Square Sum Rotation Loading		
	Sum	Variance %	Accumulative %	Sum	Variance %	Accumulative %	Sum	Variance %	Accumulative %
1	4.713	67.330	67.330	4.713	67.330	67.330	4.191	59.869	59.869
2	1.129	16.133	83.463	1.129	16.133	83.463	1.652	23.594	83.463
3	0.910	13.000	96.464						
4	0.245	3.502	99.966						
5	0.002	0.026	99.992						
6	0.001	0.008	100.000						
7	1.148	0.000	100.000						

The scree plot is displayed in Figure 3. The horizontal axis shows the component numbers, while the vertical axis shows the eigenvalues. The feasibility of the first two common factors was revealed, as the eigenvalues of the first two common factors 1 and 2 were greater than 1.

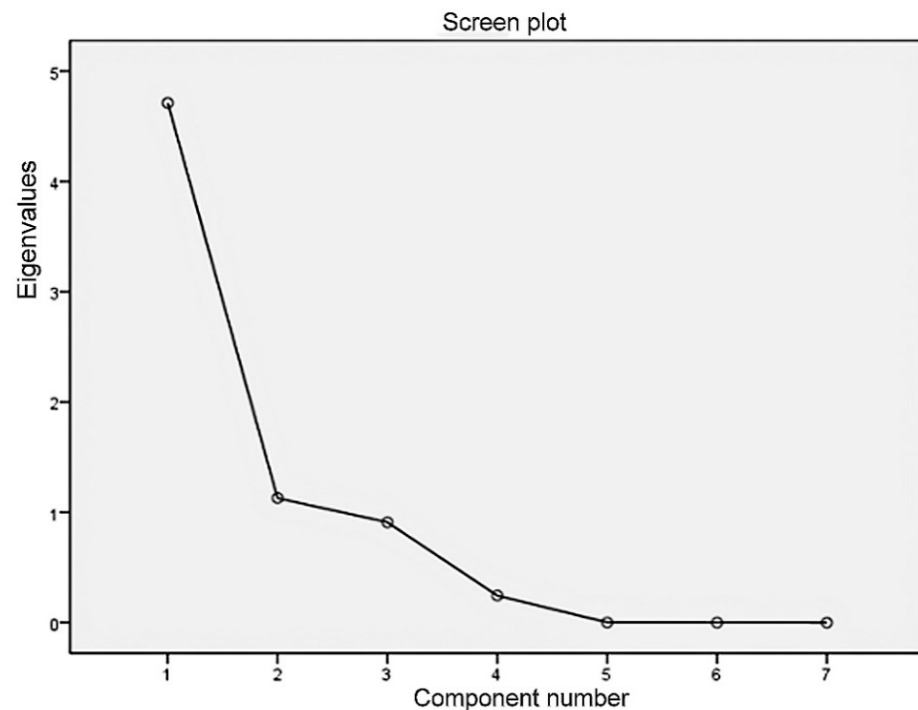


Figure 3. Scree plot.

5.2.5. Component Matrix

The component score coefficient matrix is shown in Table 12. F4 had more significant loads for the total long-distance MOU and total roaming MOU. Therefore, long-distance and roaming calls were resumed as the second factor F4. Additionally, the total monthly called MOU, total local MOU, and total local called MOU numbers showed significant loads for the first factor F3. Therefore, the first factor F3 can explain the called MOU factor.

Therefore, we confidently concluded that F3 (common factors of the called MOU) and F4 (common factors of long-distance and roaming call) characterize the call attributes. The formulas for calculation were as below, which were adapted from Zhang [6]:

$$F3 = 0.179 \times \text{Total monthly traffic MOU} + 0.073 \times \text{Total monthly caller MOU} + 0.257 \times \text{Total monthly called MOU} + 0.283 \times \text{Total local MOU} + 0.294 \times \text{Total local called MOU} - 0.119 \times \text{Total long-distance MOU} - 0.160 \times \text{Total Roaming MOU}.$$

$$F4 = 0.120 \times \text{Total monthly traffic MOU} + 0.317 \times \text{Total monthly caller MOU} - 0.100 \times \text{Total monthly called MOU} - 0.160 \times \text{Total local MOU} - 0.199 \times \text{Total local called MOU} - 0.553 \times \text{Total long-distance MOU} - 0.540 \times \text{Total Roaming MOU}.$$

Table 12. Component score coefficient matrix.

	Ingredient	
	1	2
Total monthly traffic MOU	0.179	0.120
Total monthly caller MOU	0.073	0.317
Total monthly called MOU	0.257	−0.100
Total local MOU	0.283	−0.160
Total local called MOU	0.294	−0.199
Total long-distance MOU	−0.119	0.553
Total Roaming MOU	−0.160	0.540

Principal component analysis applied to extract values.

5.3. SMS of Telecom Customers Factor Analysis

5.3.1. Selection of Variables

Factor analyses are performed to explore the factors that influence telecom customer experiences using certain variables, including customer SMS and MMS data [39]. Customer SMS data, for example relating to the SMS quantity in the telecom package, are suitable for use in factor analyses, which could help telecom companies to identify the factors that impact the customer satisfaction and loyalty [40]. The telecom sector has achieved impressive development in Bangladesh. Customer SMS data has been used in factor analyses, helping to understand the relationship between SMS data and customer loss [20]. Thus, the telecom customers’ SMS and MMS data in the data source were selected to conduct the following factor analysis. All SMS-related factors, including (1) China Unicom’ SMS quantity, (2) China Mobile’s SMS quantity, (3) China Telecom’s SMS quantity, (4) China Unicom’s MMS quantity, and (5) CRBT, were used to conduct the factor analysis and analyze the characteristics of the cost factors. Later, KMO and Bartlett tests were applied to identify whether these factors could be used to conduct the factor analysis.

5.3.2. Research Hypothesis Testing: KMO and Bartlett Tests of Sphericity

The test results of KMO and Bartlett for SMS data are shown in Table 13. The KMO measures of sampling adequacy were $0.567 > 0.5$, and the value of Sig was $0.000 < 0.05$. It was concluded that the data were suitable for factor analysis.

Table 13. Test of KMO and Bartlett.

Test of KMO and Bartlett		
KMO measures of sampling adequacy		0.567
	Value of Chi-square	636.772
Bartlett testing of sphericity	Value of df	10
	Value of Sig.	0.000

5.3.3. Common factor variance

The results of common factor variance are shown in Table 14. The common factor extracted revealed results more significant than 50%. The results were considered scientific and representative, as each variable’s loss rate was low.

Table 14. Common factor variance—adapted from Zhang [6].

	Initial Value	Extraction Value
China Unicom SMS quantity	1	0.580
China Mobile SMS quantity	1	0.594
China Telecom SMS quantity	1	0.545
China Unicom MMS quantity	1	0.556
CRBT	1	0.629

Principal component analysis applied to extract values.

5.3.4. Total Variance of Interpretation

The cumulative variance was 50.087% (Table 15), suggesting most of the observed variables were represented. Therefore, most of the original information was replaced by factors F1 and F2.

Table 15. Total interpretation variance.

Component	Eigenvalue Starting Value			Squared Sum Extraction Loading			Square Sum Rotation Loading		
	Sum	Variance %	Accumulative %	Sum	Variance %	Accumulative %	Sum	Variance %	Accumulative %
1	1.458	29.151	29.151	1.458	29.151	29.151	1.313	26.265	26.265
2	1.047	20.937	50.087	1.047	20.937	50.087	1.191	23.822	50.087
3	0.971	19.423	69.51						
4	0.815	16.291	85.801						
5	0.71	14.199	100						

The scree plot is displayed in Figure 4. The horizontal axis shows the component numbers, while the vertical axis shows the eigenvalues. The feasibility of the first two common factors was revealed, as the eigenvalues of the first two common factors 1 and 2s were greater than 1.

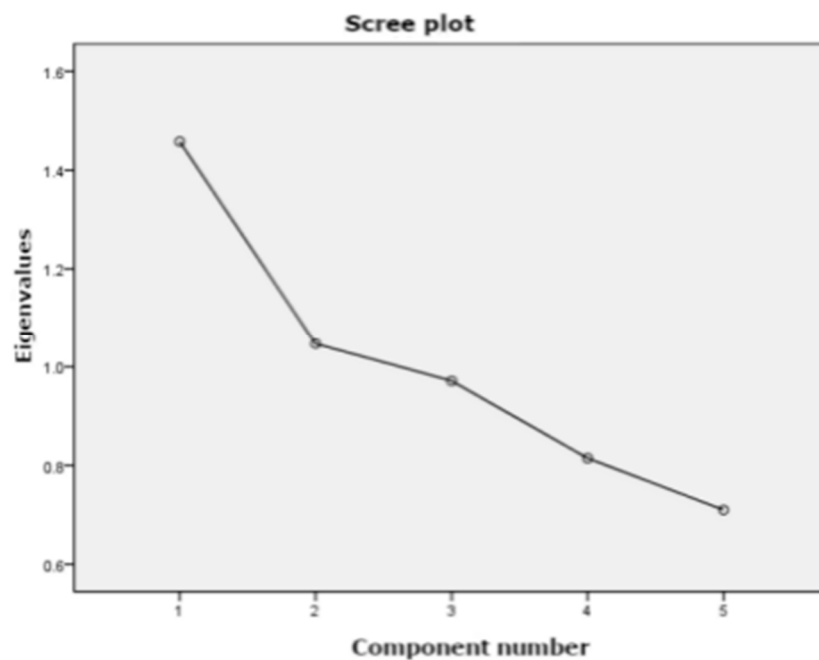


Figure 4. Scree plot.

5.3.5. Component Matrix

The component score coefficient matrix is shown in Table 16. F6 had more significant loads for China Unicom’s MMS quantity and CRBT. Therefore, MMS and CRBT were resumed as the second factor F6. Moreover, the first factor F5 showed more significant

loads for China Unicom’s SMS quantity, China Mobile’s SMS quantity, and China Telecom’s SMS quantity. Therefore, the SMS quantity can be explained by the first factor F5.

Table 16. Component score coefficient matrix.

	Ingredient	
	1	2
China Unicom’s SMS quantity	0.596	−0.106
China Mobile’s SMS quantity	0.570	0.035
China Telecom’s SMS quantity	0.295	−0.034
China Unicom’s MMS quantity	0.011	0.614
CRBT	−0.120	0.685

Principal component analysis applied to extract values.

Therefore, it can be concluded that F5 characterizes the SMS, while F6 characterizes MMS and CRBT. The used formulas were as follows:

$$F5 = 0.596 \times \text{China Unicom SMS quantity's SMS} + 0.570 \times \text{China Mobile SMS quantity's SMS} + 0.295 \times \text{China Telecom SMS quantity's SMS} + 0.011 \times \text{China Unicom MMS quantity's MMS} - 0.120 \times \text{CRBT}.$$

$$F6 = -0.106 \times \text{China Unicom SMS quantity's SMS} + 0.035 \times \text{China Mobile SMS quantity's SMS} - 0.034 \times \text{China Telecom SMS quantity} + 0.614 \times \text{China Unicom MMS quantity} + 0.685 \times \text{CRBT}.$$

6. Discriminant Telecom Customer Loss Model

6.1. Empirical Analysis for the Discriminant Model

6.1.1. Discriminant Attributes

According to the data, the discriminant analysis revealed an appropriate discriminant model. The model refers to the discrimination between the sample and the parent. First, historical data are established from the samples’ discriminant distances. Then, each sample’s data are replaced with the discriminant distance to calculate the actual distance.

6.1.2. Analysis of Discriminant Model

The discriminant model’s eigenvalues were analyzed to identify the discriminating judgment power of the function. Then, Wilks’ lambda discriminant test was applied to confirm the significance of the discriminant function, i.e., whether the discriminant function was valid or not. Afterward, Fisher’s linear discriminant function was used for the telecom customer loss prediction equation, indicating the key factors (F1, F2, F3, F4, F5, and F6) that could influence the telecom customer churn. Finally, an accuracy test was conducted for the discriminant function to investigate the accuracy of the discriminant equation.

(1) Eigenvalues of the discriminant function

The discriminant model was used in the analysis. In the table below, when the discriminant model’s eigenvalue is higher, the model’s discriminating judgment power is higher. The last column represents the canonical correlation coefficient, while the results reveal an acceptable range due to the discriminant function’s eigenvalue (0.030) and canonical correlation (0.171) (Table 17). In the Table 17, “a” means that the former canonical discriminant function was used in the analysis.

Table 17. Eigenvalues.

The Function	The Eigenvalues	Variance Percentage	Accumulative Percentage	Canonical Correlation
1	0.030a	100	100	0.171

(2) Wilks’ Lambda discriminant test

Wilks’ lambda is the ratio of the within-group sum of squares to the total sum of squares. The value is one when the group means for all observations are equal; it is close to zero when the within-group variation is small compared to the total variation. Thus, a large Wilks’ lambda value indicates that the means of each group are more or less equal; a small Wilks’ lambda value shows that the means of each group are different. It can be seen from Table 18 that the first discriminant function explained 97.1% of all variations. Moreover, the value of Sig. was $0.000 < 0.05$, meaning that this discriminant function was significantly established.

Table 18. Wilks’ lambda values.

Function Testing	Wilks’ Lambda Value	Chi-Square Value	Value of df	Value of Sig.
1	0.971	121.638	7	0

(3) Fisher’s linear discriminant function test

Y1 and Y2 represent the customer churn and customer existence, respectively (Table 19).

Table 19. Classification function coefficients—adapted from Zhang [6].

	The Loss or Retain of Customers	
	Customer Loss-Y1	Customer Retain-Y2
Factor score for F1	−1.518	−0.1
Factor score for F2	0.257	0.176
Factor score for F3	0.588	0.135
Factor score for F4	6.021	0.291
Factor score for F5	−0.712	−0.211
Factor score for F6	−1.051	−0.02
Gender	5.963	6.397
(The constant)	−16.592	−4.81
The Fisher linear discriminant equation		

The established Fisher discriminant equation was as follows, which was adapted from Zhang [6]:

$$Y1 = -16.592 - 1.518 \times F1 + 0.257 \times F2 + 0.588 \times F3 + 6.021 \times F4 - 0.712 \times F5 - 1.051 \times F6 + 5.963 \times \text{Gender}.$$

$$Y2 = -4.810 - 0.100 \times F1 + 0.176 \times F2 + 0.135 \times F3 + 0.291 \times F4 - 0.211 \times F5 - 0.020 \times F6 + 6.397 \times \text{Gender}.$$

The discriminant model indicates the top factors that could be used to forecast the telecom customer churn. The classification is considered to be Y1 if the result is one, revealing customer churn. If the result is zero, the classification is Y2, suggesting customer retention.

(4) Accuracy test for discriminant function

One hundred random samples from the dataset were chosen to conduct the accuracy test. The results are shown in Table 20. Half of them were lost customers, and half were retained customers. The one hundred random samples were imported into the telecom customer churn discrimination model. Then, the predicted customers churn results were obtained to judge the prediction accuracy rate of the model.

Table 20. Discriminant result checklist—adapted from Zhang [6].

Test Results	Customer Retain	Customer Loss	Sum
Total	50	50	100
Successful prediction	36	39	75
Failure prediction	14	11	25
The accuracy rate	0.720	0.780	0.750

From the above Table, we can see that the overall prediction accuracy rate was 75%. Among the 50 retained customers, 36 were predicted successfully. The accuracy rate was 72%. Furthermore, among the 50 churn customers, 39 of them were predicted successfully, and the accuracy rate was 78%.

7. Logistic Regression Model of Telecom Customer Churn Prediction

It can be seen from Table 21 that a total of 19 items, such as the Total fee receivable for the month, are independent variables. Moreover, filter_\$, which means the customer is lost or retained, is the dependent variable for binary logistic regression analysis to build the customer loss prediction model. When filter_\$ is one suggests that the customer is lost. When filter_\$ is 0, it suggests that the customer will be retained. Based on these results, we can estimate whether or not a customer will stay with a telecommunications service provider based on the information in the dataset. The model formula is: $\ln(p/1 - p) = -2.056 - 0.002 \times \text{Total fee receivable for the month} - 0.308 \times \text{Fixed monthly cost} - 0.077 \times \text{Local fee} + 0.023 \times \text{Roaming fee} + 0.041 \times \text{Unicom network fee} + 0.031 \times \text{Fee with China Mobile} + 0.032 \times \text{Fee with fixed-line} + 0.003 \times \text{China Unicom SMS quantity} + 0.004 \times \text{China Mobile SMS quantity} + 0.003 \times \text{China Telecom SMS quantity} + 0.009 \times \text{China Unicom MMS quantity} + 0.238 \times \text{CRBT} - 0.539 \times \text{Total monthly traffic MOU} - 0.016 \times \text{Total monthly caller MOU} - 0.057 \times \text{Total monthly called MOU} + 0.559 \times \text{Total local MOU} + 0.039 \times \text{Total local called MOU} + 0.548 \times \text{Total long-distance MOU} + 0.510 \times \text{Total Roaming MOU}$ (where p represents the probability that filter_\$ is 1, which indicates that the customer will be lost. Furthermore, $1-p$ represents the probability that filter_\$ is 0, which indicates that the customer will be retained).

According to the parameter test, it can be seen that the regression coefficient of the total fee receivable for the month was -0.002 , but this was not significant, since $z = -0.402$ and $p = 0.688 > 0.05$. This suggests that the total fee receivable for the month will not affect filter_\$. Thus, hypothesis 1 was rejected, meaning that the total monthly fee receivable does not positively impact customer loss.

The regression coefficient of the fixed monthly cost was -0.308 , which was significant, since $z = -11.564$ and $p = 0.000 < 0.05$, suggesting that the fixed monthly cost will have a significant negative impact on the customer churn. Moreover, the dominance ratio (OR value) was 0.735, suggesting that when the fixed monthly cost increases by one unit, the decrease in Y is 0.735 times. Thus, hypothesis 2 was rejected, suggesting that the monthly fixed cost does not positively impact customer loss.

The summary analysis showed that Unicom’s network fee, China Mobile’s network fee, fixed-line fee, China Unicom’s SMS quantity, China Mobile’s SMS quantity, China Unicom’s MMS quantity, CRBT, total local MOU, total long-distance MOU, and total roaming MOU have a significant favorable influence on the customer churn. On the other hand, the fixed monthly cost, local fee, total monthly traffic MOU, total monthly caller MOU, and total monthly called MOU significantly negatively impact the customer churn. However, the total fee receivable for the month, roaming fee, China Telecom’s SMS quantity, and total local called MOU do not affect the customer churn. Therefore, H1, H2, H3, H4, H8, H9, H10, and H13 were rejected, while H5, H6, H7, H11, and H12 were confirmed.

Table 21. Binary logistic regression.

Item	Regression Coefficients	Standard Error	z Value	Wald χ^2	p Value	OR Value	OR Value 95% CI
Total fee receivable for the month	−0.002	0.005	−0.402	0.162	0.688	0.998	0.988~1.008
Fixed monthly cost	−0.308	0.027	−11.564	133.734	0.000	0.735	0.698~0.774
Local fee	−0.077	0.010	−7.979	63.665	0.000	0.926	0.908~0.943
Roaming fee	0.023	0.021	1.051	1.104	0.293	1.023	0.981~1.067
Unicom network fee	0.041	0.010	3.988	15.906	0.000	1.041	1.021~1.062
Fee with China Mobile	0.031	0.009	3.639	13.243	0.000	1.032	1.014~1.049
Fee with fixed line	0.032	0.010	3.254	10.590	0.001	1.032	1.013~1.052
China Unicom SMS quantity	0.003	0.001	3.466	12.014	0.001	1.003	1.001~1.004
China Mobile SMS quantity	0.004	0.001	7.168	51.379	0.000	1.004	1.003~1.005
China Telecom SMS quantity	0.003	0.006	0.511	0.261	0.609	1.003	0.992~1.014
China Unicom MMS quantity	0.009	0.002	4.250	18.058	0.000	1.009	1.005~1.013
CRBT	0.238	0.031	7.599	57.740	0.000	1.268	1.193~1.348
Total monthly traffic MOU	−0.539	0.252	−2.143	4.592	0.032	0.583	0.356~0.955
Total monthly caller MOU	−0.016	0.006	−2.711	7.348	0.007	0.984	0.973~0.996
Total monthly called MOU	−0.057	0.023	−2.529	6.395	0.011	0.945	0.904~0.987
Total local MOU	0.559	0.252	2.217	4.916	0.027	1.749	1.067~2.867
Total local called MOU	0.039	0.022	1.812	3.284	0.070	1.040	0.997~1.086
Total long-distance MOU	0.548	0.251	2.182	4.763	0.029	1.730	1.057~2.830
Total Roaming MOU	0.510	0.254	2.010	4.041	0.044	1.665	1.013~2.736
Intercept	−2.056	0.121	−16.974	288.116	0.000	0.128	0.101~0.162

Dependent variable: filter_\$.

In Table 22, the model’s overall prediction accuracy is shown to be 93.94%, and the model’s fit is acceptable. The logistic regression analysis and hypothesis tests show that expense, SMS, and call information factors influence customer churn. Moreover, the accuracy test for the logistic regression prediction model proved that it has good prediction performance, with an accuracy rate of 93.94%. Thus, it is possible to estimate whether or not a customer will stay with a telecommunications service provider based on information from the data. This investigation indicates that the logistic regression method could be used to predict customer churn with high accuracy.

Table 22. Binary logistic regression prediction accuracy rate.

		Forecast Value		Forecast Accuracy	Forecast Error Rate
		0	1		
True value	0	3823	36	99.07%	0.93%
	1	214	53	19.85%	80.15%
Summary				93.94%	6.06%

8. Discussion

The data are mainly from three major Chinese telecom operators: China Mobile, China Unicom, and China Telecom. This study aimed to use factor analysis to investigate the business characteristics of telecom clients and to build a discriminant model and a logistic regression model to predict telecom client churn. We showed how the Fisher discriminant equations and logistic regression analysis could be applied to build a telecom customer churn prediction model and achieve better evaluation metrics results for accuracy. After comparison, we suggest that the logistic regression approach performs better when building a telecom customer churn prediction model, with an accuracy rate of 93.94%.

Today’s market is getting more competitive [41]. Telecom companies must make critical decisions and develop effective retention methods to avoid customer churn, as retaining existing customers is much less expensive [2]. In a competitive environment, retaining customers is critical. The telecom customer churn prediction model constructed

using a logistic regression approach suggests that churn can be predicted when customers are unsatisfied with the offered service.

Fisher discriminant equations and logistic regression analysis were used to build a telecom customer churn prediction model. In our preliminary study, the logistic regression approach performed better than the others, with an accuracy rate of 93.94%, as compared to Fisher's discriminant equations with 75%.

9. Conclusions

Telecom customer churn is a central issue for telecom companies, since it decreases profits [1]. Furthermore, preventing customer churn is imperative. As the global telecom industry is becoming more saturated and companies are increasingly struggling to retain customers [41]. Currently, most companies invest heavily in marketing to attract new customers. However, keeping existing customers is cheaper than acquiring new customers [2]. Thus, it is becoming more critical and a significant concern for telecommunication companies to prevent customer churn. This study inventively builds a discriminant model and a logistic regression model to predict telecom client churn using customer segmentation data from three major telecommunication Chinese companies. Moreover, the results of this study will give telecom managers the ability to predict customer behavior and loss accurately and to optimize their strategies to improve customer retention rates. Meanwhile, the findings will help companies reduce costs and optimize their budgets. Furthermore, for telecom managers, it will be possible to improve customer targeting through the results of this paper and to increase the profits of telecom companies.

There is very little knowledge about how telecom customers' opinions regarding the services provided by their telecom company impact customer churn. We aimed to cover this research gap using a Fisher discriminant analysis and a logistic regression analysis of telecom customer churn related to diverse factors. Moreover, the discriminant function and logistic regression analysis are proven to predict telecom customer churn [42]. In this study, through a Wilks' lambda discriminant test, we be concluded that the discriminant equation is valid and can explain the reasons for churn. Furthermore, through the accuracy test, the logistic regression equation was also proven to be valid and can explain the reasons for churn. Serrano et al. [43] highlighted that previous telecom customer churn studies have mainly applied factor analysis, cluster analysis, and other methods, while telecom customer churn studies conducted using Fisher discriminant analysis and logistic regression analysis remains scarce, even in top journals. This new investigation should solve this problem.

According to the results of this paper, the recommendations are for telecom companies to decrease their monthly fixed costs and local costs to increase the possibility of retaining their telecom customers. Additionally, the managers of telecom companies have already realized the value and importance of improving the service quality of the Internet, fixed-line, and CRBT products, as well as the call time for long-distant calls and the numbers of SMS and MMS in the telecom package, which has previously been proven to have a positive influence on telecom customers retention.

Research Limitations and Future Directions

The dataset includes the information for 4126 clients from 2007 to 2018. However, it has been nearly four years since then. Because of the COVID-19 pandemic, the telecom market and customer consumption habits may be significantly different from before. Therefore, more current data will be gathered to further improve the model's accuracy and move the model more in line with the current market situation. Furthermore, the model can be further improved using the repeated data testing approach.

Moreover, data were collected from three operators. Data from other operators may increase the reliability of the model. Finally, additional variables could be applied to improve its predictability.

Author Contributions: Conceptualization, T.Z.; methodology, T.Z.; software, T.Z.; validation, T.Z., S.M. and R.F.R.; formal analysis, T.Z., S.M. and R.F.R.; investigation, T.Z., S.M. and R.F.R.; resources, T.Z.; data curation, T.Z.; writing—original draft preparation, T.Z.; writing—review and editing, S.M. and R.F.R.; visualization, T.Z., S.M., and R.F.R.; supervision, S.M. and R.F.R.; project administration, S.M. and R.F.R.; funding acquisition, S.M. and R.F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Fundação para a Ciência e Tecnologia (FCT) within the following Projects: UIDB/04466/2020 and UIDP/04466/2020.

Data Availability Statement: Not Applicable, the study does not report any data.

Acknowledgments: The authors gratefully acknowledge the financial support of CICEE—Research Center in Business & Economics, UAL, Portugal.

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

References

1. Pejić Bach, M.; Pivar, J.; Jaković, B. Churn Management in Telecommunications: Hybrid Approach Using Cluster Analysis and Decision Trees. *J. Risk Financ. Manag.* **2021**, *14*, 544. [CrossRef]
2. Kim, S.; Chang, Y.; Wong, S.F.; Park, M.C. Customer resistance to churn in a mature mobile telecommunications market. *Int. J. Mob. Commun.* **2020**, *18*, 41–66. [CrossRef]
3. Xie, Y.; Li, X.; Ngai, E.W.T.; Ying, W. Customer churn prediction using improved balanced random forests. *Expert Syst. Appl.* **2009**, *36*, 5445–5449. [CrossRef]
4. Fathian, M.; Hoseinpoor, Y.; Minaei-Bidgoli, B. Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. *Kybernetes* **2016**, *45*, 732–743. [CrossRef]
5. Holtrop, N.; Wieringa, J.E.; Gijsenberg, M.J.; Verhoef, P.C. No future without the past? Predicting churn in the face of customer privacy. *Int. J. Res. Mark.* **2017**, *34*, 154–172. [CrossRef]
6. Zhang, T. Telecom customer segmentation and precise package design by using data mining (Dissertação de mestrado, Iscte-Instituto Universitário de Lisboa). *Repositório do Iscte* **2018**. Available online: <https://repositorio.iscte-iul.pt/handle/10071/17567> (accessed on 13 February 2022).
7. Asimakopoulos, G.; Whalley, J. Market leadership, technological progress and relative performance in the mobile telecommunications industry. *Technol. Forecast. Soc. Change* **2017**, *123*, 57–67. [CrossRef]
8. Aydin, S.; Özer, G. The analysis of antecedents of customer loyalty in the Turkish mobile telecommunication market. *Eur. J. Mark.* **2005**, *39*, 910–925. [CrossRef]
9. Almana, A.M.; Aksoy, M.S.; Alzahrani, R. A survey on data mining techniques in customer churn analysis for telecom industry. *Int. J. Eng. Res. Appl.* **2014**, *4*, 165–171.
10. Olle, G.D.O.; Cai, S. A hybrid churn prediction model in mobile telecommunication industry. *Int. J. e-Educ. e-Bus. e-Manag. e-Learn.* **2014**, *4*, 55–62. [CrossRef]
11. Kisioglu, P.; Topcu, Y.I. Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey. *Expert Syst. Appl.* **2011**, *38*, 7151–7157. [CrossRef]
12. Iyengar, R.; Jedidi, K.; Essegai, S.; Danaher, P.J. The impact of tariff structure on customer retention, usage, and profitability of access services. *Mark. Sci.* **2011**, *30*, 820–836. [CrossRef]
13. Mahajan, V.; Misra, R.; Mahajan, R. Review on factors affecting customer churn in telecom sector. *Int. J. Data Anal. Tech. Strateg.* **2017**, *9*, 122–144. [CrossRef]
14. Shukla, V.; Prashar, S.; Pandiya, B. Is price a significant predictor of the churn behavior during the global pandemic? A predictive modeling on the telecom industry. *J. Revenue Pricing Manag.* **2021**, *2021*, 1–14. [CrossRef]
15. Jahanzeb, S.; Jabeen, S. Churn management in the telecom industry of Pakistan: A comparative study of Ufone and Telenor. *J. Database Mark. Cust. Strategy Manag.* **2007**, *14*, 120–129. [CrossRef]
16. Kim, H.S.; Yoon, C.H. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommun. Policy* **2004**, *28*, 751–765. [CrossRef]
17. Kim, M.K.; Park, M.C.; Jeong, D.H. The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommun. Policy* **2004**, *28*, 145–159. [CrossRef]
18. Verbeke, W.; Dejaeger, K.; Martens, D.; Hur, J.; Baesens, B. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *Eur. J. Oper. Res.* **2012**, *218*, 211–229. [CrossRef]
19. Seo, D.; Ranganathan, C.; Babad, Y. Two-level model of customer retention in the US mobile telecommunications service market. *Telecommun. Policy* **2008**, *32*, 182–196. [CrossRef]
20. Al Amin, M.; Jewel, M.M.H.; Fouji. Influencing Factors of Customer Attitude towards SMS Marketing—A Case of Mobile Telecommunication Industry in Bangladesh. *Jagannath Univ. J. Bus. Stud.* **2019**, *1–2*, 65–78.
21. Huang, B.; Kechadi, M.T.; Buckley, B. Customer churn prediction in telecommunications. *Expert Syst. Appl.* **2012**, *39*, 1414–1425. [CrossRef]

22. Hansen, H.; Samuelsen, B.M.; Sallis, J.E. The moderating effects of need for cognition on drivers of customer loyalty. *Eur. J. Mark.* **2013**, *47*, 1157–1176. [[CrossRef](#)]
23. Ahmad, A.K.; Jafar, A.; Aljoumaa, K. Customer churn prediction in telecom using machine learning in big data platform. *J. Big Data* **2019**, *6*, 28. [[CrossRef](#)]
24. De Bock, K.W.; Van den Poel, D. Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Syst. Appl.* **2012**, *39*, 6816–6826. [[CrossRef](#)]
25. Idris, A.; Iftikhar, A. Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Clust. Comput.* **2017**, *22*, 7241–7255. [[CrossRef](#)]
26. De Caigny, A.; Coussement, K.; De Bock, K.W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *Eur. J. Oper. Res.* **2018**, *269*, 760–772. [[CrossRef](#)]
27. Vo, N.N.; Liu, S.; Li, X.; Xu, G. Leveraging unstructured call log data for customer churn prediction. *Knowl. Based Syst.* **2021**, *212*, 106586. [[CrossRef](#)]
28. Pustokhina, I.V.; Pustokhin, D.A.; Nguyen, P.T.; Elhoseny, M.; Shankar, K. Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. *Complex Intell. Syst.* **2021**, 1–13. Available online: <https://link.springer.com/article/10.1007/s40747-021-00353-6> (accessed on 13 February 2022). [[CrossRef](#)]
29. Jia, Y.; Chao, K.; Cheng, X.; Xu, L.; Zhao, X.; Yao, L. Telecom Big Data based Precise User Classification Scheme. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)*; IEEE: Piscataway, NJ, USA, 2019; pp. 1517–1520.
30. Xu, T.; Ma, Y.; Kim, K. Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping. *Appl. Sci.* **2021**, *11*, 4742. [[CrossRef](#)]
31. Hung, S.Y.; Yen, D.C.; Wang, H.Y. Applying data mining to telecom churn management. *Expert Syst. Appl.* **2006**, *31*, 515–524. [[CrossRef](#)]
32. Lee, E.B.; Kim, J.; Lee, S.G. Predicting customer churn in mobile industry using data mining technology. *Ind. Manag. Data Syst.* **2017**, *117*, 90–109. [[CrossRef](#)]
33. Bartholomew, D.J.; Knott, M.; Moustaki, I. *Latent Variable Models and Factor Analysis: A Unified Approach*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 904.
34. Barton, K.; Dielman, T.E.; Cattell, R.B. An Item Factor Analysis of Intrafamilial Attitudes of Parents. In *Factor Analysis*; Greenwood Press: Westport, CT, USA, 1973; Volume 90, pp. 67–72.
35. Xue, S.; Hong, Y. Earnings management, corporate governance and expense stickiness. *China J. Account. Res.* **2016**, *9*, 41–58. [[CrossRef](#)]
36. Zhang, T.J.; Huang, X.H.; Tang, J.F.; Luo, X.G. Case study on cluster analysis of the telecom customers based on consumers' behavior. In *Proceedings of the IEEE 18th International Conference on Industrial Engineering and Engineering Management, Changchun, China, 3–5 September 2011*; pp. 1358–1362.
37. Paulrajan, R.; Rajkumar, H. Service Quality and Customers preference of Cellular Mobile Service Providers. *J. Technol. Manag. Innov.* **2011**, *6*, 38–45. [[CrossRef](#)]
38. John, J. An analysis on the customer loyalty in telecom sector: Special reference to Bharath Sanchar Nigam limited, India. *Afr. J. Mark. Manag.* **2011**, *3*, 1–5.
39. Subramanian, P.; Palaniappan, S. Determinants of customer experience in the telecom industry using confirmatory factor analysis: An empirical study. *Int. J. Concept. Comput. Inf. Technol.* **2016**, *4*, 1–6.
40. Alam, N.; Rubel, A.K. Impacts of corporate social responsibility on customer satisfaction in telecom industry of Bangladesh. *ABC J. Adv. Res.* **2014**, *3*, 93–104. [[CrossRef](#)]
41. Chadha, S.K.; Kapoor, D. Effect of switching cost, service quality and customer satisfaction on customer loyalty of cellular service providers in Indian market. *IUP J. Mark. Manag.* **2009**, *8*, 23–37.
42. Alzubaidi, A.M.N.; Al-Shamery, E.S. Projection pursuit Random Forest using discriminant feature analysis model for churners prediction in telecom industry. *Int. J. Electr. Comput. Eng. (2088–8708)* **2020**, *10*, 1406. [[CrossRef](#)]
43. Serrano, Y.; Rahn, M.; Crump, E.; Venkatramanan, S.; Haas, J.D. Iron deficiency and physical activity after a dietary iron intervention in female Indian tea pickers. *FASEB J.* **2013**, *27*, 845.9. [[CrossRef](#)]