ISSN 2522-3488 Vol. 8, Issue 5: 83-91, DOI: 10.25236/IJNDES.2024.080513

Construction of Customer Satisfaction Indicator System and Analysis of Influential Factors of Communication Carriers Based on Gradient Descent LASSO

Minlian Fan[1,a,#], Shuchang Wang[1,b,#], Ying Zhang[2,c], Fuquan Qiu[1,d], Guorui Zhao[1,e,*]

[1]*School of Computer Science and Engineering, Guangdong Ocean University, Yangjiang, 529500, China*
[2]*School of Materials Science and Engineering, Guangdong Ocean University, Yangjiang, Guangdong, 529599, China*
[a]*fanminlian2021@163.com*, [b]*ayratswa408700@foxmail.com*, [c]*202143911135@stu.gdou.edu.cn*, [d]*2267794639@qq.com*, [e]*zhaoguoruimath@foxmail.com*
[#]*Co-first author*
[*]*Corresponding author*

Abstract: *The research on constructing customer satisfaction index systems and prediction models of communication operators has theoretical significance and application value. With Beijing Mobile Communication as the research case, this paper creates a screening and prediction model for communication carriers' customer satisfaction indicators through the introduction of the gradient descent LASSO algorithm and machine learning model. In this case, the total sample and sub-sample regression models were mainly built using regularized LASSO to compress the high-latitude and complex indicator system. The findings show that 7 first-level indicators and 28 second-level indicators, including the sense of use and voice call stability, have a considerable impact on customer satisfaction in voice call services and Internet access services. On this basis, machine learning algorithms such as the Bayesian network model, multi-layer perceptron, and GBDT are introduced to construct a customer satisfaction prediction model for communication carriers.*

Keywords: *Communications Operator, Customer Satisfaction, GBDT, MLP Multi-Layer Perceptron, LASSO model*

1. Introduction

Customer satisfaction hinges on the gap between post-experience results and pre-experience expectations.[1]Customer satisfaction, quantifiably assessed, becomes more critical as communication technology evolves and industry competition intensifies. Operators face challenges like poor service and network quality, necessitating a focus on improving customer satisfaction to gain trust, support, and competitiveness.[2]As a result, the study of customer satisfaction among communication operators holds social significance, and the advancement of big data technology offers new opportunities for this research.

In the field of research on customer satisfaction of communication operators, early scholars conducted qualitative research based on sociology and psychology; For example, Cardozo first studied customer satisfaction from a business marketing perspective in 1990.Clases Fornell[3]A study of the development of the Swedish Customer Satisfaction Barometer (SCSB) over the first three years was initiated in 2005 to explore the relationship between customer satisfaction and demand-supply in the industry.With the explosion in the amount of data, recent research has focused on modeling satisfaction through techniques such as data mining,e.g.Xinbei Kang[4]researched customer satisfaction in the mobile communications industry in 2006, by selecting customer types, establishing evaluation index systems, designing research plans and questionnaires, and statistically analyzing and forming customer satisfaction and loyalty reports.Hui Cao[5]studied the establishment of the customer satisfaction measurement system in the mobile communication industry. By using the factor analysis method, a comprehensive evaluation system covering internal and external customer satisfaction was constructed and implemented in enterprise operation and management.In 2008,Ying Yang[6]employed a structure-based neural network modelling approach as the foundation for a study on customer loyalty in the Chinese mobile communications market. A conceptual model on customer loyalty in this market was developed.MengZi Wang[7]2013 used the SERVPERF evaluation model and regional business halls-related information to create a questionnaire, and statistical analysis with SPSS to find the main factors affecting the quality of customer-perceived services in business halls.Linfan Cai[8]2017 collected data and information by conducting a questionnaire survey and used the hierarchical analysis method (AHP) to establish an indicator system to carry out an analysis of the concerns of users of communication carriers.Li J[9]et al. 2021 conduct customer satisfaction research with the help of computer-assisted dropped call investigation system (CATI), take China Telecom's service quality customer satisfaction index (TCSI) model as the theoretical basis, study the influencing factor variables of customer satisfaction and enhancement strategy, and construct GBDT-based satisfaction prediction model according to the satisfaction enhancement suggestion, deeply excavate potential dissatisfied customers, and finally make the Network satisfaction is improved by 2.6PP, and voice call and mobile internet satisfaction is improved by more than 3PP.GuoHong Li[10]et al. 2023 Establishing a customer satisfaction management system for communication operators through questionnaire surveys as well as expert consultation.Regarding this research problem, the integration of new techniques and methods like machine learning is relatively scarce in the existing research.

This paper takes machine learning as the main research approach and selects Beijing Mobile, a communication operator, as the research object. Through data collection, the factors or features influencing satisfaction are identified. Moreover, multi-model machine learning algorithms and gradient descent-based LASSO regression are introduced.Which aims to explore the indicators that make up the influencing factors of communication operators.On this basis, the main factors affecting customers' satisfaction with voice service and Internet service are identified and weighted. Furthermore, by taking into account the differences between customers' voice service and Internet service, the customer satisfaction prediction scoring model for communication carriers is ultimately constructed.

2. Screening Influential Factors on Customer Satisfaction of Communication Operators with Pearson Correlation Analysis and Lasso Regression Models

2.1. Pearson correlation analysis to mine the correlation of factors

First, we consider using Pearson correlation analysis to analyze the correlation coefficients between different broad asset class indices.

The Pearson correlation coefficient is calculated as follows:

$$r=\frac{\sum_{i=1}^{n}(\mathrm{x}_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}} \tag{1}$$

Using the formula, substitution solving results in the correlation heat map of each factor of voice call customer satisfaction and each factor of Internet service customer satisfaction.

Voice call satisfaction, network coverage, signal strength, call clarity, and stability are highly correlated (Pearson correlation coefficient 0.8-1.0). Similarly, Internet access service satisfaction correlates strongly (Pearson correlation coefficient 0.8-1.0) with network coverage, signal strength, mobile Internet speed, and stability. To account for potential chance error, secondary screening using Lasso's method was conducted.

2.2. LASSO Regression

Lasso model is a model optimized by *Tibshirani* [11]in 1996 based on ridge regression because ridge regression could not achieve feature screening.The Lasso method aims to compress regression coefficients of less impactful independent variables by maximizing the likelihood function, thereby screening crucial variables and enhancing model explanatory power.

The objective function of the Lasso model is as follows:

$$J(\beta)=\sum(y-X\beta)^2+\lambda\|\beta\|_1=\sum(y-X\beta)^2+\sum\lambda|\beta|=ESS(\beta)=ESS(\beta)+\lambda l_1(\beta) \tag{2}$$

Where $ESS(\beta)$ denotes the sum of squared errors and $\lambda l_1(\beta)$ denotes the penalty term.

Deriving the regression coefficients for the LASSO model, the objective function of Lasso can be expressed as follows:

$$J(\beta)=\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}\beta_i x_{ij})^2+\lambda\sum_{j=1}^{p}|\beta_j| \tag{3}$$

$$ESS(\beta)=\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}\beta_i x_{ij})^2=\sum_{i=1}^{n}(y_i^{\,2}+(\sum_{j=1}^{p}\beta_i x_{ij})^2-2y_i(\sum_{j=1}^{p}\beta_i x_{ij})) \tag{4}$$

The $ESS(\beta)$ is then partialized in the above equation:

$$\frac{\partial ESS(\beta)}{\partial\beta_i}=-2\sum_{i=1}^{n}x_{ij}(y_i-\sum_{k\neq j}\beta_j x_{ij}-\beta_j x_{ij})=-2\sum_{i=1}^{n}x_{ij}(y_i-\sum_{k\neq j}\beta_j x_{ij})+2\beta_k x_{ij})+2\beta_j\sum_{i=1}^{n}x_{ij}^2 \tag{5}$$

For ease of presentation, by making $m_j=\sum_{i=1}^{n}x(y_i-\sum_{j=1}^{p}\beta_i x_{ij})$ in the above equation, the partial derivative of $ESS(\beta)$ can be expressed as $-2m_j+2\beta_j n_j$.

$$\beta_j=\begin{cases}\dfrac{m_j-\dfrac{\lambda}{2}}{n_j},m_j>\dfrac{\lambda}{2}\\ 0,m_j\in\left[-\dfrac{\lambda}{2},\dfrac{\lambda}{2}\right]\\ \dfrac{m_j+\dfrac{\lambda}{2}}{n_j},m_j<\dfrac{\lambda}{2}\end{cases} \tag{6}$$

2.2.1. Total Sample Regression

This paper employs the likelihood function with Lasso penalty for empirical analysis, synthesizing public subscriber data of three major carriers and using Lasso for quadratic factor screening and set $\lambda=0.02$, based on prior literature,as shown in Figure 1.

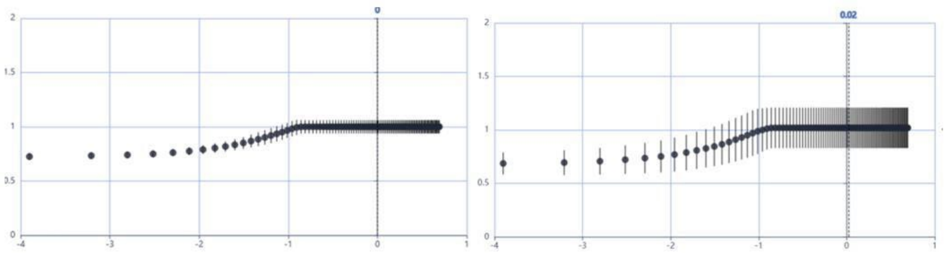

Figure 1: LASSO model λ screening results plot.

Firstly, the influence factors of overall satisfaction of voice calls and satisfaction of Internet service were divided, and the influence factors were compressed by Lasso total variable regression and combined with Pearson correlation analysis for comprehensive screening.

For overall voice call satisfaction, factors to be considered include residential, office, and commercial areas, underground settings, mobile phones with signal issues (no signal, unable to dial, call interruptions, murmurs, inaudibility, intermittency, one-way silence, dropped calls, crosstalk), as well as specific locations like colleges, universities, and high-speed railways. Rural areas' impact should also be retained.

For Internet service satisfaction, influential factors include poor network signal in residential and office areas, inability to access despite signal, intermittent network, slow mobile internet speed, video lag, slow loading of web pages/APP images, delayed mobile payments, overall slowness, and app age (e.g.,

ISSN 2522-3488 Vol. 8, Issue 5: 83-91, DOI: 10.25236/IJNDES.2024.080513

WeChat, Honor of Kings).

2.2.2. Subsample Regression

To explore the variability of certain categories of datasets, Xingxiang Zhang introduces split-sample regression,[12]which looks more closely at the relationship between different groups by grouping the samples into regression analyses.

In this paper, the overall satisfaction factor of voice calls is divided into three categories of $2G/4G/5G$, and the overall satisfaction of Internet service is divided into two categories of men and women, regression analyses are carried out to compare the difference of the influence factor, and when the different categories have different degrees of influence, it indicates that the influence factor has a more significant impact on the customer satisfaction of communication carriers,as shown in Table 1.

Table 1: Sub-sample Regression results.

Impact Factor	Category	Cardinality value	Original hypothesis	X and Y correlation
Whether or not you are experiencing network problems	2G	0.037	Accept	Irrelevant
	4G	0.000	Reject	Relevant
	5G	0.0001	Reject	Relevant
Residential Area	2G	0.047	Accept	Irrelevant
	4G	0.00016	Reject	Relevant
	5G	0.000308	Reject	Relevant
Age	Man	0.000	Reject	Relevant
	Woman	0.003	Reject	Relevant
	Gender unknown	0.346	Accept	Irrelevant

It can be seen that for different network modes, the two influencing factors of whether or not they have encountered network problems and residential neighborhoods have different results on the overall satisfaction with voice calls, indicating that such influencing factors have a more significant impact on them, and similarly, it can be seen that different ages have a significant impact on the overall satisfaction with Internet access services.

2.3. Final Selection of Customer Satisfaction Indicators for Communication Operators

Based on the influencing factors obtained from the above two methods, combined with the results of expert analyses such as Zhixin Cai and Fangping Xiong , the final indicator system of customer satisfaction of communication carriers is constructed,shown in Table 2 and Table 3.

Table 2: Selected language call satisfaction indicators.

Indicator System	Primary Indicator System	Secondary Indicator System
Overall satisfaction with voice calls	Scenarios where voice calls go wrong	Neighborhoods
		Office
		Commercial Street
		Underground
		High School
		High Speed Rail
		Rural
	Voice call stability	Sudden interruption during the call
		Noise, inaudible, intermittent calls
		One party cannot be heard during the call
		Number of missed calls
		High school
		High-speed railway
		Rural
	Network problems during voice calls	Whether encountered network problems
		No signal on mobile phone
		Can't dial with signal
		Cross talking lines

ISSN 2522-3488 Vol. 8, Issue 5: 83-91, DOI: 10.25236/IJNDES.2024.080513

Table 3: Selected indicators of satisfaction with Internet access services.

Indicator System	Primary Indicator System	Secondary Indicator System
Overall satisfaction with Internet access	Scenarios of problems with Internet access	Neighborhoods
		Office
	Mobile Phone Internet Stability	Internet access is intermittent
		Poor network signal
		It says it has a signal, but it won't go online
	Mobile Internet speed	Slow internet access on mobile phones
		Mobile phone payment is slow
		All web pages or apps are slow
		Slow to open web pages or apps
	App Usage	Age
		WeChat
		King of Glory

3. Customer Satisfaction Modelling for Communication Operators

3.1. Customer Satisfaction Modelling of Communication Operators Based on Bayesian Networks

Bayesian network models, namely models that take probability theory as the underlying theory and utilize graph theory as the presentation to illustrate the causal relationships between variables, are primarily applied in statistical inference and for solving certain uncertainty problems.

Bayesian networks, based on historical events and current changes, integrate qualitative and quantitative methods to predict future trends. Given the subjective and uncertain nature of customer satisfaction scoring for communication operators, Bayesian networks can be used for prediction.

And the formula for Bayesian networks is as follows:

$$P(A \mid B)=\frac{P(B \mid A)\times P(A)}{\sum_{i=1}^{n} P(B \mid A_i)\times P(A_i)} \tag{7}$$

$$P(\mathrm{B} \mid \mathrm{A})=\frac{P(A \mid B)\times P(B)}{P(A)} \tag{8}$$

Among them, $P(A)$ and $P(\mathrm{B})$ are determined based on existing information, objective reality, and experience as the basis for determining the possibility of the occurrence of an event, that is, a priori probability.

$P(\mathrm{A} \mid \mathrm{B})$ is the conditional probability of event A occurring if the preconditions for event B to occur are known.

$P(\mathrm{B} \mid \mathrm{A})$ is the likelihood of event A occurring based on the newly occurring conditions, based on which the conditional probability of event B occurring is derived, i.e., the posterior probability of event B.

Among the Bayesian network construction methods contain the following three:

(1)Determine the coefficients based on the Bayesian network nodes determined by the expert.

(2)Based on expert-identified Bayesian network nodes, the network structure and parameters are determined through data learning training.

(3)Based on the relevant nodes and network structure of the Bayesian network determined by the experts, the relevant parameters are determined through data learning training.

Using this algorithm, the results of the prediction set and the training set are shown in Figure 2.

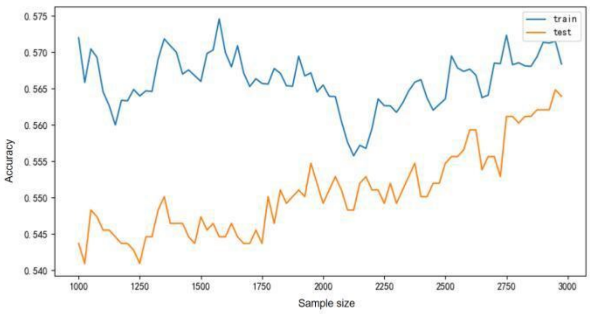

Figure 2: Bayesian training set and test set prediction effects.

As shown in Figure 2, the overall accuracy using Bayesian training and test sets is about 50%.

3.2. Customer Satisfaction Modelling for Communication Operators Based on GBDT

GBDT, an integrated learning method, is a machine learning technique for regression, classification, and sorting tasks. By setting different loss functions, the gradient approach expands its application to various learning tasks.

The GBDT is computed as follows:

$$F(\mathrm{x}, \mathrm{w})=\sum_{m=0}^{M} a_m h_m(x, w_m)=\sum_{m=0}^{M} f_m(x, w_m) \tag{9}$$

Where x is the input sample, w is the input sample, h is the input sample,and α are the weights of each tree.

Given a training datasets: $T=\{(x_1, y_1),(x_2, y_2),\cdots,(x_N, y_N)\}$, which $x_i \in x \in R^n$, x is the input space, $y_i \in Y \in R$, Y is the input space,The loss function is $L(y, f(x))$,Next, build the regression tree F_M.

Initializing the first weak learner $F_0(x)$:

$$F_0(x)=\text{argmin}\sum_{i=1}^{N} L(y_i, c) \tag{10}$$

For building M categorical regression trees $m=1,2,\cdots,M$;

For $i=1,2,\cdots N$, calculate the response value corresponding to the m tree:

$$r_{m,i}=-\frac{\partial L(y_i, F(x_i))}{\partial F(x)} \tag{11}$$

Obtain an expression for the strong learner $F_M(x)$:

$$F_M(x)=\mathrm{F}_0(x)+\sum_{m=1}^{M}\sum_{j=1}^{JM} c_{m,j} I(x \in R_{m,j}) \tag{12}$$

Using this algorithm, the results of the prediction set and the training set are shown in Figure 3.

As shown in Figure 3, using GBDT training set and test set results in about 60% accuracy.

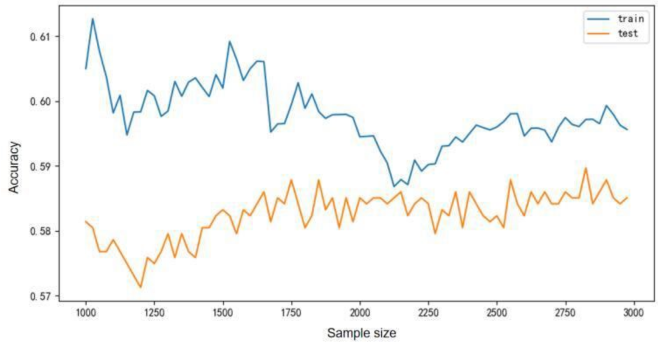

Figure 3: Prediction effect of GBDT training and test sets.

3.3. Establishment of Customer Satisfaction Prediction Model for Communication Carriers Based on MLP Multi-layer Perceptron

The MLP multi-layer perceptron, which is extended from the single-layer perceptron, is a forward-structured ANN artificial neural network that can handle nonlinear, separable problems.

The data were first normalized for $Z-score$:

$$y = (x-\mu)/\sigma \tag{13}$$

Where μ is the mean and σ is the standard deviation.

MLP neural network hidden layer activation function for hyperbolic tangent function:

$$f(y) = \frac{e^{y}-e^{-y}}{e^{y}+e^{-y}} \tag{14}$$

The loss function is the sum-of-squares error function:

$$c = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{15}$$

The factors screened by Lasso above are modeled, and the following ideas and construction steps are translated into computer language:

(1) Factor $O_m(\mathrm{m}=1,2,3,\cdots)$, which strongly correlates with communication operators' customer satisfaction T_m, was selected as a covariate and entered into the input layer, with T_m as the dependent variable. Since the covariates are on different scales (i.e., the criteria for defining them are different), the covariates were subjected to Z-score standardization to enable comparison and weighting of covariates with different scales.

(2)Seventy percent of the data is used as a training set and thirty percent as a validation set to optimize the model parameters backward. The grid search method is used to carry out certain parameter fitness optimization.

(3)Define the structure of the neural network model.The neural network model has 1 hidden layer with 4 nodes, using hyperbolic tangent activation. The output layer uses a constant function $y=x$, and the loss function is the sum of squared errors. The hidden layer's structure was determined through trial-and-error.

(4)Setting up the model training type and multiple trials to optimize the algorithm.

(5)Once the model is built, the strong correlation factor $O_m(\mathrm{m}=1,2,3,\cdots)$ is inputted into the neural network model to obtain model estimates about the customer satisfaction T_{m1} and T_{m2} of the communication operators.

Using this algorithm, the results of the prediction set and the training set are shown in Figure 4.

As shown in Figure 4, the accuracy of the training and test set effects using the MLP model is about 60%.

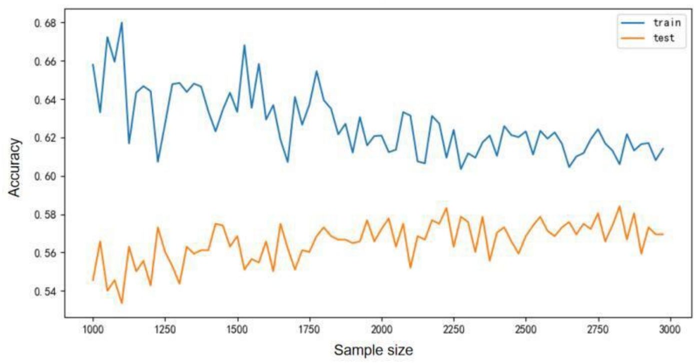

Figure 4: MLP training set and test set prediction effects.

3.4. Multi-model comparative analysis

The prediction accuracy of each model is 60%~70%. Combining the above model analysis and prediction results, this paper chooses the GBDT model for voice business prediction, and the prediction accuracy is the best value of all models at 68.76% under the huge amount of customer satisfaction data of voice business communication carriers, which is slightly higher than that of the model prediction with MLP multi-layer perceptron (67.39%) and Bayesian (66.84%) as models for prediction.

As shown in Figure 5, the MLP multi-layer perceptron is selected for the prediction of Internet access service, and the prediction accuracy is the best value of all models, 67.54%, which is much higher than the prediction of Internet access service with GBDT (65.24%) model as well as Bayesian (59.84%) model.

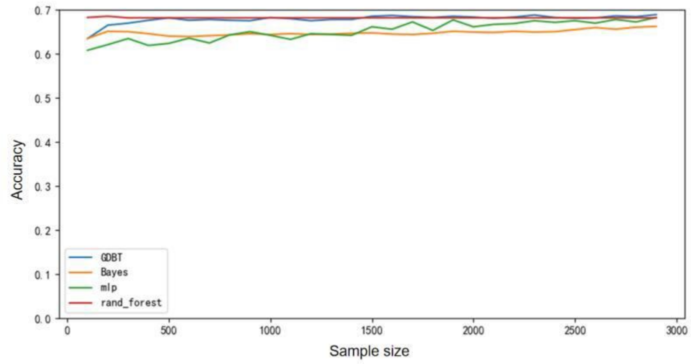

Figure 5: Model Prediction Accuracy Comparison.

Overall, the prediction rates of the four models are generally in the range of 60% to 70% Consider two possible reasons:

(1)The four machine learning models used in this paper have some prediction error.

(2)Turn to the study of the issue of customer satisfaction of communication operators.Customer satisfaction scoring for communication operators is subjective and influenced by multiple factors. The model in this paper is based on available data but may overlook other higher-level factors due to questionnaire limitations.

4. Conclusions

Aiming at the problem of improving the satisfaction of mobile customers' voice service and Internet service, based on the customer satisfaction prediction model of communication operators constructed in this paper, combined with the literature [13]-[17], it is suggested to the communication operators as follows:

Firstly, differentiated intelligent services for customer groups based on Lasso sub-sample screening

ISSN 2522-3488 Vol. 8, Issue 5: 83-91, DOI: 10.25236/IJNDES.2024.080513

show that age, gender, and network mode influence service preferences. Send questionnaires and make calls to young groups, recommend premium packages to speed-conscious customers, emphasizing brand advantages to meet their needs, enhancing customer satisfaction and loyalty.

Secondly, to enhance service levels and create a new concept, this paper uses data crawlers to analyze Micro blog discussions on communication operators. Word clouds show customers are bothered by frequent calls, even if most are not for service recommendations. Upgrading service levels and concepts can improve customer satisfaction.

Third, optimizing network and signal configurations can improve base station coverage, predicted by the MLP model to increase customer satisfaction by 12.3% and 18.6%. This underscores the importance of high-quality service and strong signal coverage for telecommunication operators' survival.

In short, customer satisfaction represents modern communication operators' commitment to service excellence and people-centered thinking. Focusing on customer satisfaction will help operators stand out in product homogenization and promote high-quality, sustainable network development.

Acknowledgements

Minlian Fan and Shuchang Wang contributed equally to this manuscript.

References

[1] Cao Hui. Analysis of the construction and application of customer satisfaction assessment system in the mobile communication industry[J]. Science and Technology Information, 2007,(17): 212-213. DOI: 10.16661/j.cnki.1672-3791.2007.17.200.
[2] Lan Wanshun, Zhang Honghong, Li Peng, et al. Strategies for Improving Customer Satisfaction of Operator Smart Home[J]. Communication Enterprise Management,2024,(09):53-55.
[3] Claes Fornell, Liu Jinlan, Kang Jian. The Swedish customer satisfaction barometer[J]. Journal of Management,2005,(03):372-378.
[4] Kang Xingbei. Research on customer satisfaction analysis of a mobile communication company [D]. Beijing University of Posts and Telecommunications,2006.
[5] Philip Kotler. Marketing Management. Asian Edition[M].Renmin University of China Press, 2004.
[6] YANG Ying. Research on customer loyalty based on structured neural network[D]. Liaoning University of Engineering and Technology,2008.
[7] Mengzi Wang. Influence factor analysis based on customer perceived service quality[D]. Dalian Maritime University,2013.
[8] Lin Fancai. Research on Improving Customer Service Capability of Communication Operators [D]. Shandong University,2017.
[9] LI Jie,WU Wei, GUAN Weijie. Research on customer satisfaction of communication operators based on TCSI model[J]. Guangdong Communication Technology,2021,41(05):10-14.
[10] Li Guohong,Wang Dongjiao. Research on customer satisfaction improvement strategy of communication operators[J]. China New Communication,2023,25(06):4-6.
[11] Qiyuan Jiang, Jinxing Xie, Jun Ye, Mathematical Modelling (Third Edition)[M], Beijing: Higher Education Press, 2003.
[12] Zhang Xingxiang, Zhong Wei, Hong Yongmiao. Indicator system construction and influence factor analysis of national happiness:A screening method based on LASSO[J]. Statistical Research, 2018, 35(11):3-13.
[13] Zhao Fabin. How telecom operators improve customer satisfaction[J]. Digital communication world, 2016(07):71.
[14] Hua Ting. Research on Customer Satisfaction Improvement Strategy of LH Telecom Branch [D]. Nanjing University of Posts and Telecommunications, 2022.
[15] Wang Zhihong, Zhou Jianbo. A study of job satisfaction based on flexible employment perception and voluntary awareness[J/OL]. Research World: 1-9[2023-01-17].
[16] Xu Zhibin. Research on service quality improvement strategy of H city Unicom based on customer satisfaction[D]. Shijiazhuang Railway University,2022.
[17] Zhu Yi-Hua. Reshaping media time: temporal representation and perception of short video communication [D]. Jinan University,2021.