ISSN: 2710-0588

Research on Identifying Important Factors and Prediction of Online Service Satisfaction for Mobile Phone Users

Xiyuan Miao*, Shi Zhang

Central University of Finance and Economics, Beijing 100081, China

Abstract

Our lives cannot do without the internet. How to improve the network quality has always been an essential problem. The paper explores the important factors of online service satisfaction and the best predicting model. Based on the data offered by Beijing Mobile Company, we identify main factors affecting online service satisfaction by calculating their mutual information values. The factors include signal problem factors, scene factors and software usage factors. Additionally, based on decision tree model and models with decision tree as base learner, we predict the online service satisfaction. The result shows that random tree model with One Vs Rest mode has the greatest accuracy among the models which offers telecommunications companies insight.

Keywords

Online Service Satisfaction; Mutual Information; Decision Tree Model; Base Learner.

1. Introduction

Internet is closely related to our lives. There are many telecommunications network users in China. As of December 2020, the number of 5G users in China exceeded 160 million, while the number of 4G users exceeded 1.2 billion. As of March 5, 2023, the number of 5G users in China surpassed 575 million. The group of telecommunications network users is enormous. Therefore, understanding and improving the influencing factors of online satisfaction among these users, as well as predicting online satisfaction, is of great significance for improving network quality [1].

Previous literature did not have much about identifying important factors about the satisfaction of online service for mobile phone users. However, there are many methods for exploring influencing factors of satisfaction in job satisfaction, accommodation satisfaction and other fields [2,3,4]. Based on the research objectives, we choose the mutual information to identify important factors of online service satisfaction. The mutual information value can reveal the nonlinear relationship between different variables [5,6]. this paper uses mutual information for judgment.

When exploring important influencing factors, we select multiple factors, including hardware facilities, scenario conditions, and software usage. The continuous development of technology and software industry implies that software needs better network support [7]. Therefore, the paper incorporates the use of networking software as influencing factors. Predicting online satisfaction is a multi-classification problem. Decision tree models is fit for the multi-classification problems and have advantages in prediction effect [8]. Therefore, we will predict satisfaction based on the decision tree and the model with the decision tree as the base learner. Further, the paper explore which model has a better effect.

According to the above analysis, this paper first calculates the importance of each factor based on the mutual information value. Further, we explore the satisfaction prediction problem based on decision trees and models with base learners as decision trees. Additionally, we explore whether the introduction of One Vs One mode and One Vs Rest mode has a significant impact

ISSN: 2710-0588

on the prediction performance. The conclusions of this paper are contributive. Firstly, important factors selected by mutual information value can help the telecom companies effectively improve their network facilities. Secondly, the model with the best predictive exploration effect also provides insights for telecommunications companies in predicting online satisfaction.

The structure of the paper is as follows: Section 2 shows the method description, Section 3 displays the data description and preprocessing, the result analysis is shown in Section 4; the last section represents conclusion.

2. Methodology

We first extract important factors of overall user satisfaction with accessing the internet based on the mutual information. Secondly, we combine machine learning algorithms including XGBoost model, Random Forests model, Decision Tree model, LightGBM model and LGBT model with One vs Rest mode and One Vs One mode to train and predict the satisfaction of online service. Finally, we analyse the accuracy of each model and give conclusions. The process is shown in Figure 1.

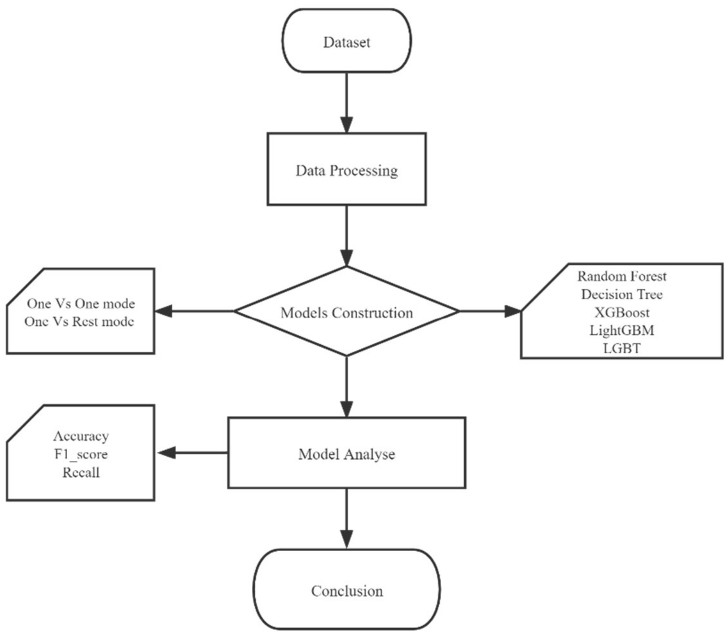

Figure 1. The process of the experiment

2.1. Extraction of Important Factors

Based on the dataset used in the paper, variables are complex and numerous which potentially affect the accuracy of models and might cause the overfitting problem. Therefore, we first extract the important factors according to mutual information of each attribute. The calculation of mutual information value depends on the distribution of data attribute values and labels.

Mutual information is an index measuring the interdependence between variables. It represents the correlation between labels and input variable through their distribution. The method has been used in many studies. [9] Different from the correlation coefficient, mutual information is not limited to real valued random variables. It is more general and determines the similarity of the joint distribution. It decomposes marginal distribution's product. The calculating method is as follows:

ISSN: 2710-0588

$$I(X,Y) = \iint p(X,Y)\log\left(\frac{p(X,Y)}{p(X)p(Y)}\right)dXdY \quad (1)$$

$I(X,Y)$ represents the information value, and $p(X,Y)$ stands for joint distribution. $p(X)$ and $p(Y)$ are behalf of the label distribution and factors distribution. Mutual information represents the increasing or decreasing degree of label caused by factors. The larger mutual information value stands for a closer relationship of the certain factor and label.

2.2. Model Construction

To predict the online service satisfaction, we construct models based on factors selected by mutual information. The target attribute considered as the online service satisfaction is a categorical variable. Additionally, dataset used in the paper has many categorical variables. Decision tree model would have a great effect on predicting in the situation. Therefore, we select decision tree model and models whose base learner is decision tree model. The models include XGBoost model, Random Forests model, decision tree model, LightGBM model and LGBT model.

2.2.1. Decision Tree Model

Decision tree model is a popular machine learning model used for classification and regression problems [10]. The main idea of decision tree is to divide the dataset into smaller subsets until each subset contains only one class or reaches a predetermined stopping condition. The core of the decision tree algorithm is to choose the best attribute to split the dataset. This process is called node splitting. There are several methods to split the dataset, including information gain, information gain ratio, Gini index, etc. At each node, the decision tree algorithm selects the best attribute to split the dataset, and this process will be repeated until all the data are classified into the same class or the predetermined stopping condition is reached. In the paper, we select Gini index. The calculation method is shown in the following formula. Gini coefficient is shown in formula (2).

$$Gini(U) = \sum_i P(u_i)\big(1 - P(u_i)\big) = 1 - \sum_i P(u_i)^2 \quad (2)$$

2.2.2. GBDT Model

GBDT (Gradient Boosting Decision Tree) is an ensemble learning method based on decision trees [11]. The core idea of the GBDT model is to gradually reduce the prediction error through multiple decision trees. During the training process, each tree is constructed to correct the residuals of the previous tree. Finally, the prediction results of all trees are accumulated to obtain the final prediction result. GBDT model has high accuracy. It can handle various types of data, which is favored by many researchers. The steps are shown as follows.

Step 1: Assume the number of categories is k. For samples 1, 2, ..., m, initialize the learner shown in formula (3):

$$f_0(x) = argmin\sum_{i=0}^{m}\sum_{k=1}^{k} L\left(y_k, f_{t-1,l}(x)\right) \quad (3)$$

Step 2: For iteration times t=1, 2, ..., T, calculate the negative gradient error of category l corresponding to sample i in round t.

ISSN: 2710-0588

$$r_{til} = -\left[\frac{\partial \mathrm{L}(y_i, \mathrm{f}(x_i))}{\partial \mathrm{f}(x_i)}\right]|_{f_k(x)=f_{l,t-1}(x_i)} = p_{il} - p_{l,t-1}(x_i) \tag{4}$$

Step3: Based on the negative gradient error of category, we adjust the parameters of model.

2.2.3. XGBoost Model

XGBoost (eXtreme Gradient Boosting) is an extension of the GBDT algorithm [12]. The algorithm uses regularization and parallel processing to improve model accuracy and efficiency. XGBoost model adds regular term on basis of GBDT model to reduce the occurrence of overfitting and enhance the accuracy of prediction. The steps of XGBoost model is shown as follows:

Step1: For the iteration times t=1, 2, ..., T, construct the loss function of the sample i (1, 2, ... m) in the iterative process shown in the formula (5).

$$L_t = \sum_{i=1}^{m} L(y_i, f_{i-1}(x_i) + h_i(x_i)) + \gamma J + \frac{\lambda}{2}\sum_{j=1}^{j} {w_{tj}}^2 \tag{5}$$

λ, γ are regularization coefficient. J represents the number of leaf nodes. w_{tj} stands for the optimal solution for each leaf node region of decision trees in iteration time t.

Step2: Expand the loss function by second-order Taylor expansion. The first derivative and second derivative of the sample i in the iteration time t are shown in formula (6) and formula (7).

$$g_{ti} = \frac{\partial L(y_i, f_{i-1}(x_i))}{\partial f_{i-1}(x_i)} \tag{6}$$

$$h_{ti} = \frac{\partial^2 L(y_i, f_{i-1}(x_i))}{\partial {f_{i-1}}^2(x_i)} \tag{7}$$

The sum of the first derivative and the second derivative of samples in each leaf node area are shown in formula (8) and formula (9):

$$G_{ti} = \sum_{x_i \in R_{tj}} g_{ti} \tag{8}$$

$$H_{ti} = \sum_{x_i \in R_{tj}} h_{ti} \tag{9}$$

Based on equation (8) and (9), the loss function is shown in formula (10):

$$L_t = \sum_{i=1}^{m} (G_{tj} W_{tj} + \frac{1}{2}(H_{tj} + \lambda){w_{tj}}^2) + \gamma J \tag{10}$$

ISSN: 2710-0588

Step3: Split the decision tree based on the current node. The default score is 0. For feature sequence numbers k=1, 2, ..., K, $G_L = H_L = 0$, sort the samples from smallest to largest based on feature k, and take out the sample i in sequence. After the current sample is placed in the left subtree, the sum of the first and second derivative of the left and right subtrees is as follows:

$$G_L = G_L + g_n, G_R = G - G_L \tag{11}$$

$$H_L = H_L + h_n, H_R = H - H_L \tag{12}$$

Based on the formula (11) and (12), try to find larger score:

$$score = \max\left(score, \frac{1}{2}\frac{{G_L}^2}{H_L + \zeta} + \frac{1}{2}\frac{{G_R}^2}{H_R + \zeta} - \frac{1}{2}\frac{(G_R + G_L)^2}{H_L + H_R + \lambda} - \gamma\right) \tag{13}$$

Step4: Split the subtree. Divide the features and their eigenvalues based on the maximum score value.

Step5: If the maximum score value is 0, the construction of decision tree is completed. Then calculate w_{tj} of all leaf node regions, and obtain weak learner $h_t(x)$. Further we update the strong learner $f_t(x)$ and proceed to the next iteration. If the maximum score is not 0, return to step 2 and continue attempting to split the decision tree.

Step6: Predict the result based on the model.

2.2.4. LightGBM Model

LightGBM is a machine learning algorithm based on gradient boosting decision tree [13]. It is improved from XGBoost model. It has advantages of efficiency, speed, and accuracy. LightGBM adopts a histogram-based decision tree learning method, which discretizes the feature values during each node splitting using histograms and uses the histograms instead of raw data, reducing the computation complexity and improving the algorithm's efficiency. The steps are similar to the steps of XGBoost model

2.2.5. Random Forest Model

Random Forest is a popular machine learning algorithm that is widely used for both classification and regression tasks [14]. The basic idea of Random Forest is to build a collection of decision trees, where each tree is trained on a randomly sampled subset of the data and a random subset of the features. This randomness helps to reduce overfitting and improves the model's ability to generalize to new data. Random Forest model is a robust algorithm which can handle noisy data and outliers well. Its powerful algorithm is favored by many researchers. The steps are shown as follows:

Step 1: For N samples in the original training set, each sample has W-dimensional features. Select x samples from the dataset that have been returned to form a training subset randomly. Conduct samples for w times and generate w training subsets;

Step 2: Each training subset forms a decision tree, which forms a total of w decision trees;

Step 3: For a single decision tree, m features are randomly selected from M features at each node of the tree, and split according to the principle of minimizing node impurity. Each tree splits down in this way until all training samples at that node belong to the same class. No pruning is required during the splitting process of the decision tree;

Step 4: Predict the test data based on multiple decision tree classifiers generated. Based on the results of each tree, we use average values to obtain the results.

ISSN: 2710-0588

2.3. One Vs X Mode

OVR (One Vs Rest mode) and OVO (One Vs One mode) are two popular modes for classification problems. The paper uses the two modes to construct the models.

2.3.1. One Vs Rest Mode

OVR (One Vs Rest mode) is a one-binary classifier, where each classifier distinguishes between that class and all the other classes combined. Based on the highest confidence score, the class is predicted. OVR is a simple and effective approach that works well when the number of classes is large or when the classes are unbalanced. The steps are as follows:

Step1: Based on the different categories, we consider a certain category as regular class and other categories as negative classes.

Step2: We train a binary classifier according to the regular class and negative classes.

Step3: Different classes are considered as input of the classification. We select the category with the highest confidence level as the prediction result.

2.3.2. One Vs One Mode

OVO (One Vs One mode) predicts whether the input belongs to one class or the other. OVO mode is computationally expensive and works well when the number of classes is small or when the data is well-balanced. OVO mode has the similar process as OVR mode.

3. Data Description

The data in this paper is provided by Beijing Mobile Company. Beijing Mobile Company is affiliated with China Mobile Communications Group Co., Ltd. The company is mainly engaged in mobile voice and traffic, wired broadband, and other network communication services. The company maintains a high level of industry-leading communication network quality.

The dataset used in this paper has many attributes. The name of attributes, their data types, meanings, and methods for handling vacancy values are presented in the appendix A. We preprocess the data based on its corresponding meaning.

4. Model Analyse

4.1. Mutual Information Value Analysis

Based on Section 2, we calculated the mutual information value for each attribute. We sorted the mutual information values in descending order and selected the top 1/3 of attributes as input variables. 37 attributes were selected. The selected variables and their corresponding mutual information values are shown in Table 1 and figure 1. The mutual information of total variables is shown in Appendix B.

According to the table 1, the main factor affecting users' mobile phone internet access is in the signal problem, including factors such as 2G resident duration and the network speed for surfing the internet with phones. It implies that communication company need to adjust the intensity of mobile phone signal. Another major factor is the scene factor. Scene factors include residential communities, offices, subways, etc. Additionally, the use of mobile phone software also influences the satisfaction. The usage of software including Tencent Video, Tiktok and games also affects mobile users' internet usage relatively significantly. With the development of technology, more and more new software is being updated. Users' experience of using these software affects how they rate their satisfaction with the internet. Communication company should improve the network quality in such factors.

ISSN: 2710-0588

Table 1. Selected influencing factors and their corresponding mutual information values

Influencing factors	Mutual Information Value
2G resident duration	0.165893
Number of redirections	0.164322
Poor network signal or no network signal	0.1492
Slow network speed for surfing the internet with phones	0.128705
Intermittent network or fluctuating network speed	0.127598
Residential community	0.115794
Slow speed for opening web pages or APP pictures	0.113065
Failing to get online with a signal	0.107199
Office	0.087986
Subway	0.08529
Standstills while watching videos	0.0782
GPRS Resource Usage of the Month (GB)	0.076922
Toutiao usage traffic	0.070585
Kuaishou usage traffic	0.067172
Tencent Video usage traffic	0.066264
Tik Tok usage traffic (MB)	0.06587
Usage traffic of video apps	0.064584
Usage traffic of Ali apps	0.064112
Usage traffic of communication apps	0.062302
Youku Video usage traffic	0.061948
Game app usage traffic	0.06191
Cumulative consumption this Year (Yuan)	0.061705
Usage traffic of web apps	0.061207
Age	0.061199
Usage traffic of video apps	0.059687
Average consumption in recent 3 months (Yuan)	0.058911
Average consumption in recent 3 months (excluding communication account payment)	0.058618
Usage times of game apps	0.057721
Usage traffic of Tmall	0.056087
Usage traffic of Tencent apps	0.055294
Disconnection times	0.053349
Slow download speed	0.052995
Wechat	0.051821
Big delay in playing games	0.050392
Usage traffic of Netease apps	0.04901
Usage traffic of music apps	0.048222
Relatively slow speed for mobile payments	0.048038

ISSN: 2710-0588

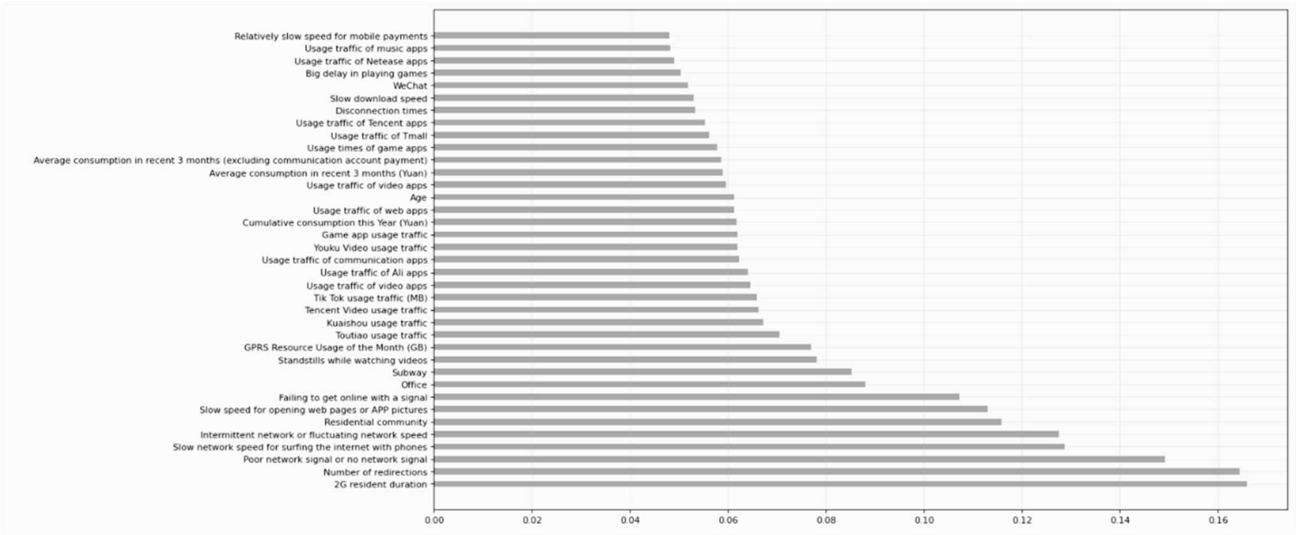

Figure 2. Selected attributes and their mutual information values

4.2. Model Analysis

We select the models based on the Random Forest model, LightGBM model, XGBoost model and GBDT model. In selecting the sample, we choose the k-fold sampling method. K is selected as 5. We calculate the scores for each model with their recall and F1 score. The results are shown as follows.

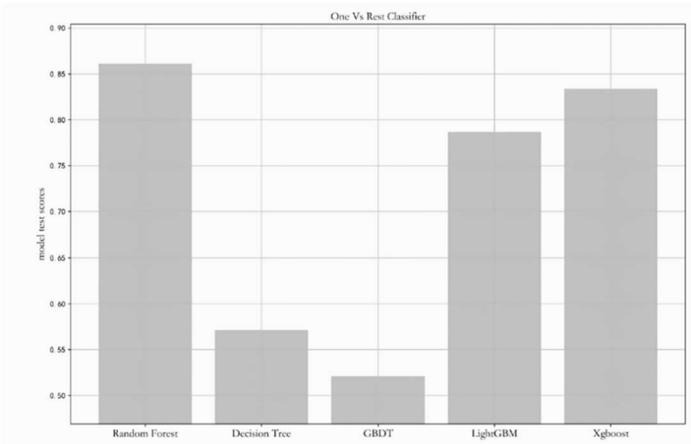

Figure 3. Models' accuracy

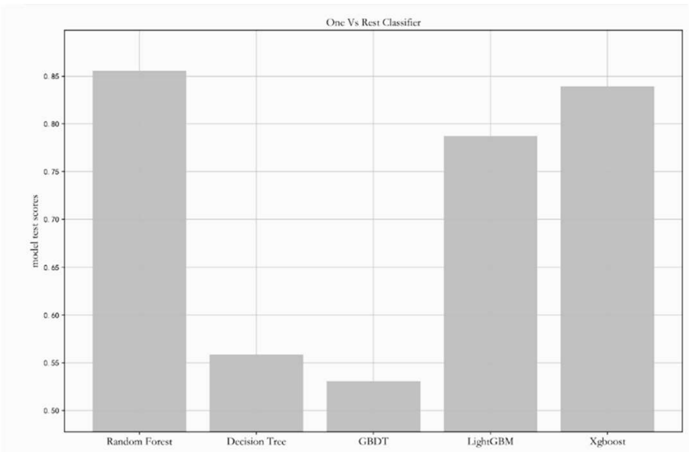

Figure 4. Models' Recall

ISSN: 2710-0588

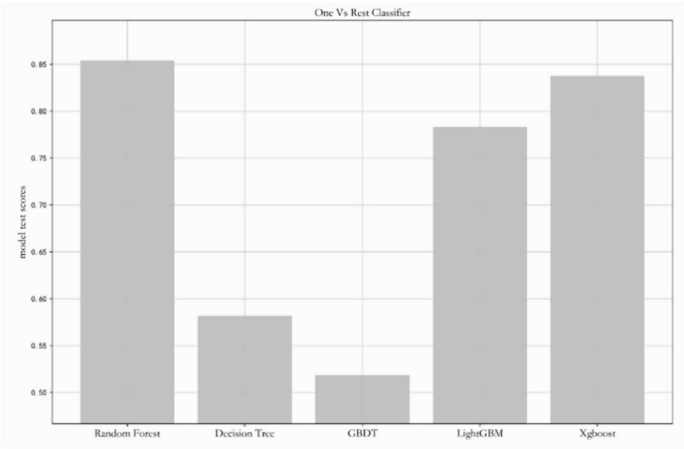

Figure 5. Models' F1 score

Based on recall, F1 score and the model score, Random Forest is more effective, followed by XGBoost. The third model is LightGBM, and the fouth is GBDT. The last model is decision tree model. Table 2 shows the result for each indicator.

Table 2. Accuracy indicators of models

	Recall	F1_score	Score
Random Forests	0.86	0.85	0.86
Decision Tree	0.56	0.58	0.57
GBDT	0.53	0.52	0.52
LightGBM	0.79	0.78	0.79
XGBoost	0.84	0.84	0.83

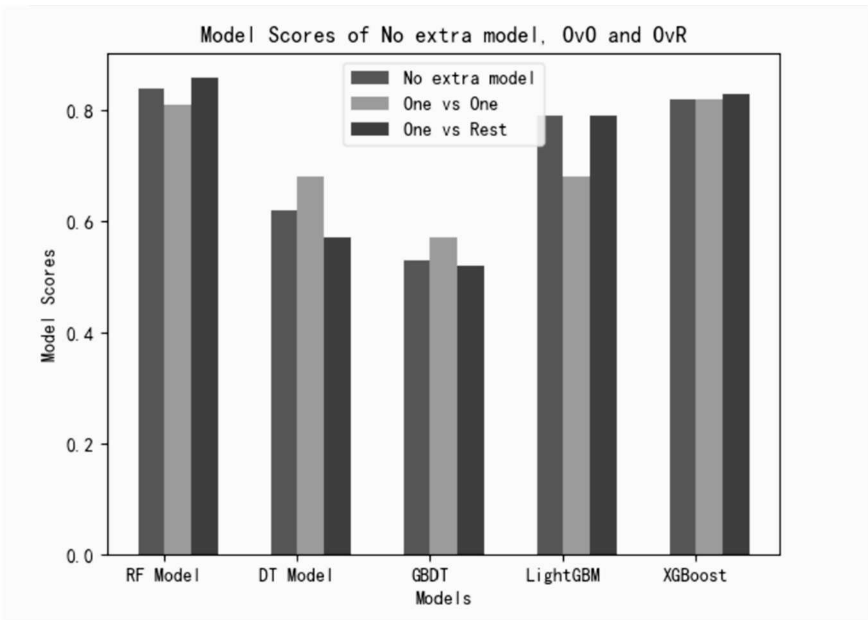

Figure 6. Models' Score with modes

To investigate the effects of OVR and OVO mode, we compare the effects of the OVO mode, OVR mode and the machine learning models respectively shown in the figure 4 and table 3. The difference in accuracy between three different situations for the same model is not significant. The effects of Random Forest, LightGBM and XGBoost models are improved in OVR mode. Decision tree and GDBT model show a decreasing trend in model scores. GBDT model, which is the least effective of models and has low reference value. The decision tree model is relatively single, and its ability to process information is not outstanding. According to the analyse above, OVR mode has an advantage in predicting satisfaction. In addition, the overall effect of the models where no mode used is greater than the group using OVO mode, which also suggests that OVO mode is not applicable in predicting the problem.

Table 3. Models' scores in diffierent modes

	Random Forest	GBDT	Decision Tree	LightGBM	XGBoost
OVR mode	0.86	0.52	0.57	0.79	0.83
OVO mode	0.81	0.57	0.68	0.68	0.82
No mode	0.84	0.53	0.62	0.79	0.82

5. Conclusion

The paper investigates the factors influencing the overall satisfaction of mobile phone users on online service and constructs prediction models. We find that the factors affecting the satisfaction mainly include signal reasons, scene reasons and cell phone software usage. For mobile companies, it is more efficient to improve network quality from these three perspectives. In addition, the paper explores the effect of decision trees and models using decision tree as base learner based on OVR as well as OVO mode. According to the model scores, we conclude that random forest model under OVR model works best in predicting users' satisfaction, which also brings insights in predicting the overall satisfaction of mobile phone users.

The research is based on the overall satisfaction of mobile phone users on online service in Beijing. Online service contains many aspects including network coverage and signal strength, cell phone Internet speed, and cell phone Internet stability. We will explore the prediction models and influencing factors of these indicators in the subsequent experiments.

Acknowledgments

I would like to express my sincere gratitude to my high school alumni Shi Zhang as the second author. Additionally, I would be appreciated to my tuition and university Central University of Finance and Economics. The road to postgraduate is tough. I hope to be worthy of three years' hard work. Wish me every success.

References

[1] S. Rahi, M. A. Ghani, A. H. Ngah: Factors propelling the adoption of internet banking: the role of e-customer service, website design, brand image and customer satisfaction, International Journal of Business Information Systems, vol. 33 (2020), 549–569.

[2] B.H. Su, P.C. Lin, J.M. Chen et al. HAPPINESS/SUFFERING factors recognition based on point-wise mutual information, 2015 International Conference on Orange Technologies (ICOT), Hong Kong, China, (2015), 14-17.

[3] R. Munir , R. A. Rahman . Determining Dimensions of Job Satisfaction Using Factor Analysis, Procedia Economics and Finance, vol. 37 (2016), 488-496.

[4] Tussyadiah, P. Iis. Factors of satisfaction and intention to use peer-to-peer accommodation, International Journal of Hospitality Management, vol. 55 (2016), 70-80.

ISSN: 2710-0588

[5] L. Song, P. Langfelder, S. Horvath, Comparison of co-expression measures: mutual information, correlation, and model based indices, BMC Bioinformatics, vol. 13 (2012), 328.

[6] C. Whitnall, E. Oswald, A Comprehensive Evaluation of Mutual Information Analysis Using a Fair Evaluation Framework, (Springer-Verlag, Germany 2011), p. 316–334.

[7] A. Hakiri , A. Gokhale, P. Berthou et al. Software-Defined Networking: Challenges and research opportunities for Future Internet, Computer Networks, vol 75 (2014), 453-471.

[8] C. Kingsford, S. Salzberg, What are decision trees?, Nature Biotechnology, vol 26 (2008) 1011–1013.

[9] Z. Zeng, H. Zhang, Z. Rui, et al., A Hybrid Feature Selection Method Based on Rough Conditional Mutual Information and Naive Bayesian Classifier[J]. Isrn Applied Mathematics, vol 2014 (2014), 36-46.

[10] J. R. Quinlan, Learning decision tree classifiers, ACM Computing Surveys, vol 28 (1996), 71-72.

[11] J. Feng , Y. Yang, Z.H. Zhou, Multi-Layered Gradient Boosting Decision Trees, Proceedings of the 32nd International Conference on Neural Information Processing Systems, (2018), 3555–3565.

[12] T. Chen , H. Tong , M. Benesty, XGBoost: Extreme Gradient Boosting, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2016), 785-794.

[13] M. Qi, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Proceedings of the 31st International Conference on Neural Information Processing Systems, (2017), 3149–3157.

[14] L. Breiman, Random Forest, Machine Learning, vol 45 (2001), 1-35.

ISSN: 2710-0588

Appendix A:

Attribute Name	Attribute Meaning	Vacancy value proportion	Processing Method
Users	Users Id	0.00000	Delete the Attribute
Satisfaction degree for surfing the internet with phones	Mobile Phone Online Service Satisfaction	0.00000	No vacancy value
Residential community	Whether users are in residential community?	0.00000	No vacancy value
Office	Whether users are in office?	0.00000	No vacancy value
Colleges and universities	Whether users are in colleges and university?	0.00000	No vacancy value
Commercial street	Whether users are in commercial street?	0.00000	No vacancy value
Subway	Whether users are in subway?	0.00000	No vacancy value
Rural area	Whether users are in rural area?	0.00000	No vacancy value
High-speed railway	Whether users are in high-speed railway?	0.00000	No vacancy value
Others, please specify	Whether users are in other situations?	0.00000	No vacancy value
Remarks for location	Other situations	0.94744	Delete the Attribute
Poor network signal or no network signal	Whether there is poor network signal or no network signal?	0.00000	No vacancy value
Failing to get online with a signal	Whether there is a failure to get online with a signal?	0.00000	No vacancy value
Intermittent network or fluctuating network speed	Whether there is intermittent network or fluctuating network speed?	0.00000	No vacancy value
Slow network speed for surfing the internet with phones	Whether there is slow network speed for surfing the internet with phones?	0.00000	No vacancy value
Others, please specify	Whether there are other situations?	0.00000	No vacancy value
Remarks for phenomena	Other situations	0.98002	Delete the Attribute
Standstills while watching videos	Whether there are standstills while watching videos?	0.00000	No vacancy value
Big delay in playing games	Whether there is big delay in playing games?	0.00000	No vacancy value
Slow speed for opening web pages or APP pictures	Whether there is slow speed for opening web pages or APP pictures?	0.00000	No vacancy value
Slow download speed	Whether there is slow download speed?	0.00000	No vacancy value

ISSN: 2710-0588

Relatively slow speed for mobile payments	Whether there is relatively slow speed for mobile payments?	0.00000	No vacancy value
Others, please specify	Other situations	0.00000	No vacancy value
Remarks for app categories	Categories of APPs	0.99046	Delete the Attribute
iQIYI	Whether users use iQIYI APP?	0.00000	No vacancy value
Youku	Whether users use iQIYI APP?	0.00000	No vacancy value
Tencent Video	Whether users use Tencent Video APP?	0.00000	No vacancy value
Mango TV	Whether users use Mango TV APP?	0.00000	No vacancy value
Sohu Video	Whether users use Sohu Video APP?	0.00000	No vacancy value
Tik Tok	Whether users use Tik Tok APP?	0.00000	No vacancy value
Kuaishou	Whether users use Kuaishou APP?	0.00000	No vacancy value
Volcano Video	Whether users use Volcano Video APP?	0.00000	No vacancy value
Migu Video	Whether users use Migu Video APP?	0.00000	No vacancy value
Others, please specify	Whether users use other software?	0.00000	No vacancy value
Remarks for video apps	Other software?	0.98433	Delete the Attribute
Standstills in all the games mentioned	Whether there are standstills in all the games mentioned?	0.00000	No vacancy value
Game for Peace	Whether users play Game for Peace APP?	0.00000	No vacancy value
Honor of Kings	Whether users play Game for Peace APP?	0.00000	No vacancy value
Cross Fire	Whether users play CrossFire APP?	0.00000	No vacancy value
Fantasy Westward Journey	Whether users play Fantasy Westward Journey APP?	0.00000	No vacancy value
Dragon Nest	Whether users play Dragon Nest APP?	0.00000	No vacancy value
Fantasy Zhu Xian	Whether users play Fantasy Zhu Xian APP?	0.00000	No vacancy value
Happy Fight against the Landlord	Whether users play Happy Fight against the Landlord APP?	0.00000	No vacancy value
Clash of Clans	Whether users play Clash of Clans APP?	0.00000	No vacancy value
Hearthstone	Whether users play Hearthstone APP?	0.00000	No vacancy value
Onmyoji	Whether users play Onmyoji APP?	0.00000	No vacancy value

ISSN: 2710-0588

Others, please specify	Whether users play other games?	0.00000	No vacancy value
Remarks for game apps	The name of other games.	0.99388	Delete the Attribute
Standstills in all games	Whether there are standstills in all the games mentioned?	0.00000	No vacancy value
WeChat	Whether users use WeChat?	0.00000	No vacancy value
QQ on mobile phones	Whether users use QQ on mobile phones?	0.00000	No vacancy value
Taobao	Whether users use Taobao APP?	0.00000	No vacancy value
JD.com	Whether users use JD.com on the phone?	0.00000	No vacancy value
Baidu	Whether users use Baidu APP?	0.00000	No vacancy value
Toutiao	Whether users use Toutiao APP?	0.00000	No vacancy value
Sina Weibo	Whether users use Sina Weibo APP?	0.00000	No vacancy value
Pinduoduo	Whether users use Pinduoduo APP?	0.00000	No vacancy value
Others, please specify	Whether users use other software?	0.00000	No vacancy value
Remarks for apps used for surfing the internet	The name of other software for surfing the internet.	0.98989	Delete the Attribute
Slow network speed for all web pages or apps	Whether there is Slow network speed for all web pages or apps above?	0.00000	No vacancy value
Times of experiencing poor internet access quality	Times of experiencing poor internet access quality	0.87379	Fill the vacancy value with 0
Disconnection times	Disconnection times	0.84388	Fill the vacancy value with 0
Number of redirections	Number of redirections	0.54188	Fill the vacancy value with 0
2G resident duration	2G resident duration	0.54188	Fill the vacancy value with 0
Times of experiencing poor internet access quality while using WeChat	Times of experiencing poor internet access quality while using WeChat	0.81311	Fill the vacancy value with 0
Times of experiencing poor internet access quality while using Honor of Kings	Times of experiencing poor internet access quality while using Honor of Kings	0.95143	Fill the vacancy value with 0
High unit price customer with consumption	Whether there is high unit price customer with consumption exceeding the package (group)	0.00000	No vacancy value

ISSN: 2710-0588

exceeding the package (group)			
Customer with frequent and massive consumption exceeding the package (group)	Whether there is customer with frequent and massive consumption exceeding the package (group)	0.00000	No vacancy value
Out-of-package traffic	Out-of-package traffic	0.00000	No vacancy value
Charges for out-of-package traffic	Charges for out-of-package traffic	0.00000	No vacancy value
Whether it is a full-month roaming user	Whether it is a full-month roaming user?	0.00000	No vacancy value
Whether it is a user of unlimited package who reaches the threshold	Whether it is a user of unlimited package who reaches the threshold?	0.00000	No vacancy value
Age	Age	0.00000	No vacancy value
Gender	Gender	0.00000	No vacancy value
Usage times of Honor of Kings	Usage times of Honor of Kings	0.05897	Fill the vacancy value with 0
Days of use of game apps	Days of use of game apps	0.05897	Fill the vacancy value with 0
Usage times of game apps	Usage times of game apps	0.05897	Lagrange polynomial
Days of use of Honor of Kings	Days of use of Honor of Kings APP	0.05897	Lagrange polynomial
Game app usage traffic	Game app usage traffic	0.05897	Fill the vacancy value with 0
Tik Tok usage traffic (MB)	Tik Tok APP usage traffic (MB)	0.00000	No vacancy value
Toutiao usage traffic	Toutiao APP usage traffic	0.05897	Fill the vacancy value with 0
Kuaishou usage traffic	Kuaishou APP usage traffic	0.05897	Fill the vacancy value with 0
Youku Video usage traffic	Youku Video APP usage traffic	0.05897	Fill the vacancy value with 0
Tencent Video usage traffic	Tencent Video APP usage traffic	0.05897	Fill the vacancy value with 0
Usage traffic of video apps	Usage traffic of video apps	0.00000	No vacancy value
Usage traffic of Ali apps	Usage traffic of Ali apps	0.00000	No vacancy value

ISSN: 2710-0588

Usage traffic of Netease apps	Usage traffic of apps developed by Netease Company	0.00000	No vacancy value
Usage traffic of Tencent apps	Usage traffic of apps developed by Tencent Company	0.00000	No vacancy value
Usage traffic of Honor of Kings	Usage traffic of Honor of Kings APP	0.00000	No vacancy value
Usage traffic of Dragonfly FM	Usage traffic of Dragonfly FM APP	0.00000	No vacancy value
Usage traffic of Ele.me	Usage traffic of Ele.me APP	0.00000	No vacancy value
Usage traffic of Meituan takeout	Usage traffic of Meituan takeout APP	0.00000	No vacancy value
Usage traffic of Tmall	Usage traffic of Tmall APP	0.00000	No vacancy value
Usage traffic of Dianping.com	Usage traffic of Dianping.com on the phone	0.00000	No vacancy value
Usage traffic of Didi Chuxing	Usage traffic of Didi Chuxing APP	0.00000	No vacancy value
Usage traffic of communication apps	Usage traffic of communication apps	0.00000	No vacancy value
Usage traffic of game apps	Usage traffic of game apps	0.00000	No vacancy value
Usage traffic of web apps	Usage traffic of web apps	0.00000	No vacancy value
Usage traffic of music apps	Usage traffic of music apps	0.00000	No vacancy value
Usage traffic of video apps	Usage traffic of video apps	0.00000	No vacancy value
Usage traffic of mailbox apps	Usage traffic of mailbox apps	0.00000	No vacancy value
Terminal type	Terminal type	0.00199	Fill vacancy values with mode
Operating system	Operating system	0.27992	Fill vacancy values with mode
Terminal system	Terminal system	0.00000	No vacancy value
Terminal brand	Terminal brand	0.00199	Fill vacancy values with mode
Terminal brand type	Terminal brand type	0.00199	Fill vacancy values with mode
Whether it is a 5G network user	Whether it is a 5G network user?	0.00000	No vacancy value
GPRS Resource Usage of the Month (GB)	GPRS Resource Usage of the Month (GB)	0.00000	No vacancy value
Whether it is a campus card package user	Whether it is a campus card package user	0.00000	No vacancy value

ISSN: 2710-0588

Whether it is a campus card user without a campus contract	Whether it is a campus card user without a campus contract?	0.00000	No vacancy value
Whether it is a campus card contracts bundle user	Whether it is a campus card contracts bundle user?	0.00000	No vacancy value
Branch of high frequency communication of the month	Branch of high frequency communication of the month	0.00228	Fill vacancy values with mode
Grade of the free package	Grade of the free package	0.00000	No vacancy value
Name of the free package	Name of the free package	0.00000	No vacancy value
Main package grade	Main package grade	0.00000	No vacancy value
MOU of the month	MOU of the month	0.00000	No vacancy value
Average consumption in recent 3 months (excluding communication account payment)	Average consumption in recent 3 months (excluding communication account payment)	0.00171	Fill vacancy values with average value
Average consumption in recent 3 months (Yuan)	Average consumption in recent 3 months (Yuan)	0.00000	No vacancy value
Cumulative consumption this Year (Yuan)	Cumulative consumption this Year (Yuan)	0.00000	No vacancy value
Code number resource - activation time	Code number resource - activation time	0.59117	Delete the attribute
Code number resource - card issuing time	Code number resource - card issuing time	0.57037	Delete the attribute
Customer star mark	Customer star mark	0.00000	No vacancy value

ISSN: 2710-0588

Appendix B:

Influencing factors	Mutual Information Value
2G resident duration	0.165893
Number of redirections	0.164322
Poor network signal or no network signal	0.149200
Slow network speed for surfing the internet with phones	0.128705
Intermittent network or fluctuating network speed	0.127598
Residential community	0.115794
Slow speed for opening web pages or APP pictures	0.113065
Failing to get online with a signal	0.107199
Office	0.087986
Subway	0.085290
Standstills while watching videos	0.078200
GPRS Resource Usage of the Month (GB)	0.076922
Toutiao usage traffic	0.070585
Kuaishou usage traffic	0.067172
Tencent Video usage traffic	0.066264
Tik Tok usage traffic (MB)	0.065870
Usage traffic of video apps	0.064584
Usage traffic of Ali apps	0.064112
Usage traffic of communication apps	0.062302
Youku Video usage traffic	0.061948
Game app usage traffic	0.061910
Cumulative consumption this Year (Yuan)	0.061705
Usage traffic of web apps	0.061207
Age	0.061199
Usage traffic of video apps	0.059687
Average consumption in recent 3 months (Yuan)	0.058911
Average consumption in recent 3 months (excluding communication account payment)	0.058618
Usage times of game apps	0.057721
Usage traffic of Tmall	0.056087
Usage traffic of Tencent apps	0.055294
Disconnection times	0.053349
Slow download speed	0.052995
WeChat	0.051821
Big delay in playing games	0.050392
Usage traffic of Netease apps	0.04901
Usage traffic of music apps	0.048222
Relatively slow speed for mobile payments	0.048038
Slow network speed for all web pages or apps	0.046843
MOU of the month	0.046157
Usage times of Honor of Kings	0.045794
Commercial street	0.041674
Usage traffic of Honor of Kings	0.040881
High-speed railway	0.039499
Standstills in all the games mentioned	0.036963

ISSN: 2710-0588

Out-of-package traffic	0.036591
Usage traffic of game apps	0.036256
Taobao	0.035423
Usage traffic of Didi Chuxing	0.034752
JD.com	0.029096
Terminal brand	0.028589
Usage traffic of Dianping.com	0.027676
Charges for out-of-package traffic	0.027458
Rural area	0.026768
Tik Tok	0.025914
Usage traffic of Ele.me	0.025445
Honor of Kings	0.024290
Tencent Video	0.023996
Baidu	0.023965
iQIYI	0.021369
Main package grade	0.020792
Youku	0.020610
Colleges and universities	0.020396
Name of the free package	0.019403
Standstills in all games	0.018751
Operating system	0.018579
Game for Peace	0.016912
Days of use of Honor of Kings	0.016662
Sina Weibo	0.016470
Days of use of game apps	0.014805
Customer star mark	0.014684
QQ on mobile phones	0.012757
Toutiao	0.011981
Times of experiencing poor internet access quality while using Honor of Kings	0.011577
Grade of the free package	0.011467
Others, please specify	0.010892
Times of experiencing poor internet access quality while using Wechat	0.009973
Kuaishou	0.009695
Usage traffic of Dragonfly FM	0.009428
Pinduoduo	0.009030
Mango TV	0.008612
Usage traffic of mailbox apps	0.007964
Times of experiencing poor internet access quality	0.007524
Migu Video	0.006472
Branch of high frequency communication of the month	0.006385
Sohu Video	0.006178
Whether it is a full-month roaming user	0.005847
Volcano Video	0.005440
Fantasy Westward Journey	0.004607
Clash of Clans	0.003958
Cross Fire	0.003893

ISSN: 2710-0588

Happy Fight against the Landlord	0.003881
Dragon Nest	0.003854
High unit price customer with consumption exceeding the package (group)	0.003705
Gender	0.003287
Onmyoji	0.003091
Hearthstone	0.002908
Whether it is a campus card package user	0.002859
Fantasy Zhu Xian	0.002809
Whether it is a user of unlimited package who reaches the threshold	0.002628
Whether it is a campus card contracts bundle user	0.002304
Whether it is a 5G network user	0.001726
Whether it is a campus card user without a campus contract	0.001545
Terminal system	0.001015
Customer with frequent and massive consumption exceeding the package (group)	0.001004
Terminal type	0.000927
Usage traffic of Meituan takeout	0.000891