2021 International Conference on Networking, Communications and Information Technology (NetCIT) | 978-1-6654-0070-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/NetCIT54147.2021.00048

Predicting Telecommunication User Favorite Package by Using Deep Neural Network

Tingshun Li*
School of Control and Computer Engineering
North China Electric Power University
Beijing, 102206, China
lits@ncepu.edu.cn

Huiyu Yang
China Power Information Technology Co., Ltd
Beijing, China
13811330386@163.com

Dadi Wang
China Power Information Technology Co., Ltd
Beijing, China
Zkht888888@163.com

Zesan Liu
R & D,
State Grid Information & Telecommunication Group Co.
Beijing 102211, China
liuzesan@sgitg.sgcc.com.cn

***Abstract*—With the popularity of mobile phone, telecom companies have launched more and more package services. It is very difficult to choose telecom user package suitable for him or her. This paper provides a method to build a model based on deep neural network (DNN) for multi-classification, which has high accuracy to help user to pick out a favorite one from dozens of packages. The model is trained out from telecom user big data. First, the telecom big data are preprocessed to be completed integrity, normalized, balanced., and then divided into training data set, validating data set and testing data set. Second, feature engineering needs to be done in order to improve prediction model. Two types of feature engineering are presented in this work to compare which is better. Manual feature engineering is a way to build new features from expertise, while auto feature engineering is from third-party library. Third, an experimental process is design out to obtain prediction models, and use them to predict the testing data set. Finally, some conclusions are obtained by analyzing the experimental results. So, this paper proposes a multi-classification model to recommend a suitable package to a telecom user, which is trained out from telecom big data with high accuracy, more practical. Moreover, this work show that feature engineering and data preprocessing are helpful to obtain better machine learning model.**

Keywords—DNN, telecom package, feature engineering, data preprocessing

I. Introduction

Telecom package is the major business of telecom company to deal with fierce market competition. In order to meet the requirements of different user groups, telecom companies have launched a great number of telecom packages [1, 2]. For example, China Mobile, as one of the three largest telecom companies in China, has 7 telecom packages for 5G individual users only in Beijing, not including for family in 2021 [3]. The similar situations also happened in many countries [4,5]. The number of Telecom packages is so great that it is very difficult for users to select one suitable and economical for him or her. The information is overload for users. Moreover, telecom package suitable for someone should be changed with the situation of him or her. On the other hand, telecom companies use so many packages to confuse users to gain more than profits, and it's so common and disgusting that recent a decree has been promulgated to improve telecom package services [6]. So, recommending and presenting a suitable package for a user is a critical issue.

Nowadays, machine learning is more and more popular, and there are so many cases in the industrial field which show that the models trained from machine learning are helpful to improve products or services [7]. Therefore, some researchers try to use machine learning to solve this issue. For instance, in 2016, Liu SS et al. [1] provides a new idea of automatic discovery and recommendation for telecom package. Jing du [8] designs and implements a package recommendation fusion model based on XGBoost and LightGBM by focusing on analysis of user behavior data, and the accuracy of the model is 91%. It is the key issue to obtain a large number of actual behavior data, and not easy. Wen Wang in [9] studies several telecom package recommendation models for different kinds of data based on deep learning using actual data provided by China Unicom, forming broad aspects to improve recommendation. [10] proposes an improved stacking algorithm for package recommendation, focusing on prediction for unbalance packages. While some researchers apply telecom data in others fields. For example, Younis, S., and A. Ahsan focus on detecting employee retention by social network analysis of telecom industry [11,12]. creatively researches on public health with the help of telecommunications networks during the COVID-19 pandemic. Tabassum, S et al. in [13] propose to profile high leverage points for detecting anomalous users by using telecom data.

This work key point is how to accurately predict a user favorite telecom package from dozens of ones by building model based on DNN, which is trained out from actual big data provided by telecom company. Therefore, this paper is organized as follows. The second section describes the data preprocessing, which involves data source description, data integrity processing, balancing data, normalizing data and dividing the data into training data set and testing data set. The third section describes feature engineering. In order to obtain a model with high accuracy, the characteristics in original data must be restructured. So, this section includes feature analysis and feature engineering. The forth section describes how to build multi-classification model with DNN, and then train the model in the training data set and check the model in the testing data set. The better model will be gained by comparing experimental results. Finally, the conclusion section describes the advantages and disadvantages of the prediction model based on DNN, and limitations. In particular, when applying this approach in the practical environment, more data preprocessing and better feature engineering are required considering the complexity of the data.

978-1-6654-0070-1/21/$31.00 ©2021 IEEE
DOI 10.1109/NetCIT54147.2021.00048

II. Data preprocessing

A. Data Source

Some researchers said that data are going to be so important a material as oil. A better model from machine learning is determined by bigger data. So, a data set corresponds a telecom in China, which includes more than 700000. The data are used as sample to train a model from machine learning algorithm. Their distributions are presented in Fig. 1.

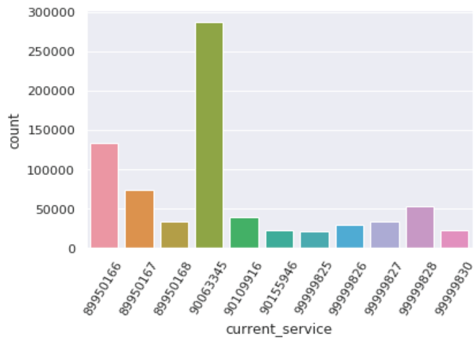

Fig. 1. Distributions of Telecom User Packages

The data sources have 27 measurements, as shown in Table I.

Table I. Original features

SN	Feature	Memo
1	USERID	User code
2	current_type	
3	service_type	0:include 2G and 3G 1:2I 2C; 2:2G; 3:3G 4:4G
4	is_mix_service	1:yes; 0:no
5	online_time	
6	1_total_fee	
7	2_total_fee	
8	3_total_fee	
9	4_total_fee	
10	month_traffic	
11	many_over_bill	1:yes; 0:no
12	contract_type	User contract type
13	contract_time	/
14	is_promise_low_consume	1.yes 0.no
15	net_service	
16	pay_times	
17	pay_num	
18	last_month_traffic	
19	local_trafffic_month	
20	local_caller_time	
21	service1_caller_time	
22	service2_caller_time	
23	gender	01.male; 02:female
24	Age	/
25	complaint_level	1:general; 2: critical; 3:Major
26	former_complaint_num	
27	former_complaint_fee	

These data are gathered by some equipment, programs or staff in the telecom company. So, there are some of the data not very complete, which need to be processed before training.

B. Data Integrity Processing

In order to obtain a good model with high accuracy, the training data are integrity. However, there are some uncomplete data in the above data for a variety of practical reasons. There are two typical kinds of incomplete data: missing data and singular data. As for those incomplete data, there are two basic methods to process: deleting or filling, and a better way is case by case.

By checking one by one, there are only 3 missing data, and they are too few. Even without them, the result will not be changed a little. So, the 3 missing data should be deleted.

As for singular data, different characteristics have different abnormal values. So, feature variable must be analyzed one by one, and then replace the singular value with average value or default value. For example, there are some data with 0 age, which are absurd. Therefore, it is a reasonable choice to replace the value of age with 0 with average age.

C. Balancing Data

The distribution of telecom user package in Fig. 1 indicates that the number of package user is very imbalance. This imbalance can have a significant impact on the model [14]. In general, the greater the quantity in data, the greater the weight in model. So, it is necessary to balance data with the number of user package.

In this work, the methods of fixed length undersampling and SMOTE algorithm oversampling are employed for large and small samples, respectively [15]. An average number is adopted by 50000. More than 50000 user package data such as 90063345 are reduced to a level of approximately 50,000 samples, while others less than 50000 ones are increased to 50000, which represents a ratio that is close to 1:1.

D. Normalizing Data

Feature normalization is a basic method in data preprocessing, which can reduce the impact of the different domains of different feature variables because normalizing can make them have the same domain. After normalizing, the sample data may be used to acquire a model which has a high accuracy.

There are two kinds of normalization [16]:

1) Standardization (z-score normalizing)

The formula is shown as below.

$$x^* = \frac{x-\bar{x}}{\sigma} \tag{1}$$

Here, μ is mean value of variable and σ is standard deviation.

2) Min-Max Scaling

The formula is presented as below

$$x^* = \frac{x - min}{max - min} \quad (2)$$

Here, min is the minimum of variable and max is the maximum.

Here, standardization is adopted in this work.

E. Dividing Data

So far, only one data set has been obtained. It means that a prediction model needs to be gained by training from the data set, which also needs to be tested by using the same one. As we known, a machine learning model can generally fit the train data set very well. Therefore, if a data set acts as the training data set and also as the testing data set, the prediction model will be like it hasn't been tested. It's better that the training data set is different from the testing data set in order to get a better model.

[17] tells how one data set is reasonably divided into training data set and testing data set. The data set in this work is divided into training set and test set by 7:3, and then the train set is divided into training set and validation set. The training data set is used to train a prediction model. When training, the validation data set is used to verify the model. After training, the testing data set is used to test the model.

III. Methods

A. Feature Engineering Method

1) Feature Analysis

There are 27 feature variables presented in above Table I. Some of them closely relate with telecom user package classifying, while the others don't do. So, it needs to analyze the features.

As we known, user must sign a contract with telecom before he/she enjoys telecom package. So, the contract types need to be firstly analyzed. There are 13 contract types as shown in Table II.

TABLE II. CONTRACT TYPES

Contract code	Contract description
0	Prepaid mobile phone
1	Call fee free in a period with purchasing mobile phone
2	Deposit fee and delivery fee
3	Single card
4	Other activities at headquarters
5	Deposit and delivery service (Complimentary voice)
6	Deposit and delivery service (Complimentary data)
7	Deposit and delivery service (Complimentary text message)
8	Deposit and delivery service (Complimentary unknow)
9	Deposit and delivery gifts
10	Deposit and delivery credits
11	Contract machine
12	Pre-purchase business and delivery mobile phone

The quantity distribution of user contract types is presented in Fig. 2.

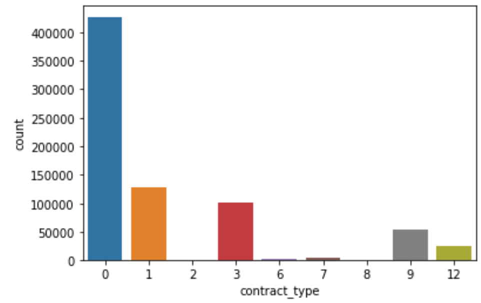

Fig. 2. Quantity distribution of contract

Fig. 2 shows that the number of No. 0 type contract is the greatest, while the quantities of No.2, 6, 7, 8, 10, 11 are the least, nearly 0, which indicates that the quantity gap of different contract is very huge.

Different generations have different requirements in mobile tele communication, which means that people of different ages cloud buy different telecom user packages. Age can affect choice of user package.

Fig. 3 is the distribution of age in the data set, which shows that the majority of user is between 20 to 40, more than 50%. There are some singular data with 0 age, which need to be processed.

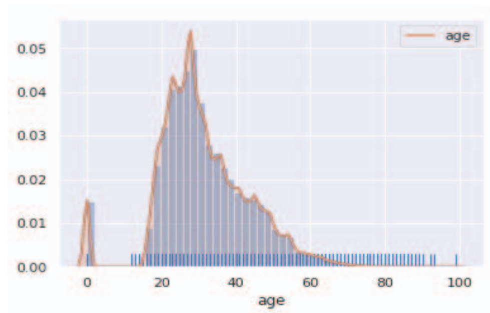

Fig. 3. Distribution of age

Different telecom packages are essentially a kind of discount for users with different consumption ability. To some extent, the consumption level of user determines their choice of package. The consumption of some month is by chance, so it is reasonable to calculate someone's average consumption in continuous 4 months to represent his or her consumption level. The distribution of user average consumption in continuous 4 months is presented in Fig. 4.

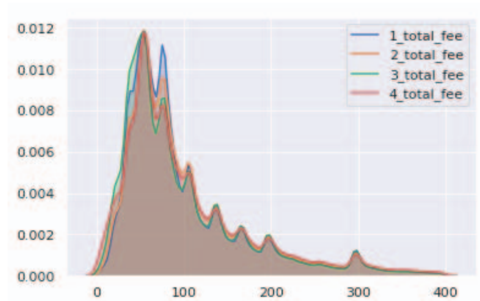

Fig. 4. Distribution of user average consumption in continuous 4 months

In Fig. 4, it is easily found that most consumptions of user are below 100.

Since someone's actual telecom services persistently overflow his/her package service, his/her package is not suitable for him/her, and it's necessary to recommend him/her new one. It is essential to complete statistics of continuous exceeding package service, and shown in Fig. 5.

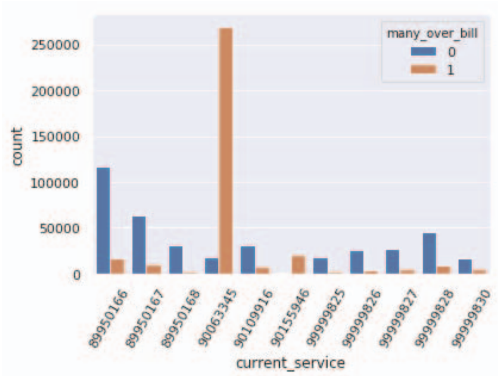

Fig. 5. Comparison between exceeding service of different packages or not

0, blue column is the number of users without exceeding package service
1, orange column is the number of users with exceeding package service

It indicates in Fig. 5 that most actual services of the users with the 90063345 and 90155946 packages exceed their package services, and need to recommend them new service packages.

Certainly, other feature variables such as gender are analyzed like the above features.

2) Feature engineering

The performance of the prediction model mainly depends on the quality of the features in the data set used to train the model. Generally, we will use various algorithms to infer the correlation of features in the data set, and then create new features to help provide more information about target variables to the model, so as to improve the training effect of the model. However, if there are not many high-quality features in our data set, feature engineering is often a good choice.

Feature engineering mainly constructs the original features in the data set into explanatory new variables (features), which can be used to train machine learning models to predict [18]. In this study, manual feature engineering and automatic feature engineering will be used to process the data.

a) Auto feature engineering: Feature engineering can be done by using third party providing function library. For example, "Featuretools" for python, which includes many functions such as "sum", "std", "max", "skew", "min", "mean", "count", "percent_true", "n_unique", "mode". At the help of the third-party library, some new features can be restructured out auto, instead of manual calculating by designing algorithm.

Three groups of new features are composed of different functions which are provided by the third-party library, named by fm1, fm2, fm3 respectively, as presented below.

i. fm1, composed of functions such as "sum", "std", "max", "skew", "min", "mean", "count", "percent_true", "n_unique", "mode".
ii. fm2, composed of functions such as "sum", "std", "max", "skew", "min", "mean".
iii. fm3, composed of functions such as "std", "max", "min", "mean".

Although auto feature engineering is very convenient, the functions is limited. Sometimes, some new features need to be built with the actual requirements, which can reflect practice or expertise, and maybe have a better effect.

b) Manual feature Engineering: When manual feature engineering, discrete data and continuous data are processed respectively. There are 8 discrete feature variables, and they are 'service_type', 'complaint_level', 'contract_type', 'gender', 'is_mix_service','is_promise_low_consume', 'many_over_bill', 'net_service'. While others are continuous, they are '1_total_fee', '2_total_fee', '3_total_fee', '4_total_fee','age', 'contract_time', 'former_complaint_fee', 'former_complaint_num', 'last_month_traffic', 'local_caller_time', 'local_trafffic_month', 'month_traffic', 'online_time', 'pay_num', 'pay_times', 'service1_caller_time', 'service2_caller_time'. They are processed with manual feature engineering, presented in Table III as below.

TABLE III. NEW FEATURES AFTER MANUAL ENGINEERING

New Feature	Meaning	Calculation method
max_total_fee	Maximum of 4 continuous months	max(1_total_fee, 2_total_fee, 3_total_fee, 4_total_fee)
min_total_fee	Minimum of 4 continuous months	min(1_total_fee, 2_total_fee, 3_total_fee, 4_total_fee)
mean_total_fee	Average of 4 continuous months	mean(1_total_fee, 2_total_fee, 3_total_fee, 4_total_fee)
diff_total_fee_1	Difference fee between current month and previous one	data['diff_total_fee_1'] = data['1_total_fee'] - data['2_total_fee']
diff_total_fee_2	Difference fee between current month and last one	data['diff_total_fee_2'] = data['2_total_fee'] - data['3_total_fee']
diff_total_fee_3	Difference fee between current month and the month before 2	data['diff_total_fee_3'] = data['3_total_fee'] - data['4_total_fee']
pay_num_1_total_fee	Difference between current month payment and fee	data['pay_num_1_total_fee'] = data['pay_num'] - data['1_total_fee']
rest_traffic_ratio	Current month data consumption	data['rest_traffic_ratio'] = \ (data['last_month_traffic_rest'] / 1024) / data['1_total_fee']
total_caller_time	Total call cost	data['total_caller_time'] = data['local_caller_time']+data['service2_caller_time'] + data['service1_caller_time']
local_caller_ratio	Proportion of local calling duration	data['local_caller_ratio']= data['local_caller_time'] / data['total_caller_time']
service1_caller_ratio	Proportion of call duration of external package in last month	data['service1_caller_ratio']= data['service1_caller_time'] / data['total_caller_time']
ervice2_caller_ratio	Proportion of call duration of external package in last 2 months	data['service2_caller_ratio']= data['service2_caller_time'] / data['total_caller_time']
1_total_fee_call_fee	Call cost of over package in last month	data['1_total_fee_call_fee'] = data['1_total_fee'] -

New Feature	Meaning	Calculation method
		data['service1_caller_time'] * 0.15
1_total_fee_call2_fee	Call cost of over package in last 2 months	data['1_total_fee_call2_fee'] = data['1_total_fee'] - data['service2_caller_time'] * 0.15

B. Classification Labeling

The codes of user package classification are composed of characters, such as "89950166", "90063345", "90155946", which is unsuitable for classifying algorithm from machine learning. These codes need to be changed into computable numbers.

One of simple methods is encoding the labels, corresponding the labels to 0-10 numbers in proper order. This way is too simple to process data in a matrix manner. For fast processing data in the manner of matrix, a 11-dimension label data need to be built up with one label corresponding to 1 dimension. That one dimension is 1 means that the classification of data is the label corresponding to the dimension, and others must be 0. So, do as this, the labels can be turned into matrix, so that the next step of training model can be done in the manner of matrix.

C. Multi-Classification Model based on DNN

Before training from the training data set, a model based on DNN need to be designed out by configuring parameters, such as the number of layers, especial the number of hidden layers, picking fit function for each layer, matching the dimensions of the data, and so on [19].

A practical model based on DNN is set up in below Fig. 6.

D. Experimental Process

The experimental process is presented as in Fig. 7. Firstly, the telecom data set are prepared. Secondly, feature engineering is done to create some new features. And then the data are processed according to new features. Thirdly, the data are preprocessed, including integrity processing, normalizing, balancing and dividing the data into training data, validating data and testing data. Fourthly, a prediction model is designed, and then gained by using training data and validation data. Finally, the prediction model is transferred to testing data to predict, and calculate the accuracy of prediction model on the test data.

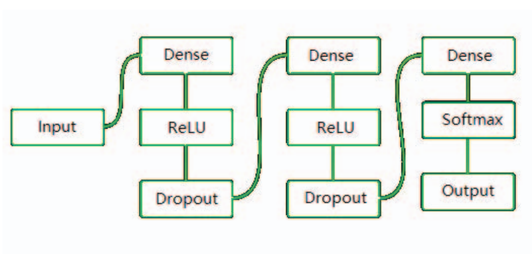

Fig. 6. Prediction model based on DNN

IV. EXPERIMENTAL RESULTS

After building the original model shown in above Fig. 6, a final prediction model should be obtained by training on the training data set. And the prediction model must be used to predict the testing data set, and then the prediction model will be measured via comparing the actual results and prediction results, one of main indicators is prediction accuracy.

The next sections are the experimental results comparisons between original features, features after engineering (including manual feature engineering and auto), different functions parameters.

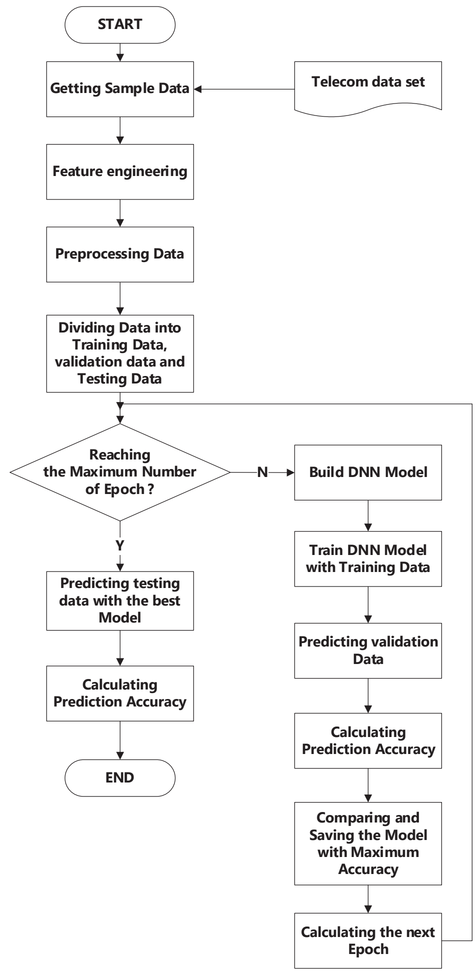

Fig. 7. Experiment process

A. Comparing between Original Features and Engineering Ones

First, comparisons between original features and engineering ones show whether feature engineering is necessary.

Two different prediction models are gained by using original training data set and processed training data set with manual feature engineering, respectively. The change curves of the two-prediction model's accuracy on the training data are shown in below Fig. 8.

From Fig. 8, it is not difficult to find that the accuracy of the prediction model with manual feature engineering is far higher than the accuracy of original one. It means that manual feature engineering can help to improve prediction model accuracy.

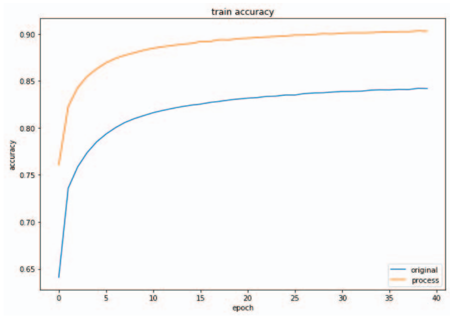

Fig. 8. Model prediction accuracy comparison curve between manual feature engineering model and original one

The label of "original" is the accuracy curve of original feature prediction model

The label of "process" is the accuracy curve of manual feature engineering prediction model

B. Comparing between Manual Feature Engineering and Auto One

There are two types of feature engineering: manual feature engineering and auto feature engineering with the assistant of the third-party library. Since manual feature engineering is helpful to improve prediction model, it is necessary to test which feature engineering is better by experiments.

Above three groups of new features via auto feature engineering (fm1, fm2, fm3) are used to build prediction respectively, and to compare with manual feature engineering prediction model, as shown in Fig. 9.

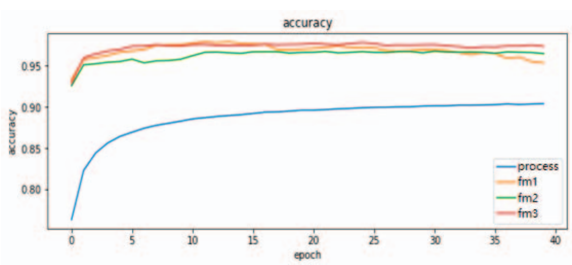

Fig. 9. Model prediction accuracy comparison curve between auto feature engineering and manual one

The label of "process" is the accuracy curve of manual feature engineering prediction model

The label of "fm1" is the accuracy curve of auto feature engineering named fm1 prediction model

The label of "fm2" is the accuracy curve of auto feature engineering named fm2 prediction model

The label of "fm3" is the accuracy curve of auto feature engineering named fm3 prediction model

Fig. 9 indicates that the accuracies of auto feature engineering prediction model are better than manual feature engineering. It means that the new features in Table III by manual engineering are not the best option. The curves of fm1, fm2 and fm3 are nearly close, which shows that fm3 owning least features can replace the other two, and the features in fm3 group play key role of training model.

C. Comprehensive Comparison

Final, all these models are applied to the testing data set to test which model will perform best. The experimental results are presented in Table IV.

TABLE IV: PREDICTION ACCURACY OF DIFFERENT MODEL

Model	Acc
Original feature model	0.8687
Manual feature engineering model	0.9102
fm1	0.9551
fm2	0.9649
fm3	0.9752

By comparing the results in Table IV, some conclusions are obtained: 1) The result of feature engineering is better than original features; 2) The result of auto feature engineering is superior to manual feature engineering, which indicates that some third-party libraries are helpful, and manual building features from expertise are not necessary the best option; 3) The result of fm3 with the least features is the best among the three groups of auto feature engineering, which means that it is not that the more features the better, but the more precision the better.

V. CONCLUSIONS

In many practical application fields, multi classification algorithm from machine learning is helpful and effective. In this study, in order to gain an effective prediction model for telecom user package, original features in the data set are processed with feature engineering., and the data are preprocessed such as normalization, balance, integrity, before training model with them. The conclusions of this work are as follows:

a) It is feasible to build a prediction model for multi-classification by using machine learning. There are many kinds of telecom user packages so that it is very difficult to recommend someone a package suitable for him or her. It is to obtain a far higher accuracy by training out a prediction model from a bigger telecom user data than by manual blind recommendation.

b) It's a better choice to preprocess data before training model. The actual data are not integrity, the size of those features is very different, the quantity varies great. These factors can affect the prediction model.

c) Feature engineering is necessary to improve prediction model. The experimental results show that feature engineering is effective to gain higher accuracy than original features. Furthermore, it is not correct that the more features the better model, but the more precision the better.

d) It is firmly admitted that a more suitable model will be obtained when a group of better features are structured out, which are more closely correlated with telecom user package, or some improved machine learning algorithm to be research out. Furthermore, the prediction model is trained out from the telecom user data, so it is necessary to be changed with the telecom user data renewed, especially when a new user package is launched by telecom company.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China under grant no. 61573138.

REFERENCES

[1] S. Liu, Y. Bo, W. Lin, et al., "Automatic Discovery and Recommendation for Telecommunication Package Using Particle Swarm Optimization", *Advances in Nature and Biologically Inspired Computing*, vol. 419, pp. 97-104, 2016.

[2] L. Xia. "Analysis on Mobile Internet Technology Innovation", *Telecom Power Technology*, vol. 38, no. 04, pp. 197-199, 2021.

[3] China Mobile, "5G Individual Package in Beijing in Mobile Shop", https://shop.10086.cn/goods/100_100_1077184_1065532.html?WT.ac=hot_search_008, 2021-12-14.

[4] A. Agarwal, "TRENDS SHAPING INDIAN TELECOM." Voice & Data, vol. 26, no. 3, pp. 37-39, 2019.

[5] P. Dinham, "MOBILE DATA BIG CONTRIBUTOR TO AUSTRALIAN TELECOM GROWTH." *Exchange AUG,* vol. 8, pp. 7-7, 2019.

[6] J. Zhang. "Telecom Package cannot Become 'Package Fraud'", http://www.nbd.com.cn/articles/2021-11-07/1984231.html, 2021-12-14.

[7] J. Lee, "Integration of Digital Twin and Deep Learning in Cyber-Physical Systems: Towards Smart Manufacturing.". *Manufacturing Letters*, vol. 27, no. 1, pp. 87-91, 2021.

[8] J. Du. "The Design and Implementation of Telecom Package Recommendation System Based on Machine Learning", *Zhongnan University of Economics and Law*, 2020.

[9] W. Wang. "Research on Telecom Package Recommendation Model based on Deep Learning", *Kunming University of Science and Technology*, 2020.

[10] Z. Bao, X. Hu, Y. Zhao, et al., "Prediction Analysis on Telecom Package based on Improved Stacking Algorithm", *Journal of Xi'an University of Posts and Telecommunications,* vol. 24, no. 02, pp. 98-104, 2019.

[11] S. Younis, and A. Ahsan. "Know Your Stars Before They Fall Apart: A Social Network Analysis of Telecom Industry to Foster Employee Retention Using Data Mining Technique." IEEE Access, vol. 99, pp. 1-1, 2020.

[12] J. Rowsell, A. Hertanto, and A. Mathur. "Telecommunications networks and public health responses during the COVID-19 pandemic: Evidence from a large national network operator in Canada." 23rd ITS Biennial Conference, Online Conference / Gothenburg 2021. Digital societies and industrial transformations: Policies, markets, and technologies in a post-Covid world International Telecommunications Society (ITS), 2021.

[13] S. Tabassum, M. A. Azad, and J. Gama. "Profiling high leverage points for detecting anomalous users in telecom data networks." annals of telecommunications-annales des télécommunications, vol. 447, 2020.

[14] N. Basurto, et al. Data Balancing to Improve Prediction of Project Success in the Telecom Sector. 2021.

[15] J. Xu, W. Tan, T. Li, "Predicting fan blade icing by using particle swarm optimization and support vector machine algorithm", *Computers and Electrical Engineering*, vol. 87, no. 1, pp. 106751(1-11), 2020.

[16] H. Li, "Statistical Learning Methods", *Tsinghua University Press,* 2012.

[17] N. Feng, D. Fang, J. Xie, "Method of data splitting for sample set based on genetic algorithms", *Computer Engineering and Applications*, vol. 44, no. 16, pp. 129-131, 2018.

[18] Y. Tang, S. Sun, "Feature Engineering Methods in Machine Learning", *Automobile Technology*, no. 12, pp. 70-72, 2020.

[19] A. A. Awan, et al. "Communication Profiling and Characterization of Deep-Learning Workloads on Clusters With High-Performance Interconnects." IEEE Micro, vol. 40, no. 1, pp. 35-43, 2020.