

New Methods of Customer Segmentation and Individual Credit Evaluation Based on Machine Learning

Zhou Yuping
Wuhan University of Technology Wuhan University of
Technology
Wuhan 430070, Peoples R China
842118147@qq.com

Chen Guanyu
Wuhan University of Technology Wuhan University of
Technology
Wuhan 430070, Peoples R China
771579266@qq.com

Petra Jílková
Masaryk Institute of Advanced Studies Czech Technical
University in Prague
Prague, Czech Republic
petra.jilkova@cvut.cz

David Weisl
Masaryk Institute of Advanced Studies Czech Technical
University in Prague
Prague, Czech Republic
david.weisl@cvut.cz

Abstract—The internet has enabled a fundamental change in consumer behaviour and their understanding of e-commerce business. The main objective of the following article is to present the latest trends in the way of client segmentation associated with individual credit evaluation based on machine learning. The first part discusses the current situation and innovations in the way people pay in an omnichannel world. We describe how the absence of physical money has affected society, how it has changed customer purchasing behaviour, and what this change means for the digital economy and marketing. In the background of the rapid development of big data and the Internet technology, the traditional personal credit evaluation method of the commercial bank faces a significant challenge in the evaluation of personal credit. Based on the limitation of the existing personal credit evaluation method, the second part discusses the necessity of the research on the personal credit evaluation based on the machine learning method and then probes into the comprehensive personal credit evaluation dimension and the advanced data acquisition method of the Internet finance company. And then, the data desensitisation and LOF test were carried out by dynamic desensitisation technique. The abnormal value of the tested data and the random forest method supplement the missing value of the data. The importance index is screened by the gradient boosting decision tree method, and the personal credit evaluation score is output through the scorecard model based on logical regression. After that, the model is tested by BP neural network, and the personal credit level is predicted. The personal credit level fosters customer market segmentation.

Keywords: customer segmentation, digital payments, BP neural network, machine learning, personal credit score

I. INTRODUCTION

The digital revolution has changed the business model and brings enormous digital opportunities and challenges in today's rapidly changing environment. Data are the new economic resource for creating economic value, brings added strength and competitive advantage. Over the past several decades, the communication technology sector and payment systems have undergone rapid development. There is an enormous shift from standard economy to the birth of the

computer in the 60s, the introduction of the internet in the 90s, and most recently, e-commerce, fintech, Artificial Intelligence (AI) and robotics or driverless cars. New technologies, especially AI, thoughtfully contribute to the objectives of sustainable development and lead to a significant shift in the labour market.

The digital economy requires several new skills and creates new risks, from cybersecurity breaches to facilitating illegal economic activities. Digital platforms are becoming increasingly important in the world economy, and some have achieved a strong market position in certain areas. According to Jirinova and Scholleova (2016), innovation activities in companies are an essential prerequisite for a competitive advantage and long-term existence of the company. The competitive advantage of the whole economy certainly depends on the competitiveness of companies. [1] For example, Google, Facebook and Amazon in Europe. In China, WeChat with Alipay (Alibaba) and their payment solutions, captured the entire Chinese mobile payments.

Today, thanks to mobile payments, China is a society that operates almost without the physical form of money. Payments have already evolved to the point that the client can only pay by scanning the face and insert the mobile number to verify the identity. Therefore, a wealth of personal information is collected about all payers, which can be further analysed. This is illustrated by the effort to create increasingly sophisticated individual credit models.

The main objective of the article is to present the latest trends in the way of client segmentation associated with individual credit evaluation based on machine learning and in the context of the omnichannel world. First, we discussed at the current situation in and innovations in the way people pay in an omnichannel world. We describe how the absence of physical money has affected society, how it has changed customer purchasing behaviour, and what this change means for the digital economy and marketing. Further parts explain the relation between new insights connection with client segmentation, clarify the individual credit evaluation and specify Individual Credit Base Model to improve the

accuracy of credit scoring. In the last part, the theoretical and practical conclusions are summarised.

II. LITERATURE REVIEW AND RESEARCH METHODS

A. *New Methods of Customer Segmentation*

The use of Internet services is growing worldwide. Shopping habits of customers in the last decade have changed. Using digital technology has changed the traditional concept of buying. Customers are increasingly using online content. They search social networks, communicate online, actively use the Internet to get news, video calls, shop online, use online banking services, and more. As a result, retail brands use new strategies to respond to customer needs, provide innovation, and improve the shopping experience. This new omnichannel behaviour is no longer limited to specific segments and technology; it is becoming widespread and mainstream across the generations. The most common tools and strategies are a responsive web page customised for each device, personalised offers, content strategy and proactivity in finding the bloggers and influencers. Alibaba is estimated to have nearly 6.72% of China's e-commerce market.

In recent years, there was an improvement in the digital position of all EU countries. The most significant progress recorded countries such as Finland, Denmark, the Netherlands, and Sweden, which are among the world leaders in digitisation. These countries are followed by the United Kingdom, Luxembourg, Ireland, Estonia and Belgium. However, the EU needs to improve in order to compete globally and to offer payment services such as China or the USA. [2]

In recent years, the scale of China's digital economy has been overgrowing, accounting for a continuing increase in the share of GDP. China's entire digital economy reached 31.3 trillion yuan in 2018, accounting for more than 1/3 of GDP. By 2019, the overall size of China's digital economy may be close to 36 trillion yuan. China's digital technology, products and services are accelerating the integration of penetration into all walks of life, and the growth and efficiency of other industrial output have been enhanced. [3]

Fintech is an emerging financial services sector that includes third-party payment, innovative financial services or products delivered via new technology. Global companies are interested in developing new technologies for more favourable settlement and payment systems and expanding their services internationally. All fintech innovations must be following CSR of these financial and non-financial companies. According to Dvořáková and Quigley [4], CSR is a strategy whereby the corporations take responsibility for the economic, social and environmental consequences of their business activities.

Giants such as Apple and Google are also moving into the mobile payment market with their payment platforms, Apple Pay and Google Pay. In the east, the internet industry giants like Alibaba and Tencent have become providers of banking services for banks without branches, such as Ant Financial and WeBank (Alibaba's online bank). These technologies have improved the quality of financial services and provided more extensive access to banking and financial services. China has become a global leader in fintech e-commerce industry. In recent years, traditional Chinese financial system has been transformed into a top financial

system where Alibaba is now one of the largest financial companies in the world. Third-party payment system (TPP), a fact which allows the development of electronic commerce.

The most significant number of internet users (97%) in Europe is young individuals from 16 to 24-year-olds and people who have a high level of formal education. In 2018, 83% of Europeans used the internet at least weekly, and 76% almost daily or daily in comparison with 81% and 72% respectively a year earlier. [2]

Profitable business creates strategies which usually start with sophisticated customer segmentation, which is a set of concepts and models that lead to profitable product offer. According to Kotler (2002) [5], e-commerce can be segmented into the following four categories: B2C (business to consumer); B2B (business to business); C2B (consumer to business), and C2C (consumer to consumer - consumers participate in the transaction), and the main variables for segmenting consumer markets are demographic, geographic, psychographic, and behavioural variables. Business-to-consumer (B2C) is a term in electronic commerce in which online transactions are made among businesses, and individual consumers and companies can sell their products directly to the consumers [6]. B2B model is a concept in which the participants in electronic commerce are often large companies. C2B is a segmentation model in which the customer contributes to the business. However, these are commonly used methods of grouping customers. Lu, and Wu [7] came up with a segmentation method based on customers' transaction patterns. According to Karimi-Majid, and Fathian [8], we can use data mining techniques to analyse customers' data and purchases in order to develop product offer. Smeureanu, Ruxanda, and Badea [9] defined customer segmentation in the private banking sector using machine learning techniques. Based on their paper, the Customer segmentation in the private banking sector is an essential step for successful business development, enabling financial institutions to address their products and services to homogeneous classes of customers. The paper approaches two of the most popular machine learning techniques, Neural Networks and Support Vector Machines, and describes how each of these perform in a segmentation process.

Retailers are increasingly converting their services to digital channels as they provide clear benefits for both retailers and customers. For customers, digital channels enable faster and more convenient services, and more beneficial offers, leading to a more informed decision-making process. The recent research has focused on understanding customer behaviour and preferences in a multichannel system. As a result, a multi-channel retailer is increasingly crucial in offering products to customers. Schoenbachler, and Gordon [10] examined the decision-making process and the implementation of a multi-channel strategy. It could be simplified if companies understand what leads users to a single channel, multiple channels, and which channels are preferred. The evolution of multi-channel thinking is omnichannel retailing. This approach considers a broader perspective on customer segmentation and how customers are influenced and move through the buying process. [11]. Following their example, Sands, Ferraro, Campbell, and Pallant [12] segmented omnichannel customers in the phase search, purchase and after-sales. Nakano, and Kondo [13] grouped customers according to channel preference and further classified these segments into

seven subgroups based on their media usage. The phenomenon of customer way model based on online marketing strategies and customer’s interaction with society is STDC matrix [14]. Many businesses make mistakes and evaluate online marketing activities only from the perspective of the so-called “Last Click”. According to [14], STDC framework consists of 4 main elements - See, Think, Do and Care. Based on this theory, we can define four customer’s segments. See customers, people who are not looking for a solution to any problem; they go through the web pages. The goal of the marketing strategy is to engage the audience with funny, entertaining and useful content. We support good customer relationship with the brand. The Think phase involves customers who have a specific problem and want to solve it. The customer usually looks for specific information, product or service. In the Do phase, there is a group of customers who make a purchase decision and choose a specific product. They are interested in specific information, specifications of specific products, brands, types, durability, parameters, and purchasing conditions. The Care phase includes customers who have purchased more than twice or who buy products and serviced regularly.

B. Individual Credit Evaluation System

The development of mobile payment platforms has revolutionised the analysis of customer behaviour. By interconnecting communication and payment platforms, it is possible to analyse interpersonal relationships in addition to purchasing behaviour. It made it possible to refine the segmentation of customer target groups significantly and at the same time, to better identify the individual buyer and knowing his or her needs. Because business activities are often associated with a lack of fund on the buyer's side. In addition to marketing purposes, it was a logical step to link the payment platform to the online lending environment.

Due to a large amount of information on individual users, their habits, groups of friends, it was possible to start approving loans virtually online through mobile applications. Below in this section, a model calls SESAME CREDIT, which evaluates all users of China's most successful payment platform ALIBABA, is described and presented in detail.

TABLE I. SESAME CREDIT SCORING DIMENSIONS

Identity trait	5 %	Basic information indicators	Age, gender, occupation, family situation, etc.
Performance capability	20 %	Payment and fund indicators	Number of credit cards, average amount of each client
User credit history	20 %	Blacklist information indicators	Record for any cheating transaction fraud behaviour
Interpersonal connections	15 %	Interpersonal connections indicators	Friends credit rating contact activity number of funs
Behavioural performance	25 %	Consumer preferences indicators	Shopping accounts, financing, payment

^a Source: Own Processing Based on Global Lighting Network, 2019

Below it is written the comparison of Individual Credit Evaluation System parameters and Commercial Bank Rating Index System parameters. Similarities are in scoring identity, compliance factors and credit history evaluation.

1) China Construction Bank:

- Gender, marriage, health, industry status, title;

- the account in China construction bank;
- balance of savings account;
- Business dealings with China construction bank.

2) China Citic Bank:

- Gender, marriage, industry status, position, working time in the unit, title;
- Other assets, insurance, loans.

3) China Minsheng Banking Corp.Ltd:

- Financial assets, other assets, insurance, accounts in Minsheng bank, whether there are bad credit records, card consumption points.

For more details, see Table. II. Comparison of Credit Evaluation Dimensions (China and Czech Republic, 2019)

TABLE II. COMPARISON OF CREDIT EVALUATION DIMENSIONS

	Central Bank China	Commercial Bank China	Sesame Credit (Ant Financial Services Group) China	Commercial Bank Czech
Basic Information Indicators	OK	OK	OK	OK
Performance Capability	OK	OK	OK	OK
User Credit History	OK	OK	OK	OK
Interpersonal Connections	OK	OK	OK	X
Behavioural Preferences	OK	OK	OK	X

^b Source: Own Processing Based on Baidu Wenku, 2019

Table II compares not only the Credit Evaluation System in the Czech Republic and China but also the Credit Evaluation System of traditional banks and online mobile payment platforms. The table shows that the credit systems used by banks in both countries are practically identical. At first glance, it is apparent that significant differences can only be observed with online mobile payment platforms. This is due to the technical capabilities of online mobile payment platforms, which can collect and further analyse users' data.

How the data is used is described in the following chapter, which describes in more detail how the personal credit is compiled.

C. Method of Machine Learning Data Analysis

Before constructing the personal credit evaluation model, it is necessary to collect and process the data. There are four main channels for collecting personal credit evaluation data. Based on the rich data collected, the collected data are processed and sorted out, such as desensitisation, replenishing missing value and so on, in preparation for further modelling.

1) Data acquisition channel

The personal credit data collected by Internet finance companies are extensive, and the sources of credit data are diversified. [15] There are four kinds of data: necessary

information, Internet data, e-commerce data and public information.

- The necessary information comes from the data provided by the user, such as the user registration information.
- The Internet data of the Internet is derived from Internet finance companies such as payment, micro-credit and other data.
- The e-commerce data is derived from the e-commerce platform of the internet finance company, Such as user trading goods, transaction amount and other transaction service data.
- Public information comes from data provided by external institutions with partnerships, such as the Learning Information Network, ID card systems, etc.

2) Data processing

The data collected by the Internet Finance Corporation for personal credit evaluation are so large that there are problems such as data loss and data exception. At the same time, attention should be paid to the protection of customer privacy and the desensitisation of sensitive information. Besides, in order to ensure the accuracy of the model, it is necessary to screen the collected variables and select the variables that have a significant influence on the final results.

- Data desensitization based on dynamic desensitization technology

The sensitivity data of the credit body sample is desensitized by the dynamic desensitisation technology, and the purpose of encrypting and hiding the sensitive data in the background database can be achieved, and the sensitivity information of the client can be protected.

- Data desensitisation mainly includes four aspects:
 - For identity data, such as email, mobile phone number, to identify.
 - For time and date data, remove its accuracy.
 - Convert text data, such as name, home address, etc, into digital data.
 - Logarithmic data, such as amount, pen, etc, positive and negative floating 5% -10%.
 - Detection of abnormal value based on the LOF test method

The desensitized data need to be detected by LOF algorithm, and the same index of different sample data is regarded as the same cluster, which is divided into n clusters (n is the index number of evaluation dimension). The average of the ratio of the local reachable density of the neighborhood point $N_i(p)$ of a certain point p of the same cluster to the local reachable density of the point p is calculated by the LOF algorithm. If the ratio is less than 1 (more than 1), the description is the average point (abnormal point) After the exception value is identified, the exception value can be deleted and regarded as a missing value, and the missing value can be processed by random forest method.

- Improve data saturation based on the random forest method

For the sample data after deleting the outliers, the missing values are supplemented by random forest method. K sample data sets were randomly selected, and d indexes were randomly selected to construct the training set.

- For data with the same category in the training set, use the force-filling method to supplement the missing value.
- The random forest model is trained by the filled training set, and the similarity matrix is counted. By analyzing the similarity degree between all the elements of the two matrices, the filling results of the missing values are determined.
- The iterations are carried out 4 - 6 times, and the final missing value processing results are output.
- Screening indexes based on gradient boosting decision tree method

After the data loss value is supplemented by the random forest method, it is necessary to filter the indicators that have an important influence on the individual credit evaluation results to increase the accuracy of the personal credit evaluation. [16] [17] [18]

In the input model of the sample data set of each index of the credit subject, the root node selects the feature of the minimum Gini index before and after the split as the feature variable, assuming that the credit card is repaid on time. Subnodes repeat the same split mode, select sub - important feature variables in turn, and build a weak learner 1. The model error is minimised by iteratively building the final strong learner.

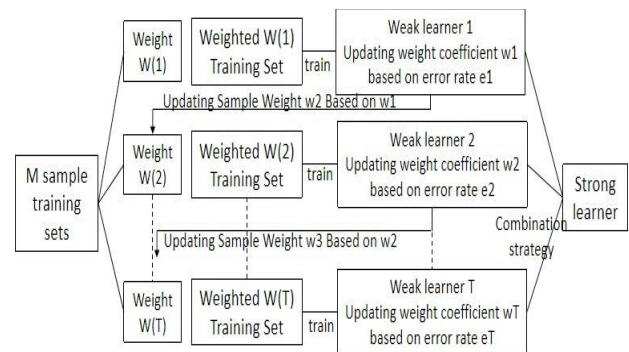


Fig. 1. Gradient boosting decision tree model (Source: Own Processing Based on Intelligent Algorithm, 2019)

According to the established strong learner, the average value of the importance of feature variables in all weak learners is taken as the importance of feature variables in the model.

III. RESULTS

A. Construct the scoring card model based on logical regression, and output the personal credit score

In the process of credit evaluation of credit subjects, banks-based financial institutions pay more attention to the use of transparent models (such as logical regression model and scorecard model) to evaluate the credit subject, so that the process of variable input to score output is transparent. [19] [20]

1) The probability of default is obtained by the logical regression model

Logical regression method is a standard method in machine learning. Through the study of the characteristic variables of the credit subject, the essential factors affecting the personal credit score are explored to predict the probability of default of the credit subject.

In the process of generating a personal credit score, after selecting the pre- λ ($\lambda < n$) indexes which have an essential influence on the production of personal credit score are selected by gradient boosting decision tree model, the credit logical regression model should obtain risk probability of credit subject.

The probability of default by the credit subject:

$$p(y = 1|x) = \frac{1}{1 + e^{-g(x)}} \tag{1}$$

The probability of non-default by the credit subject:

$$p(y = 0|x) = \frac{1}{1 + e^{g(x)}} \tag{2}$$

Of which, $g(x) = \beta_0 + \beta_1x_1 + \dots + \beta_\lambda x_\lambda$, $x_1, x_2, \dots, x_\lambda$ are the variables that have great influences on the credit score of the credit subject selected by the gradient boosting decision tree.

$\beta_1, \beta_2, \dots, \beta_\lambda$ are the coefficients in the logical regression equation. $y = 1$ or 0 indicates that the credit principal default occurred or did not occur, respectively.

2) Scorecard model outputs final credit score

After establishing the logical regression model, it is necessary to use the scorecard model to convert the probability of credit subject default into a score, which is helpful for financial institutions to quantify and manage credit risk better, to make credit decisions more objectively.

- Scorecard model

This paper investigates a credit scoring system research that combined a new perspective of client segmentation based on bank credit ratings known in Europe and the experience of new e-commerce platforms operating in China.

- Scorecard model

In the scoring card model, the personal credit score is expressed by the linear function of the ratio of the default probability of the credit subject to the non-default probability *Odds*, and the basic expression of the personal credit score can be obtained, as shown in formula (3):

$$Score = A - B * \log(Odds) \tag{3}$$

Constant A is the compensation score, which is part of the basic score of personal credit score, and constant B is called scorecard scale.

In the logical regression model, $\log(Odds)$ can be converted into a formula based on the linear combination of characteristic variables of personal credit evaluation, as shown in formula (4):

$$\log(Odds) = \log(1 - p) = \beta_0 + \beta_1x_1 + \dots + \beta_\lambda x_\lambda \tag{4}$$

Therefore, the formula (3) can be written as follows:

$$Score = A - B(\beta_0 + \beta_1x_1 + \dots + \beta_\lambda x_\lambda) \tag{5}$$

- Output scoring results

When the credit card model of the individual credit evaluation is established, the input variable $x_1, x_2, \dots, x_\lambda$ needs to be WOE-coded to reflect the effect of each characteristic variable (the loan amount, the income level, the blacklist information, the credit card number, etc.) on the default probability of the credit principal. Therefore, the form of personal credit can be further written as follows:

$$Score = (A - B\beta_0) - (B\beta_1w_{11})\delta_{11} - \dots - (B\beta_iw_{i1})\delta_{i1} - \dots \tag{6}$$

Of which, w_{ij} is the pre-i variable number j line WOE (hypothetical i variables are loan amounts, which can be divided into j paragraph. β_i is the coefficient in the logical regression equation. δ_{ij} indicates whether the variable i takes the value of row i. $A - B\beta_0$ is the basic score of the final credit score for the credit subject.

TABLE III. SCORE CARD SCORE CALCULATIONS

Variable	Number of rows (piecewise results)	Score
Datum point	—	$(A - B\beta_0)$
x_1	1	$-B\beta_1w_{11}$
	2	$-B\beta_1w_{12}$

	k_1	$-B\beta_1w_{1k_1}$
x_2	1	$-B\beta_2w_{21}$
	2	$-B\beta_2w_{22}$

	k_2	$-B\beta_2w_{2k_2}$
...
x_μ	1	$-B\beta_\mu w_{\mu 1}$
	2	$-B\beta_\mu w_{\mu 2}$

	k_μ	$-B\beta_\mu w_{\mu k_\mu}$

^c Source: Own Processing Based on Talking Data, 2019

B. Constructing BP Neural Network of Individual Credit Base Model to Improve the Accuracy of Credit Scoring

BP neural network is better at dealing with nonlinear complex classification problems. Its input layer and hidden layer are a process of feature extraction, while the output layer outputs the final results after processing the variables after feature extraction. The personal credit evaluation index is extracted through the input layer and the hidden layer, and after dimension reduction, it enters the output layer and finally outputs the personal credit score.

1) *The design of the number of nodes in BP neural network*

The BP neural network model of personal credit evaluation has three layers' structure, including input layer, hidden layer and output layer. The determination of the number of nodes in different layers will have a great impact on the output of personal credit score.

The number of input layer nodes μ is determined by the number of pre- $\mu (\lambda < \mu < n)$ important indexes selected by the gradient boosting decision tree, that is, the input layer has μ neurons.

The number of neurons in the hidden layer affects the accuracy of the output results. The initial range of the number of neurons in the hidden layer is determined by the number of neurons in the input layer and the output layer, and then the number of hidden layer nodes is determined by the cut-and-trial method.

The input layer neuron is determined by the output result, and the output result is a personal credit score, so the output node is 1.

2) *Parameter setting of BP neural network for personal credit evaluation*

There are many influencing factors on the establishment of BP neural network model for personal credit evaluation, such as the number of training times of the model, error accuracy and other parameters, which will have a great impact on the output of personal credit score. Therefore, it is necessary to set the model parameters according to the actual situation.

Training times: in the process of BP neural network training, the information is iterated continuously, and the number of iterations is the training times of BP neural network model. For the large sample data of personal credit evaluation, the number of training times is more, which is 50000 times.

Error accuracy: in the training process of BP neural network, the weight of the network is adjusted by continuous backpropagation of error, when the training error of the model reaches the maximum allowable error e , the training is stopped, and the credit maximum credit score is output.

Transfer function: in the BP neural network of personal credit evaluation, the general selection function Sigmoid is used as the transfer function of the hidden layer and the output layer to process the sample data of personal credit evaluation.

Learning function: in the BP neural network model of personal credit evaluation, the learning function selects the gradient descending momentum learning function and then adjusts the weight and threshold of the neural network.

3) *Output personal credit score results*

- Training of BP neural network model

After completing the basic construction of BP neural network model of personal credit evaluation, the model is studied and trained by inputting the sample data of personal credit evaluation. In the training process of BP neural network model, through several iterations, when the error accuracy reaches the set value or the maximum training times 5000 times, the iteration is stopped, and the credit score of the credit subject is outputted. Moreover, comparing

the output result of the credit principal credit score with the original credit score of the main credit body, and checking the accuracy of the model, and when the accuracy of the model reaches the present range, the residual sample can be input, and the model is further tested.

- Test of BP neural network model

The remaining $M - m$ samples are taken as test samples, and the sample data are inputted into the trained personal credit evaluation BP neural network model in the same way. The credit score of the test subject is outputted, the credit score of the output credit score is compared with the actual credit score of the credit subject, and the accuracy of the evaluation model is evaluated. When the accuracy of the model reaches a realistic and acceptable degree, the personal credit evaluation model is put into use. For the new sample data of the evaluation, the credit score of the credit subject is finally obtained by inputting the new sample data into the model.

4) *Relationship between credit scoring and customer segmentation*

Electronic market in recent years has grown and improve the basic functionality of the rating systems. Credit scoring is a set of decisions that can lead a bank or another decision-maker to grant the credit or another service to the customer. Depending on data, we can build not only classical credit scoring but also behavioural credit scoring. Behavioural credit scoring is in this case based on customers' behavioural parameters stored in the data warehouse.

When we think about customer segmentation model, we must consider the client basis and in the second step the segmentation model. It means that in the traditional marketing, segmentation aims to divide the client portfolio into smaller groups of clients with specific parameters. According to Siddigi (2005) [21], segmentation is obtained using statistical tools as well as data mining, machine learning or other techniques.

As a tool, any rating system provides a basic model for assessing the business relationship in the B2C market. It is also essential to mention secondary factors such as confidence building, promotion of trustworthy behaviour, and discouraging the participation of those who are dishonest. The system helps reduce transaction risks and increases the rate of successful transactions. A large number of factors, relevant financial indicators (profitability, liquidity, leverage) and other elements (activity, client behaviour) are reflected in customers' credit scoring.

IV. CONCLUSION

Changes in the way payments are made influenced not only the society but also marketing communication and customer's segmentation. Mobile payment applications have become out of payment tools, a marketing tool analysing large amounts of data. Based on the analysis, it was possible to influence customer segmentation, but also to create the personal credit score.

Through the study of the personal credit evaluation of Internet finance companies, the process of collecting, processing and using machine learning method to generate a personal credit score, it is found that the personal credit evaluation system of commercial banks can be improved from three aspects.

- In the process of personal credit evaluation, the abnormal value of sample data is tested and eliminated by LOF test method, and then the missing value of original sample data and the missing value after eliminating abnormal value are supplemented by random forest model, to replace the statistical method of replenishing missing value with average or mass number to improve the accuracy of prediction.
- The characteristic variable which influences the individual credit evaluation is screened through the gradient lifting decision tree model. And then constructing a scorecard model based on the logistic regression or a BP neural network model to evaluate and predict the filtered characteristic variable, and output the personal credit score. Make up for the inaccuracy of the credit score caused by the single use of the credit data and the logical regression model to output the personal credit score.
- To objectively and accurately output the personal credit score based on the machine learning method, and solve the one-sidedness and subjectivity defect of the traditional expert evaluation method.

REFERENCES

- [1] Jirinova, K., and Scholleova H. "Indicators of Innovation Potential". (2016). In: LÖSTER, Tomáš, PAVELKA, Tomáš (ed.). The 10th International Days of Statistics and Economics Conference Proceedings [online]. Praha, Slaný : Libuše Macáková, Melandrium, 2016, s. 770–780.
- [2] European Commission, "Digital Economy and Society Index (DESI)," 2019, Available at: <https://ec.europa.eu/digital-single-market/en/desi>.
- [3] Xinhua, "China's digital economy reaches 31.3t yuan in 2018", 2019. Available at: <https://www.chinadaily.com.cn/a/201905/07/WS5cd0fb66a3104842260ba445.html>
- [4] Dvořáková, Z., and Quigley, M. J. D. CSR strategies and practices: the case of the Czech Republic. (2014) In: LÖSTER, Tomáš, PAVELKA, Tomáš (ed.). The 8th International Days of Statistics and Economics Conference Proceedings [online]. Praha, 11.09.2014 – 13.09.2014. Slaný : Libuše Macáková, Melandrium, 2014, s. 364–373. ISBN 978-80-87990-02-5.
- [5] Kotler, P., Armstrong, G., Saunders, J., and Wong, V., "Principles of Marketing: 3rd European Edition", 2003.
- [6] Turban, E., King, D., Lee, J. K., Liang, T.-P., and Turban, D. C., "Electronic Commerce: A Managerial and Social Perspective", 8th ed., Cham, Switzerland: Springer International Publishing, 2015.
- [7] Lu, T.-C., and Wu, K.-Y., "A transaction pattern analysis system based on neural network", *Expert Systems with Applications*, 36(3), 2009, pp. 6091–6099.
- [8] Karimi-Majid, A.-M., and Fathian, M., "Extracting new ideas from the behaviour of social network users," *Decision Science Letters*, 6(3), 2017, pp. 207–220.
- [9] Smeureanu, I., Ruxanda, G., and Badea, L. M., "Customer Segmentation In Private Banking Sector Using Machine Learning Techniques", *Journal of Business Economics and Management*, 14(5), 2013, pp. 923–939. doi:10.3846/16111699.2012.749807
- [10] Schoenbachler, D. D., and Gordon, G. L., "Multi-channel shopping: understanding what drives channel choice." *Journal of Consumer Marketing*, 19(1), 2002, pp. 42–53. doi:10.1108/07363760210414943
- [11] Verhoef, P. C., Kannan, P. K., and Inman, J. J., "From Multi-Channel Retailing to Omni-Channel Retailing" *Journal of Retailing*, 91(2), 2015, pp. 174–181. doi: 10.1016/j.jretai.2015.02.005
- [12] Sands, S., Ferraro, C., Campbell, C., and Pallant, J., "Segmenting multichannel consumers across search, purchase and after-sales", *Journal of Retailing and Consumer Services*, 33, 2016, pp. 62–71. doi: 10.1016/j.jretconser.2016.08.001
- [13] Nakano, S., and Kondo, F. N., "Customer segmentation with purchase channels and media touchpoints using single source panel data", *Journal of Retailing and Consumer Services*, 41, 2018, pp. 142–152. doi: 10.1016/j.jretconser.2017.11.012.
- [14] Kaushik, A., "See, Think, Do, Care", 2013. Available at: <https://www.kaushik.net>.
- [15] Cao Jie, Shao Xiaoxiao. "Research on personal Credit Evaluation Model based on Information gain and Bagging Integrated Learning algorithm" [J] *Practice and understanding of Mathematics*, 2016, 46 (08): 90. 98.
- [16] Ju Chuanxiao. "A study on the Evaluation Mechanism of personal Credit in the Internet Age: a case study of Sesame Credit" [J]. *Modern Management Science*, 2018 (05): 109- 111.
- [17] Feng Jing. "Research on the Model of Personal Credit Risk Assessment Based on BP Neural Network" [D]. *Taiyuan University of Technology*, 2017.
- [18] Cao Jie, Shao Xiaoxiao. "Research on personal Credit Evaluation Model based on Information gain and Bagging Integrated Learning algorithm" [J] *Practice and understanding of Mathematics*, 2016, 46 (08): 90. 98.
- [19] Tan Zhongming, Xie Kun, Peng Yaopeng. "Research on the Credit Risk of the Loan of the P2P Network of the Decision Tree Model Based on the Gradient" [J]. *Soft Science*, 2018, 32 (12): 136-140.
- [20] Wang Zhongren, Han Dongmei. "Personal credit evaluation of Internet credit based on hyperparameter optimization and integrated learning" [J]. *Statistics and decision-making*, 2019, 35 (01): 87- 91.
- [21] Siddiqi, Naeem. "Credit risk scorecards: developing and implementing intelligent credit scoring." Vol. 3. John Wiley & Sons, 2012.