Churn Customer Estimation Method based on LightGBM for Improving Sales

Kohei Arai[1], Ikuya Fujikawa[2], Yusuke Nakagawa[3], Ryuya Momozaki[4], Sayuri Ogawa[5]
Information Science Dept., Saga University
Saga City, Japan[1]
SIC Co., Ltd, Hakata-ku, Fukuoka City
Fukuoka, Japan[2, 3, 4, 5]

***Abstract*—Churn customer estimation method is proposed for improving sales. By analyzing the differences between customers who churn and customers who do not churn (returning), we will conduct a customer churn analysis to reduce customer churn and take steps to reduce the number of unique customers. By predicting customers who are likely to defect using decision tree models such as LightGBM, which is a machine learning method, and logistic regression, we will discover important feature values in prediction and utilize the knowledge obtained through Exploratory Data Analysis (EDA). As results for experiments, it is found that the proposed method allows estimation and prediction of churn customers as well as characteristics and behavior of churn customers. Also, it is found that the proposed method is superior to the conventional method, GradientBoostingClassifier (GBC) by around 10%.**

Keywords*—*LightGBM (light gradient boosting machine); EDA (exploratory data analysis); churn prediction; linear regression; gradient boosting method; GradientBoostingClassifier: GBC

I. Introduction

Churn customer estimation method is very important for improving sales. By analyzing the differences between customers who churn and customers and who do not churn (returning), a customer churn analysis to reduce customer churn is conducted through taking steps to reduce the number of unique customers. By predicting customers who are likely to defect using decision tree models such as LightGBM, which is one of a machine learning method, and logistic regression, for discovering important feature values in prediction and utilize the knowledge obtained through Exploratory Data Analysis (EDA).

In order to predict churn customers, the method based on LightGBM and EDA is proposed here. LigtGBM is decision tree gradient boosting frameworks just as of XGBoost method and is convenient and fast machine learning method. Although there are differences in the details of the implementation method, there is no problem in thinking that they are almost the same framework in general. LightGBM is much faster than XGBoost method because it handles continuous values as histograms. XGBoost did not originally have this implementation, but now it is also possible to adopt a histogram-based algorithm with the parameter tree_method = hist.

The comparison between XGBoost and LightGBM is also a research topic because gradient boosting is highly practical. There is "Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms" published in September 2018 [1]. After testing XGBoost, LightGBM, and Catboost[1], it is concluded that no method is clearly superior in all situations.

The specific features and advantages of XGBoost and Light GBM is as follows,

- No need to impute missing values
- There is no problem even if there are redundant feature values (even if there are explanatory variables with high correlation, they can be used as they are)
- The difference from random forest is that trees are made in series.

On the other hand, approaches to data analysis can be broadly divided into a “hypothesis verification type” that verifies hypotheses with data and an “exploration type (EDA)” that generates hypotheses from data. Methods of data analysis are roughly divided into CDA: Confirmatory Data Analysis and EDA [2]-[6]. CDA is a general term for analytical methods aimed at hypothesis verification, while EDA is an analytical method aimed at obtaining hypotheses and knowledge from large-scale, multi-general data. EDA does not select explanatory variables in advance and performs exploratory analysis by seeking knowledge from a wide range of subjects. When we actually analyze data, we go back and forth between the hypothesis testing type and the search type to find out what we know.

Data analysis requires setting hypotheses to be verified, and there is nothing to be gained from analysis without hypotheses. However, there are times when a hypothesis cannot be obtained. Therefore, in order to create a hypothesis, it is necessary to look at the data from various angles and explore trends. Therefore, an exploratory data analysis is performed.

EDA can help by making sure stakeholders are asking the right questions. EDA helps answer questions about standard deviations, categorical variables, and confidence intervals. Once EDA is complete and insights are obtained, the features can be used for more sophisticated data analysis and modeling, including machine learning.

The cost of acquiring a new customer is higher than the cost of retaining an existing customer, up to five times as

[1] https://catboost.ai/en/docs/concepts/python-quickstart

much. Therefore, lowering the churn rate has a large positive impact on profits. Churn prediction is especially important for subscription-based services. By predicting churn, you can estimate CLTV (customer lifetime value)[2] and measure the growth potential of your business [7]-[19]. Also, customer churn is when customers cancel services such as subscriptions, and revenue churn is, for example, the loss of Monthly Recurring Revenue: MRR at the beginning of the month.

Customer profiling method with Big Data based on Binary Decision Tree: BDT and clustering for sales prediction is proposed and tested with POS: Point of Sales data [20]. Furthermore, a modified Prophet+Optuna prediction method for sales estimations is also proposed [21]. In this study, churn customer estimation method is proposed and examined with POS data for improving further sales.

In the next section, some of previous works are introduced. Then the proposed method for customer churn prediction is described followed by the experiment. Then conclusion and some discussions are described.

II. Previous Works

The 5:25 rule states that if you reduce customer churn by 5%, your profit margin will improve by 25%. From a medium- to long-term strategy perspective, it is important to implement planned measures after fully considering the balance between the customer retention rate, the defection rate, and the acquisition of new customers. Selling products to new customers requires five times the cost of selling products to existing customers (1:5 rule). Reducing the probability of customer defection and increasing sales of existing customers are important for increasing corporate profits.

It is important to maintain sales to reduce the withdrawal rate related to the top 20% of the treatment menu from the Partley's law[3]. A good way to identify the top 20% is to use a point card. With a point card, it is relatively easy to identify whether a customer is a regular customer or not.

If the new customer development cost is 100, the existing customer retention cost will be 17 to 20. The top 20% of customers account for 60-80% of total sales. Furthermore, in the bottom 30%, the degree of contribution to sales is less than 4%. The top 5% of customers with the highest loyalty often purchase related products. Reducing the defection rate (=increasing the rate of continuous purchases) has a large impact.

If the defection rate drops from 30% to 20%, the company's expected total sales now and in the future will increase by 1.5 times. A 10% increase or decrease in the attrition rate leads to a 50% increase or decrease in sales.

$$CAV = \frac{OD\ CNS\ CS}{1-CPR} \quad (1)$$

where CAV: Customer Asset Value, OD: Overall Demand, CNS: Customer Number Share, CS: Customer Share, CPR: Continuous Purchasing Rate.

where, the share of the number of customers is the ratio of customers who purchase the company's products among all customers in the relevant market, and the intra-customer share is the ratio of the company's products to all purchases of the product group by one customer. In addition, 1-continuous purchase rate: customer defection rate = the ratio of customers who purchased the company's products to no longer purchase the company's products.

For existing customers, the largest defection (=low repeat rate) occurs from the first purchase (F1, Frequency = 1) to the second purchase (F2, Frequency = 2). Also, if the purchase at F1 is not a regular purchase, the repeat rate from F1 to F2 is often about 20 to 40%. Furthermore, the repeat rate rises from F2 to F3, F3 to F4, etc., and when it exceeds F3, it rises to about 70 to 90%, and stable repeat earnings can be obtained.

Possible reasons for separation are as follows.

1) I did not get the results I wanted or could not get them.

2) I felt that the price of the treatment was higher than the benefits obtained (e.g., I was dissatisfied with the cost performance).

3) I felt dissatisfaction and anxiety about the company's response, not the treatment.

Therefore, customer defection analysis is necessary. It is necessary to calculate the "customer defection rate", the percentage of customers who did not use the service for the second time or more during a certain period of time, from Customer Relation Management (CRM) data, and to analyze the trend of "what kind of customers are defected". In particular, if the customer abandons the service after using it multiple times, it is necessary to take a customer's purchase history, frequency, and questionnaire.

For example, conducting questionnaires using Google Form, etc., and the "Frequently Asked Questions (FAQ)" page posted on the company's website have a great impact on customer satisfaction. , it is possible to avoid the risk of customers feeling dissatisfied and leaving. In addition, customer information in CRM is not just for approaching repeat customers, but it is necessary to collect and analyze data to grasp the tendency of customers who have already left, and to find out the reasons for leaving.

III. Proposed Method

First, customer churn is defined and then features of the customer churn are extracted from the customer data derived from the POS: Point of Sales data.

Customer churn prediction is performed by the following method.

- *Theme setting:* Define business problem and goals to be achieved → Define Before → After with monitorable metrics
- *Analytics design:* Define the built model and necessary data → In many cases, data such as transaction history and CRM (customer relationship management) system

[2] https://www.cccmk.co.jp/columns/hint3
[3] https://magazinn55.exblog.jp/5554516/

- *Dataset generation:* Preparing data and performing EDA, performing necessary preprocessing to create datasets suitable for machine learning algorithms.
- *Predictive model training and testing:* Train a churn prediction (departure prediction) model using various machine learning algorithms for classification problems → test the learned prediction model

After that customer churn is characterized and estimated based on LightGBM. Meantime, ROC (Receiver Operatorating Characteristic) curve evaluation method[4] is applied to the estimated churn ratio followed by feature importance is analyzed.

Some of the countermeasures are proposed for mitigation of customer churn.

IV. Experiment

A. Data Used

We used POS customer data from 1 September 2009 to December 31, 2021. The outline of the data is as follows:

1) Total number of customers (persons) 878,181 Number of unique customer IDs

2) Total number of cases (cases) 8,857,257 Number of sales item IDs (cut and color are counted as 2 cases, discounts are also counted as 1 case)

3) Cancellation of sales (number of cases) 350,017 Number of sales cancellation details

B. Definition of Customer Churn

A customer who visited the store in the previous three months did not return to the store in the next three months, and a customer who did not visit the store was defined as a churn. To give an easy-to-understand example, it was defined as "out of the customers who visited the store between January and March, the customers who visited between April and June returned, and the customers who did not visit the store were rejected."

The format of the final churn prediction output is as follows. It is a specification that predicts the probability that each customer will defect in the next three months. In other words, the customer ID and the likelihood of churn are represented as paired data as shown in Table I.

TABLE I. Format for of the Final Churn Prediction Output

Customer_ID:	Chance_of_Customer_Churn
1	5%
2	50%
3	30%

About 65,000 customers visited all stores from January to March 2021, and customers who visited between April and June returned to the store, and those who did not return to the store.

"0" in Fig. 1 represents recurrence and "1" as customer churn. The overall churn rate was about 42%.

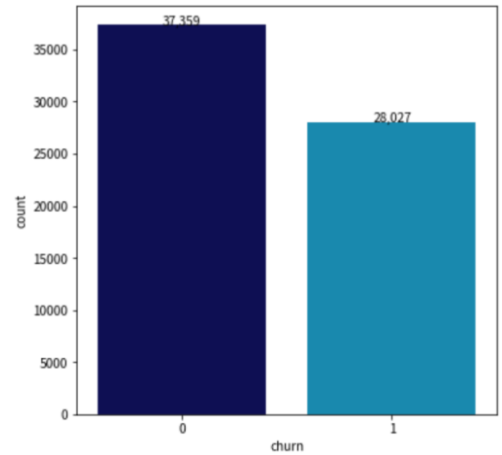


Fig. 1. Overall customer churn

The features used are shown in Table II.

TABLE II. Features Used in the POS Data Analysis

Schema_name	description
customer_id	Customer ID
visits_count	Number of visits
unit_price_ave	Average unit price per store
first_visit_date	Customer's first visit date
last_visit_date	Customer's last visit date
gender	Gender
age	Age(customers_who_do_not_enter_are_0)
distance	Distance_to_the_store_calculated_from_the_zip_code
menu	Categorization_by_menu
unit_price_per_visits_count	Average_unit_price_per_visit/number_of_visits

C. Preliminary Results

1) The difference between churn customers and non-churn customers: The difference between churn customers and non-churn customers was evaluated from the number of visits. The results are shown in Fig. 2. In the figure, orange indicates churn and blue indicates return. The lower the number of visits to the store, the higher the attrition rate, and the higher the number, the lower the attrition rate. There is a marked difference.

[4] https://zero2one.jp/ai-word/roc-curve-and-auc/

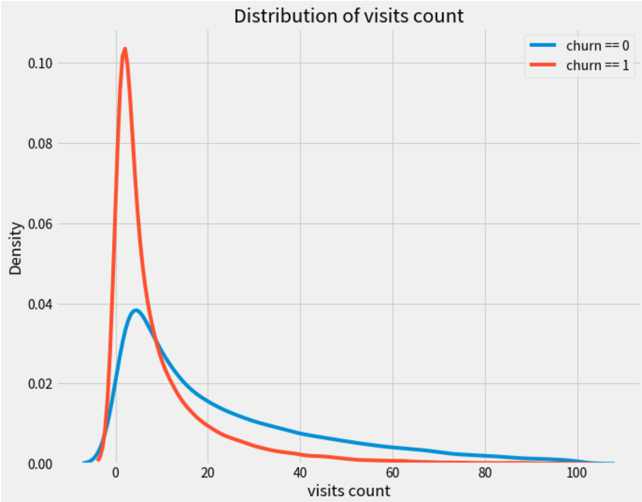


Fig. 2. The difference between churn customers and non-churn customer

2) Average cost per visit: Customers with low unit prices have a high churn rate, and customers with high unit prices have a low churn rate. However, there is no big difference depending on the unit price as shown in Fig. 3.

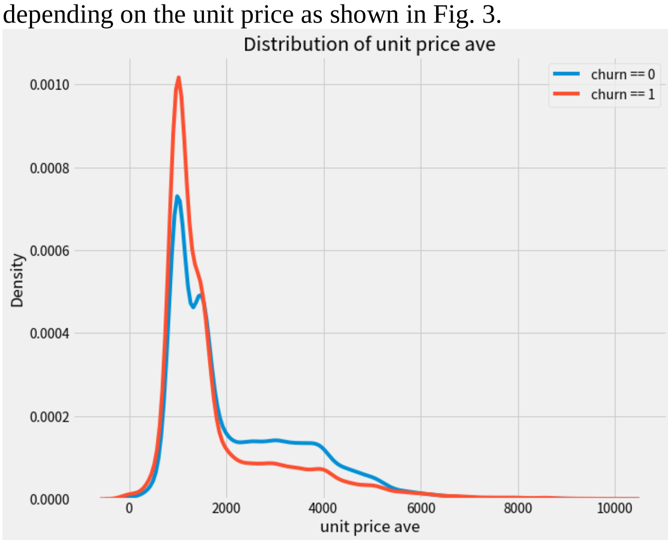


Fig. 3. Average cost per visit dependency against churn rate

3) Customer's first visit date: The horizontal axis in Fig. 4 indicates how many days before the first visit to the store from the analysis point. This time, we analyzed customers who visited the store from January to March, so March 31st was the day before. From this, we can see that the churn rate is higher for people who first visited the store recently, and the churn rate is lower for people who first visited the store a long time ago. These differences are significant.

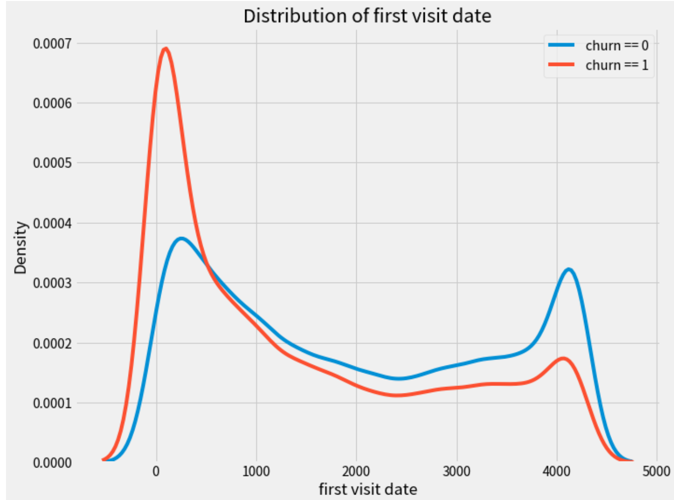


Fig. 4. Customer's first visit date dependency against churn rate

4) Customer's last visit date: The horizontal axis in Fig. 5 shows the number of days before the last visit from the point of analysis, just like the date of the first visit. This time, we analyzed customers who visited the store from January to March, so March 31st was the day before. From this result, we can see that the withdrawal rate is lower for those who last visited the store more recently, and the withdrawal rate is higher for those who last visited the store more than 50 days ago.

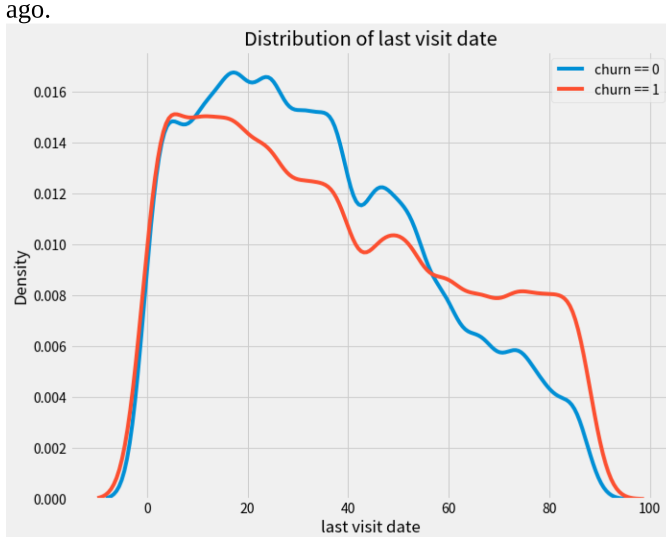


Fig. 5. Customer's last visit date dependency against churn rate

5) Gender: The churn rate is lower for men than for women (the churn rate for those entered as women exceeds 60%, but for men it is a little over 50%) as shown in Fig. 6. Customers whose gender is unknown (not entered) have a very low churn rate. The reason for that is unclear.

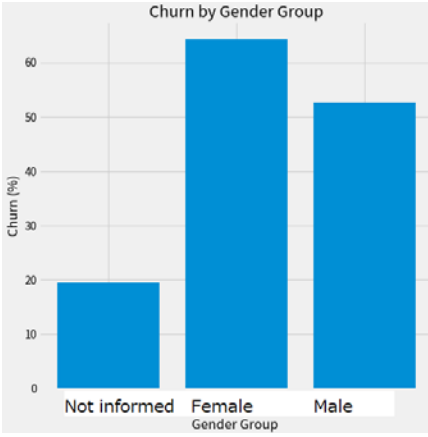


Fig. 6. Gender dependency against churn rate

6) Age: The churn rate is high for those in their 20s and 30s and decreases for those in their 50s as shown in Fig. 7.

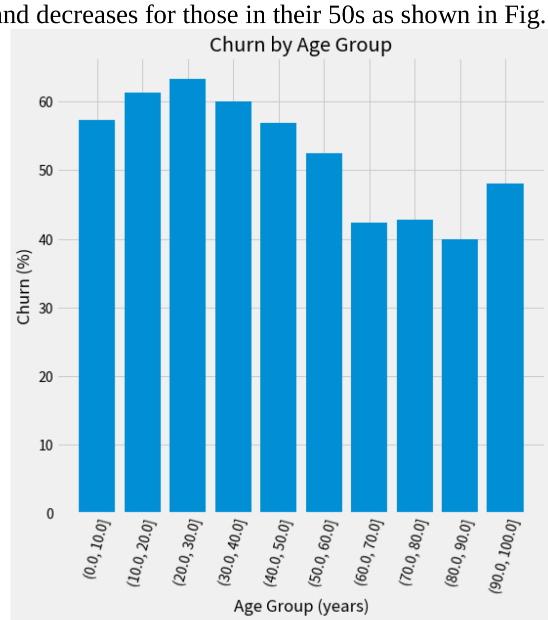


Fig. 7. Age dependency against churn rate

7) Service menu: We categorized customers according to the menu they ordered the most and investigated the churn rate. As a result, it was found that the rejection rate for dyeing white hair is very low at around 30%, while the rejection rate for child cuts and school cuts is high as shown in Fig. 8.

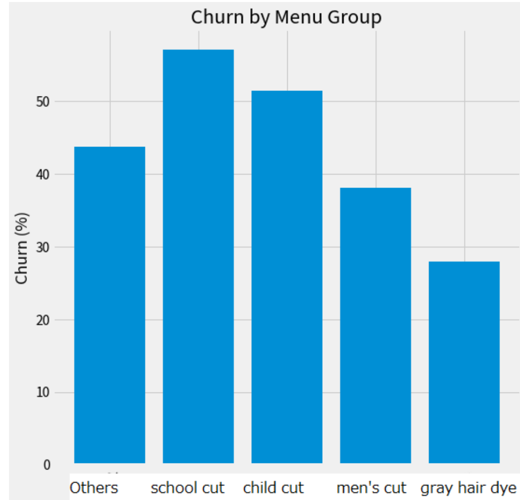


Fig. 8. Service menu dependency against churn rate

8) Average cost per visit / number of visits: Fig. 9 shows only those customers whose average unit price per visit/ number of visits is more than 2000 Yen in KDE[5] (Kernel Density Estimation). Customers with this value of 6,000 Yen or more seem to have a slightly higher churn rate. In other words, it seems that the churn rate is high for people who order expensive menus despite the fact that they visit the store less frequently.

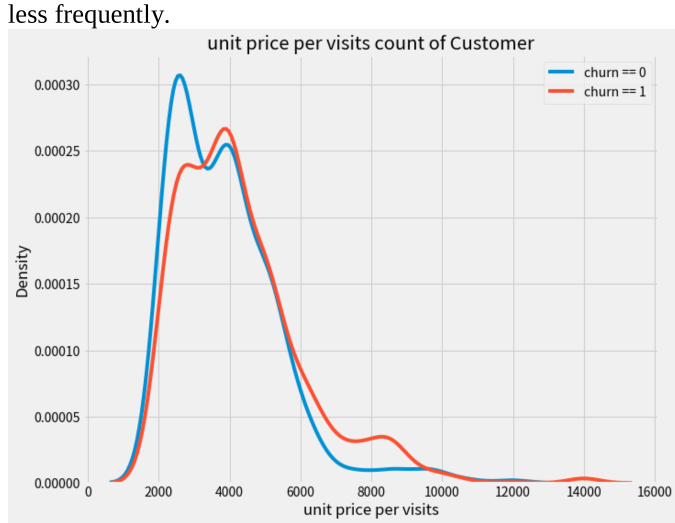


Fig. 9. Characteristics of the average cost per visit / number of visits

[5] https://cran.r-project.org/web/packages/spNetwork/vignettes/NKDE.html

D. Customer Churn Prediction

1) LightGBM based prediction of customer churn: The results of predicting customer churn using the above feature values (excluding distance to the store) are shown below. Fig. 10 shows the feature value order of customer churn prediction using LightGBM. It can be seen that the number of visits to the store on the day of the first visit has a large effect and is greatly affected to the churn.

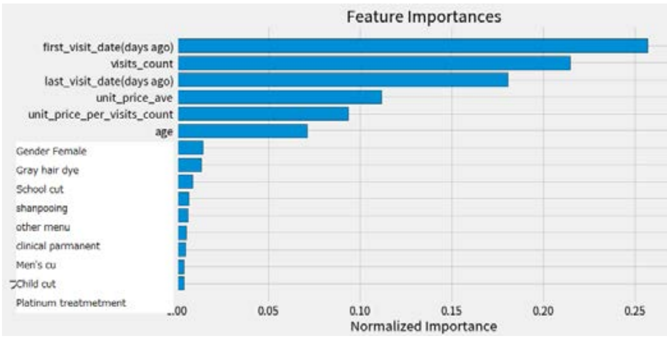


Fig. 10. Feature value order of customer churn prediction using LightGBM

2) ROC curve evaluation: ROC curve and Churn pct are evaluated. Each axis represents TPR (True Positive Rate) and FPR (False Positive Rate) and plots the TPR and FPR values when changing the threshold for classifying into Positive and Negative. As shown in Fig. 11, ROC curve and Churn pct (histogram) are seemed reasonable (not perfectly satisfied but marginal). Also, AUC (Area Under the Curve) and logarithmic function of loss are evaluated. As shown in Table III, both show reasonably satisfied values.

TABLE III. AUC AND LOGARITHMIC FUNCTION OF LOSS

AUC	0.837
Log_loss	0.496

3) Customer churn characteristics: Customer churn characteristics are summarized as follows:

a) Age: Younger customers have a higher attrition rate, and those in their 60s to 80s have a lower attrition rate.

b) Gender: Female divorce rate is high.

c) Number of visits: The lower the number, the higher the churn rate.

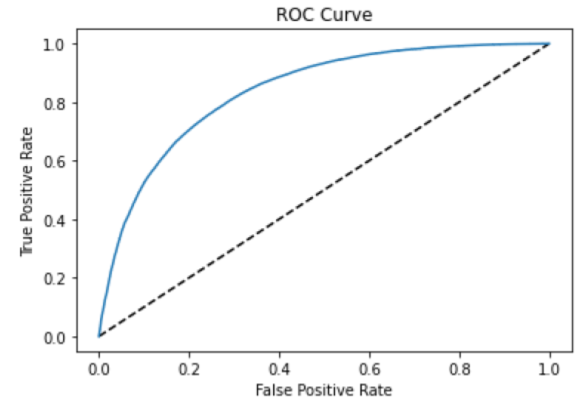
d) Unit price (1 store visit): The lower the rate, the higher the withdrawal rate.

e) Date of first visit: Most recent customers (customers who have recently visited for the first time) have a high churn rate.

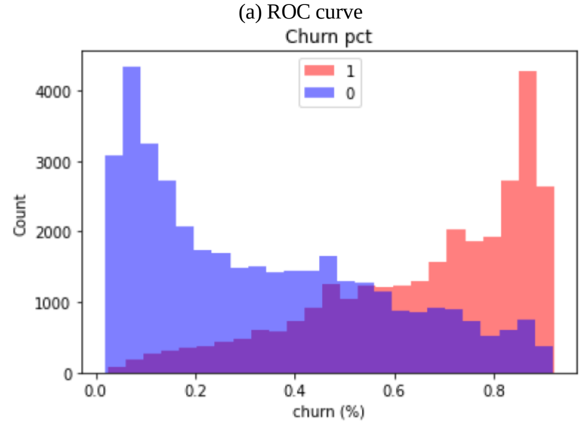
f) Date of last visit: The withdrawal rate is lower for the most recent visit.

g) Menu: Low withdrawal rate for gray hair dyeing, high withdrawal rate for school and child cuts.

h) Distance to stores: This hardly contributes to the churn rate and seems to depend on the availability of parking lots.



(a) ROC curve



(b) Churn pct (histogram)

Fig. 11. ROC curve and churn pct of the churn rate prediction based on LightGBM

V. CONCLUSION

Churn customer estimation method is proposed for improving sales. By analyzing the differences between customers who churn and customers who do not churn (returning), we conduct a customer churn analysis to reduce customer churn and take steps to reduce the number of unique customers. By predicting customers who are likely to defect using decision tree models such as LightGBM, which is a machine learning method, and logistic regression, we discover important feature values in prediction and utilize the knowledge obtained through EDA.

As results for experiments, it is found that the proposed method allows estimation and prediction of churn customers as well as characteristics and behavior of churn customers. Also, it is found that the proposed method is superior to the conventional method, GradientBoostingClassifier: GBC by around 10%.

FUTURE RESEARCH WORKS

Further investigations are required for improvement of prediction accuracy. We could be able to take measures such as sending DMs and coupons to customers with a 90% chance of churn. In order to increase the accuracy of churn prediction, not only LightGBM but also ensemble models such as Random Forest and logistic regression will be learned, and the accuracy will increase a little more. In addition, this time, we had the customers of all stores who visited the store during a specific period learn, but if we try to learn for each store without narrowing down the period, a different result may appear.

ACKNOWLEDGMENT

The authors would like to thank to Professor Dr. Hiroshi Okumura and Professor Dr. Osamu Fukuda for their valuable discussions.

REFERENCES

[1] Andreea Anghel, Nikolaos Papandreou, Thomas Parnell, Alessandro De Palma, Haralampos Pozidis, Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms, Workshop on Systems for ML and Open Source Software at NeurIPS 2018, Montreal, Canada, 2018.

[2] Hirokazu Iwasawa, Yuji Hiramatsu, "EDA (Exploratory Data Analysis)" Predictive Modeling with R: For Risk Management Using Machine Learning Tokyo Tosho pp.46-62, 2019.

[3] Yasuhito Mizoe, "Concept of Exploratory Data Analysis," Estrela, No.65, August 1999, pp.2-8, 1999.

[4] Mosteller, F. and J.W. Tukey, “Data Analysis and Regression”, Addison- Wesley, 1977.

[5] Noora Kanerva, Jukka Kontto, Maijaliisa Erkkola, Jaakko Nevalainen, Satu Männistö, “Suitability of random forest analysis for epidemiological research: Exploring sociodemographic and lifestylerelated risk factors of overweight in a cross-sectional design.” Scandinavian Journal of Public Health, Vol 46(5) pp.557-564, 2018.

[6] Tukey, J.W., “Exploratory Data Analysis”, Addison-Wesley, 1977.

[7] Stone, Merlin and Shaw, R., “Database marketing”. Aldershot, Gower. 1988.

[8] Peppers, D., and M. Rogers, Enterprise One to One: Tools for Competing in the Interactive Age. New York: Currency Doubleday, 1997.

[9] Hanssens, D., and D. Parcheta (forthcoming). “Application of Customer Lifetime Value (CLV) to Fast-Moving Consumer Goods.", 2011.

[10] Nakamura and Higa, Many studies on CLV ... Each paper has different definitions of customer lifetime value, target industries, business models, conditions for calculation, etc. 2011.

[11] Nakamura and Higa, "There are cases where COCA (Cost of customer acquisition), which is the cost of acquiring customers, is added 2011.

[12] Berger, P. D.; Nasr, N. I., “Customer lifetime value: Marketing models and applications”. Journal of Interactive Marketing 12 (1): 17–30. doi:10.1002/(SICI)1520-6653(199824)12:1<17::AID-DIR3>3.0.CO;2-K 1988.

[13] Fripp, G., "Marketing Study Guide" Marketing Study Guide, 2014.

[14] Adapted from "Customer Profitability and Lifetime Value," HBS Note 503-019, 2014..

[15] Ryals, L., Managing Customers Profitably. ISBN 978-0-470-06063-6. p.85, 2008.

[16] Gary Cokins, Performance Management: Integrating Strategy Execution, Methodologies, Risk and Analytics. ISBN 978-0-470-44998-1. p. 177, 2009.

[17] Fader, Peter S and Hardie, Bruce GS and Lee, Ka Lok, “RFM and CLV: Using iso-value curves for customer base analysis”. Journal of marketing research (SAGE Publications Sage CA: Los Angeles, CA) 42 (4): 415-430. doi:10.1509%2Fjmkr.2005.42.4.415, 2005.

[18] Tkachenko, Yegor, “Autonomous CRM control via CLV approximation with deep reinforcement learning in discrete and continuous action space”. arXiv preprint arXiv:1504.01840. doi:10.48550/arXiv.1504.01840, 2015.

[19] V. Kumar, Customer Lifetime Value. ISBN 978-1-60198-156-1. p.6, 2008.

[20] Kohei Arai, Zhang Ming Ming, Ikuya Fujikawa, Yusuke Nakagawa, Tatsuya Momozaki, Sayuri Ogawa, Customer Profiling Method with Big Data based on BDT and Clustering for Sales Prediction, International Journal of Advanced Computer Science and Applications, 13, 7, 22-28, 2022.

[21] Kohei Arai, Ikuya Fujikawa, Yusuke Nakagawa, Tatsuya Momozaki, Sayuri Ogawa, Modified Prophet+Optuna Prediction Method for Sales Estimations, International Journal of Advanced Computer Science and Applications, 13, 8, 58-63, 2022.

AUTHORS’ PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He was Adjunct Professor of University of Arizona, USA from 1998 2020. He also is Vice Chairman of the Science Commission “A” of ICSU/COSPAR from 2008 to 2018 then he is now award committee member of ICSU/COSPAR since 2018. He is now adjunct professor of Nishi-kyushu University and Kurume Institute Technology since 2018. He wrote 92 books and published 689 journal papers as well as 513 conference papers. He received 70 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA.

http://teagis.ip.is.saga-u.ac.jp/index.html