Method for Characterization of Customer Churn Based on LightBGM and Experimental Approach for Mitigation of Churn

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

***Abstract*—A method for customer churn characterization based on LightBGM (Light Gradient Boosting Machine) is proposed. Additionally, experimental approaches for mitigation of churn are conducted through churn prediction. The experiments reveal several churn characteristics such as age dependency, gender dependency (with a high divorce rate among female customers), number of visits dependency (with a higher churn rate for customers with fewer visits), unit price (per hair salon visit) dependency (with a higher withdrawal rate for lower-priced services), date of first visit dependency (with a high churn rate for recent customers), date of last visit dependency, and menu dependency (with low attrition rates for gray hair dye and high attrition rates for school and child cuts) and so on. Through the experiments, these dependencies are clarified. It is found that the first visit date is the most significant factor for churn customer character. Also, it is found that "distance to hair salon" dependency may be related to the availability of parking lots, although this factor has insignificant impact on the churn rate.**

***Keywords*—Churn; LightBGM; churn characteristics; linear regression**

I. Introduction

The method for predicting the characterization of churn customers is an important aspect of customer retention and marketing strategies. By understanding the characteristics of customers who are likely to leave or churn, businesses can take proactive measure for them. Predictive modeling techniques, such as machine learning algorithms, can be used to identify patterns and relationships in customer data, allowing businesses to predict which customers are most likely to churn and why.

Overall, the effectiveness of this method depends on the quality of data used for analysis, the accuracy of the predictive models employed, and the ability of businesses to act on the insights generated. It is important to continuously refine the models and update them with new data to ensure accuracy and relevance. Additionally, churn predictions can help develop appropriate customer retention strategies. Acquiring new customers is usually more expensive than retaining existing ones. Therefore, churn forecasting is becoming increasingly common to aim for a more profitable business.

There are two types of churns:

1) Customer churn (e.g., monthly churn customers / the number of customers at the beginning of the month).

2) Revenue churn refers to the decrease in Monthly Recurring Revenue (MRR) over a given period, typically a month. It is calculated by dividing the reduction in MRR by the MRR at the start of the month.

While customer churn specifically pertains to customers canceling a service or subscription, revenue churn encompasses not only customer churn but also instances where customers downgrade to a lower-tier service or product. Therefore, revenue churn is a more comprehensive measure than customer churn.

There are four MRR changes, each with a name:

1) New MRR: MRR from new customers.
2) Expansion MRR: MRR obtained from existing customers with increased transaction value.
3) Downgrade MRR: MRR obtained from existing customers whose transaction amount has decreased.
4) Churn MRR: MRR that would have been obtained from churned customers.

MRR changes can be classified into the following two categories:

1) Increased MRR: New MRR + Expansion MRR.
2) Decreased MRR: Downgrade MRR + Churn MRR

The Quick Ratio, which is the ratio of Increased MRR to Decreased MRR, is used as an index to measure growth potential. A Quick Ratio of less than 1 is considered bad, while a Quick Ratio of over 4 is considered great.

The churn prediction (departure prediction) model created by machine learning can be applied in two situations:

1) Complete churn (transaction amount 0).
2) Downgrade (change to lower service/product).

In the case of complete churn, it becomes a binary classification problem of "detachment or continuation." In the case of downgrading, it becomes a binary classification problem of "downgrade or maintenance." However, if there are multiple services to be downgraded, it becomes a multi-class classification problem.

When constructing a churn prediction (departure prediction) model, the process follows a similar framework to building a typical machine learning model. Firstly, the business problem and objectives are defined. Then, the model and necessary data for analysis are specified. This data is often sourced from transaction histories or CRM (Customer Relationship Management) systems. Subsequently, the dataset is generated, prepared, and subjected to EDA (Exploratory Data Analysis), as well as essential preprocessing steps like data normalization and standardization, to create suitable datasets for machine learning algorithms. Predictive models are then trained and evaluated. Churn prediction models utilize various machine learning algorithms, including deep learning, for classification problem-solving. Lastly, the trained prediction model is deployed and monitored to validate its effectiveness.

Reducing customer churn is of utmost importance. Therefore, it is crucial to analyze the differences between churned customers and those who remain (non-churned). To understand the behavior of churned customers, a method based on LightGBM (Light Gradient Boosting Machine) is proposed. While there are several linear prediction algorithms available, LightGBM demonstrates relatively high prediction accuracy. The factors contributing to churn are categorized, and churned customers are characterized based on these factors. By employing the proposed method, the importance of customer churn prediction can be assessed. LightGBM offers a functionality that identifies the key factors influencing churn prediction. Consequently, appropriate measures can be taken to reduce customer churn.

In the following section, some of the related research works are described, followed by the proposed method. Then, some of the simulation studies are described, followed by a conclusion with some discussions.

II. Related Research Works

Database marketing has been successfully introduced [1]. The book "Enterprise One to One: Tools for Competing in the Interactive Age" has also been published [2]. An attempt has been made to apply the concept of CLTV (Customer Lifetime Value) to FMCG (Fast-Moving Consumer Goods) [3]. Furthermore, several studies on CLTV have been introduced and reviewed, with each paper presenting different definitions of customer lifetime value, target industries, business models, and conditions for calculation [4].

Instances of COCA (Cost of Customer Acquisition) have been described, which refers to the cost of acquiring customers [5]. CLTV models and applications for marketing have been proposed and their applicability discussed [6]. Marketing study guides have been published and well-reviewed [7].

The book "Customer Profitability and Lifetime Value" has been published and extensively discussed [8]. Managing customers profitably has also been investigated and discussed [9]. Additionally, the analysis and discussion of "Performance management, which includes integrating strategy execution, methodologies, risk, and analytics," has taken place [10].

The paper "RFM (Recent Frequency Monetary) and CLTV: Using iso-value curves for customer base analysis" has been published, proposing and validating a method for marketing research [11]. Similarly, the paper "Autonomous CRM control via CLTV approximation with deep reinforcement learning in discrete and continuous action space" has been published, attempting to use CLTV approximation for CRM control [12].

On the other hand, CLTV has been well defined and discussed [13]. The paper "EDA of predictive modeling with “R” (a software tool for statistics) for risk management using machine learning" has been published, proposing and validating the use of EDA for predictive modeling [14]. Meanwhile, it is widely acknowledged that EDA is an important and useful technique in data science for analyzing and understanding data better. EDA involves exploring and visualizing the data to identify patterns, relationships, and anomalies.

EDA helps identify missing values, outliers, and other inconsistencies in the data, which can then be addressed before building predictive models. By visualizing the data, EDA also facilitates communicating insights to stakeholders and guiding further analysis. Furthermore, EDA is increasingly recognized as a critical step in any data analysis project as it enables a better understanding of the data, identification of potential issues, and provides insights for further analysis and decision-making. The concept of EDA has also been proposed and discussed [15]. Data analysis and regression have been well proposed for EDA analysis [16].

The paper "Suitability of random forest analysis for epidemiological research: Exploring sociodemographic and lifestyle-related risk factors of overweight in a cross-sectional design" has been published, studying and reporting on the suitability of random forest analysis for epidemiological research [17]. Additionally, EDA has been well defined, described, and investigated for its usefulness [18].

The paper "Customer Profiling Method with Big Data based on BDT and Clustering for Sales Prediction" has been published, proposing and validating a method for sales prediction using big data [19]. Meanwhile, the paper "Modified Prophet+Optuna Prediction Method for Sales Estimations" has been published, also proposing and validating a prediction method for sales using actual sales data [20].

III. Proposed Method

The objective of this paper is to clarify the behavior of churned customers, identify the reasons for churn, and propose strategies to mitigate churn. We aim to discover significant feature values for predicting customer behavior and leverage the insights obtained through Exploratory Data Analysis (EDA) to predict customers who are likely to churn using decision tree models like LightGBM and logistic regression. LightGBM is a popular open-source gradient boosting framework widely used for various machine learning tasks, including classification, regression, and ranking. Its high accuracy, efficiency, and scalability make it an excellent choice for analyzing churn customer behavior.

LightGBM utilizes a histogram-based algorithm to efficiently partition data into discrete bins, reducing memory usage and speeding up the training process. This feature, along with its gradient-based one-sided sampling technique that prioritizes data contributing the most to the loss function, further enhances training speed. Both processing speed and loss function minimization are essential factors. Additionally, LightGBM provides advanced features such as handling categorical features and missing values, early stopping, and cross-validation. These capabilities facilitate the handling of real-world datasets, prevent overfitting, and improve model performance. Overall, LightGBM is a powerful and efficient tool for building accurate and scalable machine learning models, especially for large-scale and high-dimensional datasets.

The definition of estrangement and output are as follows: The format of the final churn prediction output is shown in Table I.

TABLE I. SPECIFICATION OF CHURN PROBABILITY

Customer ID	Churn Probability
1	5%
2	50%
3	30%
…	…

We aim to predict the probability that each customer will churn in the next three months (definition of churn customer). A customer who visited the hair salon in the previous three months and did not return to the hair salon in the next three months, and a customer who did not visit the hair salon was defined as churn. The example of churn customer definition is as follows, "out of the customers who visited the hair salon between January and March, the customers who visited between April and June returned, and the customers who did not visit the hair salon were considered as having churned".

IV. EXPERIMENT

A. Data Used

The data used in the experiment targeted approximately 65,000 customers who visited all hair salons between January and March 2021. Customers who visited the hair salon between April and June were considered repeat customers, while those who did not visit the hair salon were considered as churn customer. These are illustrated in Fig. 1. In the figure, “0” is recurrence, “1” is withdrawal, and the withdrawal rate was about 42%. Table II shows the features used for churn analysis.

TABLE II. FEATURES USED

Schema name	**Description**
customer_id	Customer ID
visits_count	Number of visits
unit_price_ave	Average unit price per hair salon
first_visit_date	Customer's first visit date
last_visit_date	Customer's last visit date
gender	Male or Female
Age	Age (customers who do not enter are 0)
distance	Distance to the hair salon calculated from the zip code
Menu	Categorization by menu
unit_price_per_visits_count	Average unit price per visit / the number of visits

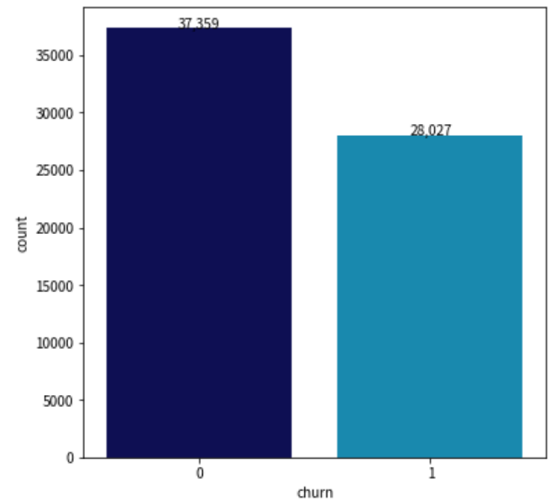

Fig. 1. Data used for churn analysis.

B. Experimental Results

Difference between churn and non-churn customers.

1) Number of visits: The total number of visits is displayed in Fig. 2. In subsequent graphs, orange color represents churn, while blue represents retention, as shown in Fig. 2. There is a notable difference in the attrition rate, with a higher number of visits resulting in lower attrition rates and vice versa for a lower number of visits.

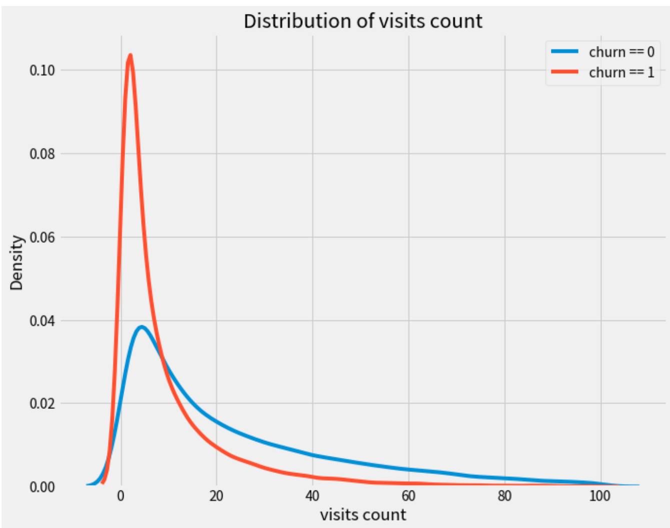

Fig. 2. The number of visits.

2) Average unit price per visit: The average unit price per visit is shown in Fig. 3. Customers with a low unit price have a high churn rate, whereas customers with a high unit price have a low churn rate. There is not much difference between the two.

3) Customer's first visit date: Fig. 4 displays the date of the customer's first visit to the hair salon. The horizontal axis represents the number of days before the first visit from the analysis point. In this analysis, we focused on customers who visited the hair salon between January and March. Hence, March 31st corresponds to the day before the analysis period's end. Notably, customers who recently made their first visit exhibit a higher churn rate, while those who visited the hair salon in the past show a lower churn rate. These differences are statistically significant.

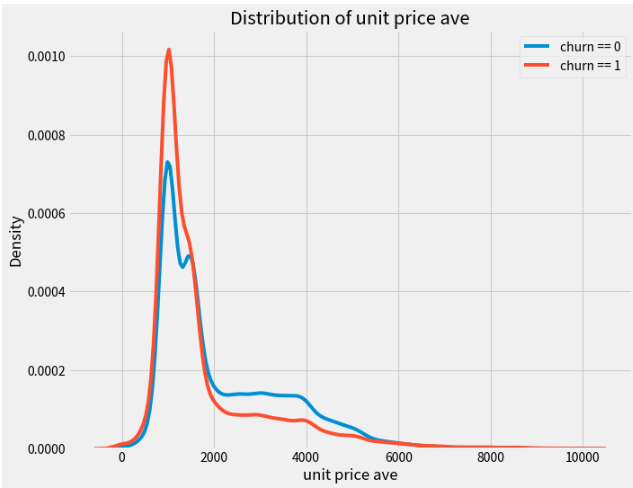

Fig. 3. Averaged unit price per visit.

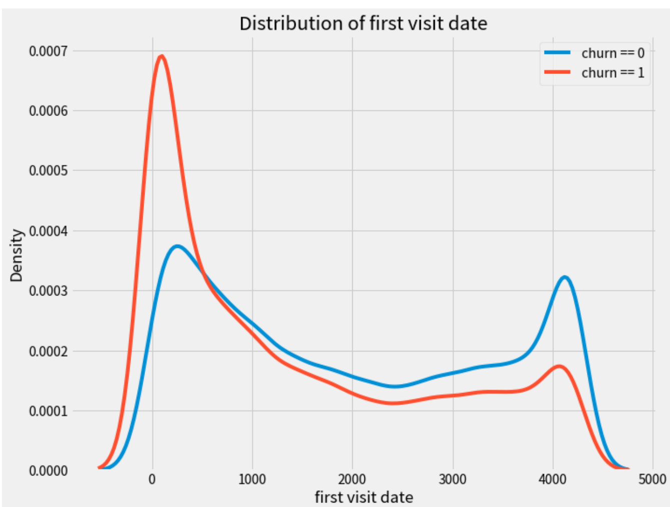

Fig. 4. Customer's first visit date.

4) Customer's last visit date: Fig. 5 illustrates the date of the customer's last visit. Similar to the first visit date, the horizontal axis indicates the number of days before the last visit from the analysis point. In this analysis, we examined customers who visited the hair salon between January and March, with March 31st being the obvious day before the analysis period's end. Interestingly, the churn rate is lower for customers who have recently visited the salon. However, customers whose last visit was more than 50 days ago exhibit a high churn rate, as depicted in Fig. 5.

5) Gender: As shown in Fig. 6, males have a lower attrition rate than females. The attrition rate for customers identified as female exceeds 60%, but it is slightly over 50% for males. Customers whose gender is unknown (not entered) have an extremely low attrition rate.

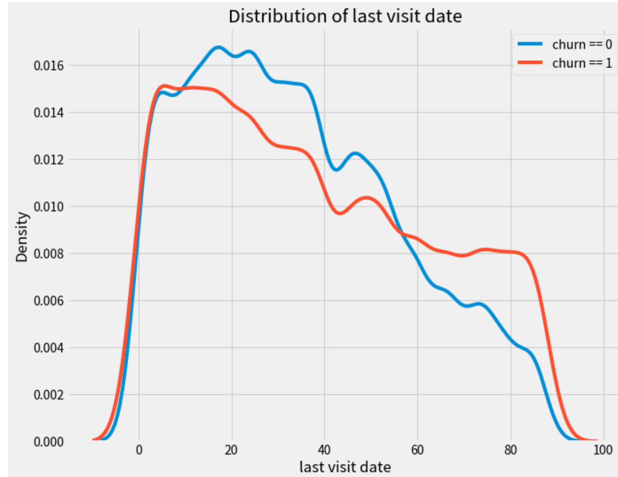

Fig. 5. Customer's last visit date.

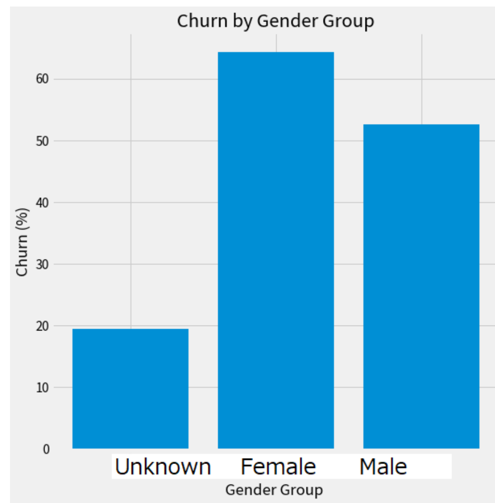

Fig. 6. Churn rate against gender.

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

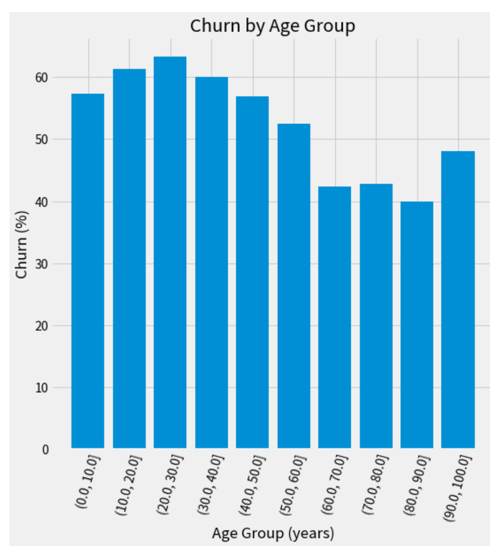

Fig. 7. Churn rate against age.

7) Survice menu: We classified customers based on their most frequently ordered menu and examined the churn rate, as presented in Fig. 8. The churn rate for customers who selected "dyeing white hair" is notably low, at approximately 30%. However, there is a high churn rate observed for "child cuts" and "school cuts".

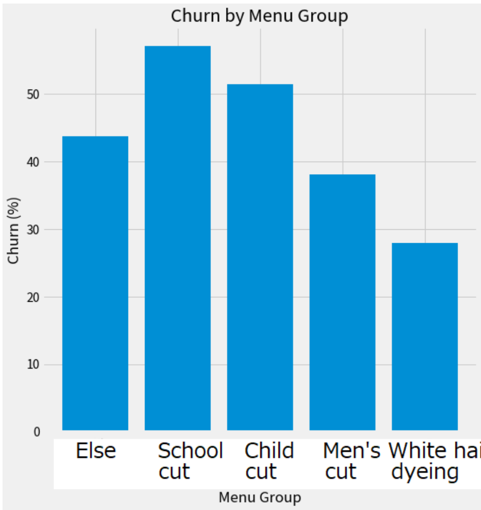

Fig. 8. Menu dependency on churn rate.

8) Average cost per visit / Number of visits: Fig. 9 displays a Kernel Density Estimation (KDE) plot for individuals with an average unit price per hair salon visit of over 2,000 yen. It is noticeable that customers with a value of 6,000 yen or more exhibit a slightly higher churn rate. This suggests that despite visiting the hair salon less frequently, individuals who opt for expensive menu options have a higher likelihood of churning.

KDE is a non-parametric method used to estimate the probability density function of a random variable based on a sample of observations (KDE in Python: Kdeplot is a KDE Plot which depicts the probability density function of the continuous or non-parametric data). KDE is in some senses an algorithm which takes the mixture-of-Gaussians idea to its logical extreme. The basic idea behind KDE present the probability density function as a weighted sum of kernel functions centered at each observation in the sample. The kernel function is a smooth, symmetric, and non-negative function that integrates to one, and its shape determines the smoothness of the estimated density.

KDE is often used in data analysis, machine learning, and statistics to visualize and estimate the probability density function of a dataset, especially when the underlying distribution is unknown or complex. It is a powerful tool for data exploration, pattern recognition, and outlier detection, and it can be applied to one-dimensional, two-dimensional, or higher-dimensional data. Some of the popular kernel functions used in KDE include Gaussian, Epanechnikov, and triangular kernels, and the bandwidth parameter determines the width of the kernel function and the smoothness of the estimated density.

KDE is a nonparametric method for estimating the probability density function of random variables in statistics. It is also known as the Parzen window, named after Emmanuel Parzen. In simple terms, kernel density estimation can be used to estimate data from a population, given data from a sample of that population.

9) Customer churn prediction: For customer churn prediction, Fig. 10 shows the results of using the above feature values (excluding distance to the hair salon). The top 6 features are displayed in the figure, with the menu features following. The feature value order of customer churn prediction using LightGBM is shown below, with the number of hair salon visits on the first day being the most influential feature.

In this churn prediction using LightGBM, the ROC curve is shown in Fig. 11 and the relationship between churn rate and its count is shown in Fig. 12 as a PCT.

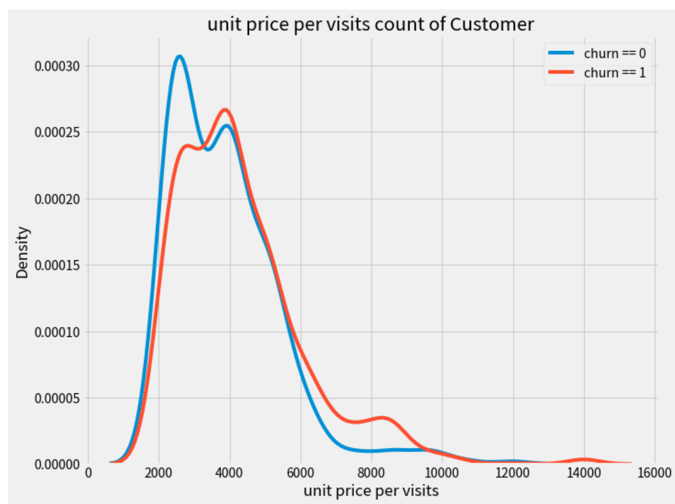

Fig. 9. Average cost per visit / Number of visits.

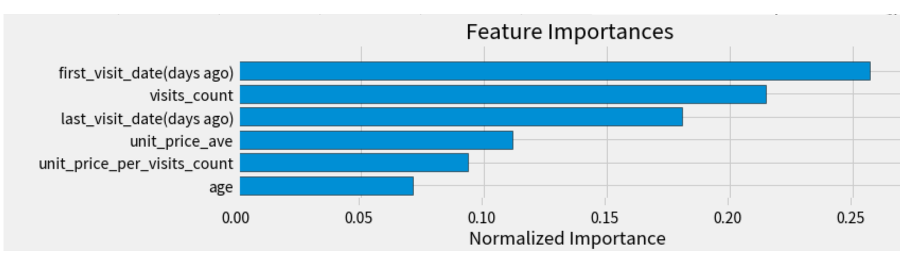

Fig. 10. Feature importance for the customer churn prediction.

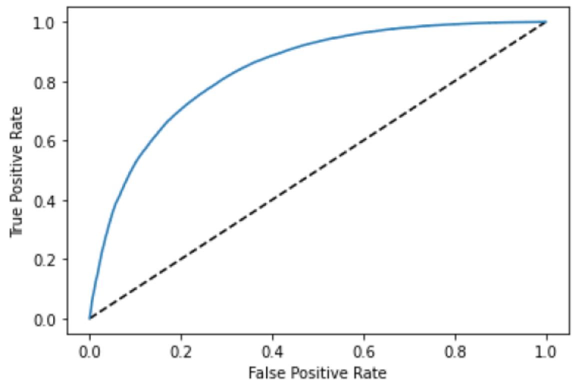

Fig. 11. ROC curve.

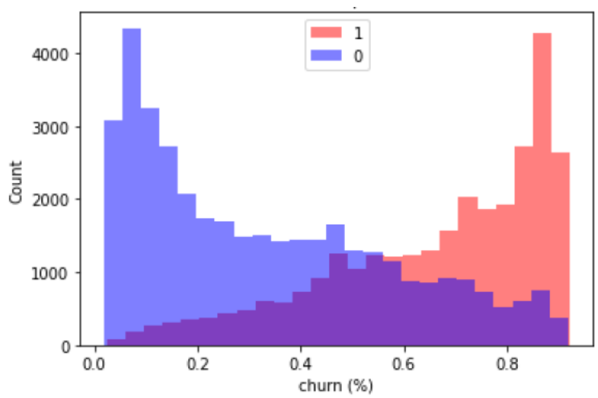

Fig. 12. PCT.

From the ROC curve, the AUC (Area Under Curve) is calculated as 0.837. ROC curve and AUC are commonly used to evaluate the performance of binary classification models. The ROC curve is a graph of the true positive rate (sensitivity) against the false positive rate (1 - specificity) for various threshold values, and it shows how well the model can distinguish between positive and negative samples. A perfect model would have an ROC curve that passes through the top left corner (100% true positive rate and 0% false positive rate), while a random guessing model would have an ROC curve that follows the diagonal line.

AUC is a single number that summarizes the ROC curve by calculating the area under the curve. AUC ranges from 0 to 1, where a score of 0.5 indicates a random guessing model and a score of 1 indicates a perfect model. A model with an AUC score of 0.7 or higher is considered to have good predictive power. ROC curve and AUC are useful tools for evaluating the performance of binary classification models, and a higher AUC score indicates better predictive power. Additionally, a logarithmic function of loss is found to be 0.496.

The results of the customer churn analysis and prediction are summarized as follows:

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

2) Gender: The churn rate is higher for female customers.

3) Number of visits: Customers with a lower number of visits have a higher churn rate.

4) Unit price (per hair salon visit): Customers with lower unit prices have a higher withdrawal rate.

5) Date of first visit: Customers who have recently visited for the first time have a higher churn rate.

6) Date of last visit: The churn rate is lower for customers who have recently visited.

7) Menu: Gray hair dye customers have a low attrition rate, while school and child cuts have a high attrition rate.

8) Distance to hair salons: The presence or absence of parking lots seems to have little impact on the churn rate.

Based on real-time database referencing (having your own database provides quicker results), the following countermeasures are implemented:

"Sending direct messages (DMs) and coupons to customers with a 90% chance of churning."

To improve churn prediction accuracy, ensemble models such as RandomForest and logistic regression will be attempted in addition to LightGBM, further enhancing accuracy. Additionally, if analyzed for each hair salon without narrowing down the period, different results may emerge, as mentioned in the previous section (churn characterization depends on the definition of churn).

V. Conclusion

In conclusion, this paper proposes a method for analyzing churn customer characteristics and conducts experimental approaches to minimize churn through churn prediction using LightGBM. The experiments revealed the following churn characteristics: age dependency, gender dependency (higher churn rate for females), dependency on the number of visits, dependency on unit price per visit, dependency on the date of first visit, dependency on the date of last visit, menu dependency, and distance to hair salons dependency. While the proposed method aids in characterizing churn customer behavior and identifying churn reasons, further considerations are required to devise effective countermeasures against churn.

VI. Future Research Works

We should conduct further investigation to identify alternative prediction methods that may lead to more accurate results. In particular, nonlinear prediction method has to be considered because the LighGBM used is essentially linear regression method and there must exist nonlinear behavior for churn customers.

Acknowledgment

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

References

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

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

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

[4] Nakamura and Higa, A Review of Past Research on Customer Lifetime Value Measurement, Japan Society for Management Information National Research Presentation Conference Abstracts, DOI https://doi.org/10.11497/jasmin.2011s.0.530.0, 2011.

[5] https://swifterm.com/how-to-calculate-cost-of-customer-acquisition-cac-or-coca/ accessed on 11 May 11, 2023.

[6] 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.

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

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

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

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

[11] 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.

[12] 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.

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

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

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

[16] Mosteller, F. and J.W. Tukey, "Data Analysis and Regression", Addison- Wesley, 1977.

[17] 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.

[18] Tukey, J.W., "Exploratory Data Analysis", Addison-Wesley, 1977.

[19] Kohei Arai, Zhang Ming Ming, Ikuya Fujikawa, Yusuke Nakagawa, Ryoya 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.

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

AUTHOR'S 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 is an Adjunct Professor of University of Arizona, USA for 1998 to 2020 and is Adjunct Professor of Nishi-Kyushu University as well as Kurume Institute of Technology pplied AI Laboratory) up to now. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 77 books and published 690 journal papers as well as 550 conference papers. He received 66 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