Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Prediction Model Using Machine Learning: Analysis Of Determinants Of Customer Churn At PT XYZ

[1]Riena Pribadi Gronloh, [2]Hendra Achmadi
[1,2]Faculty of Economy, Universitas Pelita Harapan, Jakarta

Article Info	ABSTRACT
Keywords: Customer Churn, Machine Learning, Random Forest, Logistic Regression, Purchase Retention	This study aims to identify the factors influencing customer churn at PT XYZ, a B2B application-based company selling essential goods. Machine learning algorithms such as Random Forest and Logistic Regression were used to predict churn based on demographic and behavioral variables, including age, membership duration, monthly transaction averages, spending value, and product variety. Transaction data from January 2023 to August 2024 was analyzed to understand partner behavior patterns. The results indicate that the Random Forest algorithm provides more accurate predictions than Logistic Regression, based on evaluation metrics such as accuracy, precision, recall, and ROC-AUC. This study provides strategic insights for PT XYZ to reduce churn and maintain customer purchase retention through a data-driven approach.
	Corresponding Author: Riena Pribadi Gronloh Faculty of Economy, Universitas Pelita Harapan, Jakarta rienapribadi@gmail.com

INTRODUCTION

Technological advancements have transformed business operations, especially in the Business-to-Business (B2B) sector. Many B2B companies have adopted application-based platforms to facilitate transactions with their business partners. PT XYZ, a B2B company specializing in the distribution of essential goods such as eggs, vegetables, fruits, and other household necessities, leverages mobile technology to serve its partners. Since its establishment in 2017, PT XYZ has built a customer network of 2,400 partners across Jabodetabek and Bandung as of August 31, 2024. These partners, primarily housewives, sell basic needs products to their neighbors, friends, and local communities.

Customer retention and reducing churn are crucial for sustainable business growth in the B2B industry. Churn, defined as the cessation of service use by customers or partners, presents a critical challenge (Matuszelański & Kopczewska, 2022). This issue is even more pressing in the B2B context due to the larger transaction volumes and recurring nature of business relationships, where the loss of a single partner can significantly impact revenue. According to Recurly (2024), a healthy churn rate for SaaS B2B companies ranges between 5% to 7% annually. However, data from PT XYZ revealed a churn rate of 20.96% in 2024, significantly higher than industry benchmarks. This high churn rate underscores the urgency of identifying the determinants of churn at PT XYZ and implementing predictive models to mitigate it.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Machine learning has proven effective in predicting churn by identifying high-risk customers based on historical data (Mirkovic et al., 2022). Algorithms such as Random Forest and Logistic Regression are particularly suitable for churn prediction due to their ability to handle classification tasks efficiently. Random Forest, an ensemble learning method, constructs multiple decision trees and aggregates their results to enhance prediction accuracy (Matuszelański & Kopczewska, 2022). Logistic Regression, widely used for binary classification problems like churn, is helpful for interpreting the linear relationships between independent variables and the likelihood of churn (Russo et al., 2016).

This study utilizes machine learning techniques to predict customer churn at PT XYZ, focusing on demographic and behavioral variables such as age, membership duration, transaction frequency, and spending patterns. By leveraging historical transaction data from January 2023 to August 2024, the study seeks to provide actionable insights into the most significant factors influencing churn and recommend strategies to improve customer retention. The findings will not only assist PT XYZ in addressing its churn challenges but also contribute to the broader understanding of churn management in the B2B context.

Literature Review

Customer Churn in B2B Contexts

Customer churn, defined as the cessation of service use by customers or business partners, is a significant challenge across industries, particularly in the Business-to-Business (B2B) sector (Matuszelański & Kopczewska, 2022). Unlike the Business-to-Consumer (B2C) sector, where churn often involves individual customers, churn in the B2B sector can have a disproportionate financial impact due to the typically larger transaction volumes and recurring nature of the relationships. Retaining customers is far more cost-effective than acquiring new ones, with the cost of new customer acquisition estimated to be five times that of retaining an existing customer (Jamjoom, 2021).

In the B2B SaaS industry, a healthy churn rate ranges between 5% and 7% annually (Recurly, 2024). However, PT XYZ's churn rate of 20.96% in 2024 is significantly above this benchmark, highlighting the need for urgent interventions to address this issue. Various factors, including changes in customer needs, competitive offerings, and perceived service quality, influence churn (Kriti, 2019).

Determinants of Customer Churn

The factors influencing customer churn can broadly be categorized into demographic and behavioral variables. Demographic Factors: Variables such as age, gender, marital status, and geographical location play a crucial role in predicting churn. For example, older customers or those with stable marital status are more likely to remain loyal to a service, while younger and unmarried customers exhibit a higher likelihood of churn (Jahromi et al., 2014; Mirkovic et al., 2022).

Behavioral Factors: Average transaction frequency, average spending value, average product variety, and average discount utilization are among the behavioral variables that significantly impact churn. Customers with declining transaction frequency or lower engagement are at higher risk of churning (Van Haver, 2017). Additionally, customers who purchase diverse products and regularly take advantage of promotions tend to show greater loyalty (Mirkovic et al., 2022).

Machine Learning for Churn Prediction

Machine learning (ML) has emerged as a powerful tool for analyzing and predicting customer churn. By leveraging historical data, ML algorithms can identify patterns and high-risk customers with greater accuracy than traditional methods (Matuszelański & Kopczewska, 2022).

Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees and combines their predictions to improve accuracy and reduce overfitting (Liu, Wang, & Zhang, 2012). It is particularly effective for complex classification problems like churn prediction, as it can handle large datasets with multiple variables while minimizing the impact of outliers (Couronné et al., 2018).

Logistic Regression: Logistic Regression is a simpler and widely used method for binary classification tasks, including churn prediction. It provides interpretable coefficients that quantify the relationship between independent variables and the likelihood of churn, making it a valuable tool for understanding the factors driving churn (Russo et al., 2016).

Previous Research

Previous studies have highlighted the effectiveness of ML in churn prediction. For instance, Matuszelański & Kopczewska (2022) demonstrated that Random Forest outperformed Logistic Regression in accuracy and predictive power for churn analysis in the e-commerce sector. Similarly, Mirkovic et al. (2022) showed that behavioral variables, such as transaction frequency and discount utilization, significantly influenced churn predictions in a B2B context.

RESEARCH METHOD

This study adopts a structured approach to analyze and predict customer churn at PT XYZ using machine learning techniques. The methodology incorporates the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, a widely accepted standard for data-driven projects, combined with specific steps tailored to the B2B context of PT XYZ.

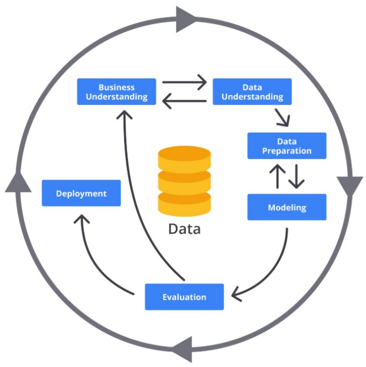

Figure 1. CRISP-DM Process Model for Data Mining

Business Understanding

The primary objective of this study is to understand and address the high churn rate at PT XYZ, a B2B application-based company. PT XYZ provides essential goods such as eggs, vegetables, and fruits to partners who sell them to their communities. The churn rate of 20.96% in 2024, significantly higher than the industry average of 5%-7% (Recurly, 2024), underscores the need for predictive analysis and intervention strategies.

Data Understanding

Data was collected from PT XYZ's transaction records spanning January 2023 to August 2024. The dataset includes both demographic and behavioral variables that are known to influence customer churn (Matuszelański & Kopczewska, 2022; Jahromi et al., 2014). The variables include:

1. Demographic Variables, Gender, age, marital status, and delivery area.
2. Behavioral Variables, Membership duration, average monthly transaction frequency, average spending, average product variety, average discount utilization, delivery method, and partner activity level.

Data Preparation

The data preparation phase included several critical steps:

1. Data Cleaning, Removing duplicates and handling missing values (Mirkovic et al., 2022).
2. Standardization, Normalizing numerical variables to ensure consistency in modeling.
3. Conversion to Factors, Transforming categorical variables into factor types to align with machine learning requirements.
4. Reliability Testing, Cronbach's Alpha was used to evaluate the reliability of variables, ensuring internal consistency (Van Haver, 2017). Variables with alpha values ≥ 0.7 were retained.
5. Multicollinearity Testing, Variance Inflation Factor (VIF) was calculated to detect and address multicollinearity. Variables with VIF values > 10 were excluded to ensure model robustness (Russo et al., 2016).
6. Correlation Analysis, Conducting a correlation matrix analysis to understand relationships between variables.

Modeling

Two machine learning algorithms were utilized to build predictive models for customer churn: Random Forest: This ensemble learning algorithm creates multiple decision trees and combines their predictions to enhance accuracy and reduce overfitting (Liu, Wang, & Zhang, 2012). Random Forest is particularly effective for classification problems involving complex and noisy data (Couronné et al., 2018). Logistic Regression: A widely used method for binary classification, Logistic Regression was employed to interpret the influence of independent variables on churn probability (Russo et al., 2016).

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Model Evaluation

The performance of the predictive models was evaluated using standard metrics:

1. Accuracy, The proportion of correct predictions out of the total predictions made.
2. Precision, The ratio of true positives to the total predicted positives.
3. Recall, The ratio of true positives to the total actual positives.
4. F1-Score, The harmonic mean of precision and recall, balancing the trade-off between the two metrics. It is particularly useful when the class distribution is imbalanced (Matuszelański & Kopczewska, 2022).
5. ROC-AUC, The area under the Receiver Operating Characteristic curve, which measures the ability of the model to differentiate between classes (Matuszelański & Kopczewska, 2022).

Implementation Framework

The modeling and analysis were conducted using RStudio, a statistical programming environment. The dataset was split into training (80%) and testing (20%) sets to ensure robust evaluation through cross-validation techniques.

RESEARCH RESULT

Descriptive Statistics

The descriptive analysis provides a comprehensive overview of the dataset, including demographic and behavioral variables.

Table 1. Descriptive Statistics for Numerical Variables

Variable	N (Obs)	Mean	Std. Dev	Median	Min	Max
Membership Duration (months)	941	30.10	22.03	23	2	82
Age (years)	941	40.37	10.32	39	18	74
Average Transactions per Month	941	4.38	9.40	2.00	0.00	192.75
Average Monthly Spending (IDR)	941	1,443,047	4,675,202	390,540	18,201	100,936,905
Average Product Variety per Month	941	7.06	10.12	4.00	0.00	124.90
Average % Discount per Month	941	1.93	4.40	1.20	0.00	92.16
% Partner Activity	941	53.65	27.97	56.00	5.00	100.00

Source: Primary Data Processed

Demographic and Behavioral Insights

The dataset reveals that the majority of partners are female (75%), with males accounting for the remaining 25%. Approximately 89% of the partners are married, while 11% are either single or widowed. These demographic characteristics highlight the predominance of family-oriented individuals within PT XYZ's partner base.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

In terms of behavioral insights, around 94% of partners prefer door-to-door delivery services, reflecting a strong inclination toward convenience, while 6% utilize pick-up facility. Additionally, most deliveries (93%) are concentrated in urban regions, with only 7% distributed across suburban and rural areas. These findings emphasize the importance of urban-focused strategies and the need to cater to varying delivery preferences among partners.

Data Preparation Result

The data preparation phase was conducted to ensure the reliability and validity of the dataset, which is crucial for building accurate machine learning models.

Data Cleaning

The dataset was cleaned by removing duplicates and handling missing values: Duplicate Data: Duplicate entries in the partner dataset were identified and removed, leaving 937 unique records from the original 941. Missing Data: Missing values were analyzed and handled using mean imputation for numerical variables and mode imputation for categorical variables

Reliability Testing

Table 2. Cronbach's Alpha Values for Variables

Variable	α (alpha)
Membership Duration (months)	0.45
Partner Age	0.37
Average Transactions per Month	0.73
Average Spending per Month	0.64
Average Product Variety per Month	0.67
Average % Discount per Month	0.28
% Partner Activity	0.30

Source: Primary Data Processed

Cronbach's Alpha was used to measure the internal consistency of numerical variables. The results indicated: Variables such as "Monthly Transactions" (α = 0.73) and "Monthly Spending" (α = 0.64) showed acceptable reliability. Variables like "Partner Activity" (α = 0.30) and "Monthly Discounts" (α = 0.28) were less reliable but retained due to their theoretical importance

Multicollinearity Testing

Table 3. Variance Inflation Factor (VIF) Values for Variables

Variable	VIF
Membership Duration (months)	1.12966
Partner Age	1.06277
Average Transactions per Month	1.52874

Variable	VIF
Average Spending per Month	1.23198
Average Product Variety per Month	1.41181
Average % Discount per Month	1.01664
% Partner Activity	1.00430

Source: Primary Data Processed

The Variance Inflation Factor (VIF) values for all variables were below the threshold of 10, indicating no significant multicollinearity issues. This confirms that the independent variables are suitable for use in the machine learning models without the risk of redundancy or inflated standard errors (Hair et al. 2019).

Correlation Analysis

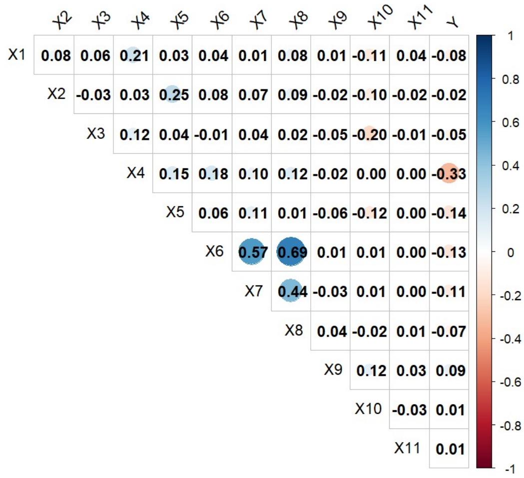

Figure 2. Correlation Plot Among Variables
Source: Primary Data Processed

A correlation analysis was conducted to evaluate the relationships among the variables in the dataset. The correlation matrix helped identify significant relationships, ensuring the most relevant variables were prioritized for the predictive models. The correlation plot (Figure 2) highlights the following key findings: Strong Positive Correlation: The highest correlation was observed between X7 (Average Transactions per Month) and X8 (Average Spending per Month) ($r = 0.69$), indicating that higher spending aligns with increased transaction frequency. Moderate Correlation: A moderate correlation was noted between X6 (Membership Duration) and X7 (Transaction Frequency) ($r = 0.57$), suggesting that longer memberships slightly influence transaction frequency. Low or Negligible Correlations: Other variables showed minimal relationships, with coefficients near 0, indicating low interdependence (Matuszelanski & Kopczewska, 2022)

Machine Learning Results

This section summarizes the results from the two predictive models: Random Forest and Logistic Regression. Both models were trained on 80% of the dataset and tested on the remaining 20%. The evaluation metrics include accuracy, precision, recall, ROC-AUC, feature importance, and p-values for interpretability.

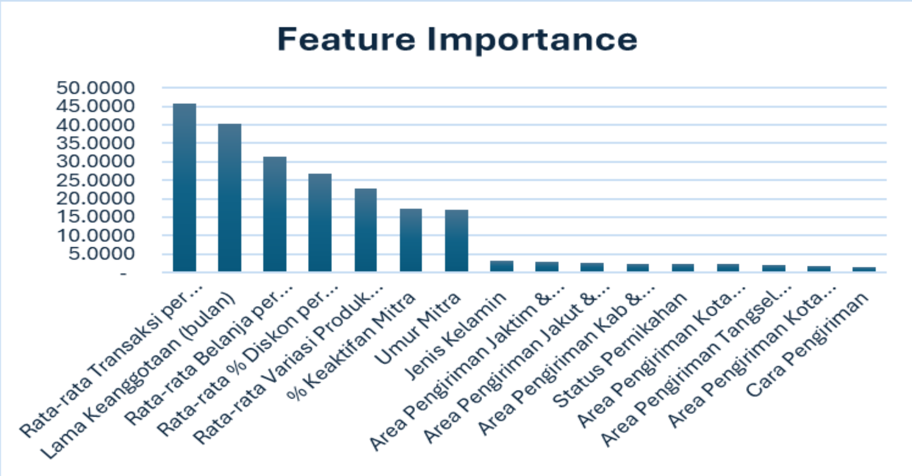

Figure 3. Feature Importance Bar Chart for Random Forest
Source: Primary Data Processed

Table 4. P-Values of Independent Variables in Logistic Regression

Variable	P-Value	Significance
Membership Duration (months)	0.00000	Highly Significant
Average Transactions per Month	0.00070	Highly Significant
Average Product Variety per Month	0.02755	Significant
Average Spending per Month	0.04358	Significant
Partner Age	0.03687	Significant
Delivery Area: Jakarta Utara	0.23038	Marginally Significant
Average % Discount per Month	0.05805	Marginally Significant
Gender	0.62735	Not Significant
Marital Status	0.67117	Not Significant
Delivery Area: Kabupaten Tangerang	0.69042	Not Significant
Delivery Area: Depok	0.75774	Not Significant
Delivery Method	0.73665	Not Significant
% Partner Activity	0.70170	Not Significant
Delivery Area: Bekasi	0.72707	Not Significant

Source: Primary Data Processed

The logistic regression equation used to calculate the probability of churn is as follows:

$$P(Y=1) = \frac{e^{\text{logit}(P(Y=1))}}{1+e^{\text{logit}(P(Y=1))}}$$

Where :

logit(P(Y=1))=−1.821631+0.174663X1+0.062792X2−0.377706X31+0.121717X32−0.648899X33−1.052401X4−0.214254X5−3.01935X6−0.524082X7+0.244668X8+0.263785X9−0.254036X10+0.001362X11

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Table 5. Summary of Model Metrics

Metric	Random Forest	Logistic Regression
Akurasi	87.17%	78.07%
Precision	71.10%	33.30%
Recall	74.40%	57.70%
F1 Score	72.70%	42.30%
AUC-ROC	91.90%	72.80%

Source: Primary Data Processed

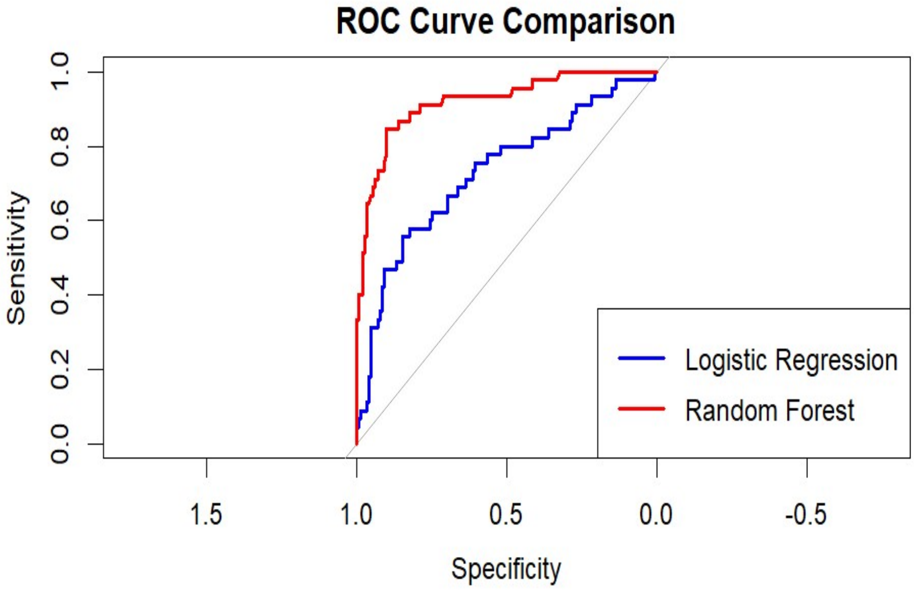

Figure 4. ROC Curve Comparison Random Forest and Logistic Regression

Interpretation of Model Performance

The comparison of the two machine learning models highlights their distinct strengths. The Random Forest model is recommended for predicting churn due to its superior performance across multiple metrics, including accuracy, precision, recall, and its capability to handle nonlinear relationships and interactions among variables. In contrast, the Logistic Regression model excels in interpretability, providing clear insights into the statistical significance of predictors and the underlying factors driving churn.

Based on the feature importance values from the Random Forest model and the p-values from the Logistic Regression model, the key variables influencing churn are as follows: Membership Duration (months): Highly significant in both models, suggesting that partners with shorter memberships are more prone to churn. Average Transactions per Month: Identified as a top predictor, with high feature importance in Random Forest and a statistically significant p-value in Logistic Regression. Average Spending per Month: A critical behavioral variable with substantial importance in Random Forest and statistical significance in Logistic Regression.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Average Product Variety per Month: Moderately important in Random Forest and statistically significant in Logistic Regression. Partner Age: Significant in Logistic Regression, though less emphasized in Random Forest. These variables underline the dominant role of behavioral factors—particularly spending, transaction frequency, and membership duration—in predicting churn. To address churn effectively, PT XYZ should prioritize monitoring these factors and implement targeted retention strategies, such as personalized campaigns or loyalty programs, aimed at partners exhibiting signs of disengagement.

CONCLUSION

Based on the research conducted on customer churn at PT XYZ, several key conclusions can be drawn. Behavioral variables, such as Average Spending per Month, Transaction Frequency, and Membership Duration, emerged as significant predictors of churn. These variables were consistently identified as critical factors in both the Random Forest and Logistic Regression models, demonstrating their importance in explaining partner behavior. Additionally, demographic variables, including Age and Delivery Area, played a lesser but noticeable role in influencing churn trends. The Random Forest model demonstrated superior predictive accuracy (87%) and a high ROC-AUC score (0.91), establishing it as the preferred tool for churn prediction. This model's ability to capture complex interactions and nonlinear relationships between variables made it highly effective for prediction. On the other hand, the Logistic Regression model provided valuable interpretability by highlighting statistically significant variables with p-values. This interpretability is critical for deriving actionable insights that can inform strategic decisions. To mitigate churn and improve partner retention, PT XYZ should enhance its operational processes. This includes investing in advanced data monitoring systems that can track key behavioral metrics such as transaction frequency and monthly spending in real time. Reliable delivery services and improved service quality in underperforming areas should be prioritized to address external factors contributing to churn. Furthermore, personalized interventions based on these insights can strengthen partner relationships and encourage loyalty. By streamlining operations and proactively addressing partner needs, PT XYZ can create a more reliable and engaging ecosystem for its partners. To further improve churn prediction and management, PT XYZ should consider incorporating additional variables into its models, such as customer satisfaction scores, feedback data, and competitor influences. These factors could provide deeper insights into partner behaviors and reasons for churn. Additionally, adopting advanced machine learning techniques like Gradient Boosting or Neural Networks may enhance predictive accuracy and uncover complex patterns within the data. Continuous improvement of the models through regular updates and retraining with new data will ensure their relevance and effectiveness over time. By embracing these enhancements, PT XYZ can stay ahead of market dynamics and foster long-term partner loyalty. By integrating these findings and recommendations, PT XYZ can effectively address churn, foster stronger partner relationships, and drive sustainable business growth.

REFERENCES

Agrawal, S., Das, A., Gaikwad, A., & Dhage, S. (2018). Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning. 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), 1–6. https://doi.org/10.1109/ICSCEE.2018.8538420

Ahmed, A., & Linen, D. M. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), 1–7. https://doi.org/10.1109/ICACCS.2017.8014605

Almahadeen, L. (2024). Evaluating Machine Learning Techniques for Predicting Customer Churn in E-Commerce: A Comparative Analysis. Journal of Logistics, Informatics and Service Science. https://doi.org/10.33168/JLISS.2024.0627

Apa itu churn pelanggan? (2024). https://www.ibm.com/id-id/think/topics/customer-churn

Banda, P. K., & Tembo, S. (2017). Factors Leading to Mobile Telecommunications Customer Churn in Zambia. International Journal of Engineering Research in Africa, 31, 143–154. https://doi.org/10.4028/www.scientific.net/JERA.31.143

Boulesteix, A.-L. (n.d.). Random forest versus logistic regression: A large-scale benchmark experiment.

Cassidy, A. P., & Deviney Jr., F. A. (2014). International Conference on Big Data. IEEE. https://doi.org/10.1109/BigData.2014.7004352

Celik, O., & Osmanoglu, U. O. (2019). Comparing to Techniques Used in Customer Churn Analysis.

Chen, M.-M., & Chen, M.-C. (2020). Modeling Road Accident Severity with Comparisons of Logistic Regression, Decision Tree and Random Forest. Information, 11(5), 270. https://doi.org/10.3390/info11050270

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. CATENA, 151, 147–160. https://doi.org/10.1016/j.catena.2016.11.032

Churn Rate Benchmarks. (n.d.). https://recurly.com/research/churn-rate-benchmarks/

Customer Lifetime Value (CLV): Pengertian, Rumus dan Contohnya. (2020). https://aksaragama.com/customer-lifetime-value-clv

Dutschmann, T., & Kinzel, L. (2023). Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation. https://doi.org/10.1186/s13321-023-00709-9

Elgeldawi, E., Sayed, A., Galal, A. R., & Zaki, A. M. (2021). Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. Informatics. https://doi.org/10.3390/informatics8040079

González-Benito, Ó. (2002). Geodemographic and socioeconomic characterization of the retail attraction of leading hypermarket chains in Spain. The International Review of Retail, Distribution and Consumer Research, 12(1), 81–103. https://doi.org/10.1080/09593960110103869

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. Emerald Publishing Limited, 31. https://doi.org/10.1108/EBR-11-2018-0203

Haver, J. V. (2017). Benchmarking analytical techniques for churn modelling in a B2B context.

Hills, W., Daniel, W., Lu, M. Y., Schaer, O., & Adams, S. (2020). Modeling Client Churn for Small Business-to-Business Firms. 2020 Systems and Information Engineering Design Symposium (SIEDS), 1–7. https://doi.org/10.1109/SIEDS49339.2020.9106673

Ishwaran, H., & Lu, M. (2018). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. John Wiley & Sons, Ltd. https://doi.org/10.1002/sim.7803

Jahromi, T. A., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. Industrial Marketing Management, 43(7), 1258–1268. https://doi.org/10.1016/j.indmarman.2014.06.016

Jamjoom, A. A. (2021). The use of knowledge extraction in predicting customer churn in B2B. Journal of Big Data, 8(1), 110. https://doi.org/10.1186/s40537-021-00500-3

Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. Taylor & Francis, 30(1), 197–215. https://doi.org/10.1080/10485252.2017.1404598

Kabut, A. S., & Windasari, N. A. (2024). A Predictive CRM Analytics Framework For Merchant Retention: Applying RFM Segmentation, Customer Profiling, and Behavioral Analytics In The B2B Payment Gateway Company. Return : Study of Management, Economic and Bussines, 3(6), 409–428. https://doi.org/10.57096/return.v3i6.246

Kamarulzaman, Y. (2010). Geodemographics of Travel E-shoppers: An Empirical Analysis of Uk Consumers. 16(2).

Kim, K.-M. (2023). Development of a prediction model for the depression level of the elderly in low-income households: Using decision trees, logistic regression, neural networks, and random forest. https://www.nature.com/articles/s41598-023-38742-1

Kirasich, K., Smith, T., & Sadler, B. (2018). Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. 1(3).

Klopotan, I., Buntak, K., & Drozdjek, I. (2014). International Journal for Quality research.

Koeslag, S. (n.d.). PREDICTION OF PARTIAL CHURNERS AND BEHAVIOURAL LOYAL CUSTOMERS THROUGH BEHAVIOURAL HISTORICAL CUSTOMER DATA PUBLIC, NON CONFIDENTIAL VERSION.

Kriti. (2019). Customer churn: A study of factors affecting customer churn using machine learning (0 ed.). Iowa State University. https://doi.org/10.31274/cc-20240624-464

Lemmens, A., & Gupta, S. (2020). Managing Churn to Maximize Profits. Marketing Science, 39(5), 956–973. https://doi.org/10.1287/mksc.2020.1229

Ling, H., Qian, C., Kang, W., Liang, C., & Chen, H. (2019). Construction and Building Materials. Elsevier Ltd. https://doi.org/10.1016/j.conbuildmat.2019.02.071

Liu, Y., & Wang, Y. (2012). New Machine Learning Algorithm: Random Forest. Springer.

Matuszelański, K., & Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. Journal of Theoretical and Applied Electronic Commerce Research, 17(1), 165–198. https://doi.org/10.3390/jtaer17010009

Mencarelli, R., & Rivière, A. (2015). Perceived value in B2B and B2C: A comparative approach and cross-fertilization. Marketing Theory, 15(2), 201–220. https://doi.org/10.1177/1470593114552581

Mirkovic, M., Lolic, T., Stefanovic, D., Anderla, A., & Gracanin, D. (2022). Customer Churn Prediction in B2B Non-Contractual Business Settings Using Invoice Data. Applied Sciences, 12(10), 5001. https://doi.org/10.3390/app12105001

Nand Kumar, C. N. (2017). Comparative Analysis of Machine Learning Algorithms for their Effectiveness in Churn Prediction in the Telecom Industry. International Research Journal of Engineering and Technology, 04(08), 485–489.

Nhu, V.-H., Mohammadi, A., Shahabi, H., Ahmad, B. B., Al-Ansari, N., Shirzadi, A., Geertsema, M., Kress, V. R., Karimzadeh, S., Kamran, K. V., Chen, W., & Nguyen, H. (2020). Landslide Detection and Susceptibility Modeling on Cameron Highlands (Malaysia): A Comparison between Random Forest, Logistic Regression and Logistic Model Tree Algorithms.

Park, W., & Ahn, H. (2022). Not All Churn Customers Are the Same: Investigating the Effect of Customer Churn Heterogeneity on Customer Value in the Financial Sector. Sustainability, 14(19), 12328. https://doi.org/10.3390/su141912328

Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification. Baltic Journal of Modern Computing, 5(2). https://doi.org/10.22364/bjmc.2017.5.2.05

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. John Wiley & Sons, Inc., 9(3). https://doi.org/10.1002/widm.1301

Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2018). The International Journal of Human Resource Management. Informa UK Limited, 1617–1643. https://doi.org/10.1080/09585192.2017.1416655

Rushi, W. A., & Pradhan, V. (2023). Factors Influencing Customer Grocery Shopping Behaviour Amid Covid-19 Pandemic. CARDIOMETRY, 25, 743–755. https://doi.org/10.18137/cardiometry.2022.25.743755

Russo, I., Confente, I., Gligor, D. M., & Autry, C. W. (2016). To be or not to be (loyal): Is there a recipe for customer loyalty in the B2B context? Journal of Business Research, 69(2), 888–896. https://doi.org/10.1016/j.jbusres.2015.07.002

Sahani, N. (2021). GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. Ecological Informatics.

Salma, N., & Aprianingsih, Ph.D, A. (2021). Customer Churn Analysis: Analyzing Customer Churn Determinants on an ISP Company in Indonesia. Buletin Pos Dan Telekomunikasi, 29–40. https://doi.org/10.17933/bpostel.2021.190103

Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning (Vol. 36). Springer US. https://doi.org/10.1007/978-1-4899-7641-3

Tamaddoni Jahromi, A., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. Industrial Marketing Management, 43(7), 1258–1268. https://doi.org/10.1016/j.indmarman.2014.06.016

Wadikar, D. (2020). Customer Churn Prediction. https://doi.org/10.21427/KPSZ-X829