ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 14 (2024) No. 12, pp. 141-156
DOI:10.33168/JSMS.2024.1209

A Comparative Study on Data Mining Algorithms for Predicting Customer Churn

Ismail Azh-Zhafir Rohaga, Gunawan Wang

Information System Management Department, BINUS Graduate Program - Master of Information System Management, Bina Nusantara University Jakarta, Indonesia 11480

ismail.rohaga@binus.ac.id; gwang@binus.edu

Abstract. This study investigates predicting customer churn in Indonesia's dynamic forestry industry through machine learning techniques. Models were developed leveraging five years of transactional data from Perum Perhutani, contrasting ten algorithms, we address the research problem by following CRISP-DM methodology. The Gradient Boosted Tree achieved the highest accuracy of 85.60% and AUC of 0.940 in discriminating churners from non-churners. Analysis of feature importance reveals transaction timing attributes as most influential, underscoring recency, frequency, and intervals as pivotal indicators of attrition. Findings provide actionable insights guiding customer retention initiatives and risk mitigation strategies for industrial suppliers in turbulent economic environments.

Keywords: Customer Churn, Data Mining, Machine Learning, Predictive Modelling, Feature Engineering.

1. Introduction

In the dynamic landscape of forestry management and wood product distribution in Indonesia, Perhutani stands as a pivotal state-owned enterprise entrusted with the stewardship of Java island's forests. As the primary custodian, Perhutani is responsible for the management and sale of various forest products, with a primary focus on catering to industrial customers. In this distinctive B2B model, Perhutani acts as a key supplier to industries that rely on a steady wood supply for their manufacturing needs. Customer churn refers to customers ending their relationship or transactions with a company (Ahn et al., 2020; Chen et al., 2015; Mirkovic et al., 2022; Ribeiro et al., 2022), holds profound consequences for Perhutani. Recognizing the value of retaining loyal customers, Perhutani's 2022 annual report outlines proactive strategies by the marketing division to retain lost customers through outreach efforts. Beyond potential revenue loss (Mirkovic et al., 2022), which can impact the company's growth and profitability (Rust et al., 2004), churn also bears implications for the company's reputation (Chandran, 2020). Predicting churn is pivotal for businesses, offering advantages such as the identification of high-risk customers and the efficient allocation of resources (Chen et al., 2015). Due to the substantial financial consequences associated with accurately predicting churn (Ahn et al., 2020; Chandran, 2020; Chen et al., 2015; Mirkovic et al., 2022; Rust et al., 2004), this study is in harmony with Perhutani's goals by developing a predictive model for customer churn through the application of data mining techniques. The study will utilize a combination of Traditional Machine Learning methods, Ensemble Learning, and Deep Learning methods (Dalli, 2022) to analyze patterns and factors influencing customer churn.

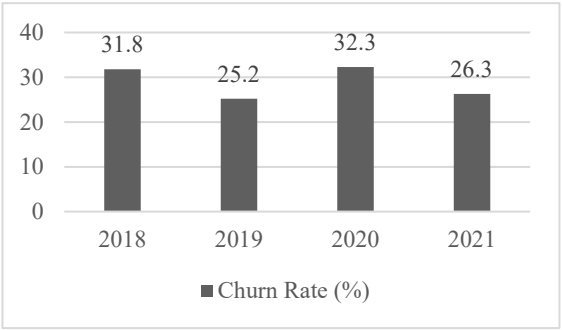

Fig. 1: Churn Rate

In Perhutani's business strategy, customer churn is defined by a customer's absence of purchases in the subsequent year. The dataset, representing the sales data of wood segment products in Perhutani, spans from 2018 to 2022, coinciding with the implementation of the e-commerce online platform in 2018. Unfortunately, data for 2023 is not available. The churn rates for the years 2018 to 2021 indicate fluctuations in customer attrition within this specified period. In 2018, the churn rate stood at 31.8%, followed by a decrease to 25.2% in 2019. However, there was a subsequent increase in 2020, reaching 32.3%, and then a decrease again to 26.3% in 2021. These variations underscore the dynamic nature of customer churn within the business.

By leveraging data mining methods and predictive modeling, it becomes possible to analyze patterns and factors influencing customer churn. This proactive approach enables businesses, such as Perhutani, to anticipate and address potential churn, thereby fostering customer retention strategies and ultimately mitigating financial repercussions associated with high churn rates. As stated previously, the primary goal of this project is to use data mining approaches to create a predictive model for customer churn at Perum Perhutani. As a company entrenched in the industrial or B2B market, this study aims to contribute to a deeper understanding of customer churn in B2B scenarios, particularly within the unique context of forest product offerings as encountered by Perhutani. In the unique context of the forestry industry, B2B customer churn presents specific challenges. Compared to B2C scenarios, a single B2B customer may possess a higher transactional value due to the nature of business-to-business interactions

(ÇALLI & KASIM, 2022; Chen et al., 2015; Figalist et al., 2019; Jahromi et al., 2014; Jamjoom, 2021; Marín Díaz et al., 2022; Mirkovic et al., 2022).

While previous studies have provided valuable insights into customer churn across various industries, a notable gap exists in the explicit examination of the forestry industry. The unique characteristics of the forestry sector, involving wood product distribution and B2B interactions, necessitate a dedicated exploration of customer churn dynamics within this specific context. Furthermore, the performance of data mining algorithms may vary across industries, and there is a distinct absence of a detailed model comparison specifically within the B2B context of the forestry industry. This study aims to fill this gap by conducting a meticulous evaluation of different algorithms. The following sections are theoretical background, research methodology, results, and discussion, then the study will be closed with a conclusion.

2. Theoretical Background

2.1. Customer Churn

Also known as customer attrition, customer turnover, and defection or departure. Customer churn refers to the state in which customers terminate their relationships or transactions with a company or brand (Ahn et al., 2020; Chen et al., 2015; Mirkovic et al., 2022). Understanding and preventing customer churn, especially among profitable customers, is imperative for the sustained survival of a business (Ribeiro et al., 2022). Causes of churn encompass factors such as pricing, service failures, the company's responses to service issues, dissatisfaction, increased costs, low quality, resource limitations, and privacy concerns (Mirkovic et al., 2022; Ribeiro et al., 2022). To identify churners, i.e., the churned customer, a common approach involves utilizing a time window to track customer activity (Ganeson et al., 2022). This process initiates a timer upon a customer's purchase, monitoring their activity within a predefined time frame. If there is no activity during this window, the customer is deemed to have churned. The duration of the churn window is usually consistent for all customers. For instance, if three months is set as the window, known as the Trimester [T]. Customers not making purchases during this period or spending less than forty percent compared to the previous T are labeled as churners. Additionally, customer churn can be tracked through three distinct methods: daily, monthly, and binary (Ahn et al., 2020). The interval in which a consumer is analyzed determines whether he or she is churning or not. For example, if a consumer does not make a purchase for three months in a row, they are considered churn. Contract continuity defines binary churn, which is important in contractual circumstances. Non-contractual scenarios involve firms establishing criteria for customer inactivity or disloyalty and classifying consumers who match these criteria as binary churn. Notably, the definition of customer churn in this study aligns with Perum Perhutani's perspective, considering a client as churn if they refrain from making any purchases for twelve consecutive months.

2.2. Data Mining

In simple terms, data mining involves uncovering or "mining" knowledge from extensive datasets (Sharda et al., 2018). Data mining or knowledge discovery is the procedure of identifying patterns, and relationships, and gaining understanding from massive data sets using techniques such as AI, machine learning, statistics, mathematics, and database systems from large data sets (Lin, 2017; Sharda et al., 2018). These patterns may manifest as business rules, affinities, correlations, trends, or prediction models. Despite the relatively recent introduction of the term "data mining," the underlying concepts have longstanding roots. Many techniques employed in data mining trace their origins to traditional statistical analysis and artificial intelligence research conducted since the early 1980s.

In general, data mining tasks can be divided into three categories: prediction, association, and grouping. Learning algorithms in data mining are classified as supervised or unsupervised based on how patterns are generated from past data. In supervised learning algorithms, training data includes descriptive features as well as output class variables or outcome variables. Unsupervised learning

training data, on the other hand, only provides descriptive features. A set method is often followed to carry out data mining initiatives in a systematic manner. Researchers and data mining practitioners have offered many procedures (workflow or step-by-step techniques) based on best practices to improve the success of data mining project execution. For successful implementation, data mining studies should be viewed as a process following standard methodologies rather than merely a collection of automated software tools and techniques. Notably, methodologies such as CRISP-DM, SEMMA, and KKD are recognized, with CRISP-DM being the most popular in the data mining process (Sharda et al., 2018).

2.3. Prediction Model

Predictive analytics forecasts future events using previous data and statistical models. Its popularity has grown as a result of technological breakthroughs like as artificial intelligence and machine learning (Ribeiro et al., 2022), enabling organizations to predict trends and behaviors accurately, ranging from seconds to years into the future (Haw et al., 2022). In line with this trend, there has been a substantial increase in the application of machine learning techniques for predicting customer churn across diverse industries. These techniques can be grouped into three categories: Traditional ML methods, Ensemble Learning, and Deep Learning (Dalli, 2022).

Table. 1: Machine Learning Techniques used in this study

Categories	ML Techniques	Description
Traditional ML methods	Logistic Regression (Log. R)	Suited for binary classification, Logistic Regression estimates probabilities using a logistic function, predicting the likelihood of an instance belonging to a particular class.
	Naïve Bayes (NB)	NB is a probabilistic algorithm based on the Bayes theorem that assumes conditional independence of features given the class label. It is frequently employed in categorization assignments.
	K-Nearest Neigbors (KNN)	Used for classification as well as regression tasks. It classifies a data point in the feature space by evaluating the prevailing class among its k-nearest neighbors. The method finds the k nearest neighbors and then uses majority voting or average to decide the class or value of the new point.
	Decision Tree (DT)	Decision Trees repeatedly partition data depending on features, resulting in a tree-like structure with nodes representing decisions and leaves denoting class labels. Both classification and regression tasks are commonly performed using DT.
	Support Vector Machine (SVM)	SVM determines the optimum hyperplane for separating data into classes while maximizing the margin between them. It is used for classification as well as regression tasks.
Ensemble Learning	Random Forest (RF)	Random Forest is an ensemble learning method that combines the predictions of several decision trees. It works well with high-dimensional data and is used for classification and regression applications where accuracy and generalization are critical.
	Gradient Boosted Tree (GB Tree)	An ensemble of regression or classification trees constitutes a gradient-boosted model. It improves predictions by applying weak classification algorithms sequentially, resulting in a succession of decision trees. It is used for both classification and regression tasks, excelling in scenarios requiring high predicted accuracy.
	AdaBoost (ADA)	AdaBoost, or Adaptive Boosting, is a meta-algorithm enhancing the performance of learning algorithms. It adapts by tweaking subsequent classifiers to favor instances misclassified

Categories	ML Techniques	Description
		by previous ones. Despite sensitivity to noise, it mitigates overfitting, and by assigning greater weight to misclassified examples, it prioritizes improvement in the classification of challenging instances during each iteration.
Deep Learning	Deep Learning (DL)	Based on a multi-layer neural network trained with stochastic gradient descent. Features advanced functions and regularization for high predictive accuracy. Nodes contribute to the global model via asynchronous model averaging
	Neural Network (NN)	Inspired by biological neural networks, artificial neurons process information through interconnected layers. Adapts and learns from data, making it versatile for diverse machine learning tasks.

2.4. Feature Development – LRFM Model

The RFM (Recency, Frequency, Monetary) framework is a useful method in churn prediction tasks; RFM variables have been shown to play a significant impact in forecasting customer churn (Jahromi et al., 2014). It simplifies consumer behavior analysis by taking into account the recency of transactions, the frequency of interactions, and the monetary value (Mitrović et al., 2017; Sun et al., 2023). The essential takeaway is that the effectiveness of churn prediction is dependent on these explanatory qualities rather than the specific modeling technique used. In the 1990s, scholar Arthur Hughes proposed the RFM model, which laid the theoretical groundwork for calculating client lifetime value (Sun et al., 2023). The RFM model is the most widely utilized way for characterizing customer interactions and, more broadly, consumer behavior. In reaction to a relevant event, the RFM model establishes specific customer-level metrics: 1) the recency of the event, indicating the time gap between the event's most recent occurrence and the reference point in time; 2) the frequency of the event, reflecting the number of occurrences within the observed timeframe; and 3) the monetary value associated with the customer's spending on the event.

The RFM model has three essential characteristics that have contributed to its success in the literature (Mitrović et al., 2017). For starters, it is based on a simple concept, making it easily comprehensible and computationally tractable. Furthermore, the only assumption that RFM variables require is that a customer's future behavior and value can be forecasted based on the customer's past behavior. Second, because the event definition is flexible and adaptable to diverse situations, the RFM model applies to a variety of domains (banking, retail, and telecommunications). Finally, the RFM variables have a high predictive value when utilized as explanatory factors. This model is extended by the LRFM (Mirkovic et al., 2022), which includes the length of the relationship, i.e. how long is the time interval between the first transaction to the latest transaction; and a new extension corresponds to the LRFMP model (Chen et al., 2015), which includes profitability.

Table. 2: Previous Study

Domain	#Instances	Features	Algorithm	Metrics	Result
B2B Retail (Sun et al., 2023)	541909; 1% churn rate	RFM	Log. R, DT, SVM, KNN, NB, RF, AdaBoost, GB Tree, Mul. Perceptron	AUC, Accuracy	Customers who have made a purchase recently and have also made a purchase in the past are more likely to make another purchase in the future. The AUC

Domain	#Instances	Features	Algorithm	Metrics	Result
					and accuracy scores of boosting fusion methods like AdaBoost and gradient-boosting decision trees are relatively high.
B2B Wholesale (Mirkovic et al., 2022)	3470; 5-38% churn rate	LRFM	Log. R, SVM, RF	AUC, TDL	The total number of invoices (Frequency) is the most important indicator of churn. In terms of overall performance, the random forest classifier came out on top in both measures.
B2B FMCG (Jahromi et al., 2014)	11021; 28% churn rate	RFM	DT, DT Cost-sensitive, Log. R, Boosting, Rand classifier	AUC, Cumulative lift curve	The likelihood of a customer churning increases with the amount of purchases (frequency) and the time since the last order (recency). The best model is the boosting technique.
Retail (Sweidan et al., 2022)	~200000; 50.69% churn rate	Customer Demographic, RFM	GB, DT, Log. R	Accuracy, Precision, Recall, AUC	The churn rate is heavily influenced by the number of previous purchases. The GB model outperforms all others.
B2B Software **(ÇALLI & KASIM, 2022)**	1951; 74.97% churn rate	Customer Demographic, services	DT, Log. R, NB, KNN, RF, Neural Network.	Accuracy	The study found that the number of customers, products, invoices (frequency), and orders are the most important factors in determining customer churn. The Random Forest algorithm achieves the best prediction accuracy.
B2B Software (Marín Díaz et al., 2022)	200615; 1.06% churn rate	RFID	XGboost, Gaussian NB, Log. R, RF, KNN, DT, SVM	AUC, Accuracy	According to the findings, the criterion importance is the most important in evaluating the churn, followed by frequency, duration, and recency. Xgboost performs the best of all.
B2B	69170;	LRFMP	C4.5, MLP,	Accuracy,	Length, recency, and

Domain	#Instances	Features	Algorithm	Metrics	Result
Logistics (Chen et al., 2015)	2% churn rate		SVM, Log. R	Precision, Recall, F1	monetary are identified as significant churn predictors, however, frequency is only a significant predictor when the variability of the three variables changes. DT (C4.5) has superior performance.

3. Methodology

CRISP-DM methodology guides the development of the predictive model in this study. The introductory section aligns with Business Understanding, and the Deployment will conclude at the Evaluation stage, this study will be utilizing RapidMiner Studio 10.2. The data for this study was derived from transactional records of log wood transactions. The data was extracted in spreadsheet format from the database of the Perhutani store's online sales platform. 472.208 records of data from 2018 to 2022 were available for this study to process.

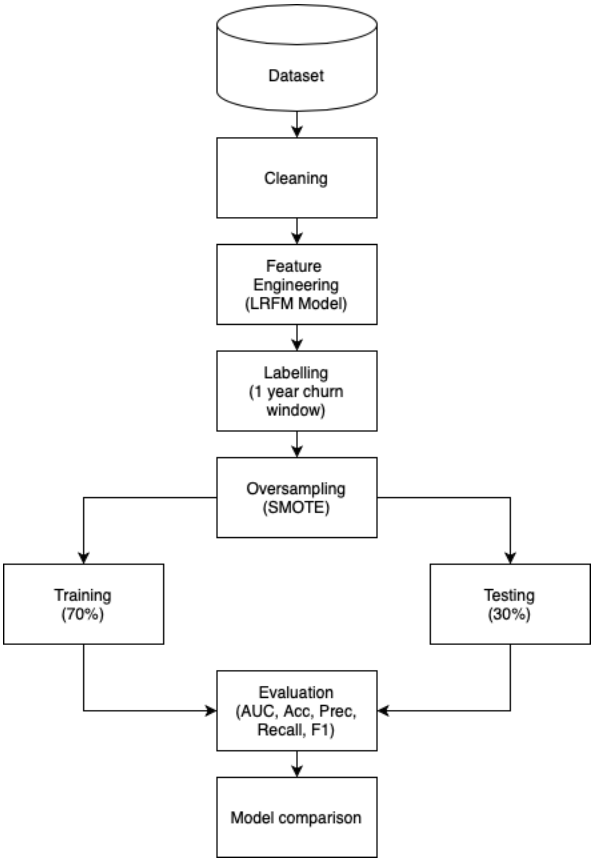

Fig. 2: The study pipeline

In the initial analysis, several issues were identified, including missing values and data duplications. These problems were addressed through a series of steps. Firstly, missing values were filled with zeros to ensure data integrity. Next, invalid values, such as #DIV/0! errors were replaced with zeros to prevent

disruption to the analysis. Additionally, data duplications were removed to ensure result accuracy. After completing the initial data cleaning steps, the next stage involves data transformation, this study will conduct feature engineering, creating relevant new variables from the available data source. Given that the data available for processing is obtained in spreadsheet format, the feature engineering process is carried out using a Pivot Table. Then, for the creation of non-existent features, namely length (L) and recency (R), special formulas will be employed. This data transformation process converts transactional data into customer profile data that reflects customer behavior and characteristics.

Table. 3: Dataset Features

Feature	Description	Statistic
ID	Buyer ID	3.680 Buyer
Min of SALE DATE	Longest transaction date within the observation period for involved customers.	Min: 02/01/18 Max: 28/12/21 Average: - Deviation: -
Max of SALE DATE	Latest transaction date within the observation period for involved customers.	Min: 02/01/18 Max: 30/12/21 Average: - Deviation: -
Count of SALE DATE (F)	Transaction frequency within the observation period for involved customers.	Min: 1 Max: 6655 Average: 128,3 Deviation: 380,5
Min of INTERVAL	The smallest time interval for involved customers.	Min: 0 Max: 1291 Average: 4,7 Deviation: 43,3
Max of INTERVAL	Largest time interval for involved customers.	Min: 0 Max: 1291 Average: 151,7 Deviation: 170,1
Average of INTERVAL	Average time interval for involved customers.	Min: 0 Max: 1291 Average: 21,8 Deviation: 52,9
StdDev of INTERVAL	The standard deviation of the interval for involved customers.	Min: 0 Max: 586,1 Average: 32,4 Deviation: 50,014
Min of AMOUNT	Smallest transaction amount in IDR during the observation period for involved customers.	Min: 27 Max: 179.018.270 Average: 2.298.341,68 Deviation: 7.999.619,84
Max of AMOUNT	Largest transaction amount in IDR during the observation period for involved customers.	Min: 4026 Max: 1.313.957.021 Average: 51.000.247,13 Deviation: 92.274.162,5

Feature	Description	Statistic
Average of AMOUNT	Average transaction amount in IDR during the observation period for involved customers.	Min: 4026 Max: 209.829.275,6 Average: 10.352.562,097 Deviation: 14.116.417,98
StDev of AMOUNT	The standard deviation of the transaction amount in IDR during the observation period for involved customers.	Min: 0 Max: 222.785.891,3 Average: 9.093.004,51 Deviation: 13.239.965,4
The sum of AMOUNT (M)	Total transaction amount in IDR during the observation period for involved customers.	Min: 4026 Max: 241.402.937.375 Average: 1.674.974.005,53 Deviation: 8.598.930.616,76
Min of VOL	The smallest volume of products involved in transactions during the observation period for involved customers.	Min: 0 Max: 667,96 Average: 1,77 Deviation: 13,01
Max of VOL	The largest volume of products involved in transactions during the observation period for involved customers.	Min: 0,008 Max: 1166,74 Average: 30,85 Deviation: 60,53
Average of VOL	The average volume of products involved in transactions during the observation period for involved customers.	Min: 0,008 Max: 667,96 Average: 7,15 Deviation: 16,75
StdDev of VOL	The standard deviation of the volume of products involved in transactions during the observation period for involved customers.	Min: 0 Max: 291,31 Average: 5,73 Deviation: 10,09
Sum of VOL	The total volume of products involved in transactions during the observation period for involved customers.	Min: 0,008 Max: 111.220,93 Average: 824,89 Deviation: 3638,68
TRX_LEN (L)	Length of the initial transaction period and final transaction period in days for involved customers.	Min: 0 Max: 1.458 Average: 584,1 Deviation: 519,1
TRX_REC (R)	Time distance from the last transaction to the prediction time in days for involved customers.	Min: 2 Max: 1.460 Average: 477,31 Deviation: 476,69
TARGET	The target variable indicates whether the customer churned or not, where a value of 0	Class 0: 1.524 Class 1: 2.156 Deviation: 0,49

Feature	Description	Statistic
	indicates that the customer did not churn, and a value of 1 indicates that the customer churned.	

The new features or attributes, namely "TRX_LEN (L)," indicate the length of the time interval between the first and last transactions within the observation period, while "TRX_REC (R)" records the time interval from the last transaction to the prediction time. Another data cleaning step was undertaken to replace #DIV/0! errors with zeros. This ensured that the data used for analysis was cleaner and free from values that could interfere with the final results. The ID, Min, and Max of SALE DATE were filtered out in the modeling stage. In this analysis, the "TARGET" attribute serves as the target variable, where a value of 0 indicates that the customer did not churn, and a value of 1 indicates a customer who churned. The labeling criteria adhere to the defined churn concept from the Theoretical Background section, which considers a client who refrains from making any transactions for twelve consecutive months. Starting from the forecast origin at the end of 2021, this study examines whether customers made purchases in the subsequent 12 months or 2022.

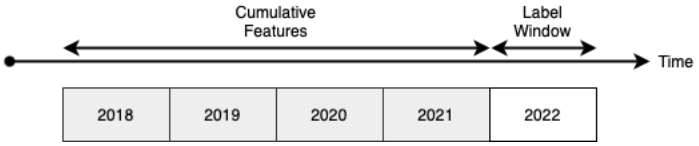

Fig. 3: Churn Window

To train and evaluate the model's performance, the provided dataset will be split into training and testing sets using a simple split approach. The training set is used to train the model, while the testing set is used to assess its performance on previously unseen data. This study will use a 70:30 split ratio for the training and testing sets, incorporating stratified sampling parameters. To address this class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) will be applied to the training set. This technique aims to generate synthetic data in the minority class (churn) to create a balance between the two classes. By applying this technique, it is expected that the model to be built will be better equipped to handle class imbalance and produce more accurate prediction results. The operator is tuned with its default configurations. In the modeling stage of this study, 10 different algorithms will be utilized, namely k-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Gradient Boosting Tree (GB), AdaBoost (ADA), Deep Learning (DL), Neural Network (NN), Logistic Regression (Log. R), and Support Vector Machine (SVM). The training set is subjected to a crucial step in model evaluation – the 10-fold cross-validation operator. It is noteworthy that during the cross-validation process, all algorithm parameters are retained in their default configurations. This decision was made to maintain consistency and simplify the initial evaluation phase.

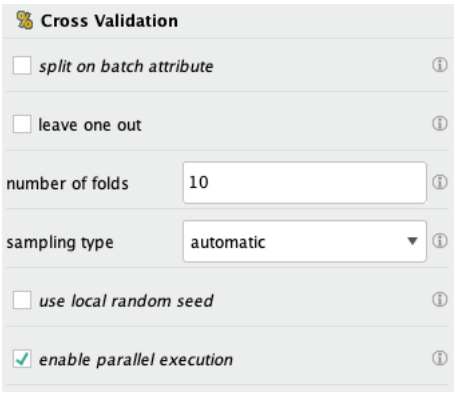

Fig. 4: Cross Validation Operator Parameter

In the Evaluation stage, an assessment will be conducted on the performance and reliability of the built predictive model. The primary objective is to measure the model's ability to predict customer churn. In selecting the evaluation metrics for our predictive model, we considered industry best practices and drew inspiration from a comprehensive review of existing literature on customer churn prediction. AUC and Accuracy emerged as key metrics due to their widespread usage across numerous studies in this domain (ÇALLI & KASIM, 2022; Chen et al., 2015; Marín Díaz et al., 2022; Mirkovic et al., 2022; Sun et al., 2023; Sweidan et al., 2022). AUC provides a robust measure of the model's ability to discriminate between positive and negative instances, while Accuracy offers an overarching view of correct predictions. Moreover, to delve deeper into the model's performance, we incorporated precision, recall, and F1 score. Precision measures the accuracy of positive predictions, focusing on the relevance of identified churn cases. Recall assesses the model's capacity to capture all actual positive instances, ensuring a comprehensive understanding of the churn population. F1 score, being the harmonic mean of precision and recall, strikes a balance between precision-oriented and recall-oriented models. The inclusion of these metrics aims to provide a nuanced evaluation, considering not only overall correctness but also the model's effectiveness in correctly identifying positive instances and minimizing false positives. This approach aligns with established practices in classification tasks and enhances the interpretability of our model's performance.

4. Result & Discussion

4.1. Evaluation Performance

The predictive modeling process was conducted using ten distinct algorithms and was evaluated on their ability to predict customer churn based on a set of relevant features derived from transactional records of logwood transactions. Several models exhibited remarkable performance across multiple metrics. Among them, are RF, GB Tree, DL, NN, Log. R and SVM emerged as top-performing models, showcasing a harmonious blend of accuracy, precision, recall, and AUC. Random Forest model showcased robust predictive capabilities, achieving an accuracy of 85.96%. This high accuracy underscores the model's proficiency in correctly identifying both churn and non-churn instances. The AUC, reaching 0.936, further solidifies its effectiveness in distinguishing between positive and negative cases. These results affirm the model's suitability for accurate customer churn predictions in the context of log wood transactions.

Table. 4: Performance Result

Metrics	KNN	NB	DT	RF	GB Tree	ADA	DL	NN	Log. R	SVM
AUC	0.755	0.890	0.867	0.936	0.940	0.900	0.937	**0.941**	0.939	**0.941**
Acc	69.02%	79.44%	84.78%	85.96%	85.60%	84.78%	**86.87%**	85.69%	84.96%	85.42%
Prec	76.43%	91.83%	92.69%	91.55%	88.73%	92.69%	90.22%	90.95%	92.12%	**92.93%**
Recall	68.16%	71.25%	80.37%	83.77%	86.40%	80.37%	**87.02%**	83.93%	81.30%	81.30%

F1	72.06%	80.24%	86.09%	87.49%	87.55%	86.09%	**88.59%**	87.30%	86.37%	86.73%

Gradient Boosted Tree model emerged as a standout performer in our analysis. With an accuracy of 85.60%, and AUC of 0.940. This signifies a robust discriminatory capacity and indicates that the model effectively balances true positive and false positive rates. The GB Tree's harmonious blend of accuracy and AUC positions it as a reliable choice for predicting customer churn in log wood transactions. Deep Learning, achieves the highest accuracy among the models, standing at 86.87%. This accuracy metric reflects the model's proficiency in correctly classifying instances. While its AUC score of 0.937 is slightly lower than NN and SVM, the balanced performance in accuracy signifies Deep Learning's robust predictive capabilities, particularly in scenarios where discerning nuanced patterns is crucial. Both Neural Network and Support Vector Machine emerge as formidable contenders, showcasing the highest AUC scores in the model lineup. Neural Network, with an AUC of 0.941, and SVM, with an AUC of 0.941 as well, demonstrate a remarkable ability to discriminate between positive and negative instances. Logistic Regression, showcased reliability with an accuracy of 84.96%. The AUC, at 0.939, reinforces its ability to effectively discriminate between positive and negative instances. While offering a simpler model structure, Logistic Regression demonstrated a well-balanced trade-off between precision and recall. This makes it an attractive option for businesses where interpretability is paramount, without compromising predictive performance.

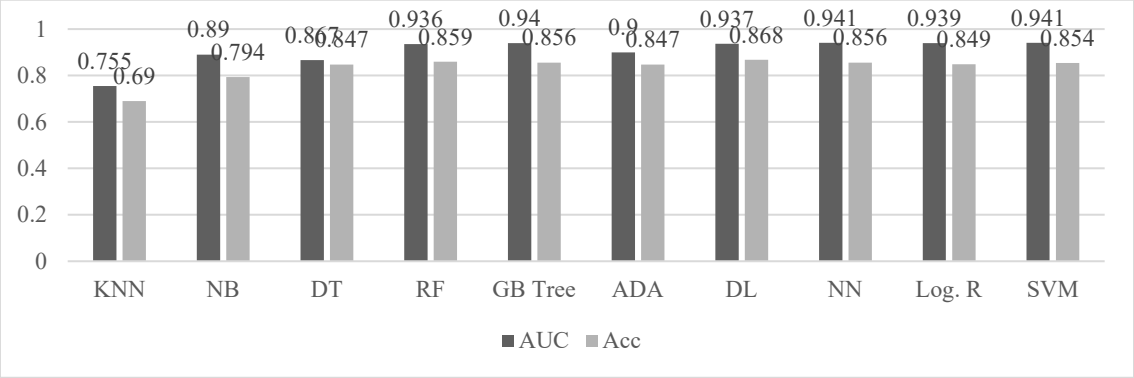

Fig. 4: AUC and Accuracy Performance

SVM (Support Vector Machine) emerges as the top performer in precision, achieving the highest score of 92.93%. This indicates that SVM excels in accurately predicting customer churn when it makes a positive prediction. A high precision score is crucial in scenarios where minimizing false positives is a priority, ensuring that the predicted churn instances are highly reliable. As for recall and F1-score metrics Deep Learning (DL) leads with a score of 87.02% and 88.59% respectively. DL showcases its capacity to identify a substantial number of customers who genuinely churned, providing businesses with a comprehensive understanding of potential attrition cases. This high recall is pivotal in customer churn prediction, as it minimizes the risk of overlooking critical churn instances and allows for proactive retention strategies. Furthermore, DL achieves an impressive F1 score indicating a well-balanced performance between precision and recall. This balanced metric underscores DL's ability to make accurate positive predictions while simultaneously capturing a significant portion of true positives.

4.2. Feature Weight

In our comprehensive evaluation of ten distinct algorithms for predicting customer churn based on logwood transaction records, five models have offered valuable insights into the significance of specific features. These insights were derived from a meticulous analysis of feature weights, shedding light on the attributes that play a crucial role in determining customer churn. The algorithms providing this feature importance information include Decision Tree, Random Forest, Gradient Boosted Tree, Deep Learning, and Logistic Regression.

Table. 5: Feature Weight Result

Feature	Weight				
	DT	RF	GB Tree	DL	LR
Count of SALE DATE (F)	0.032	**0.110**	331.63	0.70	-0.12
Min of INTERVAL	-	0.027	1.84	0.6	0.06
Max of INTERVAL	0.034	0.08	41.85	0.61	0.11
Average of INTERVAL	**0.2**	0.075	180.34	0.65	0.05
StdDev of INTERVAL	0.023	0.072	215.73	0.6	0.19
Min of AMOUNT	0.157	0.085	10.61	0.69	-0.016
Max of AMOUNT	-	0.082	27.51	0.65	0.25
Average of AMOUNT	-	0.056	107.57	0.65	-0.00
StDev of AMOUNT	-	0.027	43.45	0.66	-0.01
The sum of AMOUNT (M)	0.03	0.043	34.44	0.59	-0.38
Min of VOL	0.128	0.049	141.75	0.78	1.44
Max of VOL	0.075	0.048	47.84	0.74	-0.20
Average of VOL	-	0.030	99.69	0.68	-0.21
StdDev of VOL	0.03	0.031	6.06	0.69	0.30
Sum of VOL	-	0.039	27.91	0.57	0.01
TRX_LEN (L)	0.124	0.042	121.49	0.58	-0.44
TRX_REC (R)	0.161	0.095	**15173.42**	**1.0**	**3.14**

One striking consensus among the algorithms is the pronounced importance assigned to time-related features. Attributes such as TRX_REC (time distance from the last transaction), Count of SALE DATE (transaction frequency), and various measures of time intervals (Average, StdDev, Min, Max of INTERVAL) consistently emerge as influential. This collective emphasis underscores the significance of temporal patterns in predicting customer churn. The algorithms recognize that the recency of transactions, frequency, and intervals between transactions provide crucial insights into customer behavior and potential attrition. Transactional attributes, such as transaction amounts (Min, Max, Average, Sum of AMOUNT) and volume-related metrics (Min, Max, Average, Sum of VOL), also garner substantial importance across the algorithms. This alignment suggests a shared recognition of transactional patterns as key indicators of customer churn. The models acknowledge that the nature and volume of transactions contribute significantly to the predictive accuracy of the algorithms.

While commonalities exist, each algorithm introduces unique considerations into the feature importance landscape. Decision Tree and Random Forest accentuate the significance of the Average of INTERVAL and Count of SALE DATE, revealing their reliance on transactional frequency and timing patterns. Gradient Boosted Tree, being an ensemble model, shares similar priorities but with a notable emphasis on the TRX_REC, suggesting a focus on transactional recency. Deep Learning, as a neural network-based model, attributes the highest weight to TRX_REC, reinforcing its position as a crucial temporal factor. Logistic Regression, with its linear approach, accentuates the importance of transaction-related features (Min of VOL, StdDev of VOL) with also TRX_REC being the most important.

4.3. Discussion

The findings of our study contribute valuable insights into the predictive modeling of customer churn based on logwood transactions, corroborating and expanding upon existing research in this domain. In line with previous studies, our investigation underscores the significance of transactional features, particularly the pivotal role played by transaction frequency (Count of SALE DATE) and recency

(TRX_REC) in forecasting customer behavior (ÇALLI & KASIM, 2022; Jahromi et al., 2014; Marín Díaz et al., 2022; Mirkovic et al., 2022; Sun et al., 2023; Sweidan et al., 2022). The consistent emphasis on time-related features across studies reaffirms the notion that both the frequency, recency, and time intervals of transactions are fundamental indicators of future customer actions.

Furthermore, the examination of feature importance across diverse predictive models in our study reveals a nuanced perspective on the role of transaction frequency in predicting customer churn, adding a layer of complexity to the understanding. While the Random Forest, Gradient Boosted Tree, and Deep Learning models emphasize the significance of "Count of SALE DATE," contradicting the findings reported by Chen et al., these results diverge when considering other models such as Decision Tree and Logistic Regression. In these models, the "Count of SALE DATE" or frequency does not carry as much weight (Chen et al., 2015), suggesting a nuanced relationship with churn prediction. The discrepancy in results prompts a critical reflection on the intricacies of the data and the underlying assumptions of each model and it underscores the intricate nature of customer behavior within the logwood transactions domain, suggesting that the influence of frequency may be contingent on various contextual factors. While transaction frequency emerges as a crucial factor in some models, it may not play an equally dominant role in others. This diversity in model outcomes highlights the need for a comprehensive approach in predictive modeling, where insights from multiple algorithms are considered to obtain a more holistic understanding of feature importance. Additionally, in our study, an intriguing observation surfaced, indicating that the "length" variable (TRX_LEN) was exclusively important within the Decision Tree algorithm, aligning with the findings reported by Chen et al. This nuanced finding adds a layer of complexity to the predictive modeling landscape, suggesting that the impact of transaction length on customer churn is more pronounced in the context of decision trees compared to other algorithms employed in our study. Additionally, our results reinforce the recognition of transactional attributes, such as transaction amounts and volume-related metrics, as significant indicators of customer churn.

Ultimately, businesses seeking to enhance customer retention strategies can leverage these feature-importance insights. The emphasis on time-related features highlights the significance of temporal patterns in predicting churn. This insight enables businesses to implement timely interventions, identifying customers with prolonged inactivity for personalized re-engagement efforts. Also, the recognition of transaction frequency as a key indicator suggests the potential for loyalty programs or exclusive offers tailored to customers with high transactional engagement. Leveraging transaction amounts and volume-related metrics provides an avenue for personalized retention strategies, aligning offers with individual spending patterns. Moreover, businesses can strategically design communication initiatives based on transactional patterns, addressing specific needs or concerns to proactively prevent churn. Integrating time-related factors into customer engagement programs further enhances relationships, allowing for personalized celebrations of milestones and anniversaries. Ultimately, translating these insights into targeted retention measures empowers businesses to foster customer loyalty and mitigate churn in the dynamic landscape of logwood transactions.

Our study further aligns with previous research in evaluating the performance of predictive models. Notably, the robust predictive capabilities of our Random Forest and Gradient Boosted Tree models resonate with findings from (ÇALLI & KASIM, 2022; Mirkovic et al., 2022; Sun et al., 2023; Sweidan et al., 2022) regarding the effectiveness of ensemble and boosting techniques in predicting customer churn. Practical implications of our findings suggest that businesses can leverage Random Forest, Gradient Boosted Tree, and other identified top-performing models to enhance their customer churn prediction strategies. By considering temporal patterns and transactional attributes, organizations can proactively identify customers at risk of churn and implement targeted retention strategies. The alignment with previous studies provides a robust foundation for the adoption of these models in diverse industries, emphasizing the generalizability and reliability of the proposed approach in predicting customer churn in log wood transactions.

5. Conclusion

In conclusion, this study endeavors to predict customer churn in log wood transactions, employing a robust methodology guided by the CRISP-DM framework and leveraging advanced machine learning algorithms. The analysis, conducted on a dataset extracted from the Perhutani store's online sales platform, spans the years 2018 to 2022 and consists of 472,208 records. The key findings of this study reveal compelling insights into the predictive performance of ten distinct algorithms. Notably, the Gradient Boosted Tree emerges as a standout performer, showcasing a harmonious blend of accuracy (85.60%) and AUC (0.940). A noteworthy contribution of this study lies in the in-depth feature importance analysis provided by five algorithms. These analyses reveal the weight assigned to different features, providing valuable insights into the factors influencing customer churn. For instance, the Gradient Boosted Tree model places significant importance on features such as transaction recency, transaction frequency, and the standard deviation of the transaction interval.

This study delivers both theoretical and practical contributions through its application of data mining techniques to the pertinent issue of customer churn in industrial B2B transactions. Still, reliance on a single firm's data constrains generalizability presently. Future work, encompassing diverse corporate entities would offer more robust validation. Additionally, while accurate predictive models have value, generating actionable intelligence should remain the priority. Developing frameworks that translate churn insights into tailored intervention strategies represents fertile ground for future efforts. With disruption continually redefining markets, equipping suppliers to cultivate loyalty through data-enabled customer intimacy remains an imperative. In summary, this study contributes to the field of customer churn prediction by not only benchmarking the performance of various machine learning algorithms but also unraveling the intricate relationships between transactional features and customer attrition.

References

Ahn, J., Hwang, J., Kim, D., Choi, H., & Kang, S. (2020). A Survey on Churn Analysis in Various Business Domains. *IEEE Access*, *8*, 220816–220839. https://doi.org/10.1109/ACCESS.2020.3042657

ÇALLI, L., & KASIM, S. (2022). Using Machine Learning Algorithms to Analyze Customer Churn in the Software as a Service (SaaS) Industry. *Academic Platform Journal of Engineering and Smart Systems*, *10*(3), 115–123. https://doi.org/10.21541/APJESS.1139862

Chandran, S. (2020). Literature Survey On Customer Churn Prediction. *IJRAR*.

Chen, K., Hu, Y. H., & Hsieh, Y. C. (2015). Predicting customer churn from valuable B2B customers in the logistics industry: a case study. *Information Systems and E-Business Management*, *13*(3), 475–494. https://doi.org/10.1007/S10257-014-0264-1

Dalli, A. (2022). Impact of Hyperparameters on Deep Learning Model for Customer Churn Prediction in Telecommunication Sector. *Mathematical Problems in Engineering*, *2022*. https://doi.org/10.1155/2022/4720539

Figalist, I., Elsner, C., Bosch, J., & Olsson, H. H. (2019). Customer churn prediction in B2B contexts. *Lecture Notes in Business Information Processing*, *370 LNBIP*, 378–386. https://doi.org/10.1007/978-3-030-33742-1_30/COVER

Ganeson, S., Ling, L. S., & Razak, S. F. A. (2022). A Proposed Churn Window for Non-Contractual Purchases. *Journal of System and Management Sciences*, *12*(4), 57–68. https://doi.org/10.33168/JSMS.2022.0404

Haw, S. C., Ong, K., Chew, L. J., Ng, K. W., Naveen, P., & Anaam, E. A. (2022). Improving the Prediction Resolution Time for Customer Support Ticket System. *Journal of System and Management Sciences*, *12*(6), 1–16. https://doi.org/10.33168/JSMS.2022.0601

Jahromi, A. T., 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). https://doi.org/10.1186/S40537-021-00500-3

Lin, T. Y. (2017). Deductive Data Mining Using Granular Computing. *Encyclopedia of Database Systems*, 1–8. https://doi.org/10.1007/978-1-4899-7993-3_767-2

Marín Díaz, G., Galán, J. J., & Carrasco, R. A. (2022). XAI for Churn Prediction in B2B Models: A Use Case in an Enterprise Software Company. *Mathematics*, *10*(20). https://doi.org/10.3390/MATH10203896

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 (Switzerland)*, *12*(10). https://doi.org/10.3390/APP12105001

Mitrović, S., Singh, G., Baesens, B., Lemahieu, W., & De Weerdt, J. (2017). Scalable RFM-Enriched representation learning for churn prediction. *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017*, *2018-January*, 79–88. https://doi.org/10.1109/DSAA.2017.42

Ribeiro, H., Barbosa, B., Moreira, A. C., & Rodrigues, R. (2022). Churn in services – A bibliometric review. *Cuadernos de Gestion*, *22*(2), 97–121. https://doi.org/10.5295/CDG.211509HR

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, *68*(1), 109–127. https://doi.org/10.1509/JMKG.68.1.109.24030

Sharda, R., Delen, D., & Turban, E. (2018). Business intelligence, analytics, and data science: a managerial perspective. In *cir.nii.ac.jp* (4th ed.). Pearson. https://cir.nii.ac.jp/crid/1130003901884916608

Sun, Y., Liu, H., & Gao, Y. (2023). Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. *Heliyon*, *9*(2). https://doi.org/10.1016/J.HELIYON.2023.E13384

Sweidan, D., Johansson, U., Gidenstam, A., & Alenljung, B. (2022). Predicting Customer Churn in Retailing. *Proceedings - 21st IEEE International Conference on Machine Learning and Applications, ICMLA 2022*, 635–640. https://doi.org/10.1109/ICMLA55696.2022.00105