Churn prediction analysis of telecom customers using svm, random forest and logistic regression models using orange data mining tools

Ana Nurtriana[1*], *Devi Dwi* Rachmawati[1], *Marina* Artiyasa[1] , and *Deni Syahrudin Zapar* Sidiq[1]

[1]Electrical Engineering, Nusa Putra University, Sukabumi, Indonesia

Abstract. This research aims to apply classification algorithms to telecommunication customer churn data using Orange Data Mining. The methods used include Support Vector Machine (SVM), Random Forest and Logistic Regression. The dataset used is secondary data, the dataset is downloaded from the kaggle website with a total of 7,043 customer data and 21 variables that will be used to predict telecommunication churn and in this study Exploratory Data Analysis (EDA) was conducted to understand the characteristics of the data and identify patterns and trends that can be used to improve the performance of classification algorithms. The results of EDA show that telecommunication customer churn data has several characteristics, namely unbalanced churn data, with the number of customers who churn less than customers who do not churn. With the results of the accuracy value, namely Random Forest 76% followed by Logistic Regression 79% and SVM 74%. The best accuracy is obtained by Logistic Regression with an accuracy value of 79%. These results show that logistic regression has a better ability to classify telecommunication customer churn data compared to other classification algorithms, this research shows that Orange Data Mining can be used to classify telecommunication customer churn data.

1 Introduction

In recent years, competition among the telecommunications industry has increased significantly, Machine learning, a subset of data mining [1–3]. Various telecom service providers compete in the market to increase their customer share. Policy makers and business analysts believe that acquiring new customers is more expensive than retaining existing customers [4]. Customers have many choices of better and cheaper services. Customers switch services or telecommunications churn from one service provider to another [5,6]. Customer churn has a very high impact on the development business because of its direct link to the core engineering business [7]. Especially since it directly affects the company's sales. In the telecommunications sector, companies are trying to develop methods to predict potential customer churn [8].

Therefore, companies need data analysis to predict the possibility of customer churn by making Churn predictions in the telecommunications sector [9]. Customer churn in the telecommunications industry is certainly a serious problem. Operators need a churn prediction model to prevent customers from switching to other operators [10]. To overcome the problem of customer churn, telecommunications companies need to understand the factors that contribute to churn so that they can take appropriate action to minimize churn rates [11].

Previous research has become the basis for research on telecommunication customer churn prediction [12] but in research B. Prabadevi, R. Shalini, B.R. Kavitha with the title "Customer churning analysis using machine learning algorithms" 2023 using python implementation with the research objective to provide advice on optimal machine learning strategies for early prediction of client churn. [13] suggests the potential for preprocessing telecommunication churn datasets. The identified gaps have prompted the authors to conduct an experimental study on customer churn prediction in the telecommunications industry. The main objective in this study is to test algorithms from exploratory data analysis and from feature selection results in finding and determining trends from patterns in telecommunication datasets and preprocessing to prepare the data for further data mining processes. With that, this study will compare the performance of several machine learning algorithms namely, SVM, Logistic Regression and Random Forest, with publicly available telecommunication datasets. Metrics such as Accuracy, Precision, Recall. are used to be able to evaluate and compare the results of the performance of prediction algorithms, from classifiers and this research implements one of the data mining tools, namely orange data mining to predict telecommunication churn algorithms.

Orange Data Mining is a popular and powerful data analysis tool for building and testing classification models [14]. The advantages of Orange Data Mining

* Corresponding author: ana.nurtriana_te20@nusaputra.ac.id

include speed of model development, ease of use, and extensive support of analysis features. Orange data mining offers flexibility in data preprocessing, visualization, and data processing. [14–16] Train and test models with one software. However, Orange Data Mining has some shortcomings in performing data mining processes, such as not being suitable for large data and not having advanced features such as machine learning and deep learning [17].

The author expects by combining classification models such as SVM, Random Forest, and Logistic Regression. By using the Orange Data analysis tool, [4] the results of this study can make a practical contribution to telecommunications companies in developing more effective marketing strategies, reducing customer switching rates, and improving overall business decisions.

2 Literature study

2.1 Data mining

Data mining refers to the process of extracting added value in the form of unknown information manually from a database [18] that information will be obtained by extracting and recognizing important or interesting patterns from the data in the database. Classification, which is the process of dividing data into several classes or categories based on the attributes in the data [19].

2.2 Orange data mining

Orange Data Mining implements various algorithms and techniques related to data analysis and machine learning. Orange data mining is a useful data mining tool for visual programming and exploratory data analysis. Orange data mining consists of several components called widgets With Orange data mining software, it is easy to use because it is already in the modeling stage and contains many methods.

2.3 Exploratory data analysis

Exploratory churn in telecommunications refers to the process of analyzing customer data to identify patterns and factors that contribute to customer churn. EDA can be used to identify patterns, trends, and factors that contribute to customer churn rates in the telecommunications industry [20].

Exploratory Data Analysis (EDA) is performed to be able to describe the main characteristics of each attribute by including, minimum and maximum values, mean, standard deviation and others The main purpose of EDA in telecom churn is to extract important information from existing data to help understand customer churn behavior and identify the factors that influence it [21].

2.4 Support vector machine

Support vector machine is a supervised machine learning algorithm that can be used for regression or classification. This algorithm was introduced by. The main idea of this algorithm is to find a hyperplane to separate a data set into classes. In the literature, Support Vector Machine (SVM) can show its application to the problem of analyzing telecommunication customer churn [22,23].

2.5 Random forest

Random Forest is an ensemble learning model that is built by bagging multiple skewed trees representing independent decision trees with feature selection and generating classification results by feeding inputs into these internal trees and aggregating the results based on voting techniques, the theory of random forest is to build multiple decision trees based on sample data using only a few attributes [24].

2.6 Logistic regression

Logistic regression model is one of the machine learning algorithms that is not a black box model. Usually black box models are complicated, but logistic regression shows what it actually does. Logistic regression can be binary, multinomial or ordinal [25]

2.7 Confusion matrix

Confusion Matrix is a performance measurement for machine learning classification problems, where the output can be 2 or more classes. Confusion Matrix has a table with 4 different mixtures of predicted and actual values. Based on the table, it is explained that there are 4 designations that represent the results of the classification process in the confusion matrix, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Confusion Matrix Calculates the negative and positive values in the model. Confusion matrix can be used to display prediction results that match or do not match the actual test data class. [25,26].

$$\text{Accuracy} = ((\text{TP} + \text{TN}))/((\text{TP} + \text{TN} + \text{FP} + \text{FN})) \quad (1)$$

$$\text{Precision} = ((\text{TP}))/((\text{TP} + \text{FP})) \quad (2)$$

$$\text{Recall} = ((\text{TP}))/((\text{TP} + \text{FN})) \quad (3)$$

In this research, the churn prediction model used in the evaluation uses accuracy, precision and recall. With a number of correctly classified examples. Accuracy is the level of correlation between the predicted value and the actual value, while precision is the level of accuracy between the data expected by the user and the answer given by the system. And recall is the success rate of a system in finding back dataPrevious research.

2.8 Previous research

Jain H, Khunteta A and Srivastava S (2020) "Churn Prediction in Telecommunication using Logistic Regression and Logit Boost" uses two classification algorithms, namely logistic regression and logit boost, using the WEKA machine learning tool to classify

telecommunication customer churn data. This research shows that both algorithms can produce high accuracy, which is 85.2385% for logistic regression and 85.1785% for logit boost. [27].

Nalatissifa H, Pardede H (2021) "Prediction of Customer Decisions Using Deep Neural Network on Telco Customer Churn Data" Nusa Mandiri University This research uses Deep Neural Network (DNN), extreme gradient method boosting (XGBoost), and random forest (RF) models. Using Python random search produces DNN modeling with a performance accuracy value of 83.09%, Random Forest (RF) 80.03%, (KNN) 78.64% (DT) 74.63% using three hidden layers, the number of nodes in each hidden layer is [20, 35, 15], using an optimizer RMSprop 0.1, learning rate 0.01, with the fastest setup time of 21 seconds."[28]

Jayawiguna I, Swamardika I, Sudarma M. (2020) "Comparison of Model Prediction for Tile Production in Tabanan Regency with Orange Data Mining Tool" The purpose of this study is to predict tile production in Tabanan Regency and identify the factors that most influence tile production. The predictive modeling conducted in this study used various data mining techniques, including classification trees, AdaBost, and kNN. Because orange data mining tools are GUI-based and user-friendly. The results of predicting tile production based on the characteristics of raw materials, energy, and labor are quite accurate, with the Linear Regression prediction model getting the highest score."[29].

Thange U, Shukla V, ritu, Grobbelaar W (2021) "Analyzing COVID 19 Dataset through data mining tool orange" This study uses the Orange data mining tool to find hidden patterns in Indian Covid-19 data to predict the COVID-19 outbreak. In this study combining machine learning and SIR (Susceptible, Infectious or Recovered) / SEIR (Susceptible, Exposed, Infectious or Recovered) models is recommended to improve the current standard epidemiological models in terms of accuracy and longer processing time. This study shows that Orange can be used as an effective tool to analyze Covid-19 data and predict Covid-19 outbreaks in India. The results show that Orange can be used as a useful tool for modeling epidemic chains. Based on the results of this study, it can be seen that the number of cured cases is still small compared to the number of cured patients. The mortality rate is also very minimal [30].

Agrawai S, Das A, Gaikwad A, sudhir (2018) "Customer Churn Prediction Modeling Based on Behavioural Patterns Analysis using Deep Learning". This research uses Deep Learning methods to predict power outages on Telco datasets. A multi-layer artificial neural network is designed to build a non-linear classification model. This model predicts the churn rate based on customer features, support features, usage features, and contextual features. The probability of unsubscribing and its determinants are predicted. The trained model then applies final weights to these features and predicts the churn probability for that customer. The accuracy achieved was 80.03%. [31].

Prabadevi, R. Shalini, B.R. Kavitha "Customer churning analysis using machine learning algorithms" 2023 The purpose of this study is to provide advice on optimal machine learning strategies for early client churn prediction [32,33]. The results show that the value of each algorithm is Stochastic Gradient Booster provides the highest accuracy for predicting client churn, which is 83.9%, Random Forest 82.6% Logistic Regression 82.9% and K-Nearest Neighbor has the lowest accuracy, which is 78.1%. The data used in this study is customer data from a telecommunications company [34]. The data includes information on customer demographics, behavior, and transactions. The data was cleaned and transformed before being used for machine learning model training. The accuracy of the algorithm was measured using the area under the curve (AUC) [35]. AUC is a measure of classification model performance that calculates the area under the ROC curve. The tools used in this research are Python, a programming language used to develop machine learning models and Python's Scikit-learn Library for machine learning [36].

3 Research method

The research methodology is in the prediction process with classification techniques to predict telecommunication customer churn. telecommunication customer churn which includes several stages, starting from data collection, determining the training set data and data preparation as well as data exploration and analysis, data pre-processing Then, performing data division using the Orange application. Finally, the author analyzes and tests the algorithm obtained in this study [37].

Fig. 1 Research method

3.1 Research attributes

There are 21 research attributes of the data which include 1 target attribute with 3 numeric attributes and 16 categorical attributes and 1 text attribute [38]. The description of the data set is described in Table 1.

Table 1. Description of the data set

No	Attributes	Type	Descriptions
Identity			
1	identification number	text	Identification Number assigned to each customer
Demographic			
2	gender	categorical	customer gender
3	elderly citizens	categorical	senior citizens or not
4	partner	categorical	the customer has a partner or not

No	Attributes	Type	Descriptions
Service Linkage			
5	period of use	numeric	month the customer has used the service
6	telephone service	categorical	the customer has phone service or not
7	dependent	categorical	the customer has dependents or not
8	dual telephone service	categorical	the customer has telephone service with two lines or not
9	internet service	categorical	the type of internet service used by the customer
Additional Services			
10	online security	categorical	customer has online security service or not
11	online backup	categorical	customer has online backup service or not
12	device protection	categorical	customer has device protection or not
13	technical support	categorical	customers have technical support or not
14	TV streaming service	categorical	the customer has a TV streaming service or not
15	movie streaming service	categorical	the customer has a movie streaming service or not
Contracts and Bills			
16	contract	categorical	the type of service contract the customer has
17	paperless bills	categorical	customers choose paperless billing or not
18	payment method	categorical	customer service payment methods
19	monthly fee	numeric	the amount of monthly fees paid by the customer
20	total cost	numeric	total costs paid by customers
Target			
21	stop using the service	categorical	customers stop using the service or not

3.2 Determation of data set

To solve the problems in this research, the author starts by collecting datasets [39]. The telecommunication customer churn dataset was obtained by the author from the Kaggle Data mining website, with the amount of data obtained being 7043 rows (customers) and 21 columns (Attributes). Then from the dataset, the author began to process properly and then proceeded to the data preparation stage.

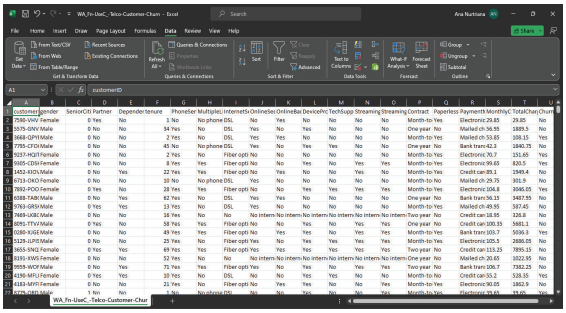

Fig 2. Collection of data in excel format

3.3 Eksploratory data analysis

In processing and exploring the Training Set data using Orange, the author uses widgets in Orange with the purpose of EDA in Orange Data Mining is to understand the properties of data and find patterns and discover patterns and relationships in the data [40]. Through EDA, the author can select the most relevant attributes to predict customer churn and create an accurate churn prediction model.

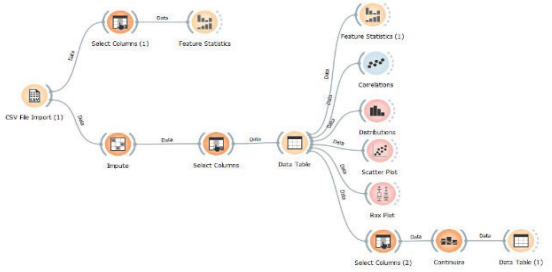

Fig. 3. Exploratory Data Analysis

The data to be used for EDA is customer data from a telecommunications company. This data can include demographic information, service usage, and customer satisfaction. Data visualization can help the authors to see patterns and trends that are not visible in the raw data.

3.4 Pra-preprocessing

In research, the datasets used to build machine learning models are usually imperfect. A dataset will likely have missing data, which is missing or incomplete data. Missing data can occur for various reasons, such as data input errors, data corruption, or uncollected data [41].

Missing data can affect the performance of machine learning models. If missing data is not handled properly, the machine learning model may produce inaccurate

results. Therefore, data preprocessing using impute widgets is necessary to fill in the missing values in the dataset [42]. The dataset preprocessing techniques in this study include data cleaning, churn dataset sharing, discretization, and reduction.

3.4.1 Cleaning data

This research uses the impute widget to replace unknown values in the dataset. there are Dataset 11 unknown or missing values in the TotalCharges column, therefore the author performs preprocessing by using the impute widget to fill in the missing values in the dataset [43].

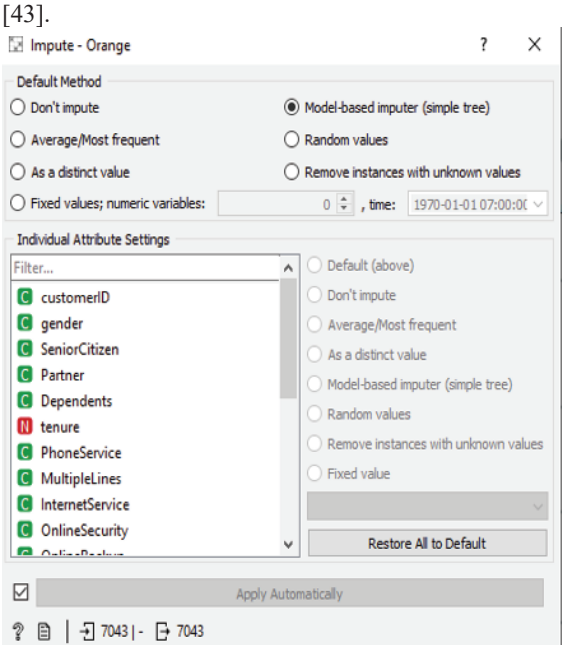

Fig. 4. Missing data import process

3.4.2 Dataset churn division

In the dataset process that uses orange data mining, data division in machine learning is the process of dividing the dataset into two subsets: training data (training set) and test data (test set).

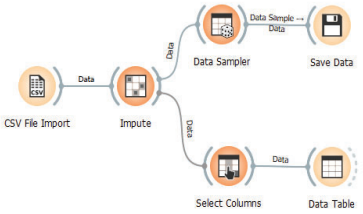

Fig. 5. Stages of Data Sharing

In Figure 5, the dataset is divided into 70% data used for training and 30% data used for testing. The purpose of this data division is to avoid overfitting. This is a condition where the model learns too much from the training data and does not generalize well to new data.

3.4.3 Data selection proces

In the telecommunication churn data selection process, data selection will be carried out on attributes, with the data selection widget used to select features that will be used to build the model. Figure 6 shows the process of selecting attribute data on 21 attribute data with 1 attribute as the target, namely churn

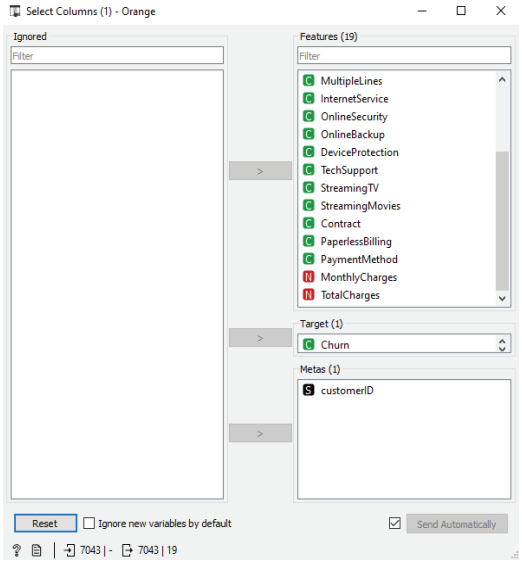

Fig. 6. Data selection process

3.4.4 Data mining process

In analyzing the performance of several classification models in tool orange, a comparison of several data mining methods was conducted to select the best method with high accuracy, in classifying the telecommunication customer churn dataset as shown in Figure 7.

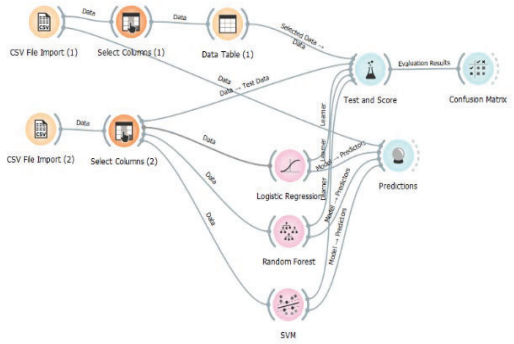

Fig. 7. Telecommunications churn widget design

The design of the data mining widget is carried out using a classification model with orange data mining, namely SVM, Random Forest and Logistic Regression (Figure 7). Furthermore, the test data will be processed by data classification with the final value of the evaluation results.

3.4.5 Model comparisontesting process

In the process of testing the classification model that has been made before, test data is needed to determine the results of the classification model as shown in Figure 8.

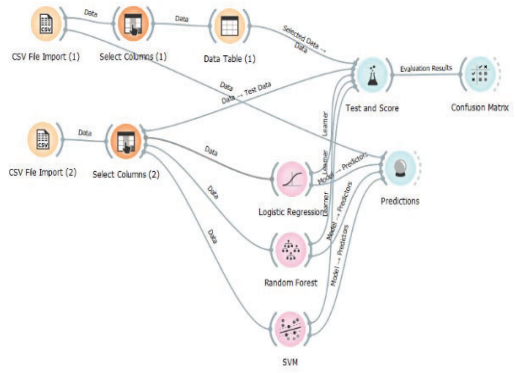

Fig. 8. Widget design for the telecommunications churn dataset classification model

In the red box, there is a model that will be tested for model comparison by entering it into the classification to find out which accuracy value is more accurate.

3.4.6 Model comparison evaluation process

The next process is to perform the classification model comparison process using Test and Score which is needed to calculate the success rate between each classification model in Orange data mining as shown in Figure 9.

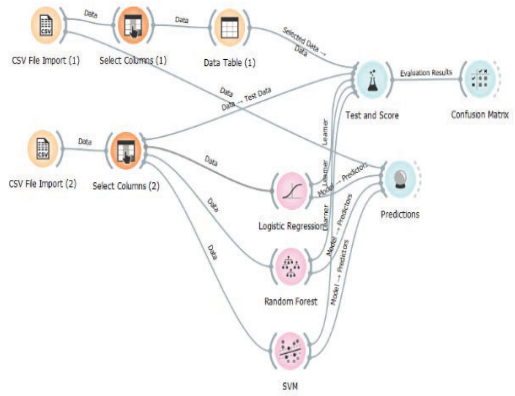

Fig. 9. Test and score widget design

The process of calculating the success rate of the classification model using the Test and Score widget is carried out, which will then be evaluated for accuracy using the Confusion Matrix.

4 Results and discussion

Research by implementing datasets using orange data mining tools, data visualization has been generated and will further understand the patterns and trends of customer churn data. By understanding the patterns and trends, companies can identify the factors that cause customer churn and take action to reduce churn.

4.1 Distribution

In orange data mining tools, distribution widgets can be used to understand how attribute values are distributed in data. This can help to identify patterns and trends that can be used to make decisions and this method can help to display the distribution values of a variable.

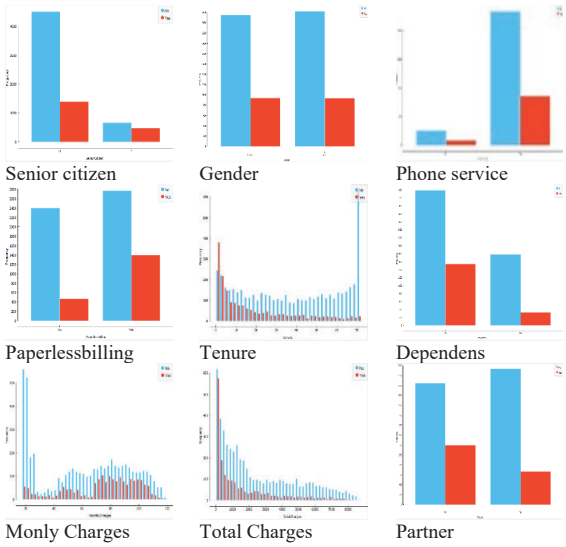

Fig. 10. Histogram Distribusi Numerik Atribut

Senior citizen customers predominantly do not churn as seen in the distribution. The most significant difference between the two distributions is that the churn distribution of retired customers has fewer non-churning customers. The histogram therefore shows that retired customers are slightly more likely to not churn than non-retired customers.

The gender distribution shows the data set displays a relatively equal amount of churn and number of male and female customers i.e. the number of female customers is 49.52% with 13.33% churn by female customers and the number of male customers is 50.48%, with 13.20 male customers churning.

The histogram of telecommunication phone service churn distribution shows that customers who have short-term or monthly contracts have a higher churn rate than customers who have one-year or two-year contracts.

The histogram (Figure 10) shows that customers who use paperless billing churn more than customers who use paper billing.

The Histogram of telecom customer churn distribution based on tenure shows that customers with shorter tenure are more likely to churn than customers with longer tenure.

From the Histogram of telecom customer churn distribution by number of dependents, it shows that customers with dependents are more likely to churn than customers without dependents. The histogram shows that about 21.91% of customers with dependents churn compared to customers without dependents churn only about 4.63%.

From the Histogram of monthly charges distribution of telecom churn shows that customers with higher

monthly charges are seen to be more likely to churn than customers with lower monthly charges to churn.

The results from the Histogram of the distribution of total charges of telecommunication churn show that customers with higher total charges tend to churn than customers with lower total charges of telecommunication churn. It can be seen that customers without partners churn more with a churn value of 17.04% compared to customers who have partners with a lower churn value of 9.50%.

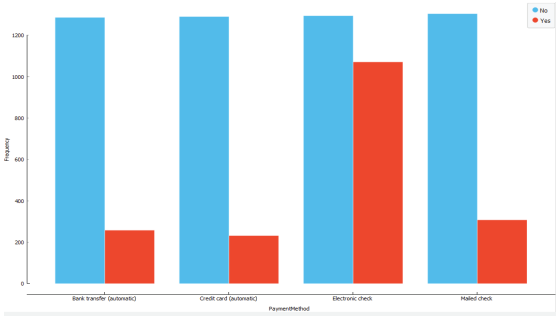

Fig. 11. Paymend methode vs churn

The distribution of telecommunication churn payment methods shows that customers tend to make payments using the electronic check method and using the automatic electronic check method has the most churn. Meanwhile, customers who use credit card payment methods have the lowest churn rate.

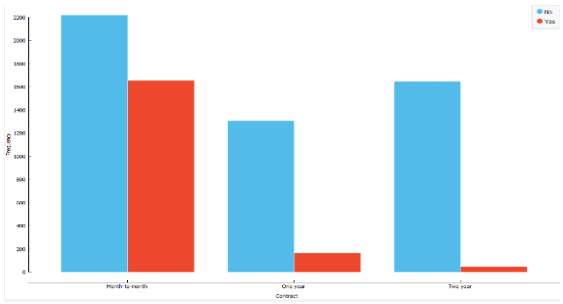

Fig. 12. Contract vs churn

From the results of the Figure 12, it shows that customers who prepay monthly tend to churn higher than customers who prepay or contract 1 or 2 years.

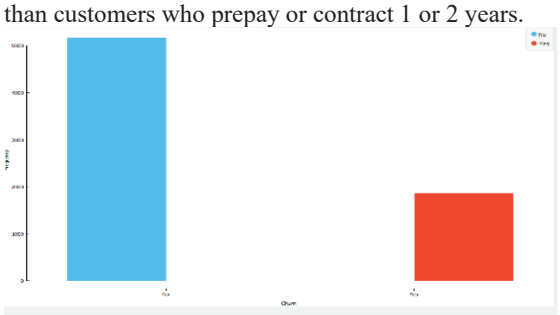

Fig. 13. Churn and Non-churn

In Figure 16 there is a graph illustrating the results of churn and non-churn with the number of churn customers being 1869 with 26.54% and the number of non-churn customers being 5174 with 73.46%. The classification of churn data is unbalanced, with the number of customers who churn is less than the number of customers who do not churn.

4.2 Scater plot

Fig. 14 shows that the monthly fee has a linear relationship with the total fee as expected. The telecom churn scatter plot shows the relationship between monthly fees and customer churn, indicating that there is a positive relationship between the two variables, i.e. the higher the monthly fees, the higher the likelihood of customer churn.

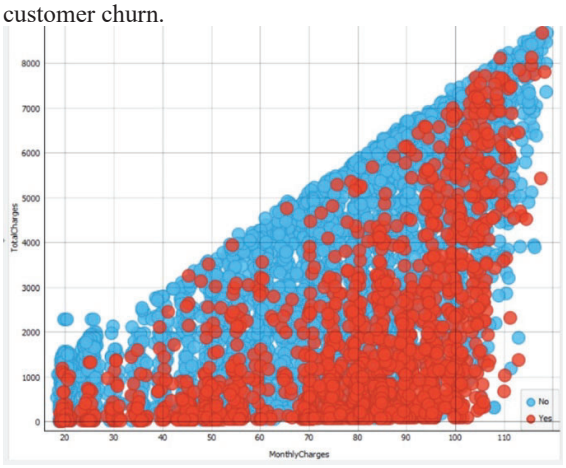

Fig. 14. Total charges vs monly charges

4.3 Evaluation result with convolution matrix

Convusion Matrix is a tool to measure the performance of machine learning classification models, with two or more classes. Confusion matrix is a table that displays the results of model predictions against actual values. The Logistic Regression model evaluation results can be seen in Figure 13, the Random Forest model confusion results can be seen in Figure 14 and the SVM model confusion value can be seen in Figure 15.

		Predicted		
		No	Yes	Σ
Actual	No	1407	166	1573
	Yes	276	263	539
	Σ	1683	429	2112

Fig. 15. Value of the Confusion Matrix Logistic Regression method

The figure shows the True Positive (TP) result value of 1407, True Negative (TN) of 264, False Positive (FP) of 276, and False Negative (FN) value of 166. So that the Accuracy, Precision and Recall values of the Logistic regression method are as follows:

$$Accuracy \quad \frac{(1407+263)}{(1407+263+276+165)} x100\% = 79\% \quad (1)$$

$$Precision \quad \frac{(1407)}{(1407+276)} x100\% = 83\% \quad (2)$$

$$\frac{Recall((1407))}{(1407+165)} x100\% = 89\% \quad (3)$$

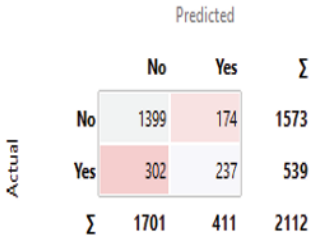

Fig. 16. Value of the Confusion Matrix Random Forest method

The figure shows the True Positive (TP) result value of 1399, True Negative (TN) of 237, False Positive (FP) of 302, and False Negative (FN) value of 174. So that the Accuracy, Precision and Recall values of the Logistic regression method are as follows:

$$Accuracy \frac{(1397+224)}{(1397+224+315+176)} x100\% = 76\% \quad (4)$$

$$Precision \frac{(1397)}{(1397+315)} x100\% = 81\% \quad (5)$$

$$Recall \frac{(1397)}{(1397+176)} x100\% = 88\% \quad (6)$$

$$Precision \frac{(1277)}{(1277+253)} x100\% = 83\% \quad (7)$$

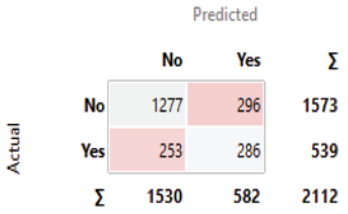

Fig. 17. Value of the Confusion Matrix SVM method

The figure shows the True Positive (TP) result value of 1277, True Negative (TN) of 286, False Positive (FP) of 253, and False Negative (FN) value of 296. So that the Accuracy, Precision and Recall values of the Logistic regression method are as follows:

$$Accuracy \frac{(1277+286)}{(1277+286+253+296)} x100\% = 74\% \quad (8)$$

$$Precision \frac{(1277)}{(1277+253)} x100\% = 83\% \quad (9)$$

$$Recall \frac{(1277)}{(1277+296)} x100\% = 81\% \quad (10)$$

The resuls of evaluation and validation using ConfusionMatrix obtained the comparison value of Accuracy, Precision and Recall from 3 methods namely SVM, Random Forest, and Logistic Regression.

Table 2. Method performance comprecision

Model	Accuracy	recall	Pression
SVM	74%	81%	83%
Logistic Regression	79%	89%	83%
Random Forest	76%	88%	81%

5 Conclusion

This study shows that, the attribute that affects telecommunication churn is the type of contract. Customers who prepay monthly tend to churn higher than customers who prepay or contract 1 or 2 years. Monthly prepaid customers have the freedom to unsubscribe at any time. Therefore, it can make it easier for customers to switch to another service. Whereas prepaid or 1 or 2 year contract customers have a longer commitment. They have to pay a monthly or annual subscription fee to use the service. This makes it harder for them to switch to another service provider. In addition to contract type, other attributes that can affect telecommunication churn are Paymend method, Paperlessbilling, Tenure and Total Charges. Telecommunication companies can use the results of this study to develop strategies to reduce customer churn by using strategies such as offering more attractive contracts for prepaid customers, improving service quality to increase customer satisfaction and providing incentives for customers who remain subscribed. By implementing the right strategy, telecommunication companies can reduce customer churn and increase revenue. The results of the algorithm comparison show that the logistic regression algorithm is the best algorithm for classifying customer data that leaves the telecommunications network, with an accuracy rate of 79%. This research can contribute to the decision to retain telecommunication customers by identifying customers who tend to move to other operators. Logistic regression algorithm is the best algorithm to classify data of telecommunication customers who switch operators. Patterns and trends identified from EDA can be used to improve the performance of the classification algorithm.

References

1. L. S. Riza, M. Ammar, F. Rahman, Y. Prasetyo, M. I. Zain, H. Siregar, T. Hidayat, K. A. Fariza, A. Samah, and M. Rosyda, Knowl. Eng. Data Sci. **6**, 231 (2023)
2. N. Sultan, Knowl. Eng. Data Sci. **5**, 101 (2022)
3. N. . Saravana Kumar, K. Hariprasath, N. Kaviyavarshini, and A. Kavinya, Sci. Inf. Technol. Lett. **1**, 52 (2020)
4. J. R. D. Arcos and A. A. Hernandez, in *Proc. 2019 7th Int. Conf. Inf. Technol. IoT Smart City* (ACM, New York, NY, USA, 2019), pp. 45–49
5. M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman, Indones. J. Electr. Eng. Comput. Sci. **13**, 521 (2019)
6. Y. Utami, I. Zuhroh, V. Prasetya, and M. Rofik, J. Inform. **15**, 1 (2021)
7. R. P. Singh, A. Turi, and D. Malerba, in *Data Min. VIII Data, Text Web Min. Their Bus. Appl.* (WIT Press, Southampton, UK, 2007), pp. 293–302
8. S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, IEEE Access **9**, 62118 (2021)
9. S. Mitrović and J. De Weerdt, Inf. Process.

Manag. **57**, 102052 (2020)
10. W. Zeng, BCP Bus. Manag. **38**, 2811 (2023)
11. Y. Huang, F. Zhu, M. Yuan, K. Deng, Y. Li, B. Ni, W. Dai, Q. Yang, and J. Zeng, in *Proc. 2015 ACM SIGMOD Int. Conf. Manag. Data* (ACM, New York, NY, USA, 2015), pp. 607–618
12. B. Prabadevi, R. Shalini, and B. R. Kavitha, Int. J. Intell. Networks **4**, 145 (2023)
13. A. Y. Saleh and F. N. Binti Mostapa, Sci. Inf. Technol. Lett. **4**, 12 (2023)
14. Z. R. Mohi, Dijlah J. **3**, 13 (2020)
15. S. Mohapatra and T. Swarnkar, in *Lect. Notes Networks Syst.* (Springer Science and Business Media Deutschland GmbH, 2021), pp. 611–620
16. M. I. Mardiyah and T. Purwaningsih, Sci. Inf. Technol. Lett. **1**, 83 (2020)
17. D. Gustian, A. Darmawan, M. I. Tohir, D. Supardi, S. Nurjanah, and A. P. Junfihrana, in *2019 Int. Conf. ICT Smart Soc.* (IEEE, 2019), pp. 1–6
18. M. Ulfah and A. Sri Irtwaty, Fidel. J. Tek. Elektro **4**, 62 (2022)
19. A. Sidik, H. Lumbantobing, A. Suryana, M. A. S. Yudono, Edwinanto, Y. Putra, Y. Imamulhak, and B. Indrawan, Int. J. Eng. Appl. Technol. **5**, 1 (2022)
20. P. Eskerod, S. Hollensen, M. F. Morales-Contreras, and J. Arteaga-Ortiz, Sustainability **11**, 5372 (2019)
21. M. M. J. Adnan, K. Hinkelmann, and E. Laurenzi, in *Commun. Comput. Inf. Sci.* (Springer Science and Business Media Deutschland GmbH, 2022), pp. 389–396
22. D. Mustafa Abdullah and A. Mohsin Abdulazeez, Qubahan Acad. J. **1**, 81 (2021)
23. M.-W. Huang, C.-W. Chen, W.-C. Lin, S.-W. Ke, and C.-F. Tsai, PLoS One **12**, e0161501 (2017)
24. M. Schonlau and R. Y. Zou, Stata J. Promot. Commun. Stat. Stata **20**, 3 (2020)
25. S. Tuba, Fidel. J. Tek. Elektro **5**, 53 (2023)
26. M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, IEEE Access **8**, 90847 (2020)
27. H. Jain, A. Khunteta, and S. Srivastava, Procedia Comput. Sci. **167**, 101 (2020)
28. H. Nalatissifa and H. F. Pardede, J. Elektron. Dan Telekomun. **21**, 122 (2021)
29. I. B. P. Jayawiguna, Int. J. Eng. Emerg. Technol. **5**, 72 (2020)
30. U. Thange, V. K. Shukla, R. Punhani, and W. Grobbelaar, in *2021 2nd Int. Conf. Comput. Autom. Knowl. Manag.* (IEEE, 2021), pp. 198–203
31. S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, in *2018 Int. Conf. Smart Comput. Electron. Enterp.* (IEEE, 2018), pp. 1–6
32. D. Purwanto, D. Damas Permadi, Novita Aprilia, and Matrunevich Oksana Viktorovna, Int. J. Eng. Appl. Technol. **6**, 1 (2023)
33. Richa Rahmalia Sunhadji, Heru Salasa, Ismail, Imanulhaq, Paikun, and Novohatko Elena Nivolaevna, Int. J. Eng. Appl. Technol. **5**, 47 (2021)
34. Paikun, Pirmansyah, Cece Suhendi, Triono, and Niels vuegen, Int. J. Eng. Appl. Technol. **4**, 101 (2021)
35. Nanda Ikhwan Khair, Utamy Sukmayu Saputri, Ardin Rozadi, Muhammad Hidayat, and Oksana Viktorovna Zadorozhnaya, Int. J. Eng. Appl. Technol. **4**, 116 (2021)
36. Paikun and H. B. Maulana, ARPN J. Eng. Appl. Sci. **15**, 2403 (2020)
37. P. Paikun, I. Iskandar, D. A. Susanto, R. F. Sunarlan, and D. Purwanto, in *2022 IEEE 8th Int. Conf. Comput. Eng. Des.* (IEEE, 2022), pp. 1–6
38. Paikun, N. D. Prastyo, R. Fadilah, R. Muhamad, and T. Kadri, in *2020 6th Int. Conf. Comput. Eng. Des.* (IEEE, 2020), pp. 1–6
39. Paikun, M. Kahpi, R. Krisnawati, A. Agustian, R. Rohimat, and Jasmansyah, in *2018 Int. Conf. Comput. Eng. Des.* (IEEE, 2018), pp. 93–98
40. Paikun, S. Rahayu, A. Selpi, A. Awalia, and Jasmanyah, in *2019 5th Int. Conf. Comput. Eng. Des.* (IEEE, 2019), pp. 1–6
41. D. A. Dewi, T. Mantoro, U. Aditiawarman, and J. Asian, in *Stud. Big Data* (Springer Science and Business Media Deutschland GmbH, 2022), pp. 41–58
42. D. T. Sartika, R. Sihotang, M. Muslih, A. Sitorus, O. Haris, D. Devianty, and R. Bulan, Acta Univ. Agric. Silvic. Mendelianae Brun. **68**, 859 (2020)
43. D. S. T., I. S. Mustakim, E. Nurachman, L. Nurpaidah, R. Ferdiansah, M. Ammar, and R. I. Sitepu, in *2018 Int. Conf. Comput. Eng. Des.* (IEEE, 2018), pp. 104–108