Customer Churn Analysis Using Machine Learning to Improve Customer Retention on Vissie Net

Ronnald Christanto Admanegara[1], Wiwik Handayani[2]

[1,2]Universitas Pembangunan Nasional "Veteran" Jawa Timur

Abstract:
The industrial transformation that positions the internet as essential for all social aspects has intensified competition among internet service providers in Indonesia, including efforts to retain customers. This study focuses on predicting customer churn at Vissie Net using the CRISP-DM method combined with supervised learning algorithms, particularly Random Forest, and utilizing a fishbone diagram to identify areas for improvement. The research analyzes a dataset of 1,119 customers, considering variables like subscription length, contract terms, bills, packages, demographics, and churn data. Results indicate that the Random Forest algorithm excels in predicting churn, with 21.8% of customers unsubscribing. The fishbone diagram highlights factors influencing churn and offers suggestions for service variety, AI integration, customer satisfaction, and staff training.

Keywords: Customer Retention, Service Science, Fishbone Diagram, Service Improvement.

1. Introduction

In order to create a more adaptable and sustainable industrial transformation, the goal of the fifth industrial revolution is the integration of digital and physical technology in a system where human cooperation is a key component (Rukmana & Handayani, 2023). The development of technology, especially the internet, has facilitated various human activities in various fields, including work, education, health, and government (Fathurrahman et al., 2022). The internet, as a network that connects people in real-time, has become a vital tool in accelerating development in various fields, even to remote areas, including in Indonesia. In Indonesia, internet penetration has reached 77% of the population, with the highest usage in Java Island covering 41% of the total population (APJII, 2024). The Indonesian Internet Service Providers Association (APJII) plays an important role in the industry, working closely with relevant state agencies such as BSSN and Kominfo to maintain cybersecurity.

The number of internet service provider (ISP) companies in Indonesia increased significantly, with 828 ISPs recorded in 2022, experiencing a 35.51% growth from the previous year (Annur, 2023). Competition among ISPs is intensifying, especially in big cities, where large companies dominate the market with a variety of products and sharp price competition. Meanwhile, in rural areas, local ISPs focus more on the market by offering products according to local preferences. To maintain customer loyalty, ISPs must maintain service quality, especially since their business relies on a subscription model that requires consumer trust and loyalty. PT. Vissie Cyber Data, an ISP in Tulungagung Regency that grew rapidly with more than 1000 customers in less than 5 years, is now facing the challenge of declining customer growth and increasing churn. Managing customers to stay subscribed is key to the company's success, considering that retaining customers is more effective than finding new ones (Yudiana et al., 2023). Knowing the factors that cause churn is important for better evaluation and determination of service policies. Accurate churn prediction can assist companies in planning more efficient loyalty strategies and marketing campaigns, saving costs and improving business sustainability (Husein & Harahap, 2021). Customer churn is the process by which customers stop using a company's products or services (Kamil et al., 2023).

Customer churn forecasting can be done using machine learning with a random forest algorithm that has high accuracy, as proven in research (Suhanda et al., 2022). However, churn analysis can also be done with the CRISP-DM method, which is more complex because it includes up to the deployment stage (Yudiana et al., 2023). Both methods have their own advantages, with random forest excelling in accuracy and CRISP-DM offering a more thorough approach. This research aims to compare the two approaches in churn

prediction at PT Vissie Cyber Data, using random forest algorithm and CRISP-DM method in analysis and deployment, in order to support more effective management decision making in retaining customers.

2. Literature Review

Service Science

Service science is a discipline that studies and integrates various fields such as information technology, management, economics, and engineering to understand and develop efficient and effective service systems (Priyambudi & Handayani, 2024). By combining organizational, human, business, and technology expertise to classify and describe service systems and their interactions in producing value, the primary objective is to generate systematic service innovation (Rukmana & Handayani, 2023). According to Beverungen in (Revika et al., 2022), service science encompasses a wide range of service scales, from automated services to direct interactions between humans, and from small-scale local service systems to large-scale systems like governments and international monetary systems. It also involves continuous technology applications and iterative interactions. According to (Mobarhantalab, 2022), service science is a subfield of applied science that examines the planning, provision, and administration of services from a holistic viewpoint, taking into account the needs of clients, staff, and other stakeholders. It draws on fields like machine learning.

Customer Churn

Customers are an important asset for companies, so customer retention is a key strategy in maintaining a position in a competitive market, which gave rise to the term customer churn (Boruhan, 2022). Customer churn is the process in which customers stop using the products or services offered by the company (Kamil et al., 2023). (Wibowo, 2024) added that customer churn occurs when customers do not continue to subscribe and move to competing companies. Therefore, predicting customer churn is important in business competition to help companies strategize prevention and avoid losses (Aziz et al., 2023). Previous research shows that the variables that influence customer churn vary depending on the context. For example, (Husein & Harahap, 2021) identified variables such as geography, tenure, number of products, and gender in the context of bank customers, while (Yudiana et al., 2023) examined internet customer characteristics such as gender, length of subscription, internet service, contract, monthly bill, and unsubscribe. This research will use data from companies for variables such as service type, address, length of subscription, contract, and bill, to predict customer churn and assist companies in retaining customers.

Supervised Learning

Machine learning is an analytical method to capture linear or nonlinear data relationships and is used in classification and system deployment (Aziz et al., 2023; Wibowo, 2024), with significant benefits in business processes (Hindrayani et al., 2021). Supervised learning builds classifiers by training models using labeled data and can utilize unlabeled data to overcome data shortages (Ouali et al., 2020). The supervised learning algorithms employed in this study include Random Forest, SVM, K-Nearest Neighbors, Decision Trees, and Logistic Regression.

Confusion Matrix

The confusion matrix, a table listing the proportion of properly and erroneously categorized test data, is used by most assessment methods for classification issues (Arifiyanti & Wahyuni, 2020). By comparing the model's predictions with the actual labels of the data, a confusion matrix is used to evaluate the effectiveness of a prediction model or classification system (Kumalasari & Handayani, 2024).

Table 1 Confusion Matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted	Positive (1)	TP	FP
Values	Negative (0)	FN	TN

Source : (Normawati & Prayogi)

True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are the four primary categories in this table (Normawati & Prayogi, 2021). Numerous assessment measures, including accuracy, precision, recall, and F1-score, may be computed using this confusion matrix. Recall is the system's effectiveness in retrieving identifying information, accuracy gauges the degree of accuracy between forecasts and facts from all observations, precision gauges the accuracy between predicted information and system responses (Fauziningrum & Suryaningsih, 2021). The accuracy and recall harmonic means are used to determine the F1-score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$F1\ Score = \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

CRISP-DM
Since 1996, five significant firms, including OHRA and NCR Corporation, have created the CRISP-DM (Cross Industry Standard methodology for Data Mining) approach, a standardized data mining methodology utilized as a problem-solving strategy for business or research (Darmawan, 2020; Pambudi & Abidin, 2023). CRISP-DM is divided into six primary stages: Understanding corporate goals and problem description is the emphasis of business understanding.; Data Understanding, which involves collecting and evaluating data quality; Data Preparation, which includes cleaning and transforming data into final datasets; Modelling, where modelling techniques are applied and data is divided and algorithms are selected (Yudiana et al., 2023); Evaluation, where models are evaluated using metrics such as confusion matrix to ensure accuracy; and Deployment, which involves reporting and visualizing results (Darmawan, 2020; Yudiana et al., 2023). This methodology has been referenced in the Indonesian National Work Competency Standards since 2020 (Yudiana et al., 2023).

Fishbone Diagram
Once the experiment is defined, finding the cause of the problem becomes very important, and one of the methods used is root cause analysis with tools such as fishbone or cause-and-effect diagrams (Ishikawa) as well as 5 Whys analysis (Sakdiyah et al., 2022). A fishbone diagram is a graphical technique that connects factors affecting a process to find significant causes in determining the quality of the output, both positive and negative (De Fretes, 2022). This diagram makes it easy to understand the problem and its causal factors, with the main problem placed on the right of the diagram. (Sakdiyah et al., 2022) explain that a fishbone diagram consists of four components: Man (the role of the individual in decision-making), Machine (the tools used in dealing with the problem), Material (machine-related data), and Measurement (measurements to ensure the effectiveness of the solution).

3. Research Method
In order to characterize events using numerical data without testing specific assumptions, this research use a quantitative descriptive technique (Sulistyawati et al., 2022). This study focuses on customer turnover at Tulungagung-based PT Vissie Cyber Data, taking into account affecting factors including service type, address, subscription duration, contract, and bill. Structured non-participant observation and a review of the literature were used to gather data (Sugiyono, 2021). Through the steps of business understanding, data understanding, data preparation, modeling, evaluation, and deployment, the data were examined utilizing the CRISP-DM approach with the Python programming language on Google Colab (Santoso et al., 2023). The confusion matrix was used to assess the final model and determine its accuracy, precision, recall, and F1-score. A fishbone diagram is created using the visualization results of feature selection as a root cause analysis for the long-term strategy of the firm (De Fretes, 2022; Sakdiyah et al., 2022).

4. Results and Discussion
4.1 Results
Business Understanding

In the early stages of business understanding, Vissie Net faced an increasing number of unsubscribed customers, so a prediction model using machine learning algorithms was needed to learn the classification of subscribed customers. The data used covers customers from 2021 to 2023 to understand their characteristics and classification. The analysis data includes the service plan chosen, demographics of customers covered by routers and OLTs, monthly billing data with post-paid or promo systems, duration of subscription to date or at the time of unsubscription, length of contract to bind the customer, and unsubscribe status. These variables are classified with machine learning to provide the company with recommendations for improving customer service and activities.

Data Understanding

After the data mining process, 1,119 customer data from Vissie Net was obtained. This data includes information on each customer's preferences and behavior towards the services provided. Some of the variables in the dataset include service preferences, customer behavior, and other relevant information for further analysis. The data characteristics can be seen in the table.

Table 2 Customer Data Description

No	Column Name	Description	Data Type	Data Content
1	Customer ID	Customer ID recorded in the company's system	Categorical	-
2	Demographics	Area or place where the customer installs the service	Categorical	Sub-districts in Tulungagung Area
3	Package	Options The package selected by the customer	Categorical	Package options offered
4	Create	Customer subscription start data	Numeric	-
5	Churn	Stop or off subscription data	Numeric	-
6	Length of subscription	Length of time the customer has subscribed	Numeric	Month
7	Contract	Company's subscription contract agreement	Numeric	Month
8	Charge	Cost that the customer has to pay every month	Numeric	Total Charge
9	Churn	Description of whether the customer unsubscribed or not	Categorical	Yes, No

Source : Company Data, 2024 (data is processed)

Data Preparation

In the data preparation stage, it is to eliminate features with information that is not needed in the research, and from the elimination results, other features are converted into numeric categories such as the top 10 data displayed.

Table 3 Label Encoding Result Data

index	DEMOGRAPHIC	PACKAGE	LENGTH OF SUBSCRIPTION	CONTRACT	CHARGE	CHURN
0	1	1	35	24	100000	1
1	1	1	35	24	100000	1
2	1	1	42	24	100000	0
3	1	1	42	24	100000	0
4	1	1	41	24	100000	0
5	1	1	41	24	100000	0
6	1	1	12	24	100000	1
7	4	1	33	24	100000	1
8	4	1	41	24	100000	0
9	4	1	41	24	100000	0

Source : Google Colaboratory, 2024 (data is processed)

Modelling

During the modeling phase, 1,119 data points were split into 70% train and 30% test data. These sets of data were then modeled using methods from Random Forest, K-Nearest Neighbors, SVM, Decision Tree, and Logistic Regression.

Table 4 Shape Data Train and Test

X_train.shape	(783, 5)	(783 row, 5 columns)
X_test.shape	(336, 5)	(336 row, 5 columns)
Y_train.shape	(783,)	(783 row)
Y_test.shape	(336,)	(336)

Source : Google Colaboratory, 2024 (data is processed)

Evaluation

In the evaluation stage, the data is displayed with a model performance matrix that displays the evaluation score of each algorithm. From the matrix, the Random Forest algorithm has the best model score with accuracy 0.964, precision 0.968, recall 0.857, and f1-score 0.909 compared to the other four algorithms.

Table 5 Algorithm Model Performance Results Table

Logistic Regression	83,0%	84,2%	22,9%	36,0%
K-Nearest Neighbors	94,0%	94,6%	75,7%	84,1%
SVM	79,2%	0,0%	0,0%	0,0%
Random Forest	96,4%	96,8%	85,7%	90,9%
Decision Tree	95,5%	98,2%	80,0%	88,2%
	Accuracy	Precision	Recall	F1-Score

Source : Google Colaboratory, 2024 (data is processed)

With 96.4% accuracy, 96.8% precision, 85.7% recall, and 90.9% F1-score, the Random Forest algorithm performs best when modeling customer turnover data, according to evaluation findings from the confusion matrix (Fauziningrum & Suryaningsih, 2021). Because of this, Random Forest outperforms other algorithms like Decision Tree, SVM, K-Nearest Neighbors, and Logistic Regression. Because it indicates the trade-off between recall and accuracy, the algorithm with the greatest F1-Score should be chosen, according to (Romadloni et al., 2022). Research by Yudiana et al. (2023) who used a ridge classifier to obtain 80.5% accuracy and research by Husein Harahap (2019) who ranked Random Forest as the best algorithm with 86% accuracy on a larger dataset also demonstrate how algorithm performance can vary depending on the characteristics of the data used.

Deployment

Following testing, the Random Forest algorithm emerged as the top performer and was prepared for deployment. Using the best model to forecast a customer's likelihood of churning or retention based on certain variables is part of the deployment stage (Nilwanda et al., 2024). Random Forest was used to make predictions on three new customer circumstances in order to determine retention and probable churn.

Figure 1 New Customer Prediction
Source : Google Colaboratory, 2024 (data is processed)

The results show that the model is able to predict accurately, especially related to changes in billing prices that can affect customer decisions. Previous research also supports the effectiveness of Random Forest in classification, as used by (Nilwanda et al., 2024) to classify citizens as potential recipients of the Family Hope Program. The more features in the dataset, the better the modeling quality.
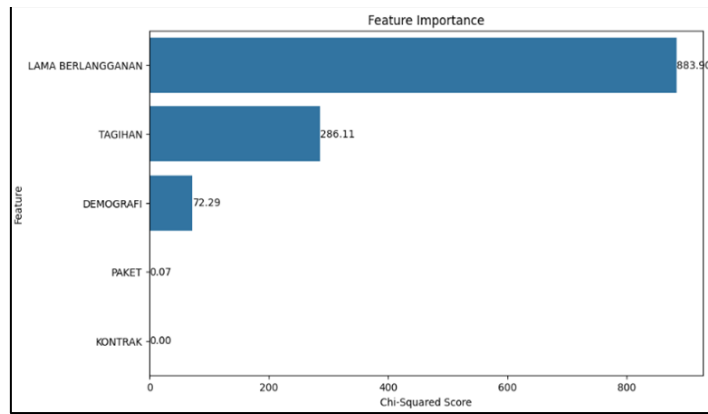
Figure 2 Feature Importance Matrix
Source : Google Colaboratory, 2024 (data is processed)

Based on supervised learning analysis, the feature selection results show that Length of Subscription is the most significant factor with the highest score of 883.90, reflecting strong relevance in the churn prediction model as it shows customer tenure is important for predicting unsubscribe decisions (Aziz et al., 2023; Nurhidayat & Anggraini, 2023). Bills are also influential with a score of 286.11, indicating that the value of a customer's bill influences prediction, although not as strongly as subscription length (Nurhidayat & Anggraini, 2023; Suhanda et al., 2022). Demographics had a score of 72.30, indicating a smaller but still significant contribution than the two main factors. Package has a very low score of 0.07, indicating that package type has little effect on the prediction model. Finally, Contract with a score of 0.00 does not contribute to the model performance as there is no variation in consistent contract data over 24 months, only differentiating between customers under or over contract with no significant impact on churn prediction.

Fishbone Diagram
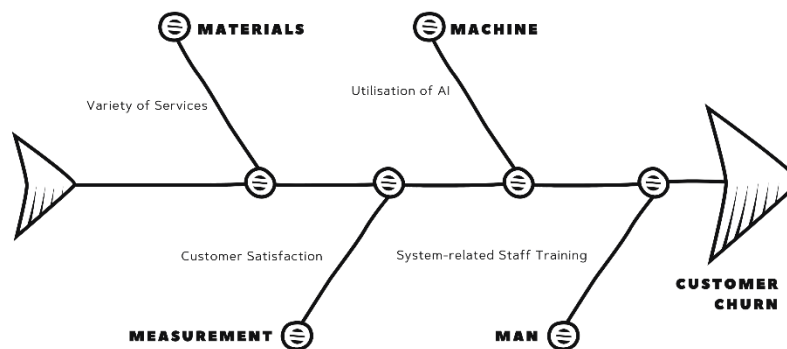
Figure 3 Fishbone Diagram
Source : Personal Document, 2024 (data is processed)

A fishbone diagram is a cause-and-effect analysis tool that describes influential factors in a process or problem, as applied in research to analyze customer churn (Sakdiyah et al., 2022). In this study, the fishbone diagram is divided into four main components-people, machines, materials, and measurement-which are adjusted to the company's conditions and the results of the supervised learning model analysis and feature importance selection. In this way, problems from root to effect can be thoroughly analyzed.

a. Materials

Materials are components that describe the services or equipment that support a company's activities, including in the context of customer churn. Although service variation may not be very significant in supervised learning algorithm modeling, it is still important in terms of customer preferences (Wibowo, 2024). In the company Vissie Net, services are divided between FTTH (Fiber To The Home) and corporate services, with packages such as Broadband and Dedicated. The Broadband plan shares internet speed between customers, while the Dedicated plan offers full speed with no sharing and a 98% SLA. Infrastructure limitations hinder the offering of additional products, such as TV, and responsive network

problem management, as well as customization based on age or customer characteristics, are aspects that need attention to improve services (Wibowo, 2024).

b. Machine

The machinery/technology component of the company also affects customers' decision to unsubscribe. In a company like Vissie Net, the recording and bookkeeping system is still manual, using books to record customer data, which can be a problem if data loss occurs. This problem is exacerbated if customers experience network disruptions and need to check location data. Instead, an integrated system based on modern technology such as a single gateway can synchronize customer data efficiently, simplifying service and complaint handling. New technologies, such as chatbots, can also improve customer service by answering common questions automatically, in accordance with company SOPs, and reduce the workload of customer service staff (Priyambudi & Handayani, 2024). Artificial intelligence and machine learning are examples of cutting edge technologies that may be used to combine data, enhance dataset modeling, and forecast consumer behavior. These capabilities make it simpler to monitor and modify services in order to lower customer attrition.

c. Measurement

Measurement or benchmarking is an important tool for assessing and planning product development, but currently companies are still not utilizing benchmarking to its full potential. Examples of benchmarks include employee performance measurements and monthly or daily targets. The company currently gives bonuses to employees based on the number of customers acquired, but this method is ineffective because there are no specific targets for marketing, who only make door-to-door visits without specific targets. With an average addition of 50 customers per month or 10 customers per week, the company can develop a marketing strategy by setting weekly targets and utilizing social media and promo days (Rukmana & Handayani, 2023). The implementation of SOPs based on Information Technology Service Management, COBIT 2019 guidelines, or ITIL is also important to ensure well-structured services, including determining downtime maintenance schedules and conducting regular internal audits and monitoring. Measurement of service quality through customer feedback can also be the basis for future service improvement and development (Wibowo, 2024).

d. Man

The human component of a company, such as the one at Vissie Net, plays a crucial role in daily operations. Currently, the company is experiencing challenges in employee management as some individuals hold multiple roles at once, such as admins who also act as secretaries, financial administrators, and customer service. This can hamper the company's performance if technical or non-technical problems occur (Wibowo, 2024). Ideally, companies should place employees according to their specializations to prevent confusion and improve customer service efficiency. In addition, employees' knowledge of technology is still limited, so companies should provide specialized training to improve their capabilities and ensure they are adaptive to the latest technological developments. This training is important so that employees can maximize the company's adaptation to new technologies, maintain service quality, and reduce the risk of customer unsubscription (Rukmana & Handayani, 2023).

4.2 Discussion

Based on the analysis of data from the company Vissie Net, which involved 1,119 data with variables such as Churn Description, Subscription Length, Contract, Package, Bill, and Demographics, and the application of machine learning algorithms, Random Forest proved to be the best algorithm (Wibowo, 2024). Feature importance analysis shows that "Length of Subscription" is the most significant factor in churn prediction, with customers who subscribe for more than 24 months are more likely to stay subscribed. The company faces challenges in keeping customers, with 244 customers unsubscribing. To overcome this, companies need to improve loyalty programs, offer additional benefits, and expand service variations, as well as optimize the use of technology (Puspitasari & Masruroh, 2024; Wibowo, 2024).

Data distribution analysis shows that customers tend to choose packages with low prices, according to the internet needs in their area. The average subscription length of customers exceeds the 24-month contract, indicating good potential in retention if managed well. The findings suggest that the company should expand customer outreach and consider opening branch offices in areas with many customers to address the problem

more efficiently. The fishbone diagram identified four main components that affect customer retention: service variation, customer satisfaction, AI utilization, and staff training (Priyambudi & Handayani, 2024; Rukmana & Handayani, 2023; Wibowo, 2024). Service variety and improved customer satisfaction, such as through feedback and service standards, as well as the utilization of AI for operational efficiency, are key steps in reducing churn. Staff training is also important to improve adaptation to technology and support customer retention. Continuous monitoring and evaluation are needed to reduce the churn rate, which currently stands at 21.8% (Wibowo, 2024).

5. Conclusion

From the analysis using the CRISP-DM method and supervised learning algorithms, especially Random Forest, it was found that "Length of Subscription" is the main factor in predicting customer churn, with the churn rate reaching 21.8%. The fishbone diagram identified four key components that affect churn-materials, machinery/technology, measurement, and people-and showed that improving infrastructure, diversifying services, training employees, and utilizing technologies such as AI can improve customer retention and company operations. For companies like Vissie Net, suggestions include focusing on loyalty programs, infrastructure and technology development, and employee training. Future research is recommended to consider more data characteristic criteria and increase data size to improve the accuracy of machine learning modeling, in order to optimize churn identification and prevention.

References

1. Annur, C. M. (2023, September 8). *Jumlah Perusahaan Penyedia Internet di Indonesia Terus Bertambah sampai 2022*. Databoks. https://databoks.katadata.co.id/datapublish/2023/09/08/jumlah-perusahaan-penyedia-internet-di-indonesia-terus-bertambah-sampai-2022
2. APJII. (2024, February 7). *APJII Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang*. APJII. https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang
3. Arifiyanti, A. A., & Wahyuni, E. D. (2020). Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining. *SCAN*, *15*(1), 34–39. https://www.cs.
4. Aziz, A. M. N., Mauliddin, A., Armanda Sintalana, V., Daryl Hafiz, R., Akbar Rismayadi, A., Studi Sistem Informasi, P., & Adhirajasa Reswara Sanjaya, U. (2023). Prediksi Customer Churn Menggunakan Logistic Regression dan Decission Tree. *E-Prosiding Teknik Informatika*, *4*(1).
5. Boruhan, G. (2022). Churn Customer Management in Retail Industry: A Case Study. *İzmir İktisat Dergisi*, *37*(4), 1094–1118. https://doi.org/10.24988/ije.1070830
6. Darmawan, T. (2020). Credit Classification Using CRISP-DM Method On Bank ABC Customers. *International Journal of Emerging Trends in Engineering Research*, *8*(6), 2375–2380. https://doi.org/10.30534/ijeter/2020/28862020
7. De Fretes, R. A. (2022). *Analisis Penyebab Kerusakan Transformator menggunakan metode RCA (Fishbone Diagram and 5-whys analysis) di PT. PLN (Persero) Kantor Pelayanan Kiandarat*. *16*(2), 117–124.
8. Fathurrahman, A. W., Thoriqulhaq, M., Dequika, F., Putra Arianto, P., Studi, P., Data, S., & Timur, J. (2022). Penerapan Machine Learning untuk Pengklasifikasian Hoaks pada Platform Media Sosial:Studi Literatur. *Seminar Nasional Sains Data 2022 (SENADA 2022) UPN "Veteran" Jawa Timur*, 66–68.
9. Fauziningrum, E., & Suryaningsih, E. I. (2021). *Evaluasi dan Prediksi Penguasaan Bahasa Inggris Maritim Menggunakan Metode Decision Tree dan Confusion Matrix (Studi Kasus Di Universitas Maritim Amni)*. http://repository.unimar-amni.ac.id/id/eprint/3837
10. Hindrayani, K. M., Anjani, A., & Nurlaili, A. L. (2021). Penerapan Machine Learning pada Penjualan Produk UMKM : Studi Literatur. *Seminar Nasional Sains Data*, 19–23.
11. Husein, A. M., & Harahap, M. (2021). Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning. *Data Sciences Indonesia (DSI)*, *1*(1), 8–13. https://doi.org/10.47709/dsi.v1i1.1169
12. Kamil, F., Salkiawati, R., & Alexander, A. * D. (2023). Analisis Churn Pelanggan Produk Fashion Campus Menggunakan Metode RFM Analysis dan Algoritma Naïve Bayes (Studi Kasus Yayasan Bakti Achmad Zaky). In *Jurnal Esensi Infokom* (Vol. 7, Issue 2).

13. Kumalasari, A. T., & Handayani, W. (2024). Sentiment Analysis To Improve The Quality Of Public Transportation Services "Suroboyo Bus." *Indonesian Interdisciplinary Journal of Sharia Economics (IIJSE)*, *7*(3).

14. Mobarhantalab, S. S. (2022). Clarifying the concept of service in service science. *11th International Conference on Management, World Trade, Economics, Finance & Social Science*. https://www.researchgate.net/publication/369529140

15. Nilwanda, L. E., Arifiyanti, A. A., & Hadiwiyanti, R. (2024). Penerapan Algoritma (Nilwanda, dkk) Madani. *Jurnal Ilmiah Multidisiplin*, *2*(1), 221–230. https://doi.org/10.5281/zenodo.10523523

16. Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. In *Jurnal Sains Komputer & Informatika (J-SAKTI* (Vol. 5, Issue 2).

17. Nurhidayat, M. M. S., & Anggraini, D. (2023). Analysis and Classification of Customer Churn Using Machine Learning Models. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(6), 1253–1259. https://doi.org/10.29207/resti.v7i6.4933

18. Ouali, Y., Hudelot, C., & Tami, M. (2020). An Overview of Deep Semi-Supervised Learning. *Cornel University*. http://arxiv.org/abs/2006.05278

19. Pambudi, A., & Abidin, Z. (2023). Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm) (Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022). *JDMSI*, *4*(1), 1–14.

20. Priyambudi, N. A., & Handayani, W. (2024). T Analysis Of Service Science-Based Academic Service Development In The Management Study Program Of UPN "Veteran" East Java. *Indonesian Interdisciplinary Journal of Sharia Economics (IIJSE)*, *7*(3).

21. Puspitasari, A., & Masruroh, N. A. (2024). Reducing Customer Churn for XL Axiata Prepaid: Factors and Strategies *Correspondence. In *Journal of Industrial Engineering and Education* (Vol. 2, Issue 1).

22. Revika, N. R., Handayani, W., Pembangunan, U., Veteran, N. ", Timur, J., Raya, J., Madya, R., & Anyar, G. (2022). Desain Service Science untuk Meningkatkan Kepuasan Mahasiswa Terhadap E-Learning Universitas Pembangunan Nasional "Veteran" Jawa Timur. *BRILIANT: Jurnal Riset Dan Konseptual*, *7*(4). https://doi.org/10.28926/briliant.v7i4

23. Romadloni, P. L., Kusuma, B. A., & Baihaqi, W. M. (2022). Komparasi Metode Pembelajaran Mesin untuk Implementasi Pengambilan Keputusan dalam Menentukan Promosi Jabatan Karyawan. *Jurnal Mahasiswa Teknik Informatika (JATI)*, *6*(2), 622–628.

24. Rukmana, S., & Handayani, W. (2023). Analysis of Service Science Design through Customer Experience to Increase PosAja Application User Satisfaction. *Kontigensi : Jurnal Ilmiah Manajemen*, *11*(2). https://doi.org/10.56457/jimk.v11i2.402

25. Sakdiyah, S. H., Eltivia, N., & Afandi, A. (2022). Root Cause Analysis Using Fishbone Diagram: Company Management Decision Making. *Journal of Applied Business, Taxation and Economics Research*, *1*(6), 566–576. https://doi.org/10.54408/jabter.v1i6.103

26. Santoso, H., Putri, R. A., & Sahbandi, S. (2023). Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest. *Jurnal Manajemen Informatika (JAMIKA)*, *13*(1), 62–72. https://doi.org/10.34010/jamika.v13i1.9303

27. Sugiyono. (2021). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (3rd ed., Vol. 1). Alfabeta.

28. Suhanda, Y., Nurlaela, L., Kurniati, I., Dharmalau, A., & Rosita, I. (2022). Predictive Analysis of Customer Retention Using the Random Forest Algorithm. *TIERS Information Technology Journal*, *3*(1), 35–47. https://doi.org/10.38043/tiers.v3i1.3616

29. Sulistyawati, W., Trinuryono, S., & Wahyudi. (2022). Analisis (Deskriptif Kuantitatif) Motivasi Belajar Siswa dengan Model Blended Learning di Masa Pandemi COVID19. *Jurnal Universitas Jember*, *13*(1), 68–73.

30. Wibowo, A. S. (2024). Analisis Churn Nasabah Bank Dengan Pendekatan Machine Learning dan Pengelompokan Profil Nasabah dengan Pendekatan Clustering. *Konstruksi: Publikasi Ilmu Teknik, Perencanaan Tata Ruang Dan Teknik Sipil*, *2*(1), 30–41. https://doi.org/10.61132/konstruksi.v2i1.43

31. Yudiana, Y., Yulia Agustina, A., & Nur Khofifah, dan. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *IJIEB : Indonesian Journal of Islamic Economics and Business*, *8*(1). http://e-journal.lp2m.uinjambi.ac.id/ojp/index.php/ijoieb