JOURNAL INNOVATIONS COMPUTER SCIENCE
E-ISSN: 2961-970X, P-ISSN: 2961-9718
Volume 4 Number 2, November 2025, pp 150-157.
DOI: https://doi.org/10.56347/jics.v4i2.310

KAWANAD Institute

Decision Tree-Based Predictive Model Development for RumahNet Customer Satisfaction Analysis in West Jakarta

Dita Tri Yuliantoro [a*], Frencis Matheos Sarimole [b]
[a*,b] Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.

ABSTRACT
The rapid growth of information technology has amplified the demand for fast and reliable internet services, particularly in urban centers such as West Jakarta. This study aims to design a predictive model of customer satisfaction for RumahNet's Fiber to the Home (FTTH) services by applying the Decision Tree (C4.5) algorithm. A survey of 250 active subscribers was conducted using a Likert-scale questionnaire distributed through Google Forms, capturing perceptions of internet speed, connection stability, pricing, and technical support. The dataset was processed and analyzed using RapidMiner Studio within the Knowledge Discovery in Databases (KDD) framework. Results show that the model achieved an accuracy of 85.33%, precision of 91.93%, recall of 90.47%, and an F1-score of 91.18%. The decision tree revealed that internet speed and connection stability were the most critical determinants of satisfaction, followed by pricing and responsiveness of customer service. These findings suggest that prioritizing technical reliability while maintaining affordability and responsive support is essential for strengthening loyalty and reducing churn. The research demonstrates that Decision Tree modeling not only provides high predictive accuracy but also offers clear interpretability, making it a valuable tool for data-driven decision-making in the ISP sector.

ARTICLE HISTORY
Received 17 July 2025
Accepted 23 August 2025
Published 30 November 2025

KEYWORDS
Customer Satisfaction; FTTH; Decision Tree; ISP; Data Mining.

1. Introduction

The rapid advancement of information technology has intensified the demand for reliable digital connectivity, particularly in urban areas such as West Jakarta. Daily activities in education, business, and entertainment increasingly depend on high-speed and stable internet connections, with Fiber to the Home (FTTH) widely regarded as the most suitable technology due to its superior performance in speed, latency, and reliability (Insani & Alijoyo, 2024). The rising demand has led to fierce competition among Internet Service Providers (ISPs), including established players such as IndiHome, MyRepublic, and First Media. As a new entrant, RumahNet—operated by PT Fiber Kerumah Indonesia—targets households and small businesses, making customer satisfaction a critical factor for building loyalty and sustaining market share. Prior studies indicate that satisfaction is strongly shaped by factors such as connection speed, service stability, pricing, and after-sales support, and that satisfied customers are more likely to remain loyal and recommend services to others (Nurus *et al*., 2023; Hanggiasyifa & Yulianto,). Nevertheless, many ISPs still rely on manual or reactive approaches to monitor satisfaction, limiting their ability to systematically identify

CONTACT Dita Tri Yuliantoro yuliandlia2700@gmail.com Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.

patterns or implement predictive mechanisms, especially among smaller providers (Zubair *et al*.). Data-driven methods such as Decision Tree classification offer a promising alternative, as they can generate interpretable rules for mapping attributes that influence satisfaction (Puspasari *et al*.,) and have been shown to outperform algorithms like Naïve Bayes in predicting customer loyalty (Dam & Dam, 2021). Moreover, while most ISP-related studies rely on simple hold-out validation, advanced validation techniques such as k-fold cross-validation and stratified sampling remain underutilized in this domain (Nugraha & Hendry, 2023). Beyond technical performance, service quality and brand perception have been identified as key mediators of satisfaction and loyalty (Deviana & Tjahjaningsih, 2022), and evidence from the mobile internet sector suggests that product quality, pricing, and service delivery exert a significant influence on long-term customer commitment (Atrisia *et al*., 2024). For RumahNet, the challenge lies in capitalizing on its existing customer data, which includes demographic profiles, subscription types, usage history, and service complaints. Leveraging predictive models such as C4.5 Decision Trees could enable early detection of dissatisfaction, reduce churn risk, and provide a competitive edge through proactive service improvements (Pirmansyah & Wahyudi, 2023; Mariati & Lestari,; Zidane *et al*., 2022). This approach aligns with broader developments in information systems that emphasize data-driven decision-making and the integration of predictive analytics into organizational strategy (Reisyer & Harman, 2024; Rahman & Sutanto, 2023). Against this backdrop, the present study develops a Decision Tree–based predictive model to classify customer satisfaction with RumahNet's FTTH service in West Jakarta, contributing both practical insights for the provider and theoretical advancement in data mining applications within the ISP sector (Kurniawan Maranto *et al*., 2024).

2. Methodology

This research focuses on developing a predictive model of customer satisfaction for RumahNet's FTTH service in West Jakarta by employing the Decision Tree (C4.5) algorithm. The model was built using survey data collected from active subscribers, with questions designed to capture perceptions of service attributes such as internet speed, stability, pricing, and technical support. The data used in this study consisted of two types: primary data gathered through an online questionnaire distributed by field technicians to customers who had been subscribed for at least one month, and secondary data obtained from academic literature, journals, and research articles that informed the design of indicators and methodological choices. Data collection was carried out in July 2025 using Google Forms, and purposive sampling was applied to select 250 respondents who met the inclusion criteria. The responses were stored in Microsoft Excel before being analyzed in RapidMiner Studio.

The research design followed the Knowledge Discovery in Databases (KDD) framework, which involves several stages: data selection, preprocessing, transformation, modeling, and evaluation (Deny Jollyta, 2023). The flow of this process is illustrated in Figure 1

Figure 1. Research Workflow (KDD-based)

In the selection stage, attributes directly related to satisfaction were retained—namely internet speed, stability, pricing, and technical service—while incomplete records were excluded. The preprocessing stage involved cleaning missing values, converting qualitative responses into numerical values (Likert scale 1–5), and classifying cases into two target labels: *Satisfied* (average score ≥ 3.5) and *Not Satisfied* (average score < 3.5). The 3.5 threshold was chosen because it lies above the neutral midpoint (3.0) and approximates the “Agree” category, a convention often adopted in satisfaction studies. Once cleaned, the dataset was transformed into formats compatible with RapidMiner Studio (.xlsx or .csv) and split into training and testing subsets using the hold-out method, with 70% of the data allocated for training and 30% for testing. Model development was performed using the C4.5 Decision Tree algorithm, which is widely recognized for producing interpretable classification rules. Several parameters were configured to optimize performance: the splitting criterion was set to Gain Ratio to reduce bias toward attributes with many values; maximum tree depth was limited to seven to prevent overfitting on the relatively small dataset; and pruning was enabled with a confidence factor of 0.1, combined with pre-pruning to halt tree growth when further splits did not improve information gain. These settings were intended to balance predictive accuracy and generalization capacity. The evaluation phase applied standard classification metrics. Accuracy measured the proportion of correct predictions relative to the total dataset, defined as:

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

Precision assessed the correctness of positive predictions:

$$Precision = \frac{TP}{TP + FP}$$

Recall, or sensitivity, measured the model’s ability to correctly identify positive cases:

$$Recall = \frac{TP}{TP + FN}$$

The F1-score, representing the harmonic mean of precision and recall, was calculated as:

$$F1 = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall}$$

In addition, a confusion matrix was generated to provide a two-by-two summary of true positives, false positives, true negatives, and false negatives, enabling analysis of classification errors. The testing design pursued three main objectives: to measure the ability of the model to categorize customers into *Satisfied* and *Not Satisfied* groups, to evaluate the efficiency and accuracy of the classification process, and to determine the most influential attributes shaping customer satisfaction with RumahNet. Validation was performed using the hold-out technique, maintaining a 70%/30% split between training and testing datasets. The tools employed included Google Forms for survey distribution, Microsoft Excel for preliminary cleaning and transformation, and RapidMiner Studio for model construction and performance evaluation. The interpretation of results relied on a minimum accuracy threshold of 80% and a balanced F1-score as indicators of a robust predictive model. The methodology outlined above ensures that the Decision Tree model is developed with methodological rigor and provides transparent logic for interpreting customer behavior. By integrating structured stages of KDD with practical tools for data mining, the approach enables both systematic analysis and applicability in real-world ISP operations.

3. Results

A total of 250 active RumahNet subscribers in West Jakarta participated in the survey. Data were collected through a digital questionnaire distributed via Google Forms, with support from field technicians during installation, maintenance, or technical visits. The questionnaire assessed customer perceptions of internet speed, connection stability, pricing, technical service quality, and overall satisfaction, using a five-point Likert scale. The responses were converted into target labels for classification. Customers with an average score of ≥ 3.5 were categorized as *Satisfied*, while those scoring < 3.5 were classified as *Not Satisfied*. The threshold of 3.5 was chosen because it lies above the neutral midpoint (3.0) and approximates the “Agree” level (4.0), ensuring that only clearly positive responses were classified as satisfied.

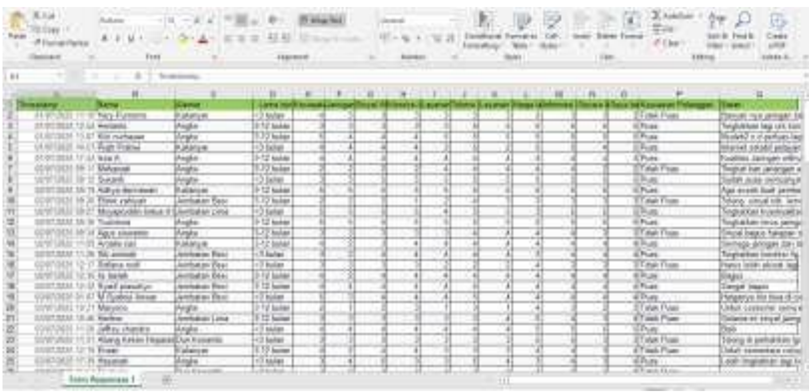

Figure 2. Data Labeling Process

The figure illustrates the classification of survey responses into two categories based on the Likert scale threshold of 3.5. This binary labeling process formed the basis for the target variable used in model training and testing.

Before modeling, the raw dataset was processed using RapidMiner Studio. The preprocessing steps included importing the dataset from Google Forms, removing missing or incomplete values, retaining only relevant attributes (internet speed, stability, pricing, and service quality), and storing the cleaned dataset for analysis.

Figure 3. Data Cleaning Workflow

The diagram shows the preprocessing stages: data import, handling missing values, attribute selection, and storing the cleaned dataset. This ensured the reliability and consistency of the input data for subsequent modeling.

The predictive model was developed using the C4.5 Decision Tree algorithm. The process included setting *Customer Satisfaction* as the target attribute, splitting the dataset into 70% training and 30% testing subsets, and generating the decision tree based on service attributes. The resulting decision tree highlighted internet speed and connection stability as the most influential attributes in classifying satisfaction levels, followed by pricing and service responsiveness. This suggests that technical performance plays a dominant role in shaping customer perceptions.

Figure 4. Decision Tree Model Output

The figure displays the decision tree structure generated by RapidMiner Studio. Internet speed appears at the top of the hierarchy, confirming its role as the strongest determinant of satisfaction, while connection stability and pricing form secondary branches in the classification rules.

The model was evaluated using a confusion matrix and standard performance metrics. From the testing data, the following values were obtained:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{57 + 7}{57 + 7 + 5 + 6} = 0.8533$$

$$Precision = \frac{TP}{TP + FP} = \frac{57}{57 + 5} = 0.9193$$

$$Recall = \frac{TP}{TP + FN} = \frac{57}{57 + 6} = 0.9047$$

$$F1 = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} = 0.9118$$

Table 1. Model Performance Metrics

Metric	Value
Accuracy	85,33%
Precision	91,93%
Recall	90,47%
F1-Score	91,18%

Table 1 summarizes the evaluation results. The high values of precision and recall demonstrate the model's ability to correctly classify satisfied customers while minimizing both false positives and false negatives.

Table 2. Confusion Matrix Output

	Predicted Satisfied	Predicted Not Satisfied
Actual Satisfied	57 (TP)	6 (FN)
Actual Not Satisfied	5 (FP)	7 (TN)

Table 2 shows the confusion matrix. The majority of satisfied customers were correctly identified (57 true positives), while only six were misclassified as not satisfied (false negatives). Similarly, five customers were incorrectly predicted as satisfied (false positives), while seven were correctly classified as not satisfied (true negatives).

4. Discussion

The results confirm that the Decision Tree (C4.5) algorithm is effective for predicting customer satisfaction in the context of RumahNet's FTTH service. With an accuracy of 85.33% and an F1-score exceeding 91%, the model meets and surpasses the acceptable threshold of 80% typically required in predictive classification studies.
The high precision value indicates that predictions of satisfied customers closely matched actual satisfaction levels, reducing the likelihood of overestimating positive experiences. Meanwhile, the strong recall score reflects the model's ability to capture the majority of genuinely satisfied customers, ensuring that the classification does not miss critical positive cases. The balanced F1-score confirms that the model performs consistently across both measures. From the decision tree structure, it is evident that internet speed and connection stability were the dominant factors influencing satisfaction. This is consistent with consumer expectations of FTTH services, where performance reliability is prioritized. Secondary factors such as pricing and technical support responsiveness also played roles, supporting previous findings that service quality and affordability jointly shape customer loyalty. For RumahNet, these insights highlight practical strategies for service improvement. Prioritizing investments in network speed and stability, while maintaining competitive pricing and responsive support, could enhance satisfaction and reduce churn. Furthermore, the interpretability of the Decision Tree model provides management with a transparent decision-support tool, allowing for proactive identification of at-risk customers and targeted service adjustments.

5. Conclusion

The conclusion of this study highlights that the predictive model for customer satisfaction with RumahNet's FTTH internet services in West Jakarta, developed using the Decision Tree (C4.5) algorithm, demonstrates reliable performance. The model achieved an accuracy of 85.33%, precision of 91.93%, recall of 90.47%, and an F1-score of 91.18%, reflecting balanced classification between satisfied and dissatisfied customers. The most influential factors in determining satisfaction include internet speed, overall satisfaction with WiFi services, alignment between price and service quality, willingness to recommend the service, responsiveness and courtesy of customer service, punctuality and problem-solving ability of technicians, and connection stability during activities such as streaming or video calls. These attributes serve as strategic indicators for service providers to continuously improve service quality. Moreover, the interpretability of the Decision Tree provides management with a clear visualization of customer behavior and satisfaction patterns. Overall, the findings confirm that applying data mining techniques with the Decision Tree algorithm is an effective approach to identifying key determinants of customer satisfaction and can serve as a solid foundation for data-driven decision-making.

Acknowledgment

The author sincerely expresses gratitude to all parties who contributed and provided support throughout the implementation of this research. Special thanks are extended to RumahNet for granting permission and facilitating the process of customer data collection. The author also conveys deep appreciation to the academic supervisor for valuable guidance, constructive feedback, and continuous support during the preparation of this study. Finally, heartfelt thanks are given to all respondents who willingly devoted their time to complete the questionnaire, thereby playing an essential role in the successful completion of this research.

References

Atrisia, M. I., Fatkar, B., Tiara, A., & Astuti, S. (2024). Assessing the impact of pricing, service, and product quality on Telkomsel customer loyalty. *Petra International Journal of Business Studies, 7*(1), 38–47. https://doi.org/10.9744/petraijbs.7.1.38-47

Dam, S. M., & Dam, T. C. (2021). Relationships between service quality, brand image, customer satisfaction, and customer loyalty. *Journal of Asian Finance, Economics and Business, 8*(3), 585–593. https://doi.org/10.13106/jafeb.2021.vol8.no3.0585

Jollyta, D., & Hendri, A. H. E. H. M. S. (2023). *Algoritma klasifikasi untuk pemula: Solusi Python dan RapidMiner*. Deepublish.

Deviana, M. E., & Tjahjaningsih, E. (2022). Pengaruh citra perusahaan dan kualitas layanan terhadap loyalitas dengan kepuasan sebagai mediasi pada pelanggan PT Indonesia Comnets Plus regional Jawa bagian tengah. *SEIKO: Journal of Management & Business, 5*(2), 454–466. https://doi.org/10.37531/sejaman.v5i2.2218

Hanggiasyifa, I., & Yulianto, A. E. (2021). *Pengaruh kualitas layanan terhadap kepuasan pelanggan PT First Media Tbk di Surabaya* (Skripsi). Sekolah Tinggi Ilmu Ekonomi Indonesia (STIESIA), Surabaya.Insani, K., & Alijoyo, F. A. (2024). Bisnis jaringan internet fiber optic di era digitalisasi untuk pemerataan infrastruktur telekomunikasi. *Jurnal Teknologi dan Sistem Informasi Bisnis, 6*(4), 799–805. https://doi.org/10.47233/jteksis.v6i4.1572

Kurniawan Maranto, A. R., Damayanti, L., & Ramadika, I. R. (2024). Perbandingan algoritma C4.5 dengan Naïve Bayes untuk menduga loyalitas pelanggan pada perusahaan internet service provider. *Bit-Tech, 7*(2), 396–405. https://doi.org/10.32877/bt.v7i2.1825

Mariati, E., & Lestari, A. (n.d.). Model klasifikasi kepuasan mahasiswa teknik terhadap sarana pembelajaran menggunakan data mining. *Jurnal Teknologi Informasi (JTI).*

Nugraha, F. T., & Hendry, H. (2023). Implementasi machine learning sebagai analisis kepuasan pelanggan terhadap penggunaan aplikasi KAI Access. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 8*(4), 1248–1255. https://doi.org/10.29100/jipi.v8i4.4185

Nurus, M., Nurdiawa, O., & Martanto, M. (2023). Analisis jaringan akses Fiber to the Home menggunakan teknologi Gigabit Passive Optical Network. *Jurnal Janitra Informatika dan Sistem Informasi, 3*(2), 56–66. https://doi.org/10.25008/janitra.v3i2.168

Pirmansyah, F., & Wahyudi, T. (2023). Implementasi data mining menggunakan algoritma C4.5 untuk prediksi evaluasi anggota satuan pengamanan: Studi kasus PT. YIMM Pulogadung. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 4*(3), 1566–1580. https://doi.org/10.35870/jimik.v4i3.370

Puspasari, A., Nurhayaty, E., Riftiasari, D., Yuniarti, P., Suharyadi, D., & Martiwi, R. (2022). The effect of service quality perception and company image on customer satisfaction and their impact on customer loyalty Indihome. *International Journal of Social and Management Studies (IJOSMAS)*, *3*(2), 220-233.

Rahman, R., & Sutanto, F. A. (2023). Data mining untuk memprediksi tingkat kepuasan konsumen Gojek menggunakan algoritma Naïve Bayes. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi, 18*(1), 8–18. https://doi.org/10.35969/interkom.v18i1.280

Reisyer, K. E., & Harman, R. (2024). Index kepuasan pelanggan Edukits Batam Centre dengan menggunakan algoritma C4.5. *CBIS Journal, 12*(1), 45–52. http://ejournal.upbatam.ac.id/index.php/cbis

Zanide, P. A. M., Zer, R. W. P. P., & Gunawan, I. (2022). Penerapan data mining Naïve Bayes dalam klasifikasi kepuasan mahasiswa berlangganan WiFi Indihome. *Jurnal Media Informatika (JUMIN)*, *3*(2), 112-118. https://doi.org/10.55338/jumin.v3i2.488

Zubair, A. (n.d.). Pengaruh kualitas pelayanan dan kualitas produk terhadap kepuasan pelanggan First Media Surabaya. *Sekolah Tinggi Ilmu Ekonomi Indonesia (STIESIA) Surabaya.*

.