Jurnal Informasi dan Teknologi

https://jidt.org/jidt

2024 Vol. 6 No. 2 Hal: 159-164 e-ISSN: 2714-9730

Prediction Analysis of Customer Satisfaction Levels at Company XXX Using the Classification Method

Evi Purnamasari[1]*, Ni Wayan Priscila Yuni Praditya[2], Dwi Asa Verano[3]

[1,3]Department of Informatics, Universitas Indo Global Mandiri
[2]Department of Computer, Universitas Indo Global Mandiri

evi.ps@uigm.ac.id

Abstract

Service in companies operating in the service system plays a very important role, including in one of the companies in the city of Palembang which we call Company XXX. The level of customer satisfaction with service at Company XXX needs to be considered in order to find out how satisfied customers are with the service system provided by Company XXX. On this occasion the researcher aims to analyze and predict the level of customer satisfaction at Company XXX using the C4.5 classification method. Customer satisfaction is an important factor in maintaining customer loyalty and improving company performance. Using historical customer data for the last 1 year, we apply the C4.5 algorithm to predict customer satisfaction levels. The research results show that the C4.5 method has quite high prediction accuracy, which reaches 83%. It is hoped that the findings from this research can help XXX Company identify the factors that influence customer satisfaction and be able to take strategic steps to improve the quality of service.

Keywords: Prediction, Customer, Service, Classification, C4.5.

1. Introduction

In this digital era, understanding and predicting customer satisfaction levels is very important for companies that want to maintain a competitive advantage. Customer satisfaction is not only related to the quality of the product or service offered, but also involves the overall experience experienced by the customer [1]. High satisfaction can increase customer loyalty, while dissatisfaction can cause a decrease in market share and company reputation [2]. The development of information technology has enabled companies to collect and analyze large amounts of customer data [3]. This data includes various aspects such as purchase history, interactions with customer service, as well as feedback through surveys and social media. By utilizing sophisticated data analysis techniques, companies can create predictive models to identify the factors that most influence customer satisfaction. One method used to predict customer satisfaction is machine learning [4]. This technique allows companies to analyze patterns in historical data and predict future levels of satisfaction. Machine learning algorithms such as linear regression, decision trees, and artificial neural networks can help in modeling the complex relationships between various variables that influence customer satisfaction. Utilizing predictions of customer satisfaction levels also allows companies to take proactive action [5] [6]. For example, if predictions indicate potential dissatisfaction, the company can immediately intervene such as offering discounts or improving service quality. This not only increases customer satisfaction but can also prevent customers from switching to competitors. Additionally, customer satisfaction predictions also provide valuable insights for product development and marketing strategies. By knowing the factors that most influence satisfaction, companies can focus on these aspects in efforts to develop new products or services. It also helps in designing marketing campaigns that are more effective and suit customer needs and expectations. However, despite advances in data analysis and prediction technology, challenges remain. One of them is ensuring that the data used is accurate and representative. Apart from that, companies must also maintain the privacy and security of customer data so as not to cause legal or ethical problems. Therefore, companies must develop good and ethical data management strategies in an effort to increase customer satisfaction through accurate predictions.

Customer satisfaction is one of the main indicators of business success, especially in the service industry where customer experience greatly influences their loyalty and retention [7] [8]. XXX Company, operating in this sector, seeks to understand and improve the level of satisfaction of its customers through predictive analytics. With advances in big data and machine learning technologies, classification techniques have become valuable

Receipt: 30-11-2023 | Revision: 02-06-2024 | Publish: 22-06-2024 | doi: 10.60083/jidt.v6i2.541

tools for analyzing customer data and predicting their behavior [9]. The classification method is a method in machine learning that aims to categorize objects into predetermined classes based on the input features [10]. This is a type of supervised learning where the model is trained using data that has been labeled. The C4.5 classification method is a decision tree algorithm that is used to predict the category or class of new data based on historical data that has been labeled [11]. This algorithm is known for its ability to handle numerical and categorical data, as well as producing easy-to-understand decision trees [12], [13] [14]. This research focuses on the application of the C4.5 algorithm to predict the level of customer satisfaction at Company XXX, with the aim of identifying the main factors that influence customer satisfaction and providing strategic recommendations for improvement.

2. Research Methods

In this research, researchers apply the C4.5 classification method which will predict the level of customer satisfaction with the service system in one of the companies in the city of Palembang to get a good accuracy value. The stages of this research are shown in Figure 1 below:

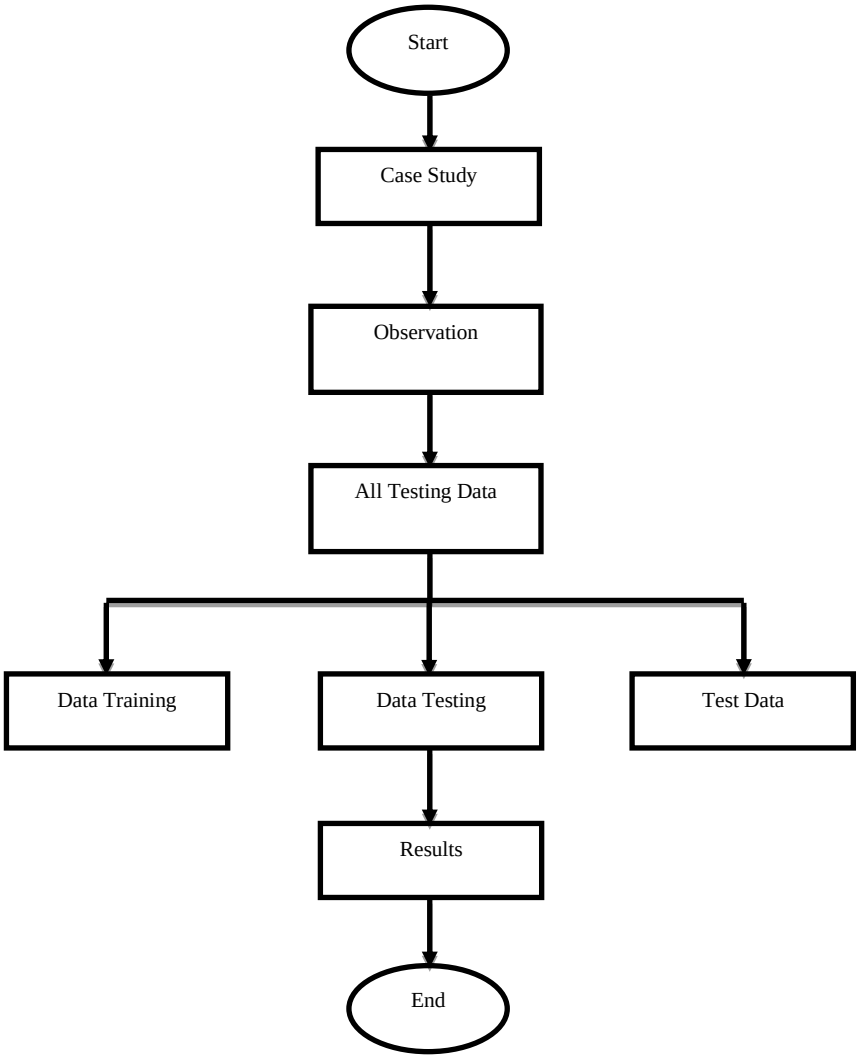

Figure 1. Research Stages

Researchers in this study use several stages for the process of solving problems which will be carried out in several stages, namely carrying out the process of searching for problems in Company XXX, then observing, then processing the testing process of all data which will be divided into several testing processes using using the C4.5 classification method including training data, test data and test data, then conclusions will be obtained from the test results.

2.1. Classification

One of the well-known classification methods and has high prediction results in the application of data mining [15]. Classification is an important component in machine learning with wide applications in various fields [16]. Advances in algorithms and techniques have improved the accuracy and efficiency of classification models, making them very useful tools in data analysis. The data mining processing carried out in this research follows the stages in Knowledge Discovery in Database (KDD) to produce information according to a predetermined sequence [17], along with the stages in this classification, including :

a. Data Preprocessing
   Removing noise, handling missing data, and feature normalization, underscore the importance of preprocessing in improving the performance of classification models [18].
b. Dataset Sharing
   Split the dataset into training and test sets for model evaluation. A common division is 70% of the data for training and 30% for testing [19].
c. Model Training
   Uses a training dataset to teach the model to recognize relevant patterns. This shows that selecting appropriate hyperparameters during training can significantly improve model performance [20].
d. Model Evaluation
   Using metrics such as accuracy, precision, recall, and F1-score to measure model performance, therefore suggests the use of multiple metrics for comprehensive evaluation of classification models [21].

2.2. Method Decision Trees C4.5

Tree-based methods that break down a dataset into smaller subsets based on features, resulting in a model that can be used for class prediction. Decision trees are easy to interpret but are susceptible to overfitting if not pruned well [22]. C4.5 is a machine learning algorithm used to build decision trees. This algorithm is a development of ID3 (Iterative Dichotomiser 3) which was developed by Ross Quinlan. C4.5 is used for data classification and is known for its ability to handle continuous data and missing attributes. C4.5 is a powerful and flexible algorithm for classification, capable of handling diverse and complex data. Although it has some disadvantages in terms of computational complexity, its advantages in dealing with continuous data and missing attributes make it a popular choice in various applications. The process stages in the C4.5 Algorithm are as follows :

a. Finding Information Gain: Calculate the information gain for each attribute and select the attribute with the highest information gain to become the dividing node. High information gain ensures that data sharing is as optimal as possible [23].
b. Building a Tree: Uses selected attributes to divide the dataset and repeats this process recursively for the resulting subsets until all data in the subset has the same class or there are no more attributes to divide. Here it is stated that this recursive process allows the formation of detailed and accurate trees [24].
c. Tree Pruning: Performs pruning to reduce the size of the tree and improve generalization by removing branches that make a small contribution to the classification. Pruning is effective for reducing overfitting [25].

3. Results and Discussion

The data produced by this research is data from one of the service companies in the city of Palembang. The data used is customer satisfaction data from the last year from researchers conducting tests, including from May 2023 to April 2024. This data has a total of 3078 customer data. The following is a table of the data:

Table 1. Research data

No	Month	PR	PS	D
1	April 2024	107	15	1
2	March 2024	70	8	7
3	February 2024	69	18	1
4	January 2024	333	15	10
5	December 2023	197	23	1
6	November 2023	49	0	1
7	October 2023	280	31	1
8	September 2023	252	56	1
9	August 2023	248	80	12
10	July 2023	353	46	2
11	June 2023	411	62	24
12	May 2023	263	69	15

Where:

PR = Promotor
PS = Pasiver
D = Detractor

PR is an assessment between 9-10, while PS is an assessment between 6-8, while D is an assessment between 1-5. From this data, the researcher then determines the attributes that will be used in this research process. There are 12 (twelve) attributes in question, namely as follows:

Table 2. Attribute

No	Attribute	Information
1	April 2024	√
2	March 2024	√
3	February 2024	√
4	January 2024	√
5	December 2023	√
6	November 2023	√
7	October 2023	√
8	September 2023	√
9	August 2023	√
10	July 2023	√
11	June 2023	√
12	May 2023	√

Data from the attributes above are used in accordance with the needs of researchers which is expected to make it easier for researchers in the calculation process using the C4.5 classification method. Researchers use a classification method where the existing data will be subjected to a data classification process so that the calculation process can be carried out using the C4.5 method. The data has been classified by combining the ratings between Types PR and PS because they can still be categorized as good ratings from customers while D remains alone which is categorized as bad ratings from customers. The classification data includes the following:

Table 3. Data Classification

No	Month	Good	Bad
1	April 2024	122	1
2	March 2024	78	7
3	February 2024	87	1
4	January 2024	348	10
5	December 2023	220	1
6	November 2023	49	1
7	October 2023	311	1
8	September 2023	308	1
9	August 2023	364	12
10	July 2023	311	1
11	June 2023	473	24
12	May 2023	332	15

In this research, the C4.5 method will produce a decision tree, which means we will determine which attribute will become the root, this can all be seen from the results of the highest gain value of all the attributes we use. Calculations on the data we use can be seen in the following table:

Table 4. Node Calculation

Month	Amount of data	Good	Bad	Entropy	Gain
	3078	**3008**	**70**	**0.095857334**	
					0.11770929
April 2024	123	122	1	0.044761789	
March 2024	85	78	7	0.182860378	
February 2024	88	87	1	0.057101874	
January 2024	358	348	10	0.104456209	
December 2023	221	220	1	0.028726138	
November 2023	50	49	1	0.084313705	
October 2023	312	311	1	0.021939172	
September 2023	309	308	1	0.022107136	
August 2023	376	364	12	0.113304317	
July 2023	312	311	1	0.021939172	

June 2023	497	473	24	0.143171968
May 2023	347	332	15	0.13490696

The results of the calculation trial process using the C4.5 classification method, here are results that have customer satisfaction values from calculation results using data from the last 1 year. March has the highest Entropy value compared to other months, from here the Team Leader can discuss what statements will be made to customers in the future.

3. Conclusion

In this research, the highest gain value was the gain in March 2024. Meanwhile, the lowest gain was in October 2023 and July 2023. Results from testing using the weka application, tsshis research shows that the C4.5 classification algorithm is effective in predicting the level of customer satisfaction at Company XXX, with a prediction accuracy level of 83%. These results show that the C4.5 method can capture complexity and variation in customer data well. By identifying the main factors that influence customer satisfaction, Company XXX can direct their efforts to improve the most influential aspects of service, thereby increasing customer loyalty and overall business performance.

Acknowledgments

Thank you to one of the Team Leaders at Company XXX, for his support who gave permission to carry out a field survey and provide data related to what data the researcher needed while the researcher was carrying out this research process.

References

[1] I. G. Ngurah, S. Wijaya, E. Triandini, E. Tifanie, and G. Kabnani, "E-commerce website service quality and customer loyalty using WebQual 4 . 0 with importance performances analysis , and structural equation model : An empirical study in Shopee," *J. Ilm. Teknol. Sist. Inf.*, vol. 7, no. July, pp. 107–124, 2021.

[2] M. Javaid, A. Haleem, and R. Pratap, "BenchCouncil Transactions on Benchmarks , Standards and Evaluations ChatGPT for healthcare services : An emerging stage for an innovative perspective," *BenchCouncil Trans. Benchmarks, Stand. Eval.*, vol. 3, no. 1, p. 100105, 2023, doi: 10.1016/j.tbench.2023.100105.

[3] A. Sestino, M. Irene, L. Piper, and G. Guido, "Technovation Internet of Things and Big Data as enablers for business digitalization strategies," *Technovation*, vol. 98, no. August, p. 102173, 2020, doi: 10.1016/j.technovation.2020.102173.

[4] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, no. January 2019, 2020, doi: 10.21275/ART20203995.

[5] N. Almumtazah, N. Azizah, Y. L. Putri, I. Negeri, and S. Ampel, "Prediksi jumlah mahasiswa baru menggunakan metode regresi linier sederhana," *J. Ilm. Mat. dan Terap.*, vol. 18, no. 1, pp. 31–40, 2021, doi: https://doi.org/10.22487/2540766X.2021.v18.i1.15465.

[6] F. Fahreni, V. Mardina, I. Indriaty, and R. Ramaidani, "Examination of Gel Hand Sanitizer from Mangrove Leaves and Patchouli Oil Against Sthapylococcus Aureus," *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 4, 2021, doi: 10.52088/ijesty.v1i4.139.

[7] C. Lovelock, *SERVICES MARKETING*. 2022.

[8] I. G. Dharma Utamayasa, "Efect Physical Activity and Nutrition During The Covid-19 Pandemic," *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 1, 2021, doi: 10.52088/ijesty.v1i1.58.

[9] E. C. Ates, G. E. Bostanci, and M. S. Guzel, "Big Data , Data Mining , Machine Learning , and Deep Learning Concepts in Big Data , Data Mining , Machine Learning , and Deep Learning Concepts in Crime Data," vol. 8(2), no. December, pp. 293–319, 2020, doi: 10.26650/JPLC2020-813328.

[10] A. S. Borodulin, "Phase-Sensitive OTDR Using Pattern Recognition Methods," *sensors Artic.*, vol. 23, no. 582, 2023, doi: https://doi.org/10.3390/s23020582.

[11] E. Purnamasari, "Prediksi Perkembangan Nilai Impor Komoditas Utama," *J. Inf. dan Teknol. Vol.*, vol. 5, no. 1, pp. 165–172, 2023, doi: 10.37034/jidt.v5i1.271.

[12] E. Purnamasari, D. P. Rini, and Sukemi, "Prediction of the Student Graduation's Level using C4.5 Decision Tree Algorithm," *ICECOS 2019 - 3rd Int. Conf. Electr. Eng. Comput. Sci. Proceeding*, pp. 192–195, 2019, doi: 10.1109/ICECOS47637.2019.8984493.

[13] M. M. S and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for International Journal of Intelligent Systems and Applications in Engineering Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. Eng.*, no. January 2019, pp. 88–91, 2021, doi: 10.1039/b000000x.

[14] Hartono, E. Ongko, and D. Abdullah, "HFLTS-DEA model for benchmarking qualitative data," *Int. J. Adv. Soft Comput. its Appl.*, vol. 11, no. 2, 2019.
[15] E. Purnamasari, D. Palupi Rini, P. Studi Magister Ilmu Komputer, F. Ilmu Komputer, and U. Sriwijaya Palembang, "Seleksi Fitur menggunakan Algoritma Particle Swarm Optimization pada Klasifikasi Kelulusan Mahasiswa dengan Metode Naive Bayes," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 1, no. 3, pp. 469–475, 2020, doi: https://doi.org/10.29207/resti.v4i3.1833.
[16] E. A. Rady and A. S. Anwar, "Informatics in Medicine Unlocked Prediction of kidney disease stages using data mining algorithms," *Informatics Med. Unlocked*, vol. 15, no. March, pp. 1–7, 2019, doi: 10.1016/j.imu.2019.100178.
[17] M. Jannah *et al.*, "Prediksi Penjualan Produk Pada PT Bintang Sriwijaya Palembang Menggunakan K-Nearest Neighbour," *J. Softw. Eng. Comput. Intell.*, vol. 01, no. 02, pp. 80–89, 2023.
[18] S. Albahra *et al.*, "Seminars in Diagnostic Pathology Artificial intelligence and machine learning overview in pathology & laboratory medicine : A general review of data preprocessing and basic supervised concepts," vol. 40, no. February, pp. 71–87, 2023, doi: 10.1053/j.semdp.2023.02.002.
[19] E. N. Boice *et al.*, "Training Ultrasound Image Classification Deep-Learning Algorithms for Pneumothorax Detection Using a Synthetic Tissue Phantom Apparatus," *J. Imaging*, vol. 8, no. 249, pp. 1–13, 2022, doi: https://doi.org/10.3390/jimaging8090249.
[20] J. Wu, X. C. Hao, Z. L. Xiong, and H. Lei, "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization," *J. Electron. Sci. Technol.*, vol. 17, no. 1, pp. 26–40, 2019, doi: 10.11989/JEST.1674-862X.80904120.
[21] I. J. Holb and V. Eva, "Classification Assessment Tool : A program to measure the uncertainty of classification models in terms of class-level metrics ´ rd Szab o," vol. 155, no. April 2023, 2024, doi: 10.1016/j.asoc.2024.111468.
[22] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," vol. 02, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
[23] S. Tangirala, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm *," vol. 11, no. 2, pp. 612–619, 2020.
[24] M. Fokkema, J. Edbrooke-childs, and M. Wolpert, "Generalized linear mixed-model (GLMM) trees : A flexible decision-tree method for multilevel and longitudinal data," *Psychother. Res.*, vol. 31, no. 3, pp. 329–341, 2021, doi: 10.1080/10503307.2020.1785037.
[25] F. Avellaneda, "Efficient Inference of Optimal Decision Trees," *Proc. AAAI Conf. Artif. Intell.*, pp. 3195–3202, 2018, doi: https://doi.org/10.1609/aaai.v34i04.5717.