



Implementation of the C4.5 Algorithm for Classification of Shipping Service Data in Shipping Companies

Siti Nur Khasanah¹, Sandra Jamu Kuryanti², Cep Adiwihardja³, Maruloh⁴

^{1,4}Information Systems, Nusa Mandiri University, Jl. Raya Jatiwaringin No 2, East Jakarta 13620

^{2,3}Information Systems, Bina Sarana Informatika University, Jl. Kramat Raya No 98, Central Jakarta, 10420

ARTICLE INFO

Article history:

Received Aug 25, 2022

Revised Aug 30, 2022

Accepted Sep 15, 2022

Keywords:

C4.5 Algorithm

Data mining

Decision Tree

Delivery service

RapidMiner.

ABSTRACT

A company engaged in the field of goods delivery services or better known as expedition services which is one of the fields that is growing rapidly with the needs of national freight forwarding services. However, from the delivery of entrustments that have been recorded, they are only stored as recording reports. With the existing problems, the current research that the author conducts aims to implement one of the algorithms in data mining, namely the C4.5 Algorithm to produce complex decisions to be simpler such as data for customers to better challenge customers with large-scale groups so as not to turn away or move to other expedition services. The results showed that the tree resulting from the C4.5 method with the C4.5 Algorithm had an accuracy percentage of 84.00%. The data processing carried out in this study used the help of the RapidMiner application to obtain the model. The results of this study can recommend marketing the sale of freight forwarding services with the criteria of what kind of customers will get a discount.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Siti Nur Khasanah,
4Information Systems,
Nusa Mandiri University,
Jl. Raya Jatiwaringin No 2, East Jakarta 13620, Indonesia
Email: siti.skx@nusamandiri.ac.id

1. INTRODUCTION

PT. Tiga Permata Ekspres are one of the freight forwarding service companies that deliver goods nationally or domestically. This company was built in 2001 in Sidoarjo which only accepts goods delivery services on the island of Java. But over time, PT. Tiga Permata Ekspres opened branches in Tangerang and Semarang, so that it can develop a wide area reached by this company covering shipping throughout Indonesia. Technological advances in the information system of shipping goods are very influential on this company in order to determine and calculate the frequency of delivery in high intensity of demand (Muhammad Akbar Ramadhan1, 2019)

One of the activities in the freight forwarding expedition business is administration in managing goods delivery data (Witten et al., 2011). The obstacles that occur are data loss due to lack of administrative order, data errors or incomplete data on the delivery of

goods will make this expedition service hampered in the delivery of goods (Listriani et al., 2016) The division of labor of each courier is complicated so that the admin must know the couriers who can carry out delivery activities (Wijaya et al., 2014). Delivery of goods can occur on time or if there are obstacles, it can take a long time to deliver. This makes consumers want to know the tracking of their goods delivery if they have not used the system, the process of tracking goods will be difficult to do. So this is what causes frequent misunderstandings between the company and couriers or with operators which can result in time and cost losses.

Based on this process, in this study a decision-making method was used with the C4.5 algorithm to analyze data patterns and classify data from databases available in PT. Tiga Permata Ekspres. In another study, one of the shipping companies increased promotional activities so as to attract shipping interests that would attract customers to be more active because of the provision of offers or promotions in the delivery of goods carried out (Damanik et al., 2020). Data management using the C4.5 Algorithm Classification resulted in a value of 96.28% in the implementation of the data analysis carried out. Other research also uses the C4.5 Algorithm method to manage freight forwarding data so that the results can be used to analyze existing markets or find new opportunities and find strategic plans to increase profits (Wulandari, 2019)

2. RESEARCH METHOD

Methods of conducting research activities based on the following figure:(Khasanah, 2017)

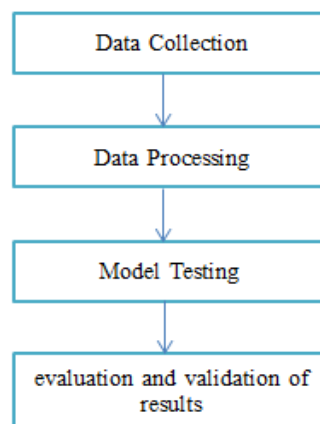


Figure1. Reseach Method

2.1 Data Collection

The first stage in the implementation of this study is data collection, in this study using primary data obtained directly from the object of study.

2.2 Data Processing

At this stage, data processing will be carried out for research, so that the variables needed to be processed and used in research will be obtained

2.3 Model Testing

At this stage, the submission of the algorithm model used, namely the C4.5 Algorithm, will be submitted. The modeling of the C4.5 algorithm is carried out by finding the highest gain value of each attribute.

2.4 Evaluation and Validation of Results

At the evaluation stage, it is called the classification stage because at this stage testing will be determined for accuracy. The testing stage is to see the results of accuracy in the C4.5 Algorithm classification process as well as evaluation with the ROC Curve.

3. RESULTS AND DISCUSSIONS

This study use customer data in.PT. Permata Express. The data collected was 93 data.

Table 1.
Attribute on The Study

Attributes	Value
Discount	Yes No
By	Land Air
Region	P. Jawa Jabodetabek Luar Pulau Jawa
Price	High Low
Publish Rate	Minimum Standar Maksimal

To select the root attribute, it is based on the highest gain value of the existing attributes. To get the gain value, it must first be determined the entropy value:

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$$

$$= \left(-\frac{56}{93} \cdot \log_2\left(\frac{56}{93}\right)\right) + \left(-\frac{37}{93} \cdot \log_2\left(\frac{37}{93}\right)\right)$$

$$= 0.969678702$$

The next step is to calculate the Entropy and Gain, Entropy and Gain values shown in table 2.

Table 2.
Entropy and Gain Calculation Results

Atribut	Value	Jumlah	Ya	Tidak	Entropy	Gain
Total		93	37	56	0.9696787	
By						0.0291138
	Land	91	35	56	0.9612366	
	Air	2	2	0	0	
Region						1.2035940
	P.Jawa	47	3	44	0.3424637	
	Jabodetabek	30	23	7	0.7837769	
	Luar P.Jawa	16	11	5	0.8960382	
Price						0.8710570
	Tinggi	56	13	43	0.7817281	
	Rendah	37	24	13	0.9352691	
Publish Rate						0.8900263
	Minimum	55	13	42	0.7889406	
	Standart	23	14	9	0.9656361	
	Maksimal	15	10	5	0.9182958	

The results of data processing using the C4.5 algorithm obtained a decision tree that was formed as follows:

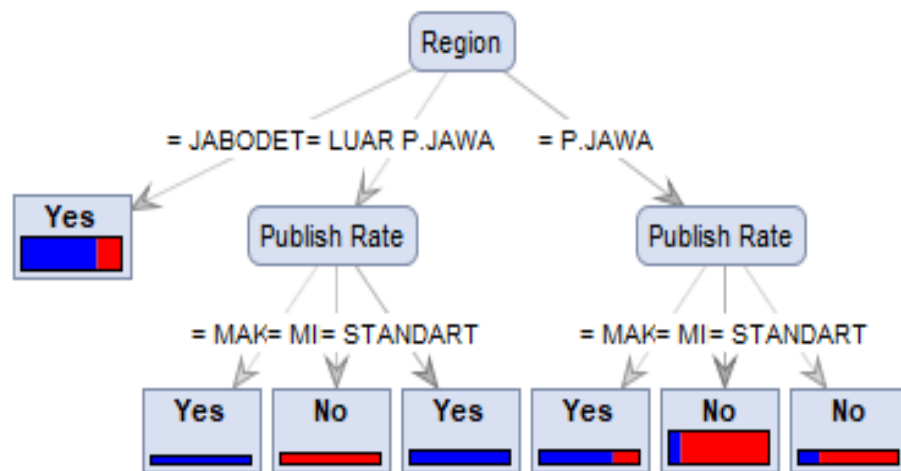


Figure 2. Decision trees formed using the C4.5 algorithm

The next step is to validate the model that has been created using Cross Validation:

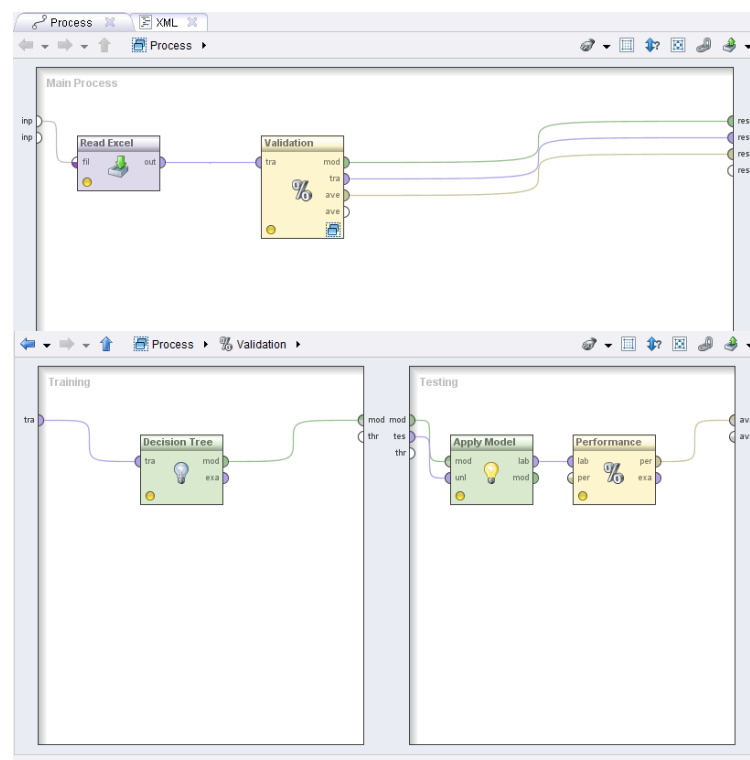
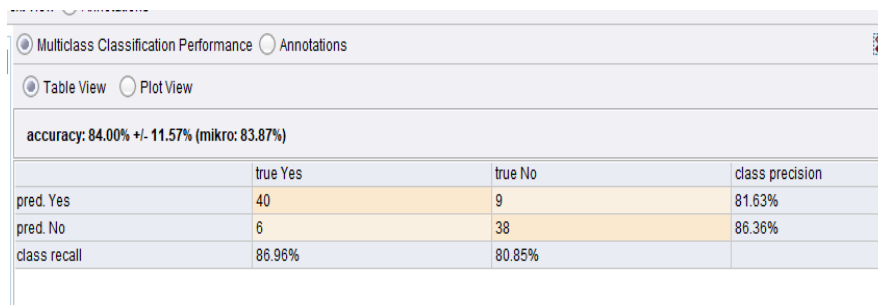


Figure 3. Evaluation of the C4.5 Algorithm by using Cross Validation



Multiclass Classification Performance			
accuracy: 84.00% +/- 11.57% (mikro: 83.87%)			
	true Yes	true No	class precision
pred. Yes	40	9	81.63%
pred. No	6	38	86.36%
class recall	86.96%	80.85%	

Figure 4. The result of data validation accuracy using the C4.5 algorithm

Based on testing using the Ten-Fold CrossValidation method, it produced an accuracy value of 84.00%, the precision class predicted NO by 86.36%. With a recall class of 80.85%. As for the precision class, the predicted value of YES is 81.63% with the recall class of 86.95%. This shows that, from the classification process carried out, it will be able to be applied to recommendations for providing discounts to customers in the delivery of goods. The highly accurate precision and recall that is still in the good category concludes that the researcher was successful in implementing the C4.5 classification algorithm well and can help PT. Three Permata Ekspres in recommending the service marketing process in the form of discounts to customers must require a lot of consideration such as looking at the amount or number of goods to be sent or the destination of the location to be sent.

4. CONCLUSION

The test results of the C4.5 algorithm using goods delivery data at PT. The three Express Gems resulted in a good accuracy of 84.00%. class precision prediction NO of 86.36%. With a class recall of 80.85%. As for the precision class, the predicted value of YES is 86.95% with the recall class of 91.89%. So it can be concluded that the researcher was successful in implementing the C4.5 classification algorithm.

REFERENCES

- Damanik, N. S., Erwansyah, K., Kom, S., Kom, M., & Halim, J. (2020). *Penerapan Data Mining Untuk Menentukan Pola Kombinasi Produk yang Dapat Diberikan Diskon pada PT . Home Center Indonesia (INFORMA) Menggunakan Algoritma Apriori. x.*
- Wulandari, R. T. (2019). Analisa Data Dan Perancangan Aplikasi Service Pelanggan Pt.Jne Untuk Peningkatan Kualitas Layanan. *Petir*, 9(1), 21–27. <https://doi.org/10.33322/petir.v9i1.184>
- Khasanah, S. N. (2017). Penerapan Algoritma C4.5 Untuk Penentuan Kelayakan Kredit. *None*, 14(1), 9–14.
- Listriani, D., Setyaningrum, A. H., & Eka, F. (2016). PENERAPAN METODE ASOSIASI MENGGUNAKAN ALGORITMA APRIORI PADA APLIKASI ANALISA POLA BELANJA KONSUMEN (Studi Kasus Toko Buku Gramedia Bintaro). *Jurnal Teknik Informatika*, 9(2), 120–127.
- Muhammad Akbar Ramadhan1, S. D. (2019). Klasifikasi Jasa Pengiriman JNE Service Menggunakan Algoritma Decision Tree. *Jurnal Ilmiah Fakultas Ilmu Komputer*, 8, 59–67.
- Wijaya, K., Jakarta, P. N., & Apriori, A. (2014). *Penerapan algoritma apriori sebagai penentu diskon paket barang berdasarkan pola pembelian konsumen.*
- Witten, I. H., Frank, E., & Hall, M. a. (2011). Data Mining. In *Data Mining* (Vol. 277, Issue Tentang Data Mining). [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)