International Journal of Advances in Data and Information Systems
Vol. 5, No. 1, April 2024, pp. 39~48
ISSN: 2721-3056, DOI: 10.59395/ijadis.v5i1.1316

Recommendation of Prospective Construction Service Providers in Government Procurement Using Decision Tree

Eva Yustina[1], Mokhamad Amin Hariyadi[2], Cahyo Crysdian[3*]
[1]Dinas Komunikasi dan Informatika Kota Batu
[1,2,3]Magister Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia

Article Info

Article history:

Received Feb 08, 2024
Revised Feb 28, 2024
Accepted Apr 24, 2024

Keywords:

Procurement of Goods/ Services;
Construction Service Provider;
Decision Tree

ABSTRACT

The determination of prospective construction service providers using the direct procurement method is the authority of the Goods/ Services Procurement Officer. Administrative requirements are an important factor in selecting prospective construction service providers. The use of the *decision tree* method in this study is to find out, determine, and analyse the variables that influence the assessment of the feasibility of prospective construction service providers, and get an accuracy value in providing an assessment of the feasibility of prospective construction service providers. The data used in this study are 153 datasets consisting of 13 variables. The existing variables are divided into basic variables and additional variables. The basic variables consist of 5 variables, namely experts, work experience, quality of work, winning tenders and contract value. While the additional variables consist of 8 variables namely business entity status, business entity form, business entity NPWP, business entity domicile, business entity qualification, type of business licence, percentage of work and construction services business licence. By using the *decision tree* method, the accuracy on the basic variable is 84.84%. The addition of additional variables to the basic variables resulted in an accuracy of 90.91%. This shows that by adding additional variables the accuracy results are higher than using only the basic variables.

Corresponding Author:

Cahyo Crysdian,
Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang
Jl. Ir. Soekarno No. 34, Dadaprejo, Kota Batu, 65233, Batu, Indonesia
Email: cahyo@ti.uin-malang.ac.id

1. INTRODUCTION

Digital technology is currently developing rapidly in almost all lines including in government. The use of an electronic-based government system in government by utilizing information and communication technology to facilitate its users. With the existence of an electronic-based government system, it is hoped that it will be able to carry out clean, effective, transparent and accountable governance. In the field of goods and services procurement, in accordance with Presidential Regulation Number 12 of 2021 concerning Amendments to Presidential Regulation Number 16 of 2018 concerning Government Goods / Services Procurement, it is described that the implementation of goods / services procurement is carried out electronically using an information system consisting of an electronic procurement system and supporting systems. In construction procurement, you can use the tender, direct appointment and direct procurement methods [1], [2]. Construction services according to the construction services law is an activity to build facilities or infrastructure which in the process includes the construction of buildings, mechanical & electrical installations, and also the construction of civil infrastructure [2]–[8].

To determine prospective construction service providers with the direct procurement method is the duty and authority of the Goods / Services Procurement Officer. In this study, the determination of the eligibility of prospective providers is taken from the assessment conducted by the Commitment Making Officer through the Provider Performance Information System application. Based on the Regulation of the Government Goods / Services Procurement Policy Agency of the Republic of Indonesia Number 4 of 2021 concerning the Guidance of Government Goods / Services Procurement Business Actors [1], [2], [6], [7], [9], [10].

Based on research that has been conducted to determine influential factors in the selection of construction service providers, the results of administrative requirements, response to inspection, work execution, commitment to maintenance responsibility, technical completeness of construction work providers, and work experience. However, the most important factor is the administrative requirements of construction work providers [4].

The number of construction service providers that exist in each region requires a study to map feasible and ineligible prospective providers. A popular approach used to make predictions is classification [5]. Classification methods that are widely used in research include: Naïve Bayes, Linear Regression, Logistic Regression, Support Vector Machines, Neural Network, Decision Tree and Random Forest [5], [6], [8], [11], [12].

2. RESEARCH METHOD

This research consists of several stages starting from data collection followed by data preposessing, system design, followed by implementation using the decision tree method [13], [14]. Furthermore, analysis and evaluation of the results obtained and the last stage of drawing conclusions. The flow of research design can be seen in Figure 1.

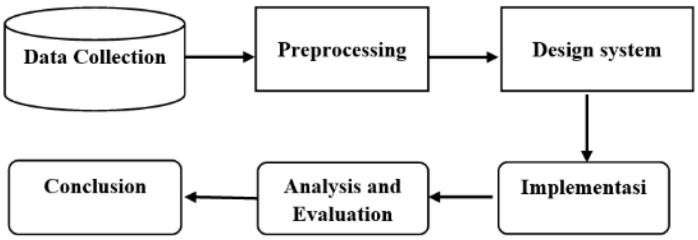

Figure 1. Research Design

2.1 Data Collections

The data used in this study is primary data, the data is collected directly from the results of the provider assessment by the Commitment Making Officer on the provider performance information system (https://sikap.lkpp.go.id/) until September 30, 2023. The specific administrative data obtained are in the form of the number of experts, the amount of work experience, the quality of the work results, the number of tenders won, the value of the contract, the status of the business entity, the form of the business entity, the suitability of the NPWP of the business entity, the domicile of the business entity, the qualifications of the business entity, the type of business license, the percentage of work and the suitability of the construction service business license.

2.2 Pre-Processing

The data obtained does not all meet the data sufficiency variable, so preprocessing is required. Preprocessing is the processing of raw data into ready-to-use data that can be understood more easily by the system. The stages in data preprocessing include:

a. Data cleaning. The raw data obtained is selected and then deletes data that is not revelant and incomplete. In data collection there are some incomplete data because the data is not published.

b. Data integration. The data obtained is combined in a unified data (dataset). The data merging process is adjusted to the same format.
c. Data transformation. At this stage the data will be normalized and generalized. Data normalization is intended so that there is no redundant data while generalization by uniforming the data.
d. Data Reduction. Reducing the amount of data is necessary if the data used is large. In this study, data reduction is not used because the data obtained is not in the large number category [14]–[19].

2.3 Design System

The system design in this study is used to assess the feasibility of prospective construction service providers using the Decision Tree method. The flow of system design can be seen in Figure 2.

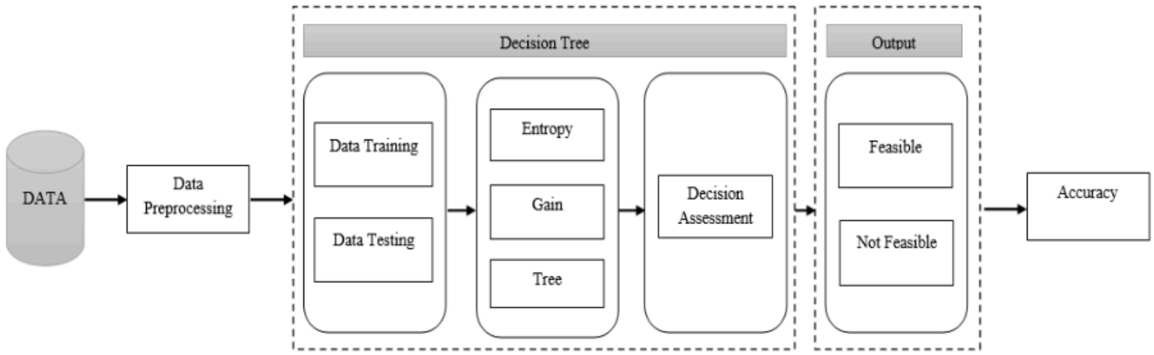

Figure 2. Design System

Figure 2 describes the flow of data processing, data processing using Decision Tree by applying certain rules to get a feasibility assessment and scenario tests to obtain output in the form of accuracy. Accuracy is the level of closeness between the predicted value and the actual value. Precision shows the level of accuracy/accuracy in classification. Recall serves to correctly identify positive actual values [16], [20], [21]. To get the accuracy value, you can use the equation:

$$Accuration = \frac{TP+TN}{TP+TN+FP+FN} X\ 100\% \tag{1}$$

$$Precitions = \frac{TP}{TP+FP} X\ 100\% \tag{2}$$

$$Recall = \frac{TP}{TP+FN} X\ 100\% \tag{3}$$

With:

$TP\ =\ The\ actual\ feasible\ value\ and\ the\ predicted\ feasible\ value$
$FP\ =\ The\ actual\ value\ is\ not\ feasible\ and\ the\ predicted\ value\ is\ feasible$
$FN\ =\ Actual\ worthy\ value\ and\ non-worthy\ prediction\ value$
$TN\ =\ The\ actual\ value\ is\ not\ feasible\ and\ the\ predicted\ value\ is\ not\ feasible$

The dataset obtained is split into two parts, namely training data and test data. Training data is data that serves to build and train classification models. Meanwhile, test data is data used for testing using classifications that have previously been built with training data [13]. The division of training data and test data generally uses 70-80% training data and 20-30% test data from the available data. Based on empirical testing, the best division for training data and test data is 80%: 20%.

2.4 Decision tree method for eligibility assessment

Decision tree is used to determine predictions using the dataset obtained. Flowchart of the prediction process using Decision Tree in this study is as follows.

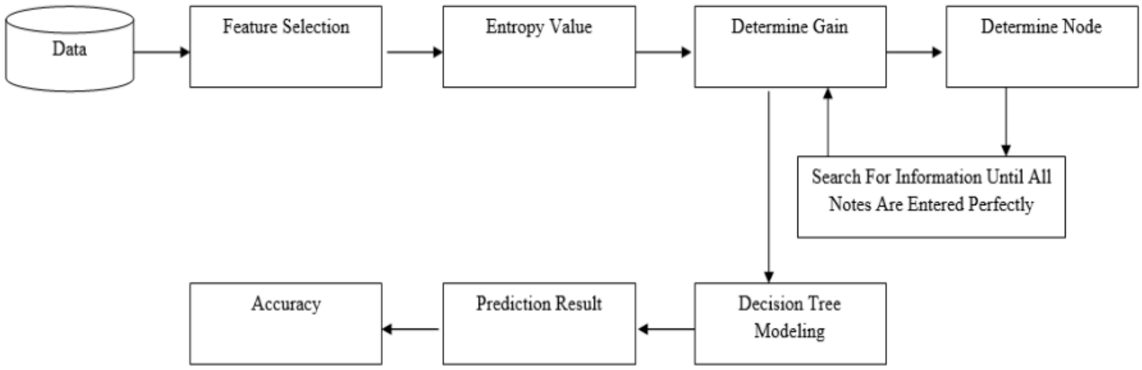

Figure 3. Flowchart *decision tree*

Figure 3 is the reference for this research in processing data. Decision tree is a tree structure where each node represents an attribute that has been tested and the leaf nodes represent certain class groups. Decision trees can be used to divide large data into smaller sets of records by applying a predetermined set of rules [7]. In general, the Decision tree method for building decision trees consists of several stages, including:

a. Determine the attribute as the root;
b. Determine the branch of each value;
c. Dividing the cases in the branches;
d. The process is repeated for each branch until all cases in the branch have the same class.

Selection of attributes that will be used as roots is by calculating the gain value of all attributes, and the gain with the highest value is used as the first root. But before determining the gain value, first calculate the entropy value. Entropy value is obtained by the following equation:

$$Entropy\ (S) = \sum_{i=1}^{n} - pi * log_2 pi \tag{4}$$

With:
$S =\ case\ set$
$n = number\ of\ partitions\ S$
$pi = proportion\ of\ Si\ to\ S$

The gain value is obtained with the following equation:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|s|} * Entropy\ (Si) \tag{5}$$

With:
$S\ =\ Case\ Set$
$A\ =\ Attributes$
$n\ =\ Number\ of\ Partitions\ of\ Attribute\ A$
$|Si|\ =\ Number\ of\ Cases\ in\ the\ i-th\ Partition$
$|S|\ =\ Number\ of\ Cases\ in\ S$

3. RESULTS AND DISCUSSION

Testing was carried out with two scenarios, the first scenario by processing data using 5 basic variables and the second scenario by using 13 variables consisting of 5 basic variables plus 8 additional variables.

Scenario 1

Before testing, first determine the rules that can be used to determine the criteria for eligibility and ineligibility of prospective construction service providers. The feasible rules used in scenario 1 are as follows:

Table 1. Feasibility assessment rules Scenario 1

No.	Variable	Critera
1.	Expertise	Have more than 3 peoples
2.	Work expererience	Have worked on at least 4 jobs
3.	Quality of work	More than 80%
4.	Winning tender	Have won the tender 3 times
5.	Contract value	More than Rp. 200.000.000,-

The rules in table 1 are used to get a feasible value if all criteria are met and not feasible if one or more criteria are not met. Based on the rules of table 1, the following results are obtained:

Table 2. Feasibility results of Scenario 1

No.	Description	Feasible	Not feasible
1.	Prospective construction service providers (number of business entities)	22	131

The next stage is modeling the decision tree by calculating the entropy and gain values for each variable using equations 1 to 4. From the calculation, the resulting gain and entropy values are as follows:

Table 3. Entropy and Gain values of scenario 1

No.	Variable	Feasible entropy	Feasible entropy	Gain
1.	Expertise	0.971	0.306	**0.179**
2.	Work expererience	0.644	0.242	0.016
3.	Quality of work	0.606	0	0.067
4.	Winning tender	0.801	0.242	0.067
5.	Contract value	0.701	0.314	0.087

Based on table 3, the highest gain value is obtained in the power variable so that the variable is used as the root of the decision tree. The resulting decision tree modeling is as follows:

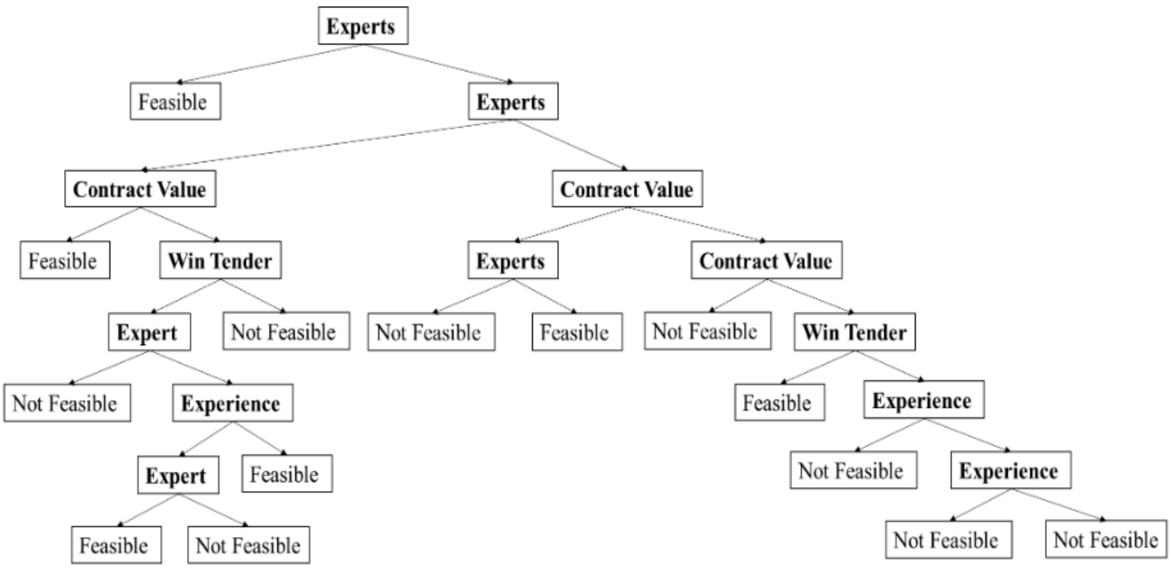

Figure 4. Scenario 1 Decision Tree

Based on the experimental results in scenario 1 using the decision tree method, the confusion matrix table is obtained as follows:

Table 4. Confussion Matrix of Scenario 1

		Actual	
Prediction		Feasible	Not feasible
	Feasible	2	2
	Not feasible	3	26

The confusion matrix results in Table 4 are used to calculate the accuracy, precision and recall values using equation 1 so that the following results are obtained:

Table 5. Accuracy Test scenario 1

Accuracy	Precision	Recall
84.85%	50%	40%

Tests using basic variables obtained good accuracy results with a value of 84.85% and an Area Under Curve (AUC) value of 0.693.

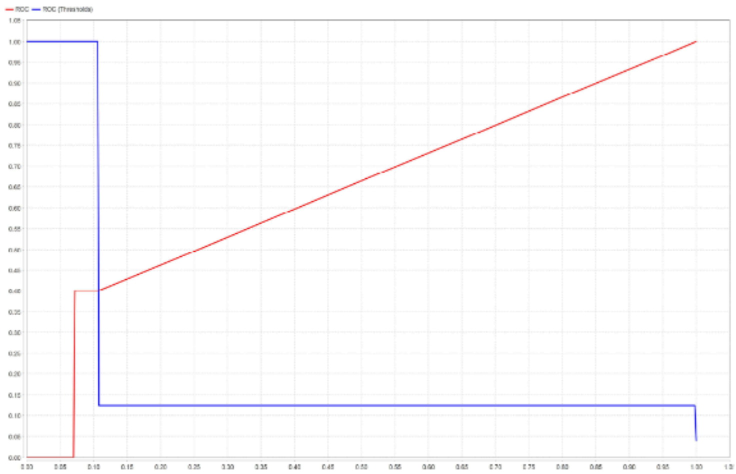

Figure 5. AUC Ccenario 1

In Figure 5, the vertical line is the true positive line and the horizontal line is the false positive line. Red lines as ROC and blue color ROC (Thresholds).

Scenario 2

In this scenario, variables were added to the basic variables. The variables added were 8 variables so that the total variables at this stage amounted to 13 variables consisting of: business entity status, business entity form, business entity NPWP, business entity domicile, business entity qualifications, business field classification, work completion, IUJK, experts, work experience, quality of work results, winning tenders and contract value.

To get prospective providers categorized as eligible or not requires underlying rules. The eligibility assessment rules used in scenario 2 are as follows:

Table 6. Eligibility Assessment Rules scenario 2

No.	Variable	Critera
1.	Business entity status	Verified
2.	Form of business entity	PT/ CV
3.	NPWP of Business entity	Existing/ Owned
4.	Business entity domicile	Head office/ branch office
5.	Business entity qualification	As per
6.	Business Licence Qualification	Compliant
7.	Expertise	Have at least 3 people
8.	Work Experience	Have worked on at least 4 jobs
9.	Quality of Work	More than 80%
10.	Winning Tender	Have won the tender 3 times

11.	Contract value	More than Rp. 200.000.000,-
12.	Work completion	More than 90%
13.	Construction service business licence (IUJK)	have

The application of eligibility rules in scenario 2 is said to be feasible if all criteria are met and not feasible if one or more criteria are not met. Based on the rules of table 6, the following results are obtained:

Table 7. Feasibility results of Scenario 2

No.	Description	Feasible	Not feasible
1.	Prospective construction service providers (number of business entities)	29	124

To produce decision tree modeling by calculating the entropy and gain values for each variable using equations 1 to 4. From the calculation, the resulting gain and entropy values are as follows:

Table 8. Entropy and Gain values of scenario 2

No.	Variable	Feasible entropy	Not Feasible entropy	Gain
1.	Expertise	0.971	0.306	**0.179**
2.	Work expererience	0.644	0.242	0.016
3.	Quality of work	0.606	0	0.067
4.	Winning tender	0.801	0.242	0.067
5.	Contract value	0.701	0.314	0.087
6.	Business entity status	0	0	0
7.	Form of business entity	0.391	0.519	0.069
8.	NPWP of Business entity	0.594	0	0
9.	Business entity domicile	0.594	0	0
10.	Business entity qualification	0.594	0	0
11.	Business Licence Qualification	0.594	0	0
12.	Work completion	0.601	0	0.004

Based on table 8, the decission tree in scenario 2 is obtained as follows:

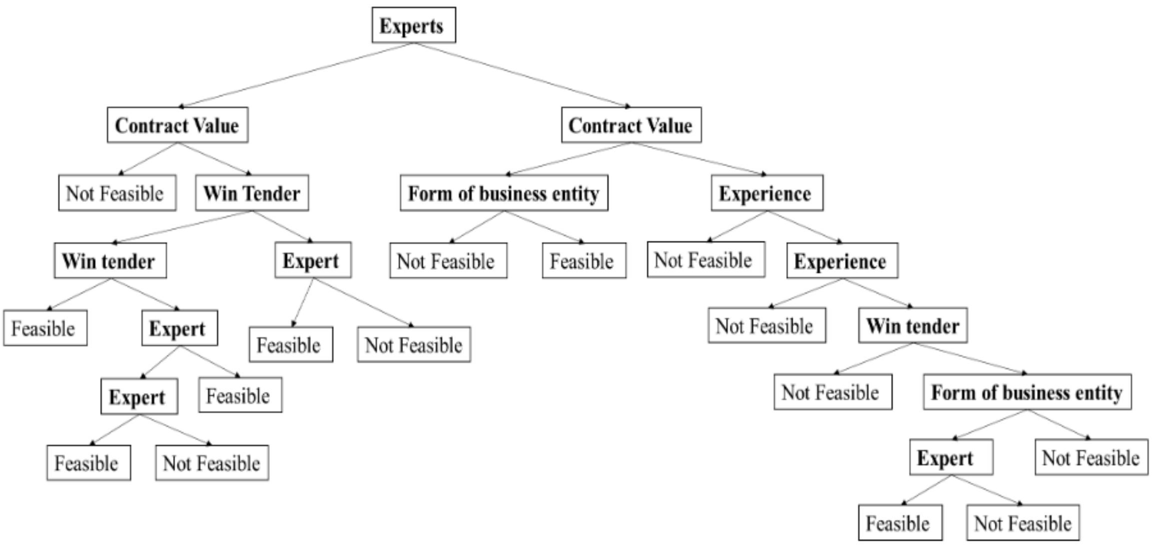

Figure 6. Scenario 2 decision tree

In Figure 6 the expert variable is used as the root because it has the highest gain value. The results of the decision tree are not pruned on each branch. The results of the experiment in scenario 2 produced the following confusion matrix:

Table 9. Confusion Matrix results scenario 2

		Actual	
		Feasible	Not feasible
Prediction	Feasible	4	0
	Not feasible	3	26

The results of the confusion matrix table 9 are used to calculate the accuracy, precision and recall values using equation 1 so that the following results are obtained:

Table 10. Accuracy test scenario 2

Accuracy	Precision	Recall
90.91%	100%	57.41%

Testing using scenario 2 obtained better accuracy results with a value of 90.91% and an Area Under Curve (AUC) value of 0.588. In Figure 7 the vertical line as a line with a true positive form and the horizontal line as a false positive line. Red lines as ROC and blue color ROC (Thresholds).

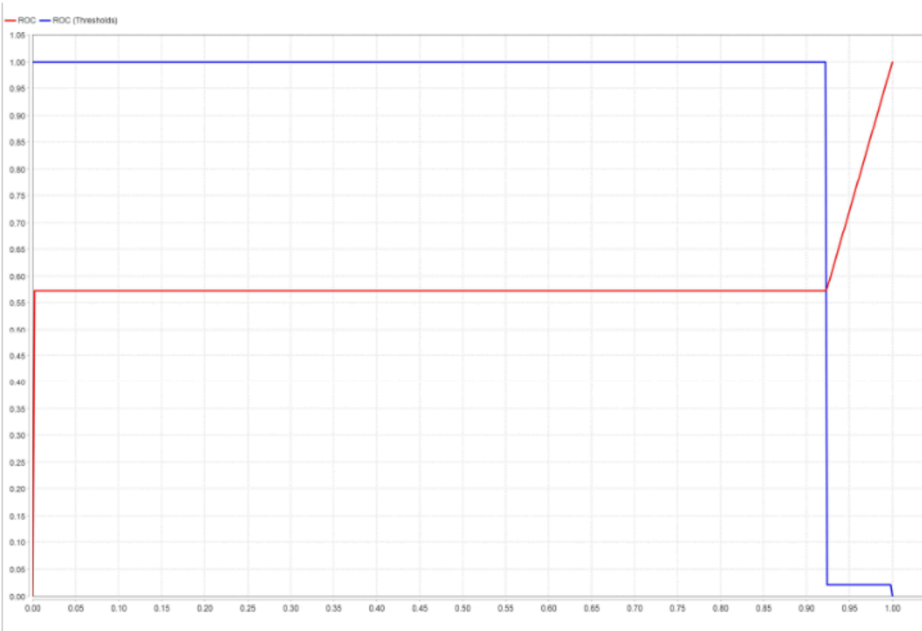

Figure 7. AUC Scenario 2

4. Conclusion

In selecting prospective construction service providers in government procurement, the decission tree method can be used as a recommendation in decision making for procurement officials. Based on testing conducted using 2 scenarios, the accuracy results have increased by 6.06%. The accuracy results in scenario 1 using the basic variables amounted to 84.85% and the accuracy results in scenario 2 by adding additional variables to the basic variables amounted to 90.91%. This shows that the addition of variables has an effect.

The results of this study can be a recommendation for Goods / Services Procurement Officials and Commitment Making Officials in the Regional Work Unit to select prospective construction service providers using the direct selection method using administrative requirements and considering additional variables.

REFERENCES

[1] JDIH LKPP, "Jaringan Dokumentasi dan Informasi Hukum Lembaga Kebijakan Pengadaan Barang/Jasa Pemerintah," *Pemerintah Republik Indonesia*, 2021. https://jdih.lkpp.go.id/regulation/peraturan-lkpp/peraturan-lkpp-nomor-12-tahun-2021

[2] L. K. P. B. J. Pemerintah, "Peraturan Pemerintah Tentang Lembaga Kebijakan Pengadaan Barang/ Jasa

Pemerintah," 2021.
[3] muhammad isa Ansori, R. Kusumawati, and m. amin Hariyadi, "View of Prediction of Service Level Agreement Time of Delivery of Goods and Documents at PT Pos Indonesia Using the Random Forest Method," *Int. J. Adv. Data Inf. Syst.*, vol. 4, no. 1, pp. 41–50, 2023, doi: https://doi.org/10.25008/ijadis.v4i2.1281.
[4] Maciej Serda *et al.*, "Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza," *Uniw. śląski*, vol. 7, no. 1, pp. 343–354, 2013, doi: 10.2/JQUERY.MIN.JS.
[5] J. Fadlil and wayan firdaus Mahmudy, "Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering," *kursor*, vol. 3, no. 1, pp. 45–66, 2007, Accessed: Feb. 08, 2024. [Online]. Available: www.movielens.org.
[6] R. Darmawan and S. Amini, "Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor pada Twitter," *SENAFTI Semin. Nas. Mhs. Fak. Teknol. Inf.*, no. September, pp. 495–501, 2022.
[7] S. Palaniappan, A. Mustapha, C. Feresa, M. Foozy, and R. Atan, "Customer Profiling using Classification Approach for Bank Telemarketing," *Int. J. Informatics Vis.*, vol. 1, no. 4, pp. 214–217, 2017.
[8] T. Hall, S. Beecham, D. Bowes, D. Gray, and S. Counsell, "A Systematic Literature Review on Fault Prediction Performance in Software Engineering," *IEEE Trans. Softw. Enginering*, vol. 38, no. 6, pp. 1276–1304, 2012.
[9] A. Florio, "2021 Last-Mile Routing Research Challenge Article XII A Machine Learning Framework for Last-Mile Delivery Optimization," no. September, 2021.
[10] J. W. Quah, C. H. Ang, R. Divakar, R. Idrus, N. L. Abdullah, and X. Y. Chew, *Timing-of-delivery prediction model to visualize delivery trends for pos laju malaysia by machine learning techniques*, vol. 937. Springer Singapore, 2019. doi: 10.1007/978-981-13-3441-2_7.
[11] R. Kitchin, *Cognitive Maps*. 2015. doi: 10.1016/B978-0-08-097086-8.72008-3.
[12] I. Report, *PREPARING FOR LIFE IN A DIGITAL WORLD Preparing for Life in a Digital World*. 2018.
[13] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mix Method Approaches*, 4th ed. United States of America: SAGE Publication Asia-Pasific Pte. Ltd., 2014.
[14] N. K. Sari, M. T. Budiarto, and R. Ekawati, "Level Kompetensi Digital Mahasiswa Pada Bidang Pemecahan Masalah," *SJME (Supremum J. Math. Educ.*, vol. 7, no. 1, pp. 45–54, 2023, doi: 10.35706/sjme.v7i1.6629.
[15] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Trans Pattern Anal Mach Intell*, vol. 7, no. 7, pp. 881–892, Jul. 2002, doi: 10.1109/tpami.2002.1017616.
[16] M. Yamamoto and H. Hwang, "A general formulation of cluster analysis with dimension reduction and subspace separation," *Behaviormetrika*, vol. 41, no. 1, pp. 115–129, Jan. 2014, doi: 10.2333/bhmk.41.115.
[17] S. Affeldt, L. Labiod, and M. Nadif, "Spectral clustering via ensemble deep autoencoder learning (SC-EDAE)," *Pattern Recognit.*, vol. 108, Dec. 2020, doi: 10.1016/j.patcog.2020.107522.
[18] M. Alkhayrat, M. Aljnidi, and K. Aljoumaa, "A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA," *J. Big Data*, vol. 7, no. 1, pp. 1–23, Dec. 2020, doi: 10.1186/S40537-020-0286-0/FIGURES/10.
[19] N. K. Sari, M. T. Budiarto, and K. M. Rozhana, "Students' Digital Thinking Skills in Solving Mathematics Problems," *Proc. 1st Int. Conf. Math. Math. Educ. (ICMMEd 2020)*, vol. 550, 2021, doi: 10.2991/assehr.k.210508.076.
[20] I. M. Al-Zuabi, A. Jafar, and K. Aljoumaa, "Predicting customer's gender and age depending on mobile phone data," *J Big Data*, vol. 6, no. 1, p. 18, Dec. 2019, doi: 10.1186/s40537-019-0180-9.
[21] M. A. Hariyadi, M. H. Purnomo, M. Hariadi, and K. E. I. Purnama, "Lung segmentation of x-ray thorax image using geometric active contour and analysis of histogram equalization filtering," *Int. J. Appl. Math. Stat.*, vol. 43, no. 13, pp. 327–335, 2013, Accessed: Feb. 20, 2024. [Online]. Available: https://scholar.its.ac.id/en/publications/lung-segmentation-of-x-ray-thorax-image-using-geometric-active-co
[22] J. Buder and F. W. Hesse, "Informational environments: Effects of use, effective designs," *Informational Environ. Eff. use, Eff. Des.*, pp. 1–298, 2017, doi: 10.1007/978-3-319-64274-1.
[23] J. Antonio, G. Gómez, J. Antonio, and G. Gómez, "Análisis de los procesos de búsqueda , acceso y selección de información digital en futuros maestros An Analysis on Information Digital Searching , Access and Discrimination Processes in Future School Teachers," no. 34, pp. 76–90.
[24] H. Chu, W. Zhang, P. Bai, and Y. Chen, "Data-driven optimization for last-mile delivery," *Complex Intell. Syst.*, no. 0123456789, 2021, doi: 10.1007/s40747-021-00293-1.

[25] O. H. Romanovskyi, V. M. Grineva, O. A. Zhernovnykova, L. A. Shtefan, and V. V. Fazan, "ФОРМУВАННЯ ЦИФРОВОЇ КОМПЕТЕНТНОСТІ МАЙБУТНІХ УЧИТЕЛІВ МАТЕМАТИКИ: КОНСТАТУВАЛЬНИЙ ЕТАП," *Inf. Technol. Learn. Tools*, vol. 65, no. 3, 2018, doi: 10.33407/itlt.v65i3.2412.