HASIL CEK5_60010383

by 60010383 Anton Yudhana

Submission date: 14-Jun-2022 11:16AM (UTC+0700)
Submission ID: 1856483535
File name: CEK 5_60010383.pdf (239.99K)
Word count: 2713
Character count: 12984

JNROnline Journal
ISSN: 2320-3358 (e)
ISSN: 0972-5547(p)

Journal of Natural Remedies
Vol. 21, No. 9(1), (2021)

DATA MINING IMPLEMENTATION USING K-NEAREST NEIGHBOR ALGORITHMTO PREDICTSENGON SAWING RESULT

Anton Yudhana[1], Sunardi[2], AgusJaka Sri Hartanta[3]
[1, 2, 3] Ahmad Dahlan University , Indonesia

ABSTRACT

The sawing result prediction that can give overview how much Sengon sawing can yield is done by manual calculation that require a long and difficult process, so there need a method to resolve this problem. This study aim is to find whether K-Nearest Neighbor is acceptable for Sengon sawing result prediction. The Research steps is collecting the data, algorithm calculation and test performance. Manual calculation method which is done by preparation of the data training (old data) with 135 data, enter the value of the new data (testing) with 10 data, checking the distance Euclidean, look for K value, determines the outcome then test performance. Test performance done by comparing result of K-NN prediction with real data and application implement. Research Result show the manual calculation prediction and application implement got an accuracy 0.7 or 7 0%, precision 1 and recall 0.7 so K-NN algorithm is good for Sengon sawing result prediction.

Keywords: K-NN, Prediction, Sengon, Euclidean

1. INTRODUCTION

Sengon is one of tree species that easy to find in tropical regions. People use Sengon in orderfulfill the needs of material for raw household material, furniture and carpentry (Ferry et al., 2018). Wood Sengon also can be used for triplex materials, plywood, and board particles and the board blocks (Krisnawati et al., 2011), before used as material Sengon wood first must through sawing process. Sengonwood sawing process either traditional or industry has the same standard.Wood sawing variations depend on the order buyers, large or small diameterSengon, defective parts in the trunk and the wood skewed level. Sengon sawing prediction can help give a picture of the final results of the pieces end that can be obtained from material raw Sengon. Thecutting process is unique where particular cut only can give certain sawing result.

First process at sawing process is wood procurement by boughtlarge number of Sengon tree then take it to sawmill and prediction is important to give result overview number of wood that can be obtained from so can reduce potential losses from cost to bought the sengon wood. The prediction process on Sengon wood is done by drawing the wood stem then doing manual calculation to determine the number of saw results that can obtained. But this manual predictions is difficult and take long time to do this manual prediction and worst thing is this process calculation must be done one by one perwood.

One of methods that can be used to speed up theSengon sawingresults prediction processis by applying prediction techniques that calculate with computer. Research that relevant one of them is design applications android smartphone to determine the pattern of sawmill Sengonone side (Yudhana et al., 2018)but the research got no algorithm yet.There also research that use Genetic Algorithm for jati wood sawing result optimalization that got accuracy of 95.83%(Wijaya & Setiyanto, 2013) but this algorithm cannot used as different kind of wood, so need to use another algorithm. To solve this problem is by use data mining algorithm. One of data miningalgorithm that can be applied is the algorithm k-nearest Neighbor (K-NN).

There also other research that use K-NN algorithm with almost identical step but different object, the research are : Predictive application of goods sales withK-NN algorithm with from the results of the application has accuracy of 94% (Hasmawati et al., 2017), Then prediction of rice yields in DI Yogyakarta province with the K-NN algorithm (Willmen TB Panjaitan, Ema Utami, 2018).K-NN algorithm for determinate study specialization (Lizarti & Ulfah, 2019). Prediksi kelulusan tepat waktu

mahasiswabarudengan k-nn(Mustafa & Simpen, 2015). Goods prediction that come out from warehouse with timeseries K-NN (Hamdi et al., 2019)
The research that use K-NN methods to predict / forecast the sawmills results is still small at number so this research goal is to find out whether the K-NN algorithm able to the predict result of Sengon sawingthen perform the test performance levels of accurate and precise.

2. PROPOSED METHODS

2.1 Data

Sengon data sawing which comes from data gathering fromwood sawmill center UD Slamet has a uniqueness in it because of the standard for the size results of cutting with major parametersinwood Sengon sawing results prediction that got diameters asprimaryattribute. The diameter 10 cm to 30 cm and length of130 cm, 260 cm and 540 cm.The sawmill worker always do rounding down the value as result there no decimal value on the data and the wood length is rounded and main weight for sawing is the wood diameter.
Data from field that use as dataset that got number is 226 data and fordatatraining on this research is135 data withdominantdata of diameter 13 - 19cm, 20 - 24cm and 25 - 29cm with length of 130 and 260. The following data is a sample of training data (old data) used for research. The training data sample is as follows on table 1.

Table 1. Training data sample table (old data)

No.	Diameter	Length	Result	Result Name
1	13	130	1	Plywood fill board
2	13	130	1	Plywood fill board
3	13	130	1	Plywood fill board
4	14	130	2	Plywood fill board
5	14	130	2	Plywood fill board
.....	.....	.....	.....	.....
132	29	260	12	2x18 board
133	29	260	13	2x18 board
134	29	260	13	2x18 board
135	30	260	13	2x18 board

2.2 K-Nearest Neighbor

K-NN is algorithm main objective isto find closest value between old data(training) and new data(trial) based on predetermined data. K-Nearest Neighbor (K-NN). KNN as a supervised algorithm predicts results by new Input data instances then classified them with old data. The K-NN algorithm classifies new objects based on basis of dataattribute and training samplebyfind nearby value range usingEuclideandistance (Lizarti & Ulfah, 2019). Euclidean Distance are used for calculating distance of classification (Saifudin & Fadlil, 2015)by search distance between twovariable points where the value is getting close and similar then distance between the two points also getting smaller(Riadi et al., 2020).
Euclidean Distance value is good if there new data has a close distance or has similarity values with training data. The results Euclidean Distancecalculation can found by calculating distance value between the training data and the new input data. Wood diameter and length from new data compared with the diameter and lengthtraining data, from thecalculation result determinate which neighbor data that closest value which is results of the sawing prediction results. Calculation of the distance is done by using the formula of Euclidean Distance: Euclidean distance between x_i and x_j If $x_i = (x_1,y_1)$ and $v = (x_2,y_2)$ aretwo points(Fadlil, 2010), then Euclidean distance are:

$$d(u,v) = \sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2} \quad (1)$$

The usage of K-NNalgorithm is done by making datasets as training data, determination of the value of k, Entering the value of data just for testing, calculating Euclidean distance value, Sorting Euclidean value from the small to the large, data is retrieved with value amount of k, then take value of the majority as a prediction result.

2.4 Algorithm Performance check

The next phase is to find Recall, precision and accuracy of the algorithm. Precision is a level of success between the requested information and the answers given. Recall is a level of success in retrieving information, accuracy is the level of closeness between the predicted value and the true value [(Sunardi et al., 2018). Precision and accuracy recall formula:

$$precision = \frac{TP}{TP + FP} \quad (2)$$

Where:
TP (True Positive) = positive tuple that in the prediction with the correct
TN (True Negative) = negative tuple who predicted the correct
TP (False Positive) = negative tuple who predicted the class of positive
FN (False Negative) = positive tuple that the negative class predicted

3. RESULT AND DISCUSSION

3.1 Manual Calculation

The sawmill data obtained is scanned into new data and old data then calculates the distance by entering new data values (testing) for testing calculated by the Euclidean formula using sample data (23,130) The amount of data used as training data (old) is 135 data then from data table 1 we use formula 1 to calculate Euclidean distance between diameter and length old data(training) and new data values (testing) the result of Euclidean distance shown on table 2.

Table 2. Euclidean Distance Calculation Result

x_1	y_1	x_2	y_2	ED
13	130	100	0	10
13	130	100	0	10
14	130	81	0	9
.....				
29	260	36	16900	130.1383879
29	260	36	16900	130.1383879
30	260	49	16900	130.1883251

The results of the Euclidean distance calculation from table 2.then sorted from the small to large, and smallest number got rank 1 while the biggest number got last rank. The sorting results from the Euclidean distance calculation are shown in the following table:

Table 3. Euclidean Distance Sorting Result

Rank	x	y	ED
1	23	130	0
2	23	130	0
3	23	130	0
4	22	130	1
5	22	130	1
...			
133	29	260	130.1383879
134	29	260	130.1383879
135	30	260	130.1383879

The result of Euclidean distance value sort thentaken based the amount of K value. Then from these result we find the majority value with K value to determine the prediction results. Here are the predictive results of diameter 23 and length 130 withvalues K=1,K=2,K=3,K=4 and K=5that show on table 4.

Table 4. Prediction Results Table

No	K	Result
1	K = 1	9 Plywood fill board
2	K = 2	9 Plywood fill board
3	K = 3	9 Plywood fill board
4	K = 4	9 Plywood fill board
5	K = 5	9 Plywood fill board

Table 4 indicate that there no difference from K=1 to K=5. Then from step above then calculate 10 new data(trial) that include K=1 to K= 5 result. The result of calculationshown below on table 5.

Table 5. Manual CalculationPrediction Results

No	Diameter	Length	K=1	K=2	K=3	K=4	K=5	Result
1	23	130	6	6	6	6	6	Plywood fill board
2	21	132	4	4	4	4	4	Plywood fill board
3	19	133	4	4	4	4	4	Plywood fill board
4	16	130	3	3	3	3	3	Plywood fill board
5	24	260	7	7	7	7	7	board 2x18
6	27	263	9	9	9	9	9	board 2x18
7	14	133	2	2	2	2	2	Plywood fill board
8	28	261	10	10	10	10	10	Plywood fill board
9	20	131	4	4	4	4	4	Plywood fill board
10	25	262	8	8	8	8	8	board 2x18

3.1 Apps Implementation Result

The next result is the application of the application using same trial data and training data from manual calculation with K= 1, K =2, K= 3, K= 4 and K =5.Theapplication prediction resultsare shown in table 6.

Table 6. Complete Prediction Results

No	K=1	K=2	K=3	K=4	K=5	Result
1	6	6	6	6	6	Plywood fill board
2	4	4	4	4	4	Plywood fill board
3	4	4	4	4	4	Plywood fill board
4	3	3	3	3	3	Plywood fill board
5	7	7	7	7	7	board 2x18
6	9	9	9	9	9	board 2x18
7	2	2	2	2	2	Plywood fill board
8	10	10	10	10	10	Plywood fill board
9	4	4	4	4	4	Plywood fill board
10	8	8	8	8	8	board 2x18

3. 2 Analysis Result

Calculation results comparison with manual calculations, calculations with implemented applications and comparisons with results from real field data show that there missed prediction is 3 points and the correct prediction is 7 points. The data detail shown below:

Table 7. Manual, Implementation and Real Prediction Results

No	Diameter	length	Real	Information	Manual	App
1	23	130	6	Plywood fill board	right	right

2	21	132	5	Plywood fill board	wrong	wrong
3	19	133	4	Plywood fill board	right	right
4	16	130	3	Plywood fill board	right	right
5	24	260	8	2x18 board	wrong	wrong
6	27	263	10	2x18 board	wrong	wrong
7	14	133	2	Plywood fill board	right	right
8	28	261	11	2x18 board	right	right
9	20	131	4	Plywood fill board	right	right
10	25	262	8	2x18 board	right	right

Results of the algorithm performance show that percentage of accuracy is 70% and error rate is30%. Precision value is 1 and Recall value is 0.7. The results of the analysis can be seen in the table below

Table 8. K-NN Algorithm Precision Recall Accuracy Table

K	Accuracy	Precision	Recall
K = 1	0.7	1	0.7
K = 2	0.7	1	0.7
K = 3	0.7	1	0.7
K = 4	0.7	1	0.7
K = 5	0.7	1	0.7

4. CONCLUSION

The conclusion from this study show that the prediction of Sengon sawing result with manual calculation and application implementation with K=1,K=2,K=3,K=4 and K=5 has an accuracy of 0.7 or 70%, precision value is 1 and recall value is 0.7so it concluded that the K-NN algorithm is good for the predicting Sengon sawing result.

5. REFERENCE

Fadlil, A. (2010). An Automatic Identification System of Human Skin Irritation. *Telkomnika*, *8*(3), 255–264.

Ferry, A., Putra, R., Wardenaar, E., & Husni, H. (2018). Analysis of the Chemical Components of Sengon Wood (Albizia Falcataria (L.) Fosberg) Based on Height of Trunk. *Jurnal Hutan Lestari*, *6*(1), 83–89.

Hamdi, A., Indriani, F., & Muliadi. (2019). Timeseries K-Nearest Neighbor Regression Method to Predicting Outgoing Goods on PT Putra Prenuer Banjarbaru Warehouse.*Seminar Nasional Ilmu Komputer (SOLITER)*, *2*, 37–45.

Hasmawati, Nangi, Jumadil, & Muchtar, M. (2017). Goods Sales Prediction Application Using the K-Nearest Neighbor (K-NN) Method (Case Study of Tumaka Mart). *SemanTIK*, *3*(2), 151–160. http://ojs.uho.ac.id/index.php/semantik/article/view/3658/2775

Krisnawati, H., Varis, E., Kallio, M., & Kanninen, M. (2011). Paraserianthes falcataria (L.) Nielsen: Ecology, silviculture and productivity. *Center for International Forestry Research (CIFOR)*, 1–16. https://doi.org/10.17528/cifor/003482

Lizarti, N., & Ulfah, A. N. (2019). K-Nearest Neighbor Algorithm Application to Determinate Study Specialties (Case Study: Informatics Engineering Study Program STMIK Amik Riau). *Fountain of Informatics Journal*, *4*(1), 1–7. https://doi.org/10.21111/fij.v4i1.2822

Mustafa, M. S., & Simpen, I. W. (2015). Graduation on Time Prediction Applications design for New Students Using Data Mining Techniques (Case Study: Academic Data of STMIK Dipanegara Students in Makassar). *Creative Information Technology Journal*, *1*(4), 270–281. https://doi.org/10.24076/citec.2014v1i4.27

Riadi, I., Fadlil, A., & Annisa, P. (2020). Handwritten Identification of Japanese Katakana Using the Euclidean Method. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *4*(1), 29. https://doi.org/10.30645/j-sakti.v4i1.184

Saifudin, S., & Fadlil, A. (2015). Wood Image Identification System Based on Texture Using Gray Level Coocurrence Matrix (Glcm) With Euclidean Distance Classification. *Sinergi*, *19*(3), 181. https://doi.org/10.22441/sinergi.2015.3.003

Sunardi, Fadlil, A., & Suprianto. (2018). Sentiment Analysis Using the Naïve Bayes Classifier Method on Student Questionnaires. *Saintekbu*, *10*(2), 1–9. https://doi.org/10.32764/saintekbu.v10i2.190

Wijaya, L. A., & Setiyanto, N. A. (2013). Calculation of Estimated Number of Teak Logs Using Genetic Algorithms in Ud. Wahyu Jaya.*Udinus Repository*, 1–9. http://eprints.dinus.ac.id/id/eprint/4807

Willmen TB Panjaitan, Ema Utami, H. A. F. (2018). rediction of Rice Harvest Using the K-Nearest Neighbor Algorithm. *Prosiding SNATIF Ke -5 Tahun*, 621–628.

Yudhana, A., Sunardi, & Hartanta, A. J. S. (2018). Designing an Android Smartphone Application for Determining One Sided Patterns of Sengon Sawmills. *Snst Ke-9*, *November*, 58–63.

HASIL CEK5_60010383

ORIGINALITY REPORT

7%	6%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Ho Chi Minh University of Technology and Education Student Paper	2%
2	www.eie.polyu.edu.hk Internet Source	2%
3	computerresearch.org Internet Source	1%
4	pure.uva.nl Internet Source	<1%
5	Dinar Ajeng Kristiyanti, Mochamad Wahyudi. "Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review", 2017 5th International Conference on Cyber and IT Service Management (CITSM), 2017 Publication	<1%
6	Soo, V.W.. "A cooperative multi-agent platform for invention based on patent	<1%

document analysis and ontology", Expert Systems With Applications, 200611
Publication

7	his.diva-portal.org Internet Source	<1%
8	pdfs.semanticscholar.org Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		