JACKA Journal of Advanced Computer Knowledge and Algorithms

Journal of Advanced Computer Knowledge and Algorithms
Vol. 2, No. 2, April, 2025, pp. 51-55
Journal homepage: http://ojs.unimal.ac.id/jacka

K Fold Cross Validation Analysis for Electricity Meter Classification at PLN Lhoksukon Using K-NN and SVM Methods

Zuboili[1*], Rozzi Kesuma Dinata[2], Irwanda Syahputra[3]

[1,2] Universitas Malikussaleh, Indonesia
[3] Universitas Samudra, Indonesia

[*]Corresponding Author Email: zuboili.200170089@mhs.unimal.ac.id

Received: 24 March 2025
Revised: 30 March 2025
Accepted: 31 March 2025
Available online: 1 April 2025

Keywords:
Electricity, Classification, K-NN, SVM, K-Fold Cross Validation

ABSTRACT

Electricity consumption continues to increase every year in line with the increase in national economic growth. Predicting current electricity demand is important to understand the overall electricity supplied to each region. The problem is that currently, the electricity supply in various areas of Lhoksukon has not matched the needs of the community. In addition, problems can arise if the power generated is less than the load power requirements, causing energy shortages in an area. To find out whether the electricity provided is appropriate or not, a classification using Supervised Learning method is used. After classification, we will use K-fold Cross Validation to measure how good the accuracy is between the methods. This study will use 200 electricity meter data consisting of 150 test data and 50 training data with a composition of 75%: 25%. The testing process where the data process that has been divided is then carried out in the testing process where the data process is obtained from manual calculations. So that in this study get results in the form of the K-NN method with 99.3% accuracy, 100% precision, 99.29% recall and the SVM method with 94.00% accuracy, 94.00% precision, 100% recall. And to find out how well the performance of the method is based on Supervised Learning method, it will be checked using K-Fold Cross Validation with the results of K-NN 99.53% and SVM 96.00%, with the conclusion that the K-Nearest Neighbor method has a better accuracy rate.

1. INTRODUCTION

PT PLN Persero is a state-owned company that has a field of work regarding electricity. At the end of the 19th century PT PLN was formed to be able to organize electricity problems in Indonesia. The development of society affects the level of electricity supply that must be met, so PT PLN makes service offices almost throughout Indonesia as well as impacting on the completeness of the workforce or the addition of large employees and is carried out annually.

Estimating current electricity demand is important to determine the total power available in each region. Electricity consumption increases from year to year along with national economic growth. The supply requirement to allocate electricity to consumers must be optimized based on the needs of the people and industries in the region. For each customer, the amount of electricity provided to new facilities and the amount of electricity consumed by existing customers will be affected.

The process to overcome these problems is to perform data mining techniques, which process data into a source of information usually using clustering methods.

With the problems mentioned above, it is important to conduct research in classifying electricity meter data in the lhoksukon area with the aim of providing appropriate information about the amount of power needed in the region with accurate calculations. using the method (K-NN) and (SVM) regulates and determines the classification of the electric power of each meter used by the local community, so that the use of electricity can be regulated and utilized evenly as well as to avoid unevenness in the distribution of electric power received by the community.

2. RESEARCH METHODS

The following is a scheme of the Electricity meter classification system in Lhoksukon shown in Figure 1.

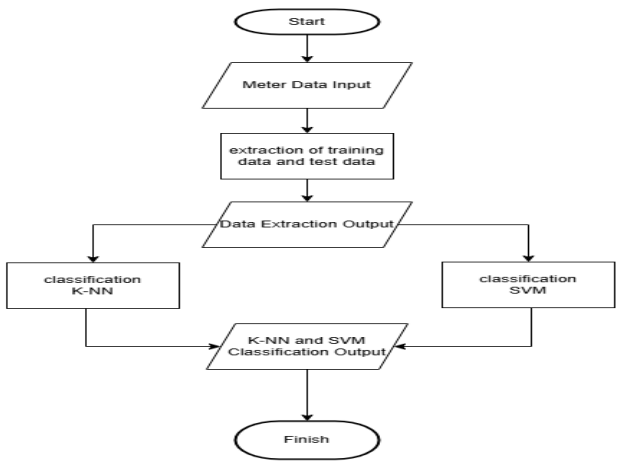

Figure 1. System Scheme

Based on Figure 1. The first step is start, is the initial initialization process of starting the application. The second step of the application will receive installed Electricity meter data as input. The third step of the data will ask to be extracted into training data and test data. The fourth step after the data is separated into training data and the next test data is to classify using the K-NN and SVM methods. The fifth step the output of the classification results appears. The sixth step is complete.

3. RESULT AND DISCUSSION

3.1 Reasearch dataset

The criteria used in this study are kwh type, power, voltage, current, group and description. The following is the sub-criteria value of each criterion which can be seen in Table 1.

Table 1. Lhoksukon electricity meter data

Criteria Code	Criteria
X1	jenis kwh
X2	daya
X3	arus
X4	golongan

The data used in this study amounted to 200 data using a data composition of 75% training data and 25% as test data which is divided sequentially, namely 150 data as training data and 50 data as test data. The data already has a description between subsidies and non-subsidies, can be seen in Table 2 and Table 3.

Table 2. Training data

No	Name	X1	X2	X3	X4	Information	Mark
1	Abdurrani	1	2	2	6	Ns	1
2	M kasim	1	2	2	2	S	0
3	Wahab abd	1	3	3	8	S	0
4	Saiful bahri	1	1	1	2	S	0
5	T marsudin	1	4	4	8	S	0
6	Hasballah	1	2	2	6	Ns	1
7	Mansur husen	1	4	4	8	S	0
8	H ibrahim malek	1	3	3	8	S	0
9	Widodo	1	2	2	8	S	0
10	Widodo	1	1	1	8	S	0
….	…..	…	…	…	…	…..	…..
150	H sulaiman	2	2	2	5	Ns	1

Table 3. Test data

No	Name	X1	X2	X3	X4	Information	Mark
151	Wahab	1	3	3	8	S	0
152	Jumadil aksah	1	2	2	2	Ns	1

Table 3. Test data

No	Name	X1	X2	X3	X4	Information	Mark
153	H ibrahim malek	1	3	3	8	S	0
154	T marsudin	1	4	4	8	S	0
155	T mahdi syafri	1	3	3	8	S	0
156	M hasan	1	4	4	8	S	0
157	ansari	1	1	1	8	S	0
158	Nazla khairina	1	2	2	6	Ns	1
159	A Rahman abdullah	1	1	1	8	S	0
160	Hasbi mawar	1	3	3	8	S	0
….	…..	…	…	…	…	…..	…..
200	widodo	1	2	2	8	S	0

3.2 K-nn calculation results

The calculation results that have been classified using the K-NN method are as follows:

3.2.1 Euclidean distance calculation results

Table 4. Euclidean distance calculation results

	Calculation results		
No	K=1	K=2	K=3
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
…	…..	…..	…..
50	0	0	0

3.2.2 Finding the average value of k

Table 5. Finding the average value of k-nn

Distance	Accuracy	Precision	Recall
K=1	100	100	100
K=2	98	100	97,87
K=3	100	100	100
Results	99,3	100	99,29

The results obtained from the classification of meter data using the K-Nearest neighbor method get the results of Accuracy 99.3%, Precision 100%, and Recall 99.29%.

3.3 Svm calculation results

The calculation results that have been classified using the SVM method are as follows:

3.3.1 Svm testing results

Table 6. Svm testing results

No	Class being tested			Result class	
	Name	Information	Mark	Information	Mark
1	Wahab	S	0	S	0
2	Jumadil aksah	Ns	1	S	0
3	H ibrahim malek	S	0	S	0
4	T marsudin	S	0	S	0
5	T mahdi syafri	S	0	S	0
6	M hasan	S	0	S	0
7	ansari	S	0	S	0
8	Nazla khairina	Ns	1	S	0
9	A Rahman abdullah	S	0	S	0
10	Hasbi mawar	S	0	S	0
…	…..	…..	…	…..	…
50	widodo	S	0	S	0

3.3.2 Knowing the accuracy value of svm

Table 7. Knowing the accuracy value of svm

Accuracy	Precision	Recall
94	94	100

The results obtained from the classification of meter data using the Support Vector Machine method get the results of 94% Accuracy, 94% Precision, and 100% Recall.

3.4 K-fold cross validation

The calculation of K Fold Cross Validation to analyze using the K-NN and SVM methods with each k value entered. In this study, the value of k used is the value of the results of the confusion matrix of each method whose calculation results can be seen in tables 5 and 7.

Table 8. K-fold cross validation testing

Testing	Accuracy	Precision	Recall	Average
K-nn	99,3	100	99,29	99,53
Svm	94,00	94,00	100	96,00

After testing using K-fold Cross Validation for the classification of KNN and SVM methods using electricity meter data. Then it shows the results using the KNN method get a better average value compared to using the SVM method. With the average results obtained are KNN = 99.53% and SVM = 96.00%.

3.5 Diagram appearance

The following is a diagram of the results of the analysis and classification of electricity meters:

3.5.1 Method results in graphical form

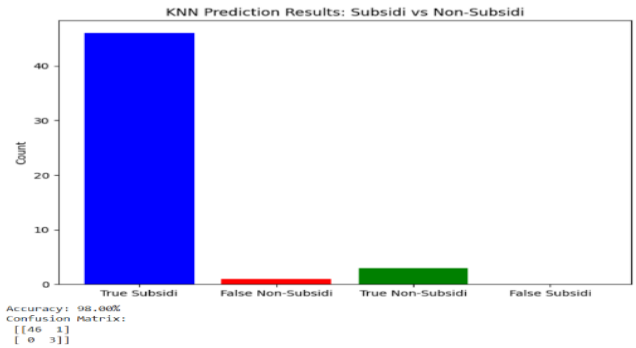

Figure 2. K-nn classification result chart

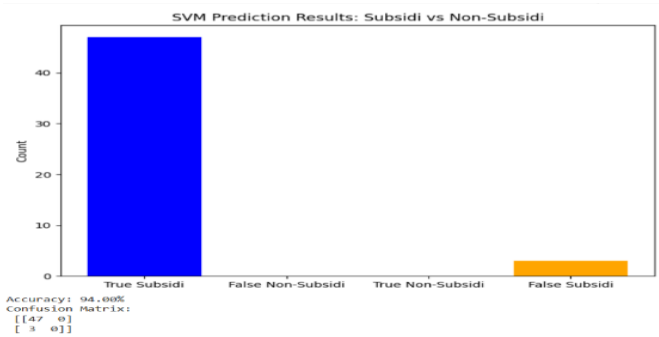

Figure 3. Svm classification result chart

3.5.2 Performance results in graphical form

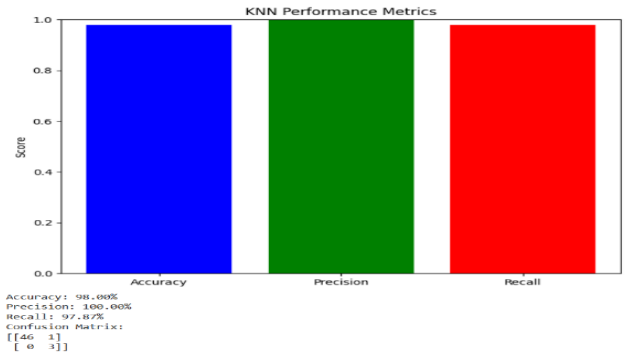

Figure 4. Graph of k-nn performance results

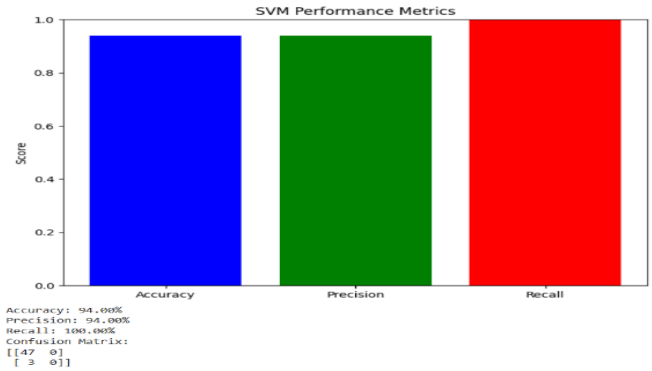

Figure 5. Graph of svm performance results

3.5.2 K fold cross validation result graph

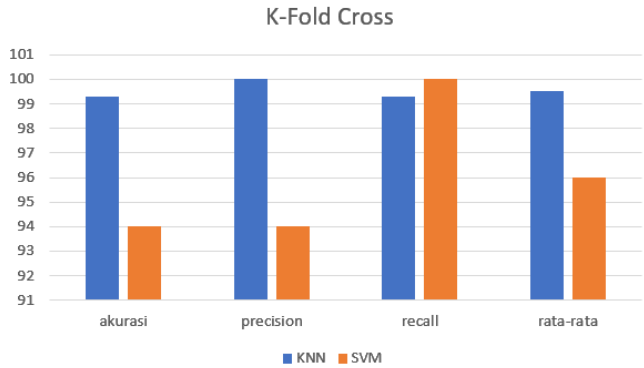

Figure 6. K fold cross validation result graph

4. CONCLUSIONS

This study wants to measure the analysis of electricity meters with K-Fold Cross Validation on electricity meter data from both methods that are better between the classification methods between the K-NN and SVM methods. Based on the test results on K Fold Cross Validation analysis using K-Nearest Neighbor (K-NN) Classification and Support Vector Machine (SVM) Classification with very high K-Fold Cross Validation optimization is using K-NN Classification with a value of 99.53% and is quite good using K-NN Classification, while using SVM Classification with a value of 96%.

REFERENCES

[1]. Abdul Kadir., (2008). Tuntunan Praktis Belajar Database Menggunakan MySQL, C.V. Andi Ofiset. Yogyakarta.

[2]. Arifin, Z. W. J., (2019). Penerapan Metode KNN (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerima Kip (Kartu Indonesia Pintar) Di Desa Pandean Berbasis Web Dan Mysql. NJCA-volume 4, 27-34.

[3]. Azizah, N., Firdaus, M.R., Resti, S., Fajar, I., (2023), *Penerapan Algoritma Klasifikasi K-Nearest Neighborpada Penyakit Diabetes*, Seminar Nasional Statistika Aktuaria II, Universitas Padjajaran, 2988-1900, https://prosidingsnsa.statistics.unpad.ac.id/?journal=prosidingsnsa&page=article&op=view&path%5B%5D=344&path%5B%5D=315

[4]. Bhavsar, H., & Panchal, M. H. (2012). A Review on Support Vector Machine for Data Classification. *Internasional Journal of Advanced Research in Computer Engineering & Technology(IJARCET).*

[5]. Daftar Wilayah dan Kode Pos Lhoksukon 2024, https://www.google.com/amp/s/kodepos.cektarif.com/amp/indonesia/desa/lhoksukon-011110, Diakses pada 2 Januari 2024.

[6]. Darma, S., Yusmartato, & Akhiruddin. (2019). STUDI SISTEM PENERAAN KWH METER. In Journal of Electrical Technology (Vol.4, Issue 3).

[7]. Dinata, R. K., Akbar, H., Hasdyna, N., (2020). Algoritma K-Nearest Neighbor Dengan Euclidean Distance Dan Manhattan Distance Untuk Klasifikasi Transportasi Bus. ILKOM Jurnal Ilmiah.106.

[8]. Eliyen, K., Tolle, H., and Muslim, M. A., (2017), *"KNearest Neighbor Untuk Klasifikasi Penilaian Pada Virtual Patient Case,"* J. Arus Elektro Indones., vol. 3, no. 1, pp. 15–18.

[9]. Fauzan, M., (2023), *Penerapan Seleksi Fitur Untuk Klasifikasi Penerima Bantuan Sosial Pangkalan Sesai Menggunakan Metode Knearest Neighbor*, Skripsi, UIN SUSKA RIAU, https://repository.uin-suska.ac.id/76139/1/repo_ojan.pdf.

[10]. Han, J., dan Kamber, M., (2006). Data Mining Concepts and Techniques Second Edition. Morgan Kaufmann Publishers, San Francisco.

[11]. Han, J., Kamber, M., Pei, J., (2011). Data Mining : Technique and Concepts 3rd ed. San Fransisco, CA, USA : Morgan Kaufmann.

[12]. Irham, L. G., Adiwijaya, A., & Wisesty, U. N. (2019). Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, *3*(4), 284.

[13]. https://doi.org/10.30865/mib.v3i4.1410.

[14]. K. A., (2017). *Support Vector Machines Succinctly.*

[15]. Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., and Haydock, M., (2012). *Analytics: The Widening Divide*, MIT Sloan Management Review, 53(2).

[16]. Larose, D. T., (2005). Discovering Knowledge in Data: An Introduction to Data Mining. Wiley, Chichester.

[17]. Larose, D. T., (2005). Discovering Knowledge in Data. New Jersey: John Willey & Sons Inc.

[18]. McLeod, Jr. R., dan Schell, G. P., (2007). Management Information System. 10th ed. Pearson Education, Inc.

[19]. Muktafin, E. H., & Luthfi, E. T., (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. 32-42. https://doi.org/10.30864/eksplora.v10il.390

[20]. Rahmiati, Irfan, D., Agustin, Hediyati, S., (2020), *Aplikasi Pengukur Tingkat Sentimen Pelanggan Berdasarkan Komplain Pelanggan Pln Menggunakan Algoritma K-Nearest Neighbor*, Jurnal Inovtek Polbeng: Seri Informatika, Vol. 5., No. 2.

[21]. Refaeilzadeh, P., Tang, L., dan Liu, H., (2009). "Cross-Validation," in *Encyclopedia of database systems*, Springer, pp. 532-538.

[22]. Sadli, M., Fajriana., Fuadi, W., Ermatita., Pahendra, I., (2018). Penerapan Model K-Nearest Neighbors Dalam Klasifikasi Kebutuhan Daya Listrik Untuk masing-masing daerah di Kota Lhokseumawe. Jurnal ECOTIPE (Vol. 5, No.2).

[23]. Sari, D. P., (2013). SISTEM PERHITUNGAN KWH METER LISTRIK PRABAYAR (LPB) UNTUK PELANGGAN BAYA 900 VA PT.PLN (PERSERO) AREA PALEMBANG. In *Jurnal Teliska ISSN* (Vol. 5, Issue 2). http://www.elib.unikom.ac.id.

[24]. Santosa, B., (2007) Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Graha Ilmu : Yogyakarta.

[25]. Simanjuntak, T. H., Mahmudy, W. F., dan Sutrisno., (2014). Implementasi Modified K-Nearest Neighbor dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai. online: http://jptiik.ub.ac.id/index.php/jptiik/article/download/15/21/

[26]. Simatupang, F. J., Wuryandari, T., & Suparti., (2016). Klasifikasi Rumah Layak Huni di Kabupaten Brebes dengan Menggunakan Metode Learning Vector Quantization dan Naive Bayes. *Gaussian*.

[27]. Suryanto., (2017). Data Mining Untuk Klasifikasi Data. Bandung : Informatika Bandung.

[28]. Tineges, R., Triayudi, A., Sholohati, I. D., (2020). Analisis sentimen terhadap layanan indihome berdasarkan twitter dengan metode klasifikasi support vector machine (SVM)., Vol. 4, No. 3, Page. 650-658.

[29]. Turban, E., (2005). *Decision Suport Systems And Intelligent Systems* Edisi Bahasa Indonesia Jilid 1, Andi, Yogyakarta.

[30]. Y. Al-Amrani, M. Lazaar., and K. El Kadiri., (2018). "Sentiment Analysis Using Hybrid Method Of Support Vector Machine And Decision Tree," *J Theor. adn Appl. Inf. Technol.*, Vol. 96, No. 7, pp. 1886-1895.