Jurnal Teknik Informatika (JUTIF)
Vol. 4, No. 3, June 2023, pp. 485-491

DOI: https://doi.org/10.52436/1.jutif.2023.4.3.725
p-ISSN: 2723-3863
e-ISSN: 2723-3871

PERFORMANCE OF K-MEANS CLUSTERING AND KNN CLASSIFIER IN FISH FEED SELLER DETERMINATION MODELS

Esmi Nur Fitri[*1], M. Hafidz Ariansyah[2], Sri Winarno[3], Fikri Budiman[4], Asih Rohmani[5], Junta Zeniarja[6], Edi Sugiarto[7]

[1,2,3,5]Information Systems Department, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
[4,5,6]Informatics Engineering Department, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
Email: [1]112201906175@mhs.dinus.ac.id, [2]1112201906146@mhs.dinus.ac.id, [3]sri.winarno@dsn.dinus.ac.id, [4]fikri.budiman@dsn.dinus.ac.id, [5]aseharsoyo@dsn.dinus.ac.id, [6]junta@dsn.dinus.ac.id, [7]edi.sugiarto@dsn.dinus.ac.id

(Article received: December 06, 2022; Revision: December 23, 2022; published: June 26, 2023)

Abstract

Feed is a crucial variable because it can determine the success of fish farming. Farmers can use two types of artificial feed, namely alternative feed and pellets. Many cultivators need pellets as the main food for the fish they are cultivating because the pellets contain a composition that has been adjusted to their needs based on the type and age of the fish. However, currently, cultivators are facing problem, namely the high price of fish pellets on the market. Therefore, an analysis of the classification of the selection of fish feed sellers is needed according to several criteria like the number of types of feed, price, order, delivery, payment, availability of discounts, and the number of assessments. This study conducted a predictive analysis to determine the criteria for selecting fish feed sellers in Kendal Regency by utilizing the K-Means Clustering and KNN Classifier methods in the classification method. This research aims to compare the fish feed seller classification method where the pattern of fish feed seller is identified by K-Means Clustering and KNN Classifier, and then the researcher conducts performance appraisal and evaluation. The results of this study are decision-making patterns to help formulate strategies for cultivators and other interested parties. For verifying the method used, measurements were made to obtain an accuracy value where K-Means was 98.6% and KNN was 86.7%.The results of this study indicate that the K-Means Clustering and KNN Classifier methods can classify the selection of freshwater fish feed sellers in Kendal Regency.

Keywords: *Fish Feed, K-Means Clustering, KNN Classifier.*

1. INTRODUCTION

Fish feed is one of the crucial variables in the process of fish farming, which can optimize fish development. The growth and survival of cultured fish are strongly influenced by the availability of feed. According to the Ministry of Maritime Affairs and Fisheries in a press release on 29 August 2022, the need for fish feed in Indonesia in 2024 is 22.65 million tons, which is calculated from the aquaculture production target [1]. Therefore, the government seeks the availability of fish feed of at least 12-13 million tons by 2024. With such a substantial need for fish feed, it proves that farmers must start having trusted fish feed sellers so they can always meet the demand for fish feed for the fish they cultivate.

Kendal Regency is one of 29 regencies and six cities in Central Java [2]. The topography of Kendal Regency consists of coastal areas (northern region), hills, and mountains (southern region) which causes the average temperature to range between 25-29ºC. With various forms of territory, Kendal Regency continues to strive to improve the fisheries sector through capture fisheries and aquaculture. According to data from the Central Statistics Agency on freshwater fish farming, in 2020 there were 228,608 cultivators [3] spread throughout Kendal Regency with various types of ponds ranging from permanent ponds, earthen ponds, and tarpaulin ponds. For freshwater fish, the majority of farmers in Kendal Regency cultivate catfish, tilapia, and pomfret with a production volume of 3,058,000, 74,800, and 18,500 respectively. This number is a substantial figure and requires collaboration between various parties to provide various facilities ranging from training, and irrigation, to the availability of medicines and feed.

In general, fish feed is divided into two varieties, namely natural feed and artificial feed. Organic feed is usually like plankton is already available in the pond, while artificial feed is still divided into two varieties, namely alternative foods and pellets. Alternative fish feed is made from the main feed staples such as tofu dregs, maggots, leaves, and others. Reciprocally with pellets, namely commercial feed with a composition that has been adjusted to the needs based on the type and age of the fish so that it can fulfill nutrients for the growth and survival of fish. For catfish, tilapia, and pomfret, most farmers

usually combine alternative feeds and pellets for the consumption of cultivated fish because these two feeds are affordable and easy to find. However, with this combination, the fish do not get enough protein and nutrition optimally, which creates a dilemma for farmers because getting maximum quality and profit requires optimal consistency of fish feed with high protein, which is found in pellets. This problem is quite disturbing for farmers because of the high price of the pellet which consumes 60-80% [4] of production costs, so farmers can only reduce consumption and replace it with alternative feeds.

Machine learning or machine learning is a method that can be used to predict future results. Machine learning is centered on improving and learning a system so that it can analyze pre-existing data which is part of artificial intelligence [5]. The use of two different methods makes the stages of data processing different, such as if K-Means Clustering performs the process of determining the centroid center to divide data into several clusters where each cluster can be occupied by several members together [6], [7] and KNN Classifier performs processing on training data and data testing to produce a classification of similarity patterns with the shortest distance. In this study, the use of K-Means Clustering and KNN Classifier was used as a method in classifying the selection of the best fish feed sellers in Kendal Regency. This study focuses on the classification criteria for determining the choice of fish feed sellers with the variables of the number of types of feed, price, order, delivery, payment, availability of discounts, and the number of ratings which are to assist freshwater fish cultivators in determining the best fish feed shop from various criteria that have been mentioned earlier.

The two studies above show that the algorithm used has a good classification result. However, no research has been found that is similar to implementing the K-Means Clustering and KNN Classifier methods in the classification in the fisheries sector, specifically determining fish feed sellers. Therefore, the goal to be achieved in this study is to analyze the criteria for selecting fish feed sellers using the K-Means Clustering and KNN Classifier methods to vary research using these methods. It is expected that the accuracy, recall, and precision values in the K-Means Clustering and KNN Classifier analysis produced are good so that they can become a reference for further research.

2. METHODS

This study utilizes data mining methods to find patterns from various data sets. The data sources used were obtained from e-marketplaces and direct interviews with fish feed farmers and sellers in Kendal Regency. The classification of the data used is then processed using the K-Means Clustering and KNN Classifier methods with 10-fold cross-validation and confusion matrix checking to determine performance as evidenced by the model accuracy value. The stages in this study consisted of 6 processes, namely data collection, data preprocessing, data transformation, modeling using the K-Means Clustering and KNN Classifier methods, and validation model as shown in Figure 1.

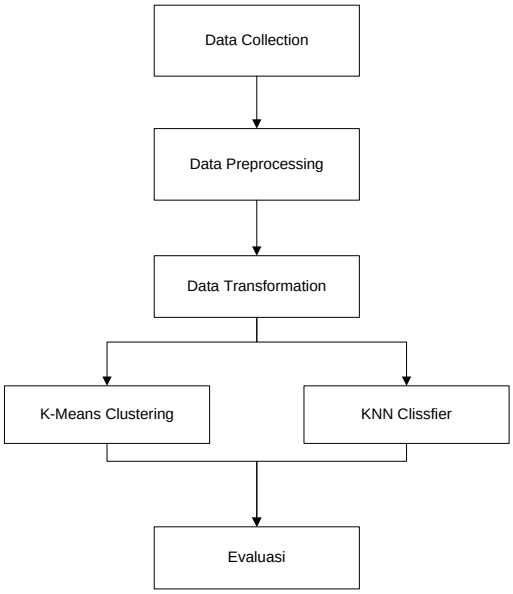

Figure 1. Research Stages.

2.1. Data Collection

In the data collection process, the variables used are the number of types of feed, price, order, delivery, payment, availability of discounts, and the number of ratings. The determination of the variable criteria used was referred to from research conducted by Dickson [8], where out of the 22 criterias presented, this study only adopted 7 variables as a reference. In addition, the criteria for selecting variables were surveyed based on direct interviews with farmers to suit their needs, although several other factors also support the decision-making of farmers in selecting fish feed sellers. The following is a description of the variables and datasets (Table 1) used in this study.

Table 1. Dataset

Features	Value	Value	Value	Value
Number of types of feed	3	1	.....	1
Price	Standard	Standard	.....	Under standard
Order	In store	In store	.....	Online
Shipping	Delivered and self collected	Self collected	.....	Delivered and self collected
Payment	Cash	Cash	.....	Cash
Discounts	Postage and discount	Discount	.....	Postage
The number of ratings	3600	1221	.....	2500
Recomm-endation	1	0	.....	0

1. Number of types of feed = number of types of feed sold in one shop
2. Price = price of each type of feed sold
3. Order = choice of ordering method offered
4. Shipping = choice of shipping method offered
5. Payment = the choice of a payment method offered
6. Discounts = availability of discounts that may be obtained
7. The number of ratings = the number of store ratings taken from the e-marketplace
8. Recommendation = recommended or not recommended.

2.2. Data Preprocessing

Data preprocessing is part of the most significant stages of data processing to prepare data into a form that is following the procedure while maintaining the performance of the data analysis being carried out [9]. Data preprocessing aims to reduce data size, find relationships between data, normalize data, remove outliers, and extract data features. Researchers carry out preprocessing to produce data that is ready to be processed by cleaning up missing data, namely by relieving values that cause information confusion [10]. Because the research dataset has a missing value of 14%, the treatment that must be given is to find the average value according to related attributes so that all data will be filled with valid values. In the next preprocessing stage, to avoid sample data that is abnormal from a data population, the researcher uses matplotlib to find out how many anomalies data are. In addition, to find out whether there is duplicate data, data is identified to detect the presence of more than one data so that it needs to be removed and will leave only one data store.

2.3. Data Transformation

Transformation is the stage of converting data into information, usually from one format to another. The format of the data source is changed to the desired information format. Data requirements are changed before the data mining analysis stage [11]. With data transformation, data analysis can be carried out more efficiently because it has a direction and the goal of unifying data.

Table 2. Dataset after transformation

Features	Value	Value	Value	Value
Price	0	0	.....	1
Order	2	2	.....	0
Shipping	1	1	.....	1
Payment	1	1	.....	1
Discounts	2	2	.....	0
The number of ratings	3600	1221	.....	2500
Recomm-endation	1	0	.....	0

This study carried out the transformation stage to change data from a categorical data type to a numeric data type, where this process was carried out with the help of one hot encoding found in the panda's library [12]. After the data has a numeric data type, then the machine learning model can process the data. Table 2 is the result of the transformation data.

2.4. Data Modelling

1) Classification

Classification [13] is called a supervision algorithm that uses training data (observations, measurements, and others) which is regulated by the existence of a label attribute. Classification seeks to classify the intended class by maximizing the gain of precision. The classification method tries to get similarities between input attributes and output attributes to produce a model that is included in the training stage [14]. Classification helps a group or categorize data based on the similarities and differences that have been analyzed before.

2) Clustering

Clustering is a method for finding and grouping data that has similar characteristics to other data. Clustering is included in unsupervised data mining, which means it does not require training or a teacher and does not require output targets [15].

3) K-Means

The K-means method is known for its speed and simplicity [16]. K-means clustering is a non-hierarchical data grouping method that groups data into one or several groups. Data with the same characteristics are grouped, and data with different characteristics are grouped, so that data in one cluster has a low level of variation [17]. The following are the clustering steps using the K-means method [15]:

1. Determine the number of clusters k
2. Start with the cluster center
3. Distribute all data/objects to the nearest cluster. The closeness of two objects is determined by their distance from each other. At this point, the distance between each data point and each cluster center must be calculated. Which data belongs to which cluster is determined by the maximum distance between a data and a particular cluster [18].

$$D(m,n) = \sqrt{(X_{1m} - X_{1n})^2 + (X_{2m} - X_{2n})^2 + \cdots + (X_{km} - X_{kn})^2} \quad (1)$$

Explanation:
$D(m,n)$ = Data distance to i to cluster center j
X_{km} = The i data on the k data attribute
X_{kn} = The jth center point on the kth attribute

With the current cluster membership, the cluster center distance is recalculated. The cluster center is

the average of all data/objects in a particular cluster. As a result, the average is not the only metric that can be used. Then the new cluster center is used to reassign each object. If the cluster center does not change, the clustering process will complete. Alternatively, return to step 3 until the cluster center no longer changes.

4) KNN

K-Nearest Neighbor (K-NN) is one of the algorithms used in classification problems. The working principle of K-NN is to find the shortest distance between data to be evaluated with the nearest neighbor in the training data. The K-Nearest Neighbor (K-NN) algorithm is one of the simplest algorithms for solving classification problems and often produces competitive and significant results. Calculate the distance using the Euclidean distance. The Euclidean distance formula is defined in Equation [19]:

$$d\_i = \sqrt{\sum_{i=1}^{p}(X_{2j} - X_{ij})^2} \quad (2)$$

Explanation:
X_{ij} = training data
X_{2j} = testing data
j = variable
d = distance
p = data dimension

5) Confusion Matrix

The confusion matrix is a performance measurement tool that can calculate accuracy in data mining concepts. This formula calculates four outputs, namely accuracy, precision, recall, and error rate. The ratio of correctly identified events to the total number of events is known as accuracy, the proportion of cases with positive results is called precision, the proportion of correctly identified positive cases is called recall, the error rate is the sum of all cases that were incorrectly identified [19]. Table 3 is a confusion matrix model.

Table 3. Model *Confusion Matrix*

	+	-
+	True positives (A)	False negatives (B)
-	False positives (C)	True negatives (D)

3. RESULT

The classification analysis process carried out in this study is divided into two, namely the process using the K-Means Clustering algorithm and the KNN Classifier. After the results are obtained, both models will be evaluated using the confusion matrix.

3.1. Classification on K-Means Clustering

This study aims to analyze to produce a classification in determining the selection of fish feed sellers for cultivators in the Kendal Regency area by utilizing the best algorithm between K-Means Clustering and KNN Classifier, more specifically, the achievement in this study is that cultivators get the lowest possible price of fish feed with superior service so that farmers can provide fish feed that is high in nutrition for the fish being cultivated and is expected to optimize the production results obtained.

The analysis of data processing is carried out by adopting data mining methods to classify datasets that have previously gone through preprocessing and data transformation. Clusters are determined into 2, namely Recommendations and Not Recommendations, where the value of K-2 is 0.355 as measured by Silhouette Score analysis. After obtaining the results of the K value, then the distribution of incoming data in clusters is visualized where the Recommended data are 27 data while the No recommendation data are 48 data as shown in Figure 2.

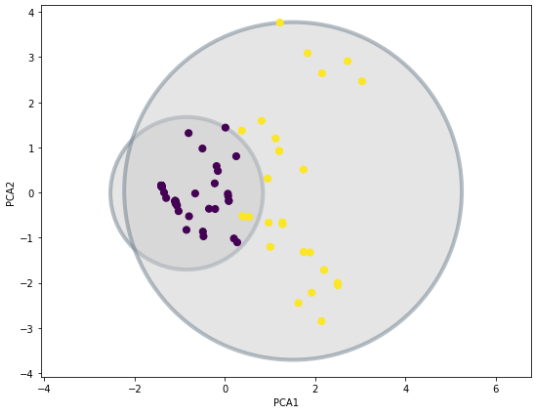

Figure 2. Fish feed seller cluster visualization

From the results of the visualization shown in Figure 2, labels are obtained for data that will be processed to the next stage, where the No Recommendation label is symbolized by 0 while the Recommendation label is concluded with 1, as shown in Table 4.

Table 4. Dataset after labeling

Features	**Value**	**Value**	**Value**	**Value**
Price	Standard	Standard	.....	Under standard
Order	In store	In store	.....	Online
Shipping	Delivered and self collected	Self collected	.....	Delivered and self collected
Payment	Cash	Cash	.....	Cash
Discounts	Postage and discount	Discount	.....	Postage
The number of ratings	3600	1221	.....	2500
Recomm-endation	1	0	.....	0
Label CLuster	1	0	.....	0

3.2. Classification on KNN Classifier

Then, the classification process uses the KNN Classifier algorithm by utilizing the clustering results obtained at the previous processing stage, namely K-Means Clustering. The distribution of labels in the data can be seen in Figure 3, where the No Recommendation label gets 63 data while the Recommendation label only gets 12. The data that is owned is less balanced so the distribution of the processed data is not fair enough.

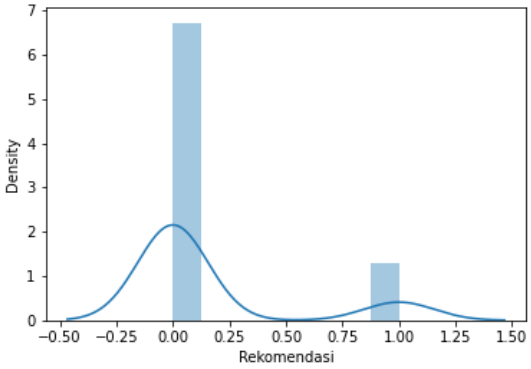

Figure 3. Data distribution

The classification modeling process utilizes the KNN algorithm using several parameters namely k_range, weight_options, and metric_options which are all taken from the sci-kit learn library with the best parameters being matrix: Minkowski, n_neighbors: 2, and weights: uniform as shown in Figure 4. Then, these parameters are applied to the entire data set to extract operational patterns and detailed statistics.

```
for param_name in sorted(param_grid.keys()):
    print('%s: %r' %(param_name,ori_clf.best_params_[param_name]))

metric: 'minkowski'
n_neighbors: 2
weights: 'uniform'
```

Figure 4. Best parameters.

3.3. Model Performance

Figures 5 and 6 show the performance results of the two models produced in the classification evaluation. Figure 5 uses the K-Means Clustering algorithm and Figure 6 uses the KNN Classifier algorithm. Both algorithms have been implemented with 10-fold cross-validation. Cross-validation is a method of data mining techniques that aims to obtain maximum accuracy results.

After getting the visualization of the confusion matrix measurements, to find out the results of accuracy, precision, and recall, you can use the equations in Table 3. The purpose of analyzing the value of the confusion matrix is to be able to know specifically the performance of the two models. The following is the result of the confusion matrix calculation for the K-Means Clustering algorithm and the KNN Classifier algorithm.

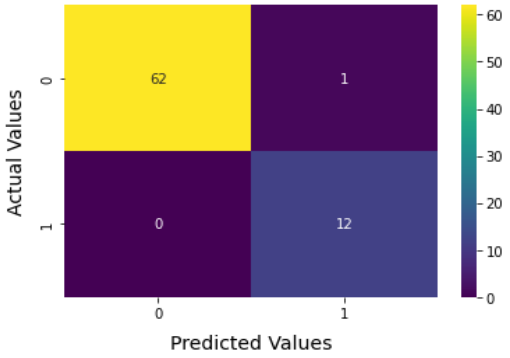

Figure 5. Confusion Matrix from the K-Means Clustering algorithm

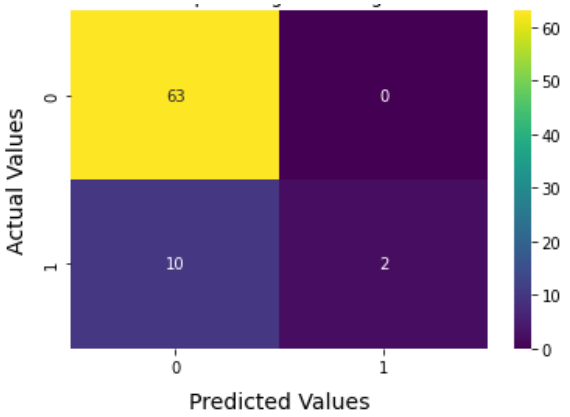

Figure 6. Confusion Matrix from the KNN Classifier algorithm

Table 5. Model Evaluation

	Akurasi	Presisi	Recall
K-Means Clustering	98,6%	92,3%	100%
KNN Classifier	86,7%	100%	16%

It can be seen in Table 5 that the overall performance of the K-Means Clustering algorithm is more stable than the performance of the KNN Classifier algorithm. It indicates that the K-Means algorithm is more optimal to be used to recommend fish feed sellers to farmers.

Because it has been proven that the K-Means Clustering algorithm is more optimal for use in classification, then to ensure the accuracy of the prediction value for fish feed seller recommendations or not, researchers tested it to find out the positive and negative predictive value. Testing is done to predict whether what is being done can work according to its function. The following is the calculation for the prediction test:

1. Positive predictive value:

$$PPV = \frac{TP}{(TP+FP)} = \frac{12}{(12+1)} = 0{,}923 \qquad (3)$$

If the prediction result is a recommendation, then a positive predictive value indicates the possibility that the fish feed seller has the characteristics indicated by that feature. The NPP of 92.3% indicated that if the prediction results in a recommendation, the probability that the fish seed seller recommend are 92.3%.

2. Negative predictive test value:

$$NPV = \frac{TP}{(TP+FN)} = \frac{12}{(12+10)} = 0{,}54 \quad (4)$$

If the prediction results are not a recommendation, then a negative predictive value indicates the possibility that fish feed sellers have similar characteristics. If the prediction result is not recommended, the chance that the fish feed seller is not recommended is 54%. Overall, the K-Means Clustering algorithm is better at identifying fish feed sellers which are expected to provide predictions of whether fish feed sellers fall into the recommendation category or not according to the needs of freshwater fish farmers in Kendal Regency. Researchers hope that with this prediction, cultivators can avoid unrecommended feed sellers that can be detrimental to the cultivators themselves.

4. DISCUSSION

In a study conducted by [20], early detection to improve the accuracy of cancer prediction results in improved quality of care and patient survival rates. This study used the K-Means Clustering and KNN Clustering methods which were then compared to determine which algorithm best fits the data. Based on these results, KNN clustering is a more efficient classification technique than cluster-based classifiers as long as the k variable values are chosen correctly, namely K=21. Furthermore, in [21] research, the performance of the KNN classifier and K-means clustering were tested in classifying the risk level of epilepsy patients from EEG signals. In applying the classification method, KNN classifier and K-means clustering are used. This study aims to design a classification algorithm with a high-performance index and a minimum rate of false alarms and missed classifications. The results show that the K-means clustering method is used in the classification to achieve the highest accuracy with a 0% false alarm rate. The two studies above show that the algorithm used has a good classification result. However, no research has been found that is similar to implementing the K-Means Clustering and KNN Classifier methods in the classification in the fisheries sector, specifically determining fish feed sellers.

5. CONCLUSION

The results of the analysis of the search for the best model between K-Means Clustering and KNN Classifier from freshwater fish feed seller data it results that the K-Means Clustering algorithm is proven to be better and optimal for predicting fish feed sellers compared to using the KNN Classifier algorithm, this is because the value the performance obtained by K-Means Clustering is more stable with an accuracy value of 98.6%, a precision of 92.3%, and a recall of 100%. The criteria used in determining the selection of freshwater fish sellers, namely the number of types of feed, price, order, delivery, payment, amount of discount and the number of ratings can be implemented by farmers to optimize to get the lowest possible price from the selected feed seller. Furthermore, this research needs to be improved on the use of data which is still too little so it requires more data that represents feed sellers throughout Kendal Regency so that it can become a reference for farmers in obtaining the right freshwater fish feed sellers.

REFERENCES

[1] Badan Riset Dan Sumber Daya Manusia Kelautan Dan Perikanan, “Kkp Gencarkan Pelatihan Guna Wujudkan Ketersediaan Pakan Ikan Mandiri,” 2022. https://kkp.go.id/brsdm/artikel/44444-kkp-gencarkan-pelatihan-guna-wujudkan-ketersediaan-pakan-ikan-mandiri (accessed Oct. 2, 2022).

[2] Badan Pusat Statistik, “Jumlah Kecamatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah 2019-2021,” 2022. https://jateng.bps.go.id/indicator/101/457/1/jumlah-kecamatan-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html (accessed Oct. 2, 2022).

[3] Statistik KKP, “Produksi Perikanan Per Kabupaten Kota,” 2022. https://statistik.kkp.go.id/home.php?m=prod_kabkota&i=2#panel-footer (accessed Oct. 2, 2022).

[4] Yunaidi, P. A. Rahmanta, A. Wibowo. "Aplikasi pakan pelet buatan untuk peningkatan produktivitas budidaya ikan air tawar di desa Jerukagung Srumbung Magelang," *Jurnal Pemberdayaan: Publikasi Hasil Pengabdian kepada Masyarakat*, vol. 3, no. 1, pp. 45-54, 2019, doi : https://doi.org/10.12928/jp.v3i1.621.

[5] A. Prasetyoi, Salahuddin, and Amirullah, "Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda," *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, vol. 6, no. 2, pp. 76-80, 2021, doi : http://dx.doi.org/10.30811/jim.v6i2.2343.

[6] Khairunnisa, J. Eska, M. F. L. Sibuea, "Application Of K-Means Clustering Method To Cluster Students’english Skill Jason English Course," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 3, pp. 479-485, 2022, doi : https://doi.org/10.20884/1.jutif.2022.3.3.167

[7] G. T. Enggiel, H. D. Purnomo, "Application Of K-Means Method In The Spread Of Positive Cases Of Covid-19 In Salatiga City," *Jurnal Teknik Informatika (Jutif)*, vol.

3, no. 5, pp. 1323-1328, 2022, doi : https://doi.org/10.20884/1.jutif.2022.3.5.356

[8] G. W. Dickson, *"An analysis of vendor selection systems and decisions," ," Journal of purchasing*, vol. 2, no.1, pp. 5-17, 1966, doi : https://doi.org/10.1111/j.1745-493X.1966.tb00818.x.

[9] I. M. A. Agastya, *"Pengaruh stemmer bahasa Indonesia terhadap peforma analisis sentimen terjemahan ulasan film,"* *Jurnal Tekno Kompak*, vol. 12, no. 1, pp. 18-23, 2018, doi : https://doi.org/10.33365/jtk.v12i1.70.

[10] P. S. Raja, K. Sasirekha, and K. Thangavel, *"A novel fuzzy rough clustering parameter-based missing value imputation," Neural Computing and Applications*, vol. 32, pp. 10033-10050, 2020, doi : https://doi.org/10.1007/s00521-019-04535-9.

[11] J. Muschelli, *"ROC and AUC with a binary predictor: a potentially misleading metric," Journal of classification*, vol. 37, no. 3, pp. 696-708, 2020, doi : https://doi.org/10.1007/s00357-019-09345-1.

[12] N. A. Obuchowski and J. A. Bullen, *"Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine," Physics in Medicine & Biology*, vol. 63, no. 7, 2018, doi : 10.1088/1361-6560/aab4b1.

[13] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, *"Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," Jurnal Khatulistiwa Informatika*, vol. 6, no. 1, 2018, doi : https://doi.org/10.31294/jki.v6i1.3799.

[14] S. Faisal, "Klasifikasi data minning menggunakan algoritma c4. 5 terhadap kepuasan pelanggan sewa kamera cikarang," *Techno Xplore: Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 4, no.1, pp. 38-45, 2019, doi : https://doi.org/10.36805/technoxplore.v4i1.541.

[15] A. Bastian, "Penerapan algoritma k-means clustering analysis pada penyakit menular manusia (studi kasus kabupaten Majalengka)," *Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 28-34, 2018, doi : https://doi.org/10.21609/jsi.v14i1.566

[16] A. Aziz, A. Siregar, C. Zonyfar, "Penerapan Algoritma K-Means dan Fuzzy C-Means untuk Pengelompokan Kabupaten Kota Berdasarkan Produksi Padi di Provinsi Jawa Barat," *Scientific Student Journal for Information, Technology and Science*, vol. 3, no.1, pp. 1-8, 2022.

[17] R. A. Indraputra, R. Fitriana,"K-Means Clustering Data COVID-19," *Jurnal Teknik Industri*, vol. 10, no. 3, pp. 275-282, 2020, doi: https://doi.org/10.25105/jti.v10i3.8428.

[18] Z. Nabila, A. R. Isnain, Permata, Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *Jurnal Teknologi Dan Sistem Informasi*, vol. 2, no. 2, pp. 100-108, 2021, doi : https://doi.org/10.33365/jtsi.v2i2.868

[19] Nikmatun, I. Alvi, I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 10, no. 2, pp. 421-432, 2019, doi : https://doi.org/10.24176/simet.v10i2.2882.

[20] K. Mittal, G. Aggarwal, P. Mahajan, "Performance study of K-nearest neighbor classifier and K-means clustering for predicting the diagnostic accuracy," *International Journal of Information Technology*, vol. 11, no. 3, pp. 535-540, 2019, doi : 10.1007/s41870-018-0233-x.

[21] M. Manjusha, R. Harikumar, "Performance analysis of KNN classifier and K-means clustering for robust classification of epilepsy from EEG signals," *In 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), March* 2016, pp. 2412-2416, doi : 10.1109/WiSPNET.2016.7566575.