

## Clustering Potato Seeds Using DBSCAN Algorithm in Optimizing Sales

N P Dharshinni<sup>1</sup>, Andre Putra Persada Tarigan<sup>2</sup>

Universitas Prima Indonesia, Indonesia

Email : [priyadharshinninaidu@gmail.com](mailto:priyadharshinninaidu@gmail.com)

### Abstract

#### Article Info

Received : 29 November 2021

Revised : 18 December 2021

Accepted : 27 December 2021

Potato seed shop CV. ASAP sells various types of potato seeds which are quite trusted by potato farmers. The payment methods used so far are cash and credit payments, but the shop owner never has a balanced stock so that customer demand is always not in accordance with the stock of seeds, resulting in the shop being at a loss. Based on the problems experienced by potato farmers, it is necessary to group potato seed sales data using the DBSCAN clustering algorithm. The DBSCAN algorithm groups sales data into several cluster groups based on epsilon values and minimal points, besides that this algorithm can detect noise in data grouping so that the grouping results obtained are better than other algorithms. The purpose of this study was to apply the DBSCAN algorithm to potato seedling sales data to obtain results in knowing the types of potato seeds that were most purchased by customers from potato seed sales with cash payment methods and credit payments. The results of the application of the DBSCAN algorithm found that he most purchased by customer using cash payment method were the type of spread seeds, class D potato seeds, class A potato seeds and class C potato seeds with the test value of epsilon 1,2 and Minimum points 1 while the sale of seeds Potatoes, the most purchased by customer using credit payment method are the type of potato produced by class C and potato spread seeds with the test value of Epsilon 2 and Minimum points 1.

Keywords : Potato Seed Sales, DBSCAN, Clustering

### 1. Introduction

Grouping sales data is one of the techniques needed as a support in increasing the productivity of potato seed sales, especially in areas that are one of the largest potato producers of agricultural commodities. This is also one of the triggers for the increase in consumers who need potato seeds. In addition, knowing the characteristics of the sale of potato seeds is useful in increasing sales (1). Data processing of potato seed sales to produce information requires data mining techniques. Data mining is a study that collects, cleans, processes, analyzes, and obtains new information from data (4). Data mining can also be used to carry out the process of classifying, predicting, forecasting and obtaining useful information from large data sets (5). One of the most frequently used data mining analysis techniques in classifying sales data is clustering (2). DBSCAN (Density-based Spatial Cluster of Applications with Noise) is a clustering algorithm based on data density (6) and the grouping results will produce 3 kinds of status for each data, namely core, border, and (noise)(7).

Research (Ni Made Anindya Santika Devi, I Ketut Gede Darma Putra, I Made Sukarsa, 2015) explains that the DBSCAN method can assist in data grouping and result in the determination of potential

buyers based on sales datasets (8). (Fitri Indriyani & Eni irfani, 2019) Explains that clustering was developed to produce good clustering performance and has the ability to estimate volume from sales (9). The DBSCAN algorithm has a different way of grouping data compared to other clusters so as to produce accurate data by performing a broad application performance on data grouping (10). (Betha Nurina Sari, & Aji Primajaya 2019) explained applying DBSCAN clustering to sales data to find out the characteristics of each consumer who made a purchase of one item (11). DBSCAN Clustering is an efficient algorithm in clustering large datasets(12). DBSCAN has been widely used in other fields because of its simplicity and good and accurate clustering effect (13).

This study applies the DBSCAN algorithm method in grouping potato seeds based on the potato seed sales dataset to find out what kind of potato seed sales are needed by farmers as consumers of potato seeds and to know the type of payment in cash or credit that is widely used by consumers in optimizing the sale of potato seeds.

## 2. Literature Review

Several studies that support this research can be seen from the research conducted by previous authors by applying the DBSCAN algorithm, namely research (Sharma Dr.AK, 2017) testing the performance value of EPS DBSCAN and E-DBSCAN, producing the best performance value at EPS value: 1.33 and E-DBSCAN : 0.9 which has the highest level of accuracy. (JEBARI,S.,2019) examined the performance of the DBSCAN algorithm and showed the most superior performance based on the AF-DSCAN value of 2,562, FN-DBSCAN: 2,185 and AF-DBSCAN: 2,562. (Zhu,QX,2020) examines the comparison of K-Means and DBSCAN in their clustering speed which shows the k-means algorithm has a faster execution time than DBSCAN with a K-Means computation time of 0.278 and DBSCAN of 1.956 but a low rand index indicates better clustering quality. poor based on its value with the ratio of K-Means: 0.28 and DBSCAN: 1.00. (Safitri.D,Wuryandari.T,RR,2017) applies the DBSCAN algorithm in classifying rice-producing areas in Central Java province from 35 districts/cities in Central Java and the results show cluster 3 is the highest rice-producing area group with 2 noise in the data cluster, namely Cilacap Regency and Wonogiri Regency . (Isnarwaty, D P, Irhamah, 2019) researching based on the silhouette coefficient value shows the DBSCAN method is better than K-Means in classifying tweets aimed at JNE, J&T, and Pos Indonesia expedition services because it produces a higher silhouette coefficient . (M, P. Dewi, C. Siam, Wijayanti, AuliaN, Raeni, 2020) comparing the K-Means and DBSCAN algorithms, getting the results that DBSCAN is better in terms of data grouping as seen from the silhouette coefficient value. (Soni.D, 2017) conducted a study to increase the accuracy of the DBSCAN algorithm based on the back propagation algorithm in calculating the Euclidean distance dynamically. (Li X, Zhang P, Zhu G.2019) aims to identify groups of users of public services using the DBSCAN algorithm based on the density of existing data on public service users. abnormal passengers and the results of this study are input in determining attitudes in emergencies that occur in public services. (Lazarevic A, Xu X, Fiez T, Obradovic b, 2020) applies information gain to evaluate the DBSCAN algorithm in integrating data so that each cluster gets a group in each area . (Li H,a,Zhang A,Pei X,2019) apply the DBSCAN algorithm and BP Neural Network to optimize the CNS machine tool temperature sensor measurement point AND the results of the research get the Pearson correlation coefficient method reducing the temperature measurement point from 16 to 5.

## 3. Method

This study uses a dataset of potato seed sales for the period February 2020 to January 2021 with a total of 131 records. The potato seed sales dataset is divided based on cash payments and credit payments made by customers to see which types of potato seeds are in demand according to the payment method.

Furthermore, the preprocessing stage is carried out by cleaning data to remove duplicate and incomplete data, then proceeding with normalization using the Z Transformation method, data selection selecting attributes used in clustering and data transformation to prepare data into an appropriate form so that it can be processed with the DBSCAN algorithm. After going through the preprocessing stage, then the DBSCAN algorithm is applied and the evaluation of the results of DBSCAN clustering is carried out as seen from the epsilon (eps) and minimum points (Min.pts) values as well as noise in grouping data. The final result was the grouping of potato seeds that were purchased the most by potato farmers. The framework in this research can be seen in Figure 1.

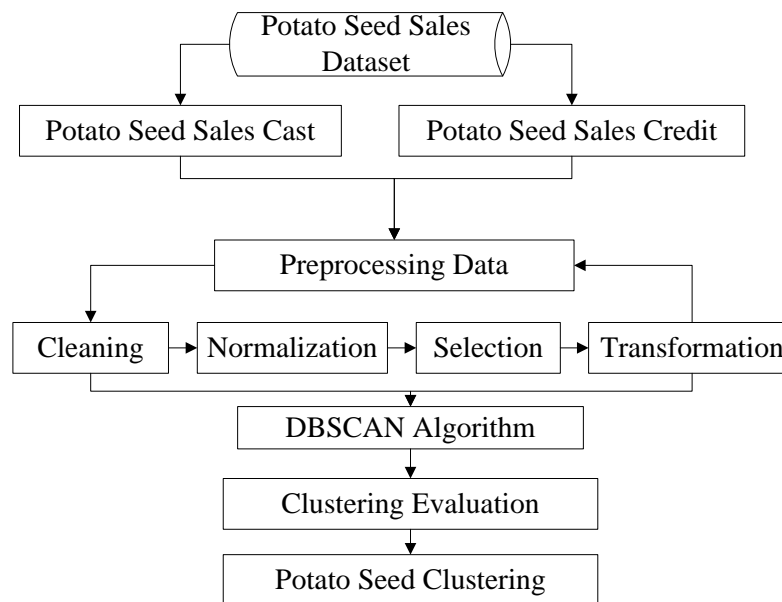


Figure 1. Research Framework

#### 4. Results and Discussion

In this study, the variables used for grouping potato seeds were G-0 (Production Potato Seeds Class C), G-2 (Production Potato Seeds Class D), G-3 (Production Potato Seeds Class A), G4 (Sprouted Potato Seeds)., Price and Type of Payment (Paid/Credit). The dataset is divided into 2 parts based on the type of payment, namely cash or credit payments which can be seen in Table 1 and Table 2.

Table 1. Sales of potato seeds using cash payment method

No.	G – 0 ( Kg )	G -2 (Kg)	G – 3 (Kg)	G – 4 (Kg)	Seed	Price	Payment Method
1	10	20	200	30	160	20000	Cash
2	40	40	200	10	344	20000	Cash
3	50	10	100	20	320	20000	Cash
4	10	20	50	40	260	20000	Cash
5	10	40	90	20	245	28000	Cash
	...	...	...	...	...	...	...
102	10	10	20	513	553	18000	Cash

Table 2. Dataset Penjualan Bibit Kentang Kredit

No.	G – 0 (Kg)	G – 2 (Kg)	G - 3 (Kg)	G - 4 (Kg)	Seed	Price	Payment Method
1	100	50	50	200	400	28000	Credit
2	10	30	50	50	140	28000	Credit
3	20	10	50	35	115	28000	Credit
4	50	40	100	10	200	36000	Credit
5	50	10	200	10	270	36000	Credit
...	...	...	...	...	...	...	...
29	100	30	30	50	210	36000	Credit

The application of the DBSCAN algorithm in grouping potato seeds with cash and credit payment methods can be seen in figure 2 and figure 3. After getting the clustering results, the next step is to test the cluster results to get the best clustering results by setting the epsilon value and minimum points.

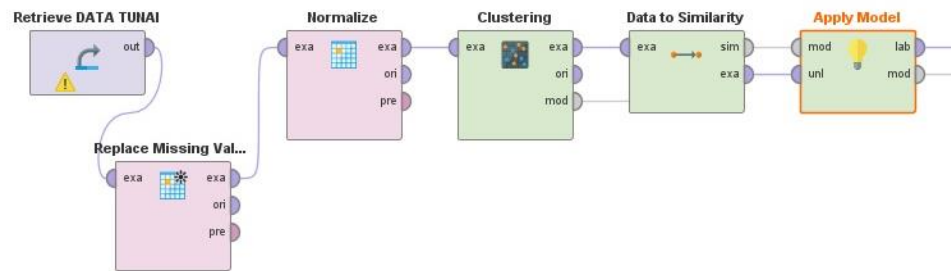


Figure 2. DBSCAN Testing of Potato Seed Sales Data (Cash Payment )

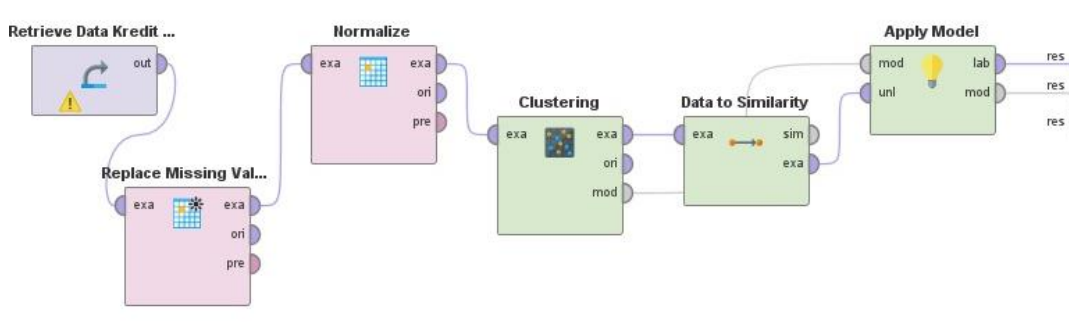


Figure 3. DBSCAN Testing of Credit Potato Seed Sales Data ( Credit Payment)

The results of the DBSCAN clustering test on the potato seed sales dataset with the cash payment method resulted in 4 Clusters based on the epsilon 1.2 value and minimum points 1 with the number of cluster members in Cluster 0 = 33 members, Cluster 1 = 52 members, Cluster 2 = 2 members, Cluster 3 = 2 members and Noise data as many as 13 members. While the credit payment method with an epsilon

value of 2.0 and minimum points 1 produces 2 Clusters with 23 members in Cluster 0, 2 in Cluster 1 and 4 Noise members.

Table 4. Potato Seed Sales Cluster (Cash )

score(clust...	cluster	G - 0( Kg )	G - 2(Kg)	G - 3(Kg)	G - 4(Kg)	Benih (Kg)	Harga Satuan	Total
3.225	cluster_0	-0.891	-0.729	-0.587	-0.377	-0.816	-0.689	-1.067
3.186	cluster_0	-0.441	-0.729	-0.303	-0.377	-0.626	-0.689	-0.886
4.498	cluster_0	0.009	-0.599	1.825	-0.296	0.511	-0.689	0.203
6.233	cluster_2	4.281	-0.729	-0.303	0.003	-0.026	0.734	1.194
6.086	cluster_2	4.281	-0.729	-0.729	0.003	-0.216	0.734	0.940
4.525	cluster_1	0.908	-0.003	-0.729	-0.215	-0.879	0.734	-0.584
4.534	cluster_1	-0.216	-0.210	-0.005	-0.054	-1.037	0.734	-1.007
4.533	cluster_1	0.908	-0.210	-0.445	-0.215	-0.879	0.734	-0.584
4.520	cluster_1	-0.216	-0.210	-0.161	-0.215	-1.037	0.734	-1.007
4.600	cluster_1	-0.441	0.438	-0.303	-0.256	-0.974	0.734	-0.965
4.561	cluster_1	-0.441	-0.210	-0.729	-0.377	-0.721	0.734	-0.626
4.605	cluster_1	0.009	-0.210	-0.005	-0.215	-0.689	0.734	-0.499
2.345	Noise	3.157	-0.210	-0.303	0.003	-1.005	-3.712	-1.312

Table 5. Seed Sales Cluster (Credit )

score(clust...	cluster	G - 0( Kg )	G - 2(Kg)	G - 3 (Kg)	G - 4 (Kg)	Benih (Kg)	Harga Satuan	Total
3.462	cluster_0	-0.816	0.302	-0.806	-1.074	-0.694	0.356	-0.465
4.593	cluster_0	-0.816	-1.215	-0.806	0.002	-1.168	-0.677	-0.983
2.554	cluster_0	-0.013	0.302	-0.380	0.986	0.254	-0.677	-0.120
4.656	Noise	-0.118	0.302	3.536	2.497	2.908	0.356	2.595
0	cluster_1	-0.816	2.199	-0.380	0.986	0.822	0.356	0.824
3.399	cluster_0	-0.816	-0.836	0.684	-0.159	0.064	0.356	0.179
3.254	cluster_0	-0.816	-0.456	-0.380	-0.159	-0.315	0.356	-0.143
0	cluster_1	-0.816	2.199	0.684	0.528	1.107	0.356	1.065

In Table 1 shows the results of the cluster on the sale of potato seeds using the cash payment method to produce 4 clusters with potato seeds G-4, G-2 , G-3 and G-0 which were purchased the most by potato farmers with a price range of Rp.20,000 to Rp. 28,000. While in Table 2 shows the cluster results on the sale of potato seeds using the credit payment method, there are 2 clusters with the most purchased potato seeds being G-0 and G-4 with a seed price range of Rp.20000 to Rp.28000.

Visualization of cluster results on potato seed sales data with cash payment methods and credit payment methods can be seen in Figure 2 and Figure 3.

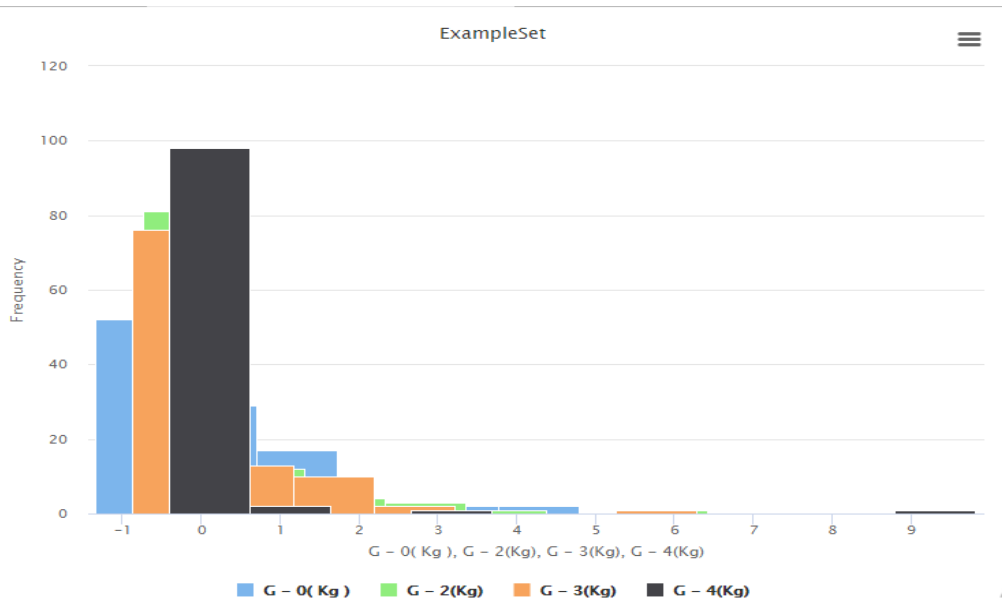


Figure 2. Visualization of Potato Seed Cluter With Cash Payment Method

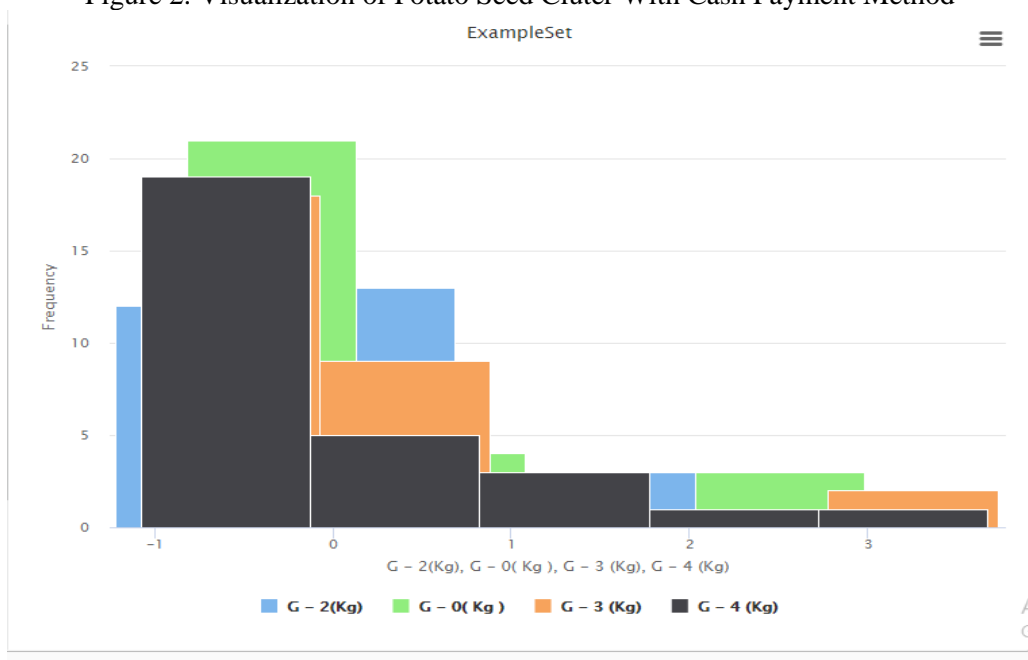


Figure 3. Visualization of Potato Seed Cluter With Credit Payment Method

## 5. Conclusions

The grouping of potato seeds with the DBSCAN algorithm found that the potato seeds purchased by potato farmers using the cash payment method were the type of spread seeds, class D potato seeds, class A potato seeds and potato seeds produced C. Meanwhile, sales of potato seeds used the credit payment method, the most purchased by potato farmers are the types of potato seeds produced by class C

and potato seeds that are spread. The development of further research can perform performance comparisons with other clustering algorithms such as K-Means, K-Medoids and Hierarchical clustering and develop applications to make it easier for seed shops to sell potato seeds to potato farmers.

## Reference

- [1] Metisen. BM., & Sari. HL., "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila", *Jurnal Media Infotama*, 11(2), 110–8, 2015.
- [2] Bu'ulolo. E., & Purba. B., "Algoritma Clustering Untuk Membentuk Cluster Zona Penyebaran Covid-19", *Jurnal Teknologi Informasi dan Komunikasi*, 12(1), 59–67, 2021.
- [3] Izhari. F., "Analisis Algoritma Dbscan Dalam Menentukan Parameter Epsilon Pada Clustering Data Numerik", (Studi kasus menentukan parameter Epsilon pada Clustering), *Seminar Nasional Teknologi Komputer Sains*, 156–8, 2020.
- [4] Hidayati. QR., & Suroho. S., "Implementasi Algoritma Spectral Clustering Untuk Analisis Sentimen", *Jurnal Ilmiah Pendidikan Matematika*, 9(1), 27, 2021.
- [5] Zheng. H., et al., "Clustering algorithm based on characteristics of density distribution", *Proc - 2nd IEEE Int Conf Adv Comput Control ICACC 2010*, 2, 431–5, 2010.
- [6] Huang. TQ., et al., "Reckon the parameter of DBSCAN for multi-density data sets with constraints", *Int Conf Artif Intell Comput Intell AICI 2009*, 4(1996), 375–9, 2009.
- [7] Ashari. BS., Otniel. SC., Rianto., "Perbandingan Kinerja K-Means Dengan DSCAN Untuk Metode Clustering Data Penjualan Online Retail". *Jurnal Siliwangi [Internet]*, 5(2), 72–7, 2019.
- [8] Devi. AS., et al., "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan", *Lontar Komputer JITI*, 6(3), 185, 2015.
- [9] Indriyani. F., & Irfiani. E., "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means" *JUITA Jurnal Informatika*, 7(2), 109, 2019.
- [10] Jain. D., Singh. M., Sharma. AK., "Performance enhancement of DBSCAN density based clustering algorithm in data mining" *Int Conf Energy, Commun Data Anal Soft Comput ICECDS 2017*, 1559–64, 2018.
- [11] Sari. BN., & Primajaya. A., "Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang", *JIKO Jurnal Informatika dan Komputer*, 4(1), 28–34, 2019
- [12] Jebari. S., Smiti. A., Louati. A., "AF-DBSCAN: An unsupervised Automatic Fuzzy Clustering method based on DBSCAN approach", *IWOBI 2019 - IEEE Int Work Conf Bioinspired Intell Proc*, 117–22, 2019.
- [13] Zhu. Q., Tang. X., Liu. Z., "Revised DBSCAN Clustering Algorithm Based on Dual Grid. Proc 32nd Chinese Control Decis Conf CCDC", 3461–6, 2020.
- [14] Safitri. D., Wuryandari. T., Rahmawati. R., "Metode Dbscan Untuk Pengelompokan Kabupaten / Kota Di Provinsi Jawa Tengah. Statistika", *STATISTIKA*, 5(1), 8–13, 2017.
- [15] Isnarwaty. DP., & Irhamah., "Text clustering pada akun twitter layanan ekspedisi JNE , J&T, dan Pos Indonesia menggunakan metode Density-Based Spatial Clustering of Applications with Noise ( DBSCAN )", *Jurnal Sains dan Seni*, 8(2), 2–9, 2019
- [16] M. MP., et al., "Comparison of DBSCAN and K-Means clustering for grouping the village status in Central Java 2020", *Jurnal Matematika Statistika dan Komputasi*, 17(Vol. 17 No.3 (2021), 394–404, 2021.
- [17] Sharma. S., Sharma. AK., Soni. D., "Enhancing DBSCAN algorithm for data mining", *International Conf Energy, Commun Data Anal Soft Comput ICECDS*, 1634–8, 2018.

- 
- [18] Li. X., Zhang. P., Zhu. G., "DBSCAN clustering algorithms for non-uniform density data and its application in urban rail passenger aggregation distribution", *Energies*, 12(19), 1–22, 2019.
19. Lazarevic. A., et al., "Clustering-regression-ordering steps for knowledge discovery in spatial databases" *Proc Int Jt Conf Neural Networks*, 4, 2530–4, 2019.
- [20] Li. H., Zhang. A., Pei. X., "Research on Thermal Error of CNC Machine Tool Based on DBSCAN Clustering and BP Neural Network Algorithm", *Proc 2019 IEEE Int Conf Intell Appl Syst Eng ICIASE*, (100), 294–6, 2019.
21. Mulyono D, Syah MJA, Sayekti AL, Hilman Y. Kelas Benih Kentang (*Solanum tuberosum* L.) Berdasarkan Pertumbuhan, Produksi, dan Mutu Produk. *J Hortik*. 2018;27(2):209.
22. Baharuddin MM, Azis H, Hasanuddin T. Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *Ilk J Ilm*. 2019;11(3):269–74.
23. Pingping H, Yu X, Yan Z, Kui L. Equivalent model of wind farm based on DBSCAN. 2017 IEEE Innov Smart Grid Technol - Asia Smart Grid Smart Community, ISGT-Asia 2017. 2018;1–6.
24. Yu H, Zhang W. Stabilizing System. 2013;1297–301.