JURNAL INFOTEL
Informatics - Telecommunication - Electronics

Website Jurnal : http://ejournal.st3telkom.ac.id/index.php/infotel
ISSN : 2085-3688; e-ISSN : 2460-0997

Performance of the K-Nearest Neighbors method on identification of maize plant nutrients

Bain Khusnul Khotimah [1*]
[1] Universitas Trunojoyo Madura
[1] Madura, Indonesia
*Corresponding email: bain@trunojoyo.ac.id

Received 25 October 2021, Revised 17 November 2021, Accepted 26 February 2022

Abstract — Maize is one type of commodity for domestic consumption and export that has high economic value. However, the low productivity is one of the reasons for declining soil fertility due to nutrient deficiency. Symptoms of nutrient deficiency that arise because the nutritional needs are not met properly from the soil or from the application of fertilizers. Finally, the plant becomes deficient in certain nutrients with the appearance of specific deficiency symptoms. Features in the form of good visual symptoms with damage to stems, leaves, corn cobs there are 17 features from observations to build a diagnostic system using the KNN method. This study uses KNN with minkowski distance calculation by adjusting the change in the value of K to produce the best performance. The final performance result is compared with the Naive Bayes (NB) method to determine the accuracy and computational value. The results of research testing using K = 75 to get an accuracy of 92.40%. Comparative analysis of the K-nearest neighbor (K-NN) and NB methods by looking for the closeness between the new case criteria and the old case criteria based on the closest case criteria. KNN classification performance is able to detect nutrients to anticipate the need for maintenance during planting to support the growth of maize plants. The results show that the K-Nearest Neighbor (K-NN) algorithm has a better accuracy value than Naïve Bayes.

Keywords – K-Nearest Neighbor, maize nutrient deficiencies, classification, category data

I. INTRODUCTION

The advantage of local Madura maize has a longer shelf life compared to other types of maize. So, maize can be used as a food reserve for one year. The local Madura maize had high protein and fat content of 11.24% and 4.96%, respectively, compared to hybrid maize such as Pioneer 7 (8.22% protein content and 3.24% fat content, respectively), Pioneer 11 (8.70% protein content and 3.34 percent fat content), Bisi 2 (9.51% protein content and 3.95% fat content) and Bisi 7 (10.09% protein content and fat content 3.27%) [1].

In Madura, the maize grains are relatively small very suitable for bird food, the relative price of maize is more expensive than others. The result of processing into Madura maize rice is more expensive than other maize. Thus, the reluctance of the Madurese community to change the type of farming from local maize to hybrid maize. They are closely related to the consumption patterns, customs and culture of the Madurese community. Furthermore, Madurese farming is still subsistence, providing for themselves and families [2][3]. The evaluating soil fertility are based on observations or measurements of the growth parameters of growing plants. The plant functions as an indicator of all growth factors is the productivity. Symptoms of nutrient deficiency that arise are caused by the need for nutrients, that are not met either from the soil or from the application of fertilizers. Plants lacking certain nutrients cause specific deficiency symptoms to appear, which can be seen from the yellowing of the leaves and the plants becoming stunted. Nutrient deficiency can be detected early by adding fertilizer [4][5].

K-Nearest Neighbor (KNN) is a method for looking for groups of objects in the training data, that are most similar to the objects in the new data or data testing. This research includes measurements of

performance (accuracy, precision, recall and F-measure) KNN method to value K = 3 to 9 on the object in 1000 of heart disease patient data obtained from the central dataset UCI Machine Learning Repository. The results of the performance measurement obtained that the best K value is 6 where the accuracy value is 85%, precision is 78%, recall is 93% and f-measure is 85% [6-8]. The difference in the results of the test methods on Naïve Bayes, Random Forest, Decision Tree and K-Nearest Neighbor are a disease classification algorithm that generally gets a relatively high accuracy value. This study indicates Neural Network algorithm for the highest accuracy value 89.71%. The second highest accuracy value is the Logistic Regression algorithm which gets an accuracy value of 89.32%, then the Decision Tree algorithm gets an accuracy value of 89.10%. Furthermore, the K-Nearest Neighbor algorithm with an accuracy value of 87.79% and the last one the Naïve Bayes algorithm with an accuracy value of 84.70% [9]. K-Nearest Neighbor can make a decision to classify training data, that using large amounts of data [10] [11].

Dataset of elements of pests lacking nutrients in the form of categorical data. Classification statistical methods for categorization include logistic regression, Naive Bayes, k-nearest neighbor (KNN). Logistics Regression is wrong one classic method that is usually used in classification. Advantages of the method. KNN is a simple but capable algorithm produces high accuracy [2]. Besides, another method that can be used is SVM. SVM is able to work very well on data with many dimensions and avoid dimensionality problems [3]. Based on the two previous studies, the researcher only used the assumption-free method in his research. In this study, the researcher wanted to compare how the results of the classification using assumption-free method and conventional method. Therefore, in this study several methods will be used. Researchers to classify the corn nutrient dataset are logistic regression, decision tree, nave Bayes, k-nearest neighbor, and support vector machine.

This study will use the KNN classification method to detect nutrient deficiencies by comparing the performance measures (accuracy, recall, and precision) with NB. The KNN method has several advantages, namely being strong against noisy training data and effective when the training data is large. The rest of this paper is organized as section 1: introduction, section 2: Research Method, section 3: Result and Discussion Performance the class based KNN and finally section 4 as Conclusion.

II. RESEARCH METHODS

A. Maize Nutrient Deficiency

Maize (Zea mays L.) is the third largest food crop in the world after wheat and rice. In Indonesia, maize is the second largest food crop after rice. Maize can be grown in hot and cold areas with sufficient rainfall and irrigation. The maize crop in Madura is widespread in all four districts namely Bangkalan, Sampang, Pamekasan and Sumenep with the proportion of maize use being dominated by food consumption, with maize generally grown on dry land but rainfed rice fields. Especially in Sumenep Regency, maize can be said to be a staple food for some people, especially for remote or rural communities [10][13].

Problems productivity of maize in Madura development is highly dependent on the existing soil nutrient and fertilization process. The failure of the maize planting process causes public unrest because many farmers make mistakes in recognizing the condition of soil nutrient levels when planting. Symptoms of nutrient deficiency that arise are caused by nutrient needs that are not met either from the soil or from fertilizer application. When certain plant nutrient deficiencies, then the specific deficiency symptoms will appear in the form of a visual sighting of the plant. Each region especially in Madura have fertility levels vary and depend on the type of soil and the geographical location of a region as one of the critical success factors of farming [13][14].

Fig. 1. Maize crop due to nutrient deficiency: a. Nitrogen deficiency; b. Lack of micro elements (iron); c. Lack of phosphorus;

Expert system research determines nutrient deficiency and fertilizer use in post-planting maize plants using the Forward Chaining (Fc). Method is used by the extension council as a consultation medium, which is about determining the type of nutrient deficiency that often attacks maize plants. The information generated is based on symptom criteria for all existing symptoms both on leaves, stems, roots and maize cobs, types and how to overcome them [12][13].

Nutrient detection research is used as a consultation medium for determining the suitability of agricultural land for plant cultivation. The information generated is based on the existing land criteria, so that the decisions made by the system for the types of plants refer to the land criteria. Soil fertility data acquisition technology indispensable. Spatial variability of nutrients in soil and physical properties of the soil are the elemental content Nutrients in the

soil that always fluctuate as one of the supporting factors to meet agricultural requirements [14].

Maize Planting Media does not require special soil requirements. In order to grow optimally, the soil must be loose, fertile and rich in humus. Types of soil that can be planted with maize include andosol (derived from volcanoes), latosol, grumosol, sandy soil. On soil with a heavy texture (grumosol) maize can still be planted with good results with good soil management. Meanwhile, soil with a dusty loam/clay (latosol) texture is the best for its growth. plants absorbing micronutrients is a useful starting point when examining plant requirements. Nutrients removed in harvested crops do not need to be replaced by application of fertilizers, as they may be supplied from sources such as atmospheric deposition or very large uptakes in the soil. Since the land supply low and not replenished from other sources, found for all nutrition and further work needed to fix this knowledge gap crop requirement. Soil acidity is closely related to the availability of plant nutrients [15] [22]. Here are the pH ranges for all soil types and plants.

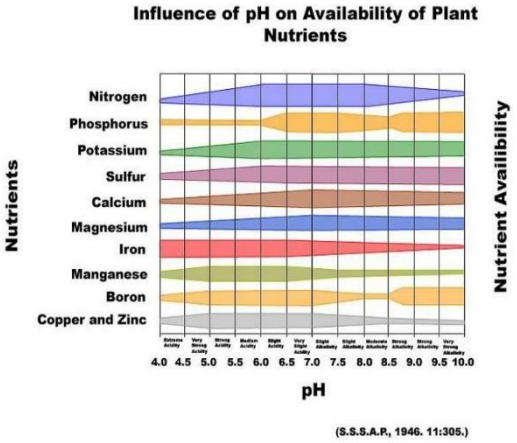

Fig. 2. Graph of the Effect of pH on Plant Nutrition, Data from Van Paemel et al., (2010) [15]

Soil minerals are the main elements needed by all plants to grow. The more soil minerals available, the better the plant will grow. The graph above shows soil mineral content based on soil pH for all soil types and plants. The availability of soil minerals depends on the pH of the soil. Soil pH indicates the degree of soil acidity or the balance between the concentrations of H+ and OH in the soil solution. If the concentration of H+ in the soil solution is more than OH- then the soil solution atmosphere becomes acidic, and if the OH concentration is more than the H+ concentration, the soil atmosphere becomes alkaline. Soil is too acidic, that take effect bad for plant growth. It can cause a decrease in the availability of nutrients for plants [20] [21].

B. Classification

Data mining is a series of processes to solve problems by analyzing data. One method that is often used in processing data mining is the K-Nearest Neighbor (KNN), which improving decision making [2]. The training stage, a training data set with known classes are analyzed and built a model of each class with classification algorithm. This rule is tested data for estimate accuracy, that independent and not used in the process training. That rules can be used for data classification new whose class is unknown [16][17].

C. K-Nearest Neighbor (KNN)

Methods for calculating distances, such as Euclidean distance functions, Manhattan, Mahalanobis. KNN is a non-parametric lazy learning algorithm that does not make any assumptions on the distribution of the underlying data. It does not use generalizations and the training phase is very fast. KNN produces decisions based on the entire training data set [6]. Technique measure the similarity between two samples with n attribute values. Each measure is assumed if dist(x, y) is the distance between two points (x,y) then the shortest distance between the two points is a straight line. If the data is for continuous variables, standardization of Z scores and min max normalization are used first [7][8]. The calculation of the distance near or far from the location (distance) can be calculated through one of the predetermined distance quantities, namely the Euclidean distance, Minkowski distance, and Mahalanobis distance. Minkowski concept treats all uncorrelated independent variables. The transformation performed for the diversity of the data for all variables will contribute equally to the distance [18]. Minkowski distance formula is as follows.

$$d(x_i, x_j) = \left(\sum_{k=1}^{K} |x_{ik} - x_{jk}|^r\right)^{1/r} \quad (1)$$

with

d : the distance between points on the data training x and testing data points y to be classified,
x : sample of data (x_1, x_2,…,x_i)
x_{ik}: the i-th test data on the k-th variable
x_{jk}: the j-th training data on the k-th variable,
i : 1,2,…., C attribute values
k : dimension of independent variable data,

The value r of Minkowski distance is 1, then this distance is equal to Manhattan distance. So, If the value of r is 2 then the Minkowski distance is the same as the Euclidean distance.

$$g_j(X) = \min g_i(X) \quad (2)$$

The nearest neighbor rule is one of the common methods of class reasoning. Ide The nearest neighbor rule is one method for class reasoning. For all examples in the set E, if x_{ik} is the nearest neighbor of x_{jk} , then the category of g_j is the result of the decision, which is the nearest neighbor rule [19].

D. K-Fold Cross-Validation

The usage cross validation predicts the error rate of using 10-fold. Cross validation is used to find the best value from one model by testing the number of errors in the test data. The data is divided into k samples of the same size, where k-1 is the training data sample, and the remaining sample is for test data. The measurement are sensitivity, specificity, and accuracy which calculated. Sensitivity is the proportion of positive events that are classified correctly. Specificity is the proportion of negative events that are correctly classified as negative. Accuracy is a correctly classified proportion [17][18].

E. Data Transformation

The research data includes category data types regarding nutrient deficiency, namely 17 symptoms with two classes. Features in the form of symptoms require modification process by transformation. Data transformation aims to change the scale of measurement of the original data into another form so that the data can meet the assumptions underlying the analysis of variance. data transformation to change it in the form of actual data because the classification only accepts ordinal data input. In this study, the generalization technique was used because changing the low-level data was replaced with a higher-level concept. Data changes are done manually by sorting attribute classes from the smallest to the largest [19].

F. Measurement Performance

Measurement accuracy is used for testing the level of performance of a dataset based on a certain algorithm. The confusion matrix is based on the level of accuracy with units of percent (%) for performance measurement, by comparing the data set from the actual data classification results with the test data. The values of (TN), (FP), (FN), and (TP) will be calculated for accuracy, precision and recall values. Precision values provide information about the number of data that is categorized as positive, namely classified correctly divided by the total data classified as positive. Finally, recall for prediction is the number of correctly predicted for cases in the dataset which obtained by Equation (3), (4), and (5):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} x100\% \quad (3)$$

$$Precision = \frac{TP}{TP+FP} x100\% \quad (4)$$

$$Recall = \frac{TP}{FN+TP} x100\% \quad (5)$$

III. RESULT AND DISCUSSION

K-NN is one of the most famous classification algorithms used to predict the class of records with an unspecified class based on the class of neighboring records. This algorithm is made of three steps as follows [6]:

1. Calculating the distance of the input record from all training records.
2. Determine Parameter K (Number of closest neighbors).
3. Calculate the square of the Euclidean distance (query instance), each object to the given sample data.
4. Then sort the objects into groups that have the smallest Euclidean distance.
5. Collect Yes and No categories (Nearest Neighbor Classification).
6. The results of the classification using the KNN method are the majority, so it can be predicted the value of the query distance to calculate its accuracy.

The process of detecting maize disease caused by nutrient deficiency in the system is divided into three stages, namely pre-processing, design and measurement with the flow of the system as shown in Fig. 3.

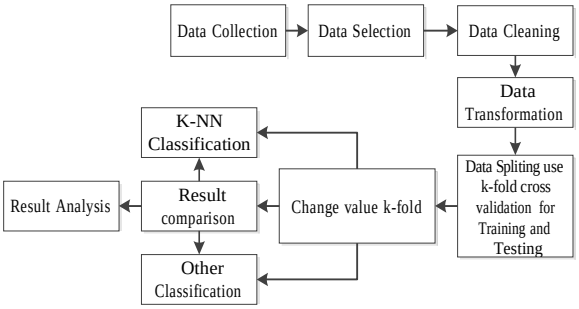

Fig. 3. Steps of Analysis Performance K-NN

G. Nutrient Deficiency in Maize Dataset

The data mining process is collecting data as an analytical material that is influenced by attribute factors, labels and parameters to test the performance of the method. The dataset was taken from observations of leaves, stem conditions in general on maize. This study uses experimental data from the observation of symptoms stated in the questionnaire, consisting of 17 question variables related to the condition of the plant. Symptoms of lack of soil nutrients in corn plants were taken in Rubaru District, Sumenep Regency. The analysis technique is carried out by calculating the performance of the KNN method, namely accuracy, precision, recall and f-measure. The initial stage of data collection, the data obtained as many as 400 data, namely the lack of nutrients that cause maize disease. In the last stage, the performance calculation of all the test data is carried out with various neighbor simulations using the KNN method.

Data transformation aims to change the scale measurement of the original data into another form, that the data can meet the underlying assumptions analysis of variance. Table 1 shown transformation technique data generate special data format to be applied. The usage of generalization technique that changes the low-level data are replaced with higher-

level concepts, or sort attribute classes from higher-level ones smallest to largest.

Table 1. The result of General transformation

Criteria	Scala	Value
Level of colour	Low	1
	Medium	2
	High	3
Leaf spot condition	Low	1
	Medium	2
	High	3
Tree length level	Low	1
	Medium	2
	High	3

The attributes used are symptoms based on observations of nutrient deficiency growth in maize plants totaling 17 symptoms with 2 labels. The form of deficiency or disease is shown by Table 2.

Table 2. Describes attributes associated with nutrient deficiencies

Feature	Description	Criteria
1	Light green plants	Nitrogen deficiency
2	Lower leaves are yellow	Nitrogen deficiency
3	Leaves dry to brown	Nitrogen deficiency
4	Dark green plants interspersed with reds and purples	Phosphorus deficiency
5	Leaves are yellow and dry dark green	Phosphorus deficiency
6	The edges of the leaves look older and yellowed, but the middle is still green	Magnesium deficiency
7	Old leaves wilt or dry up	Magnesium deficiency
8	Old leaves wither but the tips die and dry up	Potassium deficiency
9	The ends of the leaves die with spots	Potassium deficiency
10	Corrugated leaves with dead shoots	Potassium deficiency
11	The leaves are about to die with the new leaves wavy but the edges of the leaves are nicrotic	Calcium deficiency
12	Dwarf plants	Copper deficiency
13	Greenish blue leaves, small and wavy	Copper deficiency
14	The whole nicrotic leaf spreads all over the plant	Sulfur deficiency
15	Stem size short and notched	Zink deficiency
16	Stems are not shortened and not notched, with leaves without necrotic spots	Iron deficiency
17	Leaves develop necrotic patches	Manganese deficiency

H. Performance K-NN

The process of calculating the KNN algorithm needs to prepare class data (labels), attribute data, training data and new data as test data. The measurement results showed the sensitivity, specificity, and accuracy of the diagnosis of nutrient deficiency. In this study, the concept of K-Nearest Neighbor measure distance with using Minkowski. So, many neighbors use the optimal number will be analyzed the performance with distance of 1 to 100. The performance of method chooses good testing data using all variables, the best accuracy test results with k of 1 to 100 is found at K=75 which is 92.40%, Can see in Fig. 4.

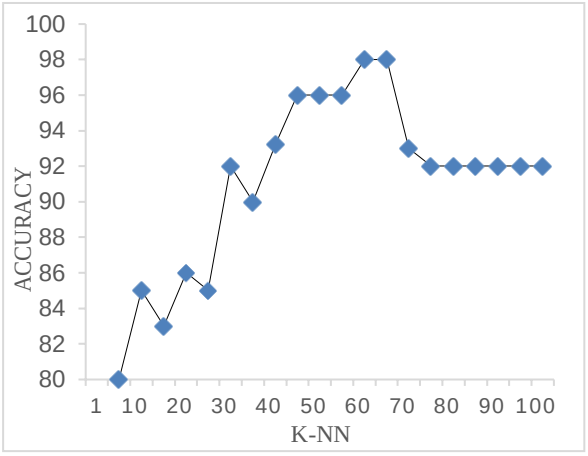

Fig. 4. Performance of changes to the nearest K parameter value on KNN

In Table 2 there are several uses of the K parameter starting from k = 1 to K = 100 nearest to see the best performance in the calculation of the KNN algorithm. The higher the value of k, the more stable the accuracy will not change. From several experiments using the Rapid Miner tool, k-fold cross-validation test on the K-NN classifier divides the data using k = 10 for k-fold crossvalidation as shown in Table 3. Comparison of KNN and Naïve Bayes (NB) classification tests using k-fold cross-validation by entering test data one by one, then recording the classification results. The test carried out on the system was a k-fold cross-validation test with k=10, with details of 200 cases of severe nutrient deficiency and cases of low nutrient deficiency. Then the randomized data was divided into 10 folds, with each fold containing 20 data, shown in Table 3.

Table 3. performance of k-cross validation at k=10 on the K-Nearest Neighbor method

k-fold	KNN			Naïve Bayes		
	recall	prec	acc	recall	prec	acc
1	80.04	82.52	84.79	78.63	92.59	86.47
2	82.01	81.76	90.68	79.45	84.67	84.32
3	88.64	87.84	91.59	82.58	82.16	82.91
4	84.53	82.38	89.56	76.51	86.48	79.75
5	82.51	82.62	92.03	81.54	87.39	81.83

k-fold	KNN			Naïve Bayes		
	recall	prec	acc	recall	prec	acc
6	79.97	86.91	90.57	79.87	90.86	74.92
7	85.95	82.38	92.40	82.53	86.29	85.73
8	87.32	92.10	92.09	84.28	67.60	89.81
9	88.65	92.29	92.32	82.37	76.71	92.29
10	78.21	92.00	91.02	85.59	86.93	81.01

Based on Table 3, the best accuracy, precision and recall are obtained when the number of K-Nearest is being modified repeatedly. While the best recall is obtained when K-Nearest are 75. System test results using K-NN classification method is able to classify 200 data with accuracy, precision, recall. In comparison, the Nave Bayes classification is correct with accuracy using k-fold cross-validation test with k = 10. The classification KNN method has better accuracy than the Naïve Bayes classification method on nutrient deficiency data. The K-NN method gets higher accuracy because the weakness of the Naïve Bayes Algorithm is a parametric algorithm which assumes that each data attribute is independent. When the value of k varies, the KNN accuracy achieved by KNN ranges from between 84.79% and 92.40%. Highest accuracy results on k value is equal to 7 with 92.40%.

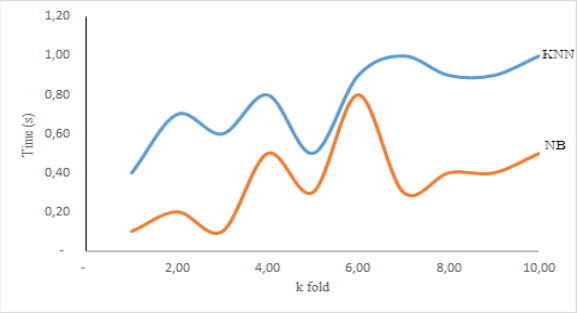

Fig. 5. Comparison of performance on the KNN method with NB based on computational time

In comparison, Naïve Bayes is right classifying 200 data using k-fold cross-validation test with k = 10. This shows that K-NN the classification method has better accuracy than the Naïve Bayes classification method on the data used. The K-NN method gets higher accuracy because Naïve Bayes algorithm is a parametric algorithm that assumes that each data attribute is independent, which is a rare property in the real world. NB of average length time required the K-NN to do the classification is a bit slower because The K-NN algorithm will calculate each training data distance to the test data. On the other hand, the Naïve Bayes algorithm is enough to calculate each probability of each feature for all test data.

System accuracy testing is done by entering the value of k-fold which varies with the values. At the time of testing influenced by the parameters of the method. Where the computational time of KNN is influenced by changes in the value of K-Nearest. The higher the value of K, the longer the computation. Meanwhile, NB is influenced by data variance to find the probability.

IV. Conclusion

System test results using k-fold cross validation for k = 10, K-NN classification method compared to NB able to classify 200 data. The study resulted in the average accuracy of K-NN with the best k difference at k = 7 with an accuracy of 92.40%, and k = 9 with an accuracy of 92.32%. In comparison, Naïve Bayes is right classifying 200 data using the k-fold cross-validation test with k = 9 with accuracy of 92.29%. The K-NN classification method has better accuracy than the Naïve Bayes classification method because of the parametric algorithm used, which assumes that each data attribute is independent, this is a rare property in the real world. During the learning process, the average computation time required for the K-NN method is performing slower classification compared NB, because K-NN algorithm will calculate each training data distance to the test data. On the other hand, Naïve Bayes the algorithm needs to calculate the probability once for all test data for categorical data.

REFERENCES

[1]. Roesmarkam, S., Arifin, F., Sa`adah, SZ., Abu dan Robi`in. 2006. Usulan Pemutihan Varietas Lokal Jagung Madura Manding, Talango dan Guluk-guluk. Balai Pengkajian Teknologi Pertanian Jawa Timur

[2]. Sugiarti, T., dan Hayati, M., 2009. Persepsi Petani Madura dalam Menolak 86 Jurnal Rekayasa Komoditas Jagung Varietas Baru. Embryo. 6(1) : 35-46.

[3]. Zaed, A.S.M, Muhsoni, F.F, Amzeri, A dan Hasan, F., 2009. Pengembangan Pola Tanam dan Diversifikasi Tanaman Pangan di Madura : Suatu Upaya Peningkatan Produksi dan Pendapatan Petani. Laporan Penelitian : Hibah Kompetitif Penelitian Sesuai Prioritas Nasional

[4]. W. Aimrun, M. S. M. Amin, M. Rusnam, D. Ahmad, M. M. Hanafi, A.R. Anuar, "Bulk soil electrical conductivity as an estimator of nutrients in the maize cultivated land," European Journal of Scientific Research, 31 (1), pp. 37-51, 2009.

[5]. L. T. Ngoune, C. S. Mutengwa, "Estimation of Maize (Zea mays L.) Yield Per Harvest Area: Appropriate Methods," Agronomy, 10 (1), 29, December 2019.

[6]. M. Shouman, T. Turner, and R. Stocker, Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients, International Journal of Information and Education Technology, Vol. 2, No. 3, pp.220-223, June 2012.

[7]. L Lei L, C. Xiu-Min, J. Zhong-Lian, "Ship trajectory classification algorithm based on KNN", Journal of Dalian Maritime University, vol.44, no.3, pp.15-21, 2018.

[8]. S. Zhang, X. Li, M. Zong, et al. "Efficient kNN Classification With Different Numbers of Nearest Neighbors", IEEE Transactions on Neural Networks & Learning Systems, vol.29, no.5, pp.1774-1785, 2018.

[9]. A. Kamilaris, F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and

Electronics in Agriculture, 2018.
[10]. Syarief, M., Prastiti, N., Setiawan, W., (2019). Maize Leaf Disease Image Classification Using Bag of Features, *Jurnal Infotel* 11(2), May, 48-54.
[11]. D. Chauhan, "Detection of Maize Disease Using Random Forest Classification Algorithm, Turkish Journal of Computer and Mathematichs Education (Turcomat)," Turkish Journal of Computer and Mathematics Education (Turcomat), 12 (9), 2021.
[12]. K. T. Aliyu, J. Huising, A. Y. Kamara, J. M. Jibrin, I. B. Mohammed, G. Nziguheba, A, M. Adam, B. Vanlauwe, "Understanding nutrient imbalances in maize (Zea mays L.) using the diagnosis and recommendation integrated system (DRIS) approach in the Maize belt of Nigeria," Scientifc Reports, 11:16018, 2021.
[13]. A. Amzeri, Tinjauan Perkembangan Pertanian Jagung Di Madura Dan Alternatif Pengolahan Menjadi Biomaterial, Rekayasa, Vol.11, No. 1, pp.74-86, April 2018.
[14]. K. B. L. Ma Liu, L. Luan, and Chaohai Li. 2011. Nitrogen, phosphor, and potassium nutriunt effect on grain filling and yield of high-yielding summer corn. Journal of Plant Nutrition. Vol (34): 1516-1531.
[15]. M. V. Paemel, N. Dierick, G. Janssens, V. Fievez, S. De Smet, 2010. Selected trace elements and ultra trace elements: Biological role, content in feed and requirements in animal nutrition – Elements for risk assessment. European Food Safety Authority report.
[16]. S. K. Sarangi, V. Jaglan, Y. Dash, A Review of Clustering and Classification Techniques in Data Mining, International Journal of Engineering, Business and Enterprise Applications (IJEBEA), 4(2), pp.140-145, March-May, 2013
[17]. N. Maurya, N. Kumar, V, Maurya, A Review On Machine Learning (Feature Selection, Classification And Clustering) Approaches Of Big Data Mining In Different Area Of Research, August 2020, Journal Of Critical Reviews 7(19):2610-2626
[18]. M. Bramer, Principles of data mining. Springer, 2007
[19]. H. A. Ur Rehman, C. Lin & Z. Mushtaq, Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Disease, Journal of the Chinese Institute of Engineers, 14 November 2020
[20]. R.P. Chaudari R.P., D. V., Ahire, M. Chkravarty, S. Maity. 2014. "Electrical conductivity as a tool for determining the physical properties of indian soils," International Journal of Scientific and Research Publications, 4 (4), pp. 1-4, 2014.
[21]. G. B. Tesfahunegn, T. A. Gebru, "Variation in soil properties under different cropping and other land-use systems in Dura catchment," Northern Ethiopia, PLOS ONE. February 5, 2020.
[22]. H. F. M. ten Berge, R. Hijbeek, J. Rurinda, K.Tesfaye, S.Zingore, P.Craufurd, "Maize crop nutrient input requirements for food security in sub-Saharan Africa," Global Food Security, 23, pp. 9-21, December 2019.