Analysis of Online Shopping Addiction Level Using the K-nearest Neighbor Algorithm at SMK Negeri 1 Tanjung Pura

Zulham Sitorus[1], Alviona Marsya[2], Desy Ramatika[3], Ramli S Siburian[4]
[1,2,3,4]Master of Information Technology, Universitas Pembangunan Panca Budi

ABSTRACT

The rapid development of technology has significantly had an indirect impact on all aspects and dimensions of human life. Online shopping is a form of technological progress. Online shopping is all activities related to online transactions that take place via the internet or other electronic networks. The online shopping process no longer requires face-to-face contact but can be done simply by communicating via the internet. Based on the researcher's preliminary study, at SMK Negeri 1 Tanjung Pura many students spend time shopping online. This online shopping is not only done during break times or after school but is also done during class hours. Based on researchers' observations, students at SMK Negeri 1 Tanjung Pura who like online shopping have an impact on their education. Their study time is reduced, causing their grades to drop. Apart from that, online shopping affects its users, it can cause relational and social problems which have made children rarely socialize with their surroundings, withdraw from social interactions and ultimately make their lives uncontrollable because online shopping takes over their minds. Therefore, through this task, an analysis of the level of online shopping addiction was made using the K-Nearest Neighbor method. The data used is data from students at SMA Negeri 1 Tanjung Pura and the results of this data will be classified using the K-Nearest Neighbor algorithm to find out whether someone is addicted to online shopping based on the level of addiction. The results of the analysis of the level of online shopping addiction in these students, whether they have low addiction or high addiction.

Keyword : K – Nearest Neighbor ; Algoritm; Prediction.

Corresponding Author:	Article history:
Zulham Sitorus,	Received Jun 20, 2024
4Master of Information Technology	Revised Jun 25, 2024
Universitas Pembangunan Panca Budi	Accepted Jun 30, 2024
Jl.gatot Subroto km. 4,5 medan, Indonesia.	

1.   INTRODUCTION

In today's era, technological advances follow the times and develop very rapidly. The rapid development of technology has significantly had an indirect impact on all aspects and dimensions of human life. Online shopping is one form of technological progress. Online shopping is all activities related to online transactions that take place via the internet or other electronic networks. This online shopping offers many changes in buying and selling transactions. The Online Shopping process no longer requires face-to-face meetings but can be done simply by communicating via the internet. There are many applications that can already carry out these online Shopping transactions such as Shopee, OLX, Tokopedia, Blibi, Lazada, Kaskus and others. Based on the researcher's preliminary study, that at SMK Negeri 1 Tanjung Pura many students spend time shopping online. This online shopping is not only done during breaks or after school but also during class hours. Based on the researcher's observations, students at SMK Negeri 1 Tanjung Pura who like to shop online have an impact on their education. Their study time is reduced, causing their grades to drop. In addition, online shopping affects its users and can cause relational and social problems that have made children rarely socialize with their surroundings, withdraw from socializing and ultimately make their lives uncontrollable because online shopping takes over their minds. Therefore, through this assignment, an analysis of the level of online shopping addiction was made using the K-Nearest Neighbor method. The data used is data from students at SMA Negeri 1 Tanjung Pura and the results of the data will be classified using the K-Nearest Neighbor algorithm to find out whether someone is addicted to online shopping based on their level of addiction. The results of the

analysis of the level of online shopping addiction in these students, whether they have low addiction or high addiction.

2.   RESEARCH METHOD

In this study, we apply the KNN (K-Nearest Neighbor) algorithm to analyze the level of online shopping addiction at SMK Negeri 1 Tanjung Pura. Based on the analysis process data, the addiction level consists of Wasteful, Social Interaction and Fraud. The addiction level process is analyzed from training data using an algorithm. Furthermore, what is used to ensure the level of accuracy of the analysis process is data testing. K-Nearest Neighbor is one of the algorithms used in data classification.

K-Nearest Neighbor

One of the methods used in data classification is the K-Nearest Neighbor algorithm or commonly called the KNN algorithm. compared to other classification methods, the way this algorithm works is simpler. One of the drawbacks of the K-Nearest Neighbor algorithm is the low final result of the accuracy value obtained. Here are the steps to calculate the K-Nearest Neighbor algorithm method to classify new unknown data:

1. Determine the number of k neighbors, the number of k taken is generally odd because it avoids the same distance in the classification process.
2. From the data that has been given, then continue by calculating the Euclidean distance of each object.
3. The results of the distance calculation are then sorted from the smallest Euclidean distance.
4. The classification results are collected according to the categories on the label
5. After the calculation is carried out, the prediction results will appear from the category determined by the K-Nearest Neighbor method.

There are several calculations in the KNN method, the most commonly used calculation is the calculation based on the closest distance. The function in determining the most commonly used distance is the Euclidean distance

$$dis(x1, x2) = \sqrt{\sum (xtraining - xtesting)} \, 2$$

Xtraining   = data training to i `
Xtesting   = data testing
N   = amount of data training
i   = record row i of table

Euclidean Distance

Eunclidean distance is the most frequently used metric to calculate the similarity of two vectors. Euclidean distance calculates the root of square differences between 2 vectors. The equation of Euclidean distance is (Putra, 2009)

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Confusion matrix

The method of calculating accuracy in data mining can be done with a confusion matrix, this method can calculate the values of accuracy, precision and recall and f-Measure. Confusion matrix is also a tool whose usefulness has a function to analyze whether the classification recognizes tuples from different classes well. Confusion Matrix testing represents the results of model evaluation using a matrix table, Evaluation using a confusion matrix produces values of accuracy, precision, recall.

Table 1. *Confusion Matrix*

Correct Clasification	*Clasification*	
	Positif	**Negatif**
Positif	TP	TN
Negatif	FP	FN

Description:
TP = True Positive, which is the number of data with positive actual values and positive predicted values.
FP = False Positive, which is the number of data with negative actual values and positive predicted values.
TN = True Negative, which is the number of data with negative actual values and negative predicted values.
FN = False Negative, which is the number of data with positive actual values and negative predicted values

Research Overview
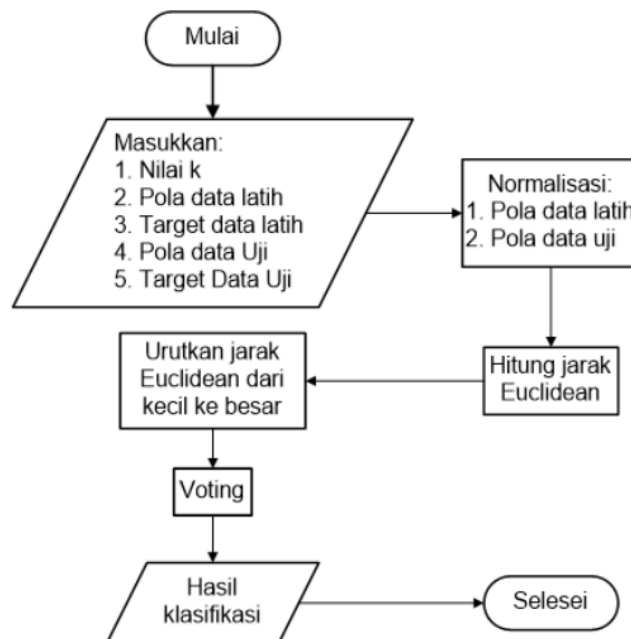The following is an overview of the research process.

Figure 1. Research Overview

The research process begins with determining the topic and formulating the problem, followed by data collection using literature studies and field studies. The analyzed data will then be implemented with the K-Nearest Neighbor algorithm. Testing is carried out on training data and test data that have been determined to determine the results of the accuracy calculation of the K-Nearest Neighbor algorithm..

Data Collection
The type of data used in this study is quantitative data. The data was obtained from students of SMK Negeri 1 Tanjung Pura. The data in the form of a questionnaire used to analyze the level of addiction, namely Wasteful, Social Interaction and Fraud in online shopping. This value data uses data from class X students with a total of 30 people. The research method used is the literature study method and field study with the questionnaire method.

In this Questionnaire Method to find the high or low value of the level of online shopping addiction in students of SMK Negeri 1 Tanjung Pura, the Numeric Rating Scale (0-10) will be used for 3 Dimensions, namely:
Extravagance is measuring how much students spend on online shopping compared to their income or pocket money. Excessive spending can be an indicator of shopping addiction. Numeric Rating Scale (0-10), namely:
a. 0-3: Not Extravagant
b. 4-6: Quite Extravagant

c. 7-10: Very Extravagant
1. Social Interaction is measuring how much time students spend interacting with others online and offline. Online shopping addiction is often associated with reduced social interaction because more time is spent browsing and shopping online. Numeric Rating Scale is (0-10), namely:
a. 0-7: Easy to Interact
b. 8-10: Not Easy to Interact
2. Fraud is measuring the frequency or tendency of students to be involved in or become victims of online fraud. Online shopping addiction can increase the risk of being trapped in online fraud due to impulsive shopping decisions and lack of vigilance. Numeric Rating Scale is (0-10), namely:
a. 0-7: Not Easily Fooled
b. 8-10 : Easily fooled

Data Transformation
In the data transformation process, a process will be carried out to change an attribute into a form/format that is in accordance with the program to be run.
Data transformation will be carried out on the questionnaire data of SMK Negeri 1 Tanjung Pura students for the level of online shopping addiction to be used as a label/class. Where the determination of the label/class is determined by the principal or school with the Numeric Rating Scale (0-10), namely:
a. 0-7.4 = Low
b. 7.5-10 = High
Based on the raw data in the students' data, the label/class will be determined so that it can be drawn to determine the level of addiction of the students in online shopping. The data needs to be transformed first to adjust to the needs of the data mining method. In the Transformation Stage, the Training Data and Test Data are determined. Can be seen in Table 1. Training Data and Table 2 Test Data.

Table 2. Training Data

Boros	Interaksi Sosial	Penipuan	Tingkat Kecanduan
8	9	7.5	Tinggi
5	6	4	Rendah
7.5	8.5	9.5	Tinggi
8.1	8.2	8.3	Tinggi
7.6	7.7	7.8	Tinggi
9	9.1	9.2	Tinggi
2.8	3.5	5.5	Rendah
7.9	8	8.8	Tinggi
8.6	7.6	9.6	Tinggi
7.5	8.8	8	Tinggi
6.5	6	5	Rendah
6.6	5.7	2.7	Rendah
9.9	8.4	8.1	Tinggi
9.7	8.1	7.6	Tinggi
6.4	4.7	2.9	Rendah
9.8	7.5	8	Tinggi
6.7	4.9	6.8	Rendah
8.9	9.8	8.6	Tinggi
8.8	7.8	9	Tinggi
6.3	5.8	6.5	Rendah
8.4	7.9	7.5	Tinggi

7.7	8.9	9.9	Tinggi
8.7	9.9	9.5	Tinggi
4.4	3.8	5.6	Rendah
7.8	8.6	9.3	Tinggi
8.3	8.7	9.9	Tinggi
8.6	9.6	8.8	Tinggi
3.5	6.6	6.6	Rendah
7.9	9.7	7.7	Tinggi
8.9	8.7	8.5	Tinggi

Table 3. Test Data

Boros	Interaksi Sosial	Penipuan	Tingkat Kecanduan
8.8	7.7	9.9	?
6.7	6.6	6.5	?
7.9	8.9	9.5	?

3. RESULTS AND DISCUSSION

Algorithm Implementation Results Before carrying out the calculation process, first divide the data into training data and test data. Training data is taken from old data while new data is taken from the 31st data. After dividing the data, then determine the value of k by conducting a test by increasing the value of K until the prediction becomes more stable, thus making it more likely to make a more accurate prediction up to a certain point.

Before calculating the distance, it is necessary to determine the value of the nearest neighbor k that will be used. In this problem, the value of the parameter k that will be used is k = 1 to k = 10. The next stage is to calculate the distance based on the Euclidean Distance. To get better results in the data mining method, data that has been previously transformed needs to be normalized first. In this study, the normalization used is min-max.

Table 4. Distance calculation

Max	9.9	9.9	9.9
Min	2.8	3.5	2.7

Table 4.1 Training Data Normalization

Boros	Interaksi Sosial	Penipuan
0.732394366	0.859375	0.666666667
0.309859155	0.390625	0.180555556
0.661971831	0.78125	0.944444444
0.746478873	0.734375	0.777777778
0.676056338	0.65625	0.708333333
0.873239437	0.875	0.902777778
0	0	0.388888889
0.718309859	0.703125	0.847222222
0.816901408	0.640625	0.958333333
0.661971831	0.828125	0.736111111
0.521126761	0.390625	0.319444444

0.535211268	0.34375	0
1	0.765625	0.75
0.971830986	0.71875	0.680555556
0.507042254	0.1875	0.027777778
0.985915493	0.625	0.736111111
0.549295775	0.21875	0.569444444
0.85915493	0.984375	0.819444444
0.845070423	0.671875	0.875
0.492957746	0.359375	0.527777778
0.788732394	0.6875	0.666666667
0.690140845	0.84375	1
0.830985915	1	0.944444444
0.225352113	0.046875	0.402777778
0.704225352	0.796875	0.916666667
0.774647887	0.8125	1
0.816901408	0.953125	0.847222222
0.098591549	0.484375	0.541666667
0.718309859	0.96875	0.694444444
0.85915493	0.8125	0.805555556

Table 4.2 Test Data Normalization

Max	8.8	8.9	9.9
Min	6.7	6.6	6.5

Boros	Interaksi Sosial	Penipuan
1	0.47826087	1
0	0	0
0.571428571	1	0.882352941

Table 4.3 Euclidean Distance		
Data Uji 1	Data Uji 2	Data Uji 3
0.57268828	1.31124801	0.30365432
1.07492494	0.860883359	0.68728215
0.45817116	1.393030357	7.95907745
0.85720807	8.189446329	5.83090652
8.20000137	5.966527662	0.79425886
5.71242604	1.530744861	0.96006381
1.1747204	0.388888889	1.21563365
0.92070555	1.314587127	0.90966718
0.98908543	1.41284503	1.05252557
0.88234207	1.29068094	0.76131378
0.58227517	0.725400403	0.68986442
0.48386128	0.636093675	0.65724863
0.80316757	1.465838204	0.89504475
0.72234644	1.387160009	0.83820017

0.57299287	0.541312943	0.81552036
0.75072652	1.380041277	0.92427484
0.77119853	0.820880251	0.96701003
0.97338484	1.542280951	0.86863099
0.90945841	1.389670837	0.97374066
0.74146819	0.806664189	0.83373104
0.72997223	1.240644786	0.76767288
1.10887105	1.479259426	1.01907147
1.09213222	1.607019882	0.97946175
0.97385987	0.463908157	1.09107567
1.0145321	1.404001741	0.94824708
1.07819264	1.503407995	1.03752318
0.98833523	1.514450617	0.88331176
1.05165458	0.733309084	0.88478628
0.89564614	1.391653226	0.71049543
0.88344376	1.430812076	0.87570683

After the calculation results of the Euclidean distance value are obtained, then they will be sorted based on the smallest value of the distance value. Next, a classification is carried out whether it is included in the High or Low addiction level by showing the distance obtained from the test data to each training data, the next stage is to sort each distance value from the smallest (closest) to the largest distance. Then an examination is carried out on each class based on the input value of k (nearest neighbor). Because in this study the value of the k parameter used is k = 1 to k = 29, then the 11 smallest Euclidean Distances are taken.

Table. 5 Ranking calculation results

Euclidean Distance			Rangking			Label / Kelas		
Data Uji 1	Data Uji 2	Data Uji 3	Data Uji 1	Data Uji 2	Data Uji 3	Data Uji 1	Data Uji 2	Data Uji 3
0.45817116	0.388888889	0.303654321	1	1	1	Tinggi	Rendah	Tinggi
0.48386128	0.463908157	0.657248625	2	2	2	Rendah	Rendah	Rendah
0.57268828	0.541312943	0.687282154	3	3	3	Tinggi	Rendah	Rendah
0.85720807	0.636093675	0.689864418	4	4	4	Rendah	Rendah	Rendah
0.58227517	0.725400403	0.710495434	5	5	5	Rendah	Rendah	Tinggi
0.72234644	0.733309084	0.761313776	6	6	6	Tinggi	Rendah	Tinggi
0.72997223	0.806664189	0.767672877	7	7	7	Tinggi	Rendah	Tinggi
0.74146819	0.820880251	0.794258864	8	8	8	Tinggi	Rendah	Tinggi
0.75072652	0.860883359	0.815520357	9	9	9	Rendah	Rendah	Rendah
0.77119853	1.240644786	0.833731038	10	10	10	Tinggi	Tinggi	Rendah
0.80316757	1.29068094	0.838200167	11	11	11	Tinggi	Tinggi	Tinggi

Based on Table -- in Test Data 1 the class that mostly appears when k = 1 is High, k = 3 is High, k = 5 is High, k = 7 is High and so on. So that the classification results for test data 1 student online shopping addiction are included in the High level classification category, In Test Data 2 the class that mostly appears when k = 1 is Low, k = 3 is Low, k = 5 is Low, k = 7 is Low and so on. So that the classification results for test data 2 student online shopping addiction are included in the Low level classification category, While in Test Data 3 the class that mostly appears when k = 1 is High, k = 3 is High, k = 5 is High, k = 7 is High and so on. So that the classification results for test data 3 student online shopping

addiction are included in the High level classification category using the KNN Algorithm. Can be seen in the table below.

**Table 6. Test Data Calculation Results Using the KNN Algorithm**

Boros	Interaksi Sosial	Penipuan	Tingkat Kecanduan
8.8	7.7	9.9	Tinggi
6.7	6.6	6.5	Rendah
7.9	8.9	9.5	Tinggi

Based on the table above, if calculated actually/original data, the online addiction class gets the same results, because the provisions of the school with the Average Value of the Numeric Rating Scale (0-10) are:

    a. a. 0-7.4 = Low
    b. b. 7.5-10 = High

Testing and Accuracy Using Rapid Miner
In this study, 30 test data were used with parameter values k = 1
to k = 29 with 10-fold cross validation testing using the Rapid Miner tool for evaluating the K-NN algorithm, in Figure ---. So that a comparison of the accuracy results of the K-NN algorithm is obtained based on parameter values k = 1 to k = 29. Table --- Comparison of KNN Parameter Accuracy

Parameter K	Akurasi %
K1-K14	100 %
K15-K16	90 %
K17-K29	70 %
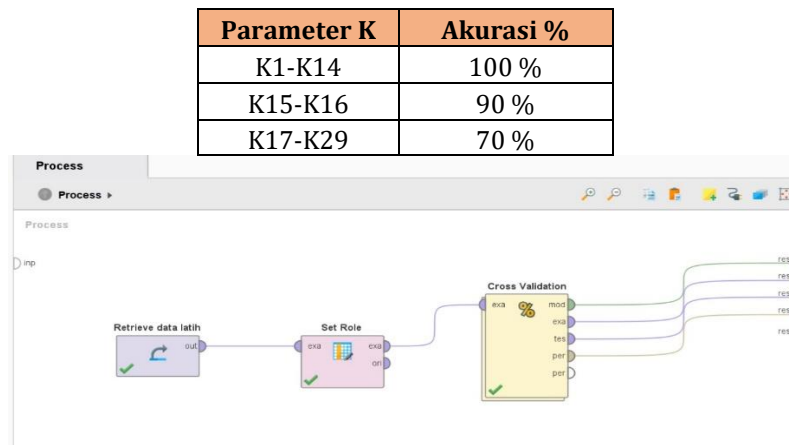
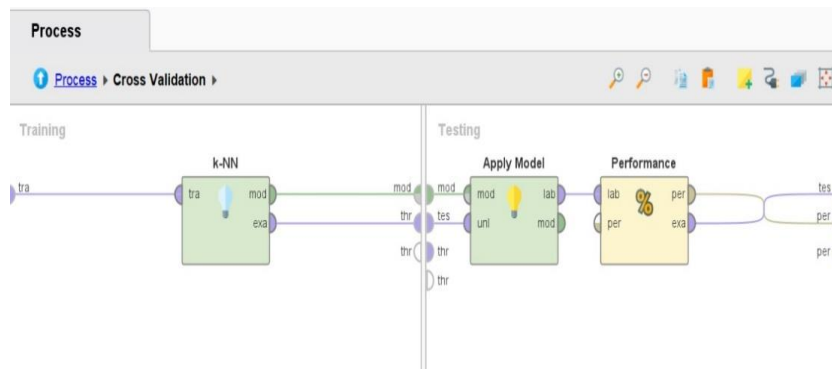**Figure 2. KNN Algorithm Classification Process**

Figure 3. KNN 10 Fold Cross Validation Algorithm Classification Process

In the comparison table of knn parameter accuracy, it can be seen that the best accuracy is produced by the parameter value k = 1 to k = 14 with the resulting accuracy value of 100%. To determine the performance of the K-Nearest Neighbor algorithm in classifying a predetermined class/label, a test will be carried out on the accuracy results. In this test, there are 30 data. The results of the accuracy test will appear in the form of a 3 x 3 confusion matrix in the test image on the accuracy results

accuracy: 90.00% +/- 22.50% (micro average: 90.00%)

	true Tinggi	true Rendah	class precision
pred. Tinggi	21	3	87.50%
pred. Rendah	0	6	100.00%
class recall	100.00%	66.67%	

Figure 4. Testing on accuracy results

PerformanceVector

```
PerformanceVector:
accuracy: 90.00% +/- 22.50% (micro average: 90.00%)
ConfusionMatrix:
True:    Tinggi  Rendah
Tinggi: 21       3
Rendah: 0        6
```

Figure 5. PerformanceVektor

For the accuracy value based on the confusion matrix table, the results of the KNearest Neighbor classification obtained a classification accuracy of 90.00%. Because in the original class, namely 21, it has a High value, and 9 Low, while based on the K parameter value of 15, then for the results of the accuracy of other k parameters can be seen in the Table --- accuracy comparison. The confusion matrix table shows that it can classify 21 True Highs that are predicted correctly and 6 True Lows that are predicted correctly. Then 0 True Highs that are predicted Wrong and 3 True Highs that are predicted wrong. So we can see that there are some differences. The results of measuring data accuracy from the confusion matrix for the KNN algorithm can be calculated with the accuracy level value. *Accuracy* :

$$\frac{TP+TP}{Total\ Data} \times 100\ \% = \frac{21+6}{30} \times 100\% = \frac{27}{100} \times 100\% = 90,00\ \%$$

From the results of the table --- it shows that the level of accuracy produced is 90.00% or 1.00, so it can be classified into the following classification:
a. 0.90-1.00 = *excellent classification*
b. 0.80-0.90 = *good classification*
c. 0.70-0.80 = *fair classification*
d. 0.60-0.70 = *poor classification*
e. 0.50-0.60 = *failure classification*

So it can be concluded that the Level of Online Shopping Addiction at SMK Negeri 1 Tanjung Pura is in accordance with the K Parameter grouping, namely K1 to K 14 Reaching 100%, K15 to K16 Reaching 90%, can be classified in the 0.90-1.00 classification, which is included in the excellent classification, while K17 to K29 Reaching 70%, can be classified in the 0.70-0.80 classification, which is included in the good classification.*

4. CONCLUSION

Based on the results and discussions in this study, the conclusion obtained from the rules formed can be concluded that students at SMK Negeri 1 Tanjung Pura must improve the level of online shopping addiction in these students by improving the three attributes that are the level of addiction. And with this KNN Algorithm, it can help SMK Negeri 1 Tanjung Pura to find out how much addiction to online shopping in these students. with the rapidminer tool the best accuracy is produced by the parameter value k = 1 to k = 14 with the resulting accuracy value of 100%. and K15 to K16 Reaching 90%, can be classified in the classification of 0.90-1.00 which is included in the excellent classification, while K17 to K29 Reaching 70%, can be classified in the classification of 0.70-0.80 which is included in the good classification.

REFERENCES

Dwynne, Z. C., Aini, D. N., & Pertiwi, T. A. (2023). *Cluster Level of Online Game Addiction in Students of the Faculty of Science and Technology and its Correlation with Learning Interest Cluster Tingkat Kecanduan Game Online Pada Mahasiswa Fakultas Sains dan Teknologi dan Korelasinya Terhadap Minat Belajar*. 126–132.

Hidayat, A., Mustakim, Fauzi, M. Z., & Syukra, I. (2019). Implementasi Algoritma K-Nearest Neighbor dan Probabilistic Neural Network untuk Analisis Opini Masyarakat Terhadap Toko Online di Indonesia. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 11*, November, 254–261.

Manurung, A., Santoso, H. G., Yustanto, R., Susiani, T., & Afrisawati, A. (2023). Decision Support System Dalam Pemilihan Buah Kelapa Sawit Terbaik Menggunakan Metode Moora. *J-Com (Journal of Computer)*, *3*(2), 78–84. https://doi.org/10.33330/j-com.v3i2.2487

Neighbor, K., & Chasanah, D. N. (2022). *Klasifikasi Kelayakan Siswa dalam Menentukan Kelas Unggulan Menggunakan Algoritma*. *III*, 51–58.

Panjaitan, D. L., & Hasugian, P. M. (2021). Implementation of K-Nearest Neighbor Algorithm for Classification of Class Placement At Junior High School , Padang Month Issn : 2302-9706. *Jurnal Infokum*, *10*(1), 43–49.

Rachma, C. A. (2022). *Implementasi Algoritma K-Nearest Neighbor Dalam Penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur.*

Rosanensi, K. M., Jayadi, M., Syafitri, D., Studi, P., & Komputer, I. (2023). *ANALISIS TINGKAT KECANDUAN GAME ONLINE MENGGUNAKAN METODE CERTAINTY FACTOR DAN ALGORITMA K-NEAREST NEIGHBOR LOMBOK*. *7*(2), 473–482. https://doi.org/10.52362/jisamar.v7i2.1040

Sebastian, D. (2019). Implementasi Algoritma K-Nearest Neighbor Untuk Melakukan Klasifikasi Produk dari beberapa E-marketplace. *Jurnal Teknik Informatika dan Sistem Informasi*, 5(1), 51-61. https://journal.maranatha.edu/index.php/jutisi/article/view/1581/1246

Sianturi, F. A., Simangunsong, A., Simanjorang, R. M., & Sijabat, P. I. (2020). Implementasi Algoritma Modified Nearest Neighbourt (M-KNN) Untuk Klasifikasi Buku. *Jurnal Media Informatika [Jumin]*, *1*(2), 45–51.

Widiana, S. A., Firdaus, I., Tenda, E., & Ketaren, E. (2023). Sistem Informasi Prediksi Penjualan Produk Thrift Di Toko Manado Menggunakan Algoritma K-Nearest Neighbor (K-Nn). *Jurnal TIMES*, *XII*(2), 52–57. http://ejournal.stmik-time.ac.id/index.php/jurnalTIMES/article/view/708%0Ahttps://ejournal.stmik-time.ac.id/index.php/jurnalTIMES/article/download/708/271

Yunus, M., Pratiwi, N. K. A (2023), Prediksi Status Gizi Balita dengan Agoritma KNN di Puskesmas Cakranegara. *Jurnal Teknologi Informasi dan Multimedia,* 4(4), 222-231. *https://journal.sekawan-org.id/index.php/jtim/article/view/328/197*