Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13377
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Clothing Recommendation System Using the K-Nearest Neighbor Method

Arya Maghrizal Putra[1)*], Muhamad Irsan[2)], Muhammad Fariz Fathoni[3)]
[1,2,3)] Telkom University, Indonesia
[1)]aryamaghrizal@email.com, [2)]irsanfaiz@telkomuniversity.ac.id, [3)]mfariswork@telkomuniversity.ac.id

Submitted :Jan 10, 2024 | **Accepted** : Mar 28, 2024 | **Published** : Apr 1, 2024

Abstract: The world of fashion and the way we interact with it has been transformed by advances in information and communication technology. Clothing recommendation applications have become increasingly common, helping people choose clothes that suit their style and preferences. This study suggests using the KNN Method as a basis for building a more intelligent and personalized clothing recommendation system. To address the growing need for accurate clothing recommendations that match users' preferences, The goal of this research is to create a clothing recommendation system that can help users choose more appropriately because advances in technology have made it possible to gather and examine user data more thoroughly. In this study, the clothing recommendation system was implemented using the KNN Method. We ran simulations by setting the clothing dataset's parameter K value from 3 to 11. The simulation results show that the system's performance reaches its peak at parameter value K=8. We measured the system's accuracy, precision, and recall at this K value in order to assess its performance. The results show that the clothing recommendation system uses the KNN Method. A clothing recommendation system based on the KNN Method with the parameter K=8 has proven successful in classifying clothes with an accuracy of 83,67%.

Keywords: Clothing Recommendation System, K-nearest Neighbor, Classification, Accuracy, System Performance

INTRODUCTION

Technology-based applications have become an important part of everyday life in the modern era, which is influenced by rapid advances in information and communication technology. The rapid development of technology and communication has brought many changes in many fields, one of the aspects that has received a lot of development is in the world of style and clothing. Amidst the many clothing options and ever-changing style trends, consumers often face difficulties in choosing clothes that suit their style and preferences. In situations like this, recommendation systems seem to be able to provide more accurate and personalized clothing to customers. Recommender systems aim to suggest the most suitable items for a particular type of consumer, such as goods or services, by estimating consumer tastes based on information about the item, the consumer, and the interaction between the item and the consumer (Faturrahman et al., 2017) online learning recommendation system needs to have the capability to offer customized recommendations (Murad et al., 2023).

Customers often face difficult problems when they shop for clothes online, namely, many customers feel disappointed when buying clothes because the size does not fit the color of the clothes or the customer's body Their decisions about clothing are influenced by things like color preference, style, brand, body shape, and purchase history. Decision support systems are interactive information platforms that enable data modeling, information, and manipulation to support decision-making (Hayuningtyas, 2019). The problem is how to create a system that can understand these preferences and make suggestions that help customers make better decisions.

This research focuses on creating a clothing recommendation system that uses the K-Nearest Neighbor method. KNN is a method that is often used in the development of recommendation systems by classifying based on the proximity of data to other data (Kafil, 2019). The KNN approach works by comparing the level of similarity between existing data and finding the closest data in terms of attributes or characteristics. Using the KNN method, new objects are grouped or classified according to their attributes and the training examples that already exist. (Herianto & Cahyaningrum, 2020). This approach involves finding and delivering recommendations for an item based on the similarity of information characteristics provided by consumers, such as ratings or ratings (Behera et al., 2021) . The findings in this research are an inventive approach that uses user preference data, the latest fashion

*name of corresponding author

trend data, and the K-Nearest Neighbor algorithm to provide more accurate and personalized clothing recommendations. As well as in processing a very large amount of test data and training data when using KNN is considered to have a low level of complexity (Cholil et al., 2021).

In the previous study (Normah et al., 2022) entitled Recommendation System on Tokopedia Using the K-Nearest Neighbor Algorithm Using the K-Nearest Neighbor method to test the review data on Tokopedia yielded the highest accuracy rate of 73.53% in this study. The findings of the subsequent investigation (Baharuddin et al., 2019) entitled Analysis of the K-Nearest Neighbor Method for Identifying Glass Types in the study analyzed the KNN method to identify glass types, this study obtained the highest accuracy of 64%. With the problems in previous studies that show accuracy results that do not reach the desired level, the researchers conducted this research with a focus on developing and testing a clothing recommendation system that uses the KNN algorithm to have a better level of accuracy of the clothing recommendation system to produce accurate data classification.

LITERATURE REVIEW

The research is entitled Recommendation System on Tokopedia Using the K-Nearest Neighbor Algorithm. Written by Rubangi, 2022. The background in the study aims to develop a recommendation system for E-Commerce using the K-NN method and measure the success of the algorithm in providing the right recommendations to users. This research was conducted because of the difficulty for buyers to find certain products on the internet and to improve buyers' decision-making in choosing products. The test results show that the Content Based method is superior in providing product recommendations on E-Commerce that has sparse rating data. In addition, this research also describes the stages of research conducted, the data sources used, and the design of the K-NN algorithm to classify data (Normah et al., 2022).

Next study is titled Sentiment Analysis on User Reviews Bibit and Bareksa Apps with KNN Algorithm. Written by Adhi Putra, 2021. The background of the problem of this research is to find out how much accuracy is generated from the application of the K-Nearest Neighbors algorithm in classifying sentiment from a user review related to online investment applications, namely reviews related to online investment applications, namely Bibit and Bareksa. Bareksa. This research uses data obtained data obtained through Google Play Store. The results showed that by using a split data model with a ratio of 60:40 on the dataset of and Bareksa apps, the accuracy, precision, and recall values generated were 85.14%, 91.91%, and 76.44%, respectively. recall are 85.14%, 91.91%, and 76.44% for Bibit, while for Bareksa are 81.70%, 87.15%, 75.73%. Conclusion from conclusion of this research is that the K-Nearest Neighbors algorithm can be used to classify sentiment from classify sentiment from user reviews on online investment applications, such as Bibit and Bareksa with a fairly high level of accuracy. In addition, the results of of the sentiment analysis process can be used as a consideration for the people in choosing online investment apps available on Google Play Store. Store. However, the researcher also suggested that further research or development be carried out with other methods and use newer and more datasets than previous research. more than previous research.

One technique for employing supervised learning in decision-making is K-Nearest Neighbor. learning where new input data results are categorized using the data that is closest in value (Yanosma et al., 2017). KNN will evaluate the results using long-term data that has a high number of neighbor in the selected range (Baharuddin et al., 2019).

METHOD

This research was conducted in several stages starting from collecting data. Then the data that has been collected will be pre-processed. After that, the process of designing algorithms and processing data with the Google Collab Tool using the K-Nearest Neighbors method to get results from the classification of data accuracy values. Figure 1 illustrates several stages in creating a clothing recommendation application that uses the K-Nearest Neighbor algorithm.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13377
e-ISSN : 2541-2019
p-ISSN : 2541-044X

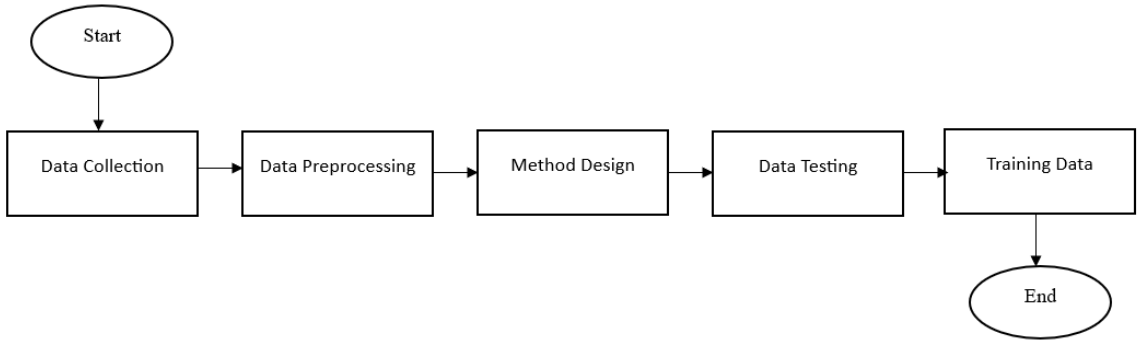

Fig. 1 Research Stages

RESULT

Data collection

The research data source used in the study involves the clothing dataset available from Kaggle.com. The primary records from the gathered dataset include the user's reference to their attire. (Patro et al., 2020), the amount of data contained in the dataset is 44,000 images from many clothing categories such as clothes, pants, accessories, shoes, and many more. Then the data that will be used is only from the category of clothes and pants. In this study, the data used for clothes and pants were 2,500 images. Data collection seen in Figure 2 are as follows.

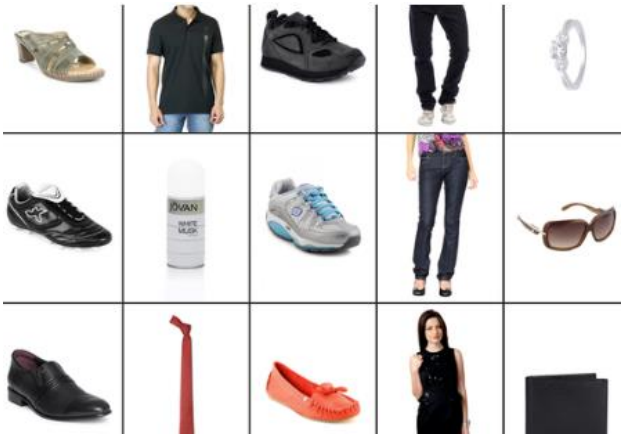

Fig. 2. Data Sources

Data preprocessing

Following the successful acquisition of all data, data preprocessing is the next research step. Pre-processing is the process of transforming the data into a format that will be easier to handle and more efficient. After that, the data moves on to the classification stage, which improves the accuracy of the outcomes. Three processes comprise the processing stage in this study: segmentation, data label attribute reduction, and data filtering. The preprocessing flow is shown in Figure 3.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13377
e-ISSN : 2541-2019
p-ISSN : 2541-044X

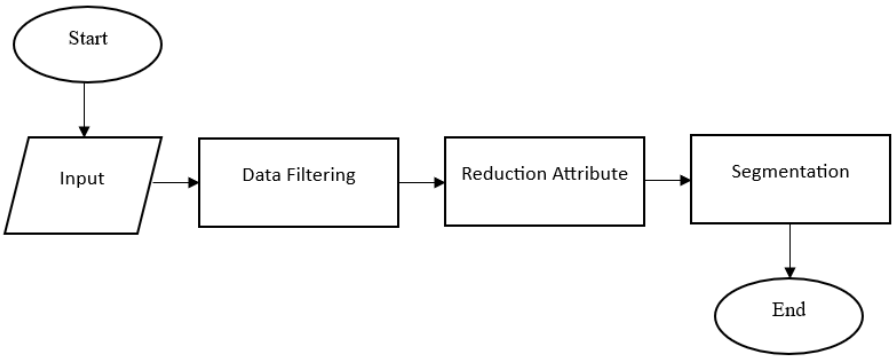

Fig. 3. Preprocessing Flow

Data Filtering

Data filtering is the process of sorting data from a dataset consisting of many categories such as clothes, pants, accessories, shoes, and many more. So, a category reduction is carried out to expedite the testing process by leaving only the shirt and pants categories. As erroneous and abnormal data can substantially impact the model's accuracy, data filtering is a crucial phase in the KNN process. (Somehsaraei et al., 2020), leading to a decline in prediction performance, the results of the data filter in Figure 4 are as follows.

Fig. 4. Clothing Data Filter

Reduction Attribute

Attribute reduction aims to reduce attributes in clothes such as year, season, and product name attributes. At this stage, the attributes that exist in each outfit are gender, color, and usability. Finding the smallest subset of attributes that can meet certain constraints is the goal of attribute reduction. attributes that can meet certain constraints is the goal of attribute reduction. Variations in constraints lead to variations in evaluation criteria, resulting in variants in attribute reduction. Constraint variation gives rise to variation in evaluation criteria, resulting in variance in attribute reduction. Despite the fact that there have been numerous approaches to attribute reduction, (Chen et al., 2020).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13377
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Segmentation

Segmentation is done because some of the data collected still has photos of models wearing their clothes, so segmentation separates the clothes and pants from the model's body. Then the results in all data will display images of clothing. Two common constraints of image segmentation data sets are weak annotations, for which the training data has only sparse, noisy, or image-level annotations, and rare annotations, for which there are only sparse annotated data available for training. (Tajbakhsh et al., 2020). The results of clothing segmentation from the model's body can be seen in Figure 5.

Fig. 5. Clothing Segmentation Results

Design Methods

One method that is often used to classify data is the KNN Method. The KNN method is a technique for classifying data that relies on category labels in training documents, like test documents, rather than developing explicit and declarative category representations from basic examples (Putra & Juanita, 2021). Due to the fact that the K-Nearest neighbor algorithm groups objects according to their characteristics and training sets (Normah et al., 2022). The KNN method used in this study is divided into 2 parts, namely training data and test data so that it is based on predictions. There are steps from the KNN method, namely determining the parameter k, calculating the k-training data that is closest to the data in the test, sorting the distance parameters based on the smallest value, and determining the test data group based on the majority label k. The distance calculation that is often used in the KNN method is the Euclidean distance calculation (Normah et al., 2022). The value of k in KNN is a variable number of nearest neighbors that will be taken for the classification process (Yanosma et al., 2017). While a high value of k will lessen the impact of noise on classification, it will also cause more blurring of the boundaries between each classification. (Baharuddin et al., 2019). The following is the concept of the KNN method shown in Figure 6.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13377
e-ISSN : 2541-2019
p-ISSN : 2541-044X

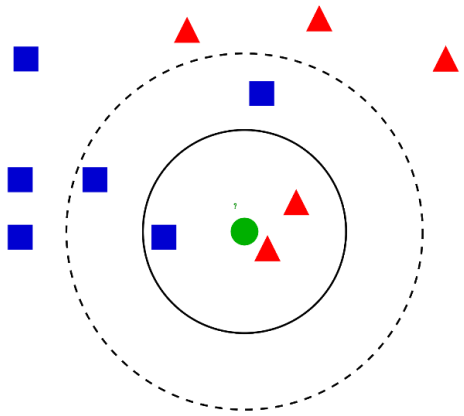

Fig. 6. KNN Method Concept

Training Data

In this research, training is done using data on the dataset to train the K-Nearest Neighbors method. Training datasets are used from clothing data that has been pre-processed using KNN before. The data that will be used for training is divided into two, namely 70% for training datasets and 30% for testing datasets (Rubangi & Rianto, 2022). In the training stage, a total of 1750 data, or 30% of the dataset is used.

Data Testing

This study employs the K Nearest Neighbors method for testing, which is carried out on datasets to measure the accuracy, precision, and recall of the KNN method. The testing phase describes how to use the KNN method to determine the accuracy value, and the performance obtained from the KNN method will demonstrate the accuracy of the method model. The testing is carried out on 30% or 750 data from the entire dataset. Table 1 below shows the source code for the KNN method implementation.

Table 1
Source Code of KNN Method Implementation

Desc	Source Code
Load	dataset = pd.read_csv('baju_transparan_full.csv')
Split	x = dataset.iloc[:, 0:7]) y = dataset.iloc[:, 7] x_train, x_test, y_train, y_test = train_test_split(x, y, random_state=0, test_size=0.3)
Train	classifier = KNeighborsClassifier (n_neighbors=3 , metric='euclidean') classifier.fit (x_train, y_train)
Test	y_pred = classifier.predict (x_test)
Result	accuracy = accuracy_score (y_test, y_pred) precision = precision_score (y_test, y_pred, average=None) recall = recall_score (y_test, y_pred, average=None) print("Accuracy:", accuracy) print("Precision:", precision) print("Recall:", recall)

In this chapter, the data used for classification has successfully evaluated a clothing recommendation system based on the KNN algortihm. This framework undergoes testing to evaluate its performance in providing accurate clothing recommendations according to user preferences using the Google Collab tool.

According to the outcome of the classification of clothing using the KNN algorithm with testing data totaling 1750 data, testing methods in this study were carried out with various values of the KNN method with parameter values K = 3 to K = 11 as shown in table 2 below.

Table 2

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13377
e-ISSN : 2541-2019
p-ISSN : 2541-044X

KNN Method Testing

No	K	Accuracy	Precision	Recall
1	3	75,49%	82,36%	87,16%
2	4	80,61%	82,92%	93,76%
3	5	81,27%	82,31%	95,92%
4	6	81,83%	84,39%	96,77%
5	7	82,34%	83,46%	96,83%
6	8	83,67%	83,82%	99,58%
7	9	83,28%	84,28%	99,34%
8	10	83,06%	83,58%	98,29%
9	11	82,93%	84,19%	97,35%

The test results for each parameter value K3 to K11 show the highest accuracy value, which is 83.67% for parameter k = 8, and the lowest accuracy value, which is 75.49% at k = 3. With the accuracy results obtained, it is known that the high level of closeness between the values obtained and the actual value. Testing with parameter k = 11 shows the highest precision value of 84.19%, and testing with parameter k = 5 shows the lowest precision value of 82.31%. This result proves that the k11 experiment has a high match of clothing data in retrieving the required information. Testing with parameter k = 8 shows the highest recall value of 99.58%, and testing with parameter K = 3 shows the lowest recall value of 87.16% so the experiment on the parameter value k8 has an almost perfect system success rate in finding information.

The best value is obtained when the KNN method is tested for its performance on the classification of clothing recommendations, with a parameter value of K = 8, accuracy of 83.67%, precision of 83.82%, and recall of 99.34%. A graph similar to the one in Figure 8 is produced after the clothing recommendation system has run all of its classification tests. The overall performance results of parameter K can be seen in the graph in Figure 7.

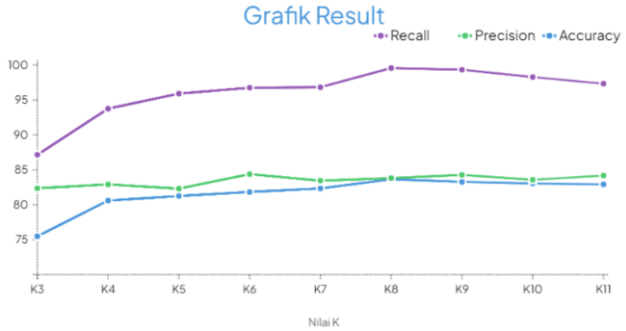

Fig. 7. K-Nearest Neighbor Performance

With an 83.67% classification performance for clothing recommendations, the KNN method is widely used in recommendation systems based on this data. This shows that the KNN method can provide accurate results. presented insights from a study by Rubangi (2022) on a recommendation system for E-Commerce, which employed the Content-Based method. The research indicated that the KNN method, as implemented in the clothing recommendation system, outperformed the Content-Based method, particularly in scenarios with sparse rating data. This underscores the effectiveness of the KNN algorithm in providing relevant product recommendations.

DISCUSSIONS

In the study, the parameter K was varied from 3 to 11 while simulating the K-Nearest Neighbor (KNN) algorithm on a clothing dataset. Based on the analysis, the accuracy level of the system was 83.67%, and the system's performance peaked at K = 8. This shows that the system for recommending clothes has a high degree of accuracy when it comes to categorizing items according to user preferences. Although the KNN method proved successful in the clothing recommendation system, there are a few limitations that should be considered. First off, there may be differences in the system's efficacy amongst various clothing datasets, indicating the need for additional study to evaluate the system's performance in a range of datasets. Furthermore, the study did not delve into specific algorithms or examine potential synergies among them, even though it suggested combining different algorithmic methods for clothing recommendation systems, thereby identifying opportunities for future research. Finally, the study mainly concentrated on algorithmic performance metrics; user feedback and satisfaction metrics were not included, which could have improved the evaluation's comprehensiveness. By addressing these issues, a more resilient and flexible system for recommending clothes could be created that works with a variety of datasets,

*name of corresponding author

takes advantage of algorithmic synergies, and integrates user-centric viewpoints for a more comprehensive assessment.

CONCLUSION

The study's findings, which included a clothing recommendation system utilizing the K-Nearest Neighbor method, led the authors to make several deductions. One of these was to simulate KNN on the clothing dataset using values ranging from K = 3 to K = 11, the highest performance value is obtained at the parameter value K = 8, where the accuracy level reaches 83.67%, this shows the high level of closeness between the data obtained and the actual data, precision 83.82% aims to see a high match of clothing data in retrieving the information needed, recall 99.34%. The clothing recommendation system uses the KNN method to classify clothes that can be implemented in clothing product applications to be recommended according to user preferences, so further research can be done by combining various algorithm methods for clothing recommendation systems to make better clothing product recommendations.

REFERENCES

Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, *11*(3), 269–274. https://doi.org/10.33096/ilkom.v11i3.489.269-274

Behera, D. K., Das, M., Swetanisha, S., & Sethy, P. K. (2021). Hybrid Model For Movie Recommendation System Using Content K-Nearest Neighbors And Restricted Boltzmann Machine. *Indonesian Journal of Electrical Engineering and Computer Science*, *23*(1), 1–8. https://doi.org/10.11591/ijeecs.v23.i1.ppab-cd

Chen, Y., Liu, K., Song, J., Fujita, H., Yang, X., & Qian, Y. (2020). Attribute group for attribute reduction. *Information Sciences*, *535*, 64–80. https://doi.org/10.1016/J.INS.2020.05.010

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(2), 118–127. https://doi.org/10.31294/ijcit.v6i2.10438

Faturrahman, M. I., Nurjanah, D., & Rismala, R. (2017). *Sistem Rekomendasi pada Buku dengan Menggunakan Metode Trust-Aware Recommendation*. https://doi.org/https://doi.org/10.33369/rekursif.v4i2.894

Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita. *JURNAL INFORMATIKA*, *6*(1), 18–22. https://doi.org/https://doi.org/10.31294/ji.v6i1.4685

Herianto, & Cahyaningrum, N. (2020). Implementasi K-NN dan AHP Untuk Rekomendasi Model Pakaian Toko Online. *Jurnal Sains & Teknologi Fakultas Teknik Universitas Darma Persada*, *10*(2).

Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *3*(2), 59–66. https://doi.org/10.36040/jati.v3i2.860

Murad, D. F., Murad, S. A., & Irsan, M. (2023). the Effect of Contextual Information As an Additional Feature in the Recommendation System. *Journal of Educators Online*, *20*(1). https://doi.org/10.9743/JEO.2023.20.1.10

Normah, Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer AMIK BSI*, *8*(2), 174–180. https://doi.org/10.31294/jtk.v4i2

Patro, S. G. K., Mishra, B. K., Panda, S. K., Kumar, R., Long, H. V., Taniar, D., & Priyadarshini, I. (2020). A Hybrid Action-Related K-Nearest Neighbour (HAR-KNN) Approach for Recommendation Systems. *IEEE Access*, *8*, 90978–90991. https://doi.org/10.1109/ACCESS.2020.2994056

Putra, aluisius D. A., & Juanita, S. (2021). Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *8*(2), 636–646. https://doi.org/10.35957/jatisi.v8i2.962

Somehsaraei, H. N., Ghosh, S., Maity, S., Pramanik, P., De, S., & Assadi, M. (2020). Automated data filtering approach for ANN modeling of distributed energy systems: Exploring the application of machine learning. *Energies*, *13*(14). https://doi.org/10.3390/en13143750

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, *63*, 101693. https://doi.org/10.1016/J.MEDIA.2020.101693

Yanosma, D., T, A. J., & Anggriani, K. (2017). Implementasi Metode K-Nearest Neighbor (KNN) Dan Simple Additive Weighting (SAW) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka. *Pseudocode*, *3*(2), 98–112. https://doi.org/10.33369/pseudocode.3.2.98-112

*name of corresponding author