Institute of Research and Publication Indonesia (IRPI)
Public Research Journal of Engineering, Data Technology and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/predatecs
Vol. 2 Iss. 1 July 2024, pp: 18-25
ISSN(P): 3024-921X | ISSN(E): 3024-8043

Classifications of Offline Shopping Trends and Patterns with Machine Learning Algorithms

Muta'alimah[1*], Cindy Kirana Zarry[2], Atha Kurniawan[3], Hauriya Hasysya[4], Muhammad Farhan Firas[5], Nurin Nadhirah[6]

[1,2,3]Department of Information Systems, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[4]Department of Human Resource Development, Faculty of Social Science and Humanities,
University Technology Malaysia, Malaysia
[5]Department of Theology and Philosophy, Faculty of Islamic Studies,
University Kebangsaan Malaysia, Malaysia
[6]Department of Malay Studies, Faculty of Malay Studies, University Malaya, Malaysia

E-Mail: [1]12150323778@students.uin-suska.ac.id, [2]12150322263@students.uin-suska.ac.id,
[3]12150311424@students.uin-suska.ac.id, [4]hauriya@graduate.utm.my,
[5]127340@siswa.ukm.edu.my, [6]23004004@siswa.um.edu.my

Received Dec 27th 2023; Revised Mar 11th 2024; Accepted Apr 18th 2024
Corresponding Author: Muta'alimah

Abstract

Advancements in technology have made online shopping popular among many. However, the use of offline marketing models is still considered a profitable and important way of business development. This can be seen in the 2022 Association of Retail Entrepreneurs of Indonesia (APRINDO), which states that 60% of Indonesians shop offline, and in 2023, more than 75% of continental European consumers will prefer to shop offline. This is because many benefits can be achieved through offline marketing that cannot be obtained from online marketing. Therefore, classification of patterns and trends is performed to compare the results of the algorithms under study. Furthermore, this research was conducted to help offline retailers understand consumption patterns and trends that affect purchases. The algorithms analyzed in this study are K-Nearest Neighbor (K-NN), Naive Bayes, and Artificial Neural Network (ANN). As a result, the ANN algorithm obtained the highest confusion matrix results with an Accuracy value of 96.38%, Precision of 100.00%, and Recall of 100.00%. Meanwhile, when the Naive Bayes algorithm was used, the lowest Accuracy value was 57.39%, the Precision value was 57.86%, and when the K-NN algorithm was used, the Recall value was as low as 92.00%. These results indicate that the ANN algorithm is better at classifying offline shopping image data than the K-NN and Naive Bayes algorithms.

Keyword: Artificial Neural Network, Classification, K-Nearest Neighbor, Naive Bayes, Offline Shopping

1. INTRODUCTION

Technological development has changed various aspects of people and life, which has affected daily life. One of the habits of people to fulfill their needs is shopping offline or online. Although online shopping has been in high demand among the general public, especially the younger generation, in recent years, the use of offline shopping models is still considered useful and important for business development [1]. In 2022, the Association of Indonesian Retail Entrepreneurs (APRINDO) reported that 60% of Indonesians will shop offline [2], and in 2023, more 75% of continental European consumers will prefer to shop offline [3], [4].

Offline shopping is a functional activity that is done completely in person, it is a meeting between merchants and buyers where buying and selling transactions take place [5]. Offline shopping in malls or stores has long been an activity that dominates customers' lifestyles. This experience is not only about getting the desired product but also about creating fun things, one of which is the opportunity to negotiate with the seller [6].

Offline retailers face significant challenges in the face of changing consumer trends. The growing trend of shopping is forcing retailers to find ways to delight consumers when they shop offline, with rapid competition among offline retailers and a shift in consumer interest to online stores that offer convenience and efficiency [7]. Over the past decade, the number of online shopping portals, the range of products available online, and high-speed internet access have continued to grow [8]. Therefore, offline retailers must

DOI: https://doi.org/10.57152/predatecs.v2i1.1099

be able to understand consumer buying habits and trends by analyzing product customizations, marketing strategies, and overall customer experience [9].

Examining patterns and trends in traditional marketing is crucial for obtaining precise outcomes [10]. Numerous algorithms are at one's disposal, and in this investigation, three were scrutinized: K-Nearest Neighbor (K-NN), Naive Bayes, and Artificial Neural Network (ANN) [11], [12]. Some previous research related to analysis and case studies namely, [13] in this study using the integration of iForest Outlier Detection, ADASYN data balancing, and Multilayer Perceptron (MLP) with an accuracy value of 97.778% these results can help store owners in understanding customer preferences in offline shopping. Research [14] shows the Artificial Neural Network (ANN) algorithm produces the highest accuracy value of 88% compared to the SVM and LR algorithms in processing customer perception data on online and offline shopping. Research [15] also shows the accuracy value of the K-NN algorithm of 80.16% which is included in the good accuracy category in processing offline and online shopping trip data. Therefore, this research has a more specific focus on classifying offline shopping trends and patterns using three algorithms that have not been explored by previous researchers.

Referring to prior studies, it's established that K-NN, Naive Bayes, and ANN exhibit notable proficiency in addressing the classification challenges relevant to the discussed subject. Notably, there's a dearth of research concerning the classification of offline shopping trends and patterns utilizing these algorithms. Hence, the research is titled "Classification of Offline Shopping Trends and Patterns with Machine Learning Algorithms." This endeavor seeks to assist offline merchants in innovating to sustain appeal, align with diverse product types, and meet consumer preferences and satisfaction [16]. Additionally, it aims to discern the highest confusion matrix results among the three classification algorithms [17].

2. MATERIAL AND METHOD

By performing the steps in the figure above, the data becomes more structured, clean, and ready to use, improving the accuracy and performance of the model, and avoiding potential distortions or errors in the analysis results [21]. This research uses K-NN, Naive Bayes, and ANN classification algorithms. Each step in Figure 1. is performed in each research process.

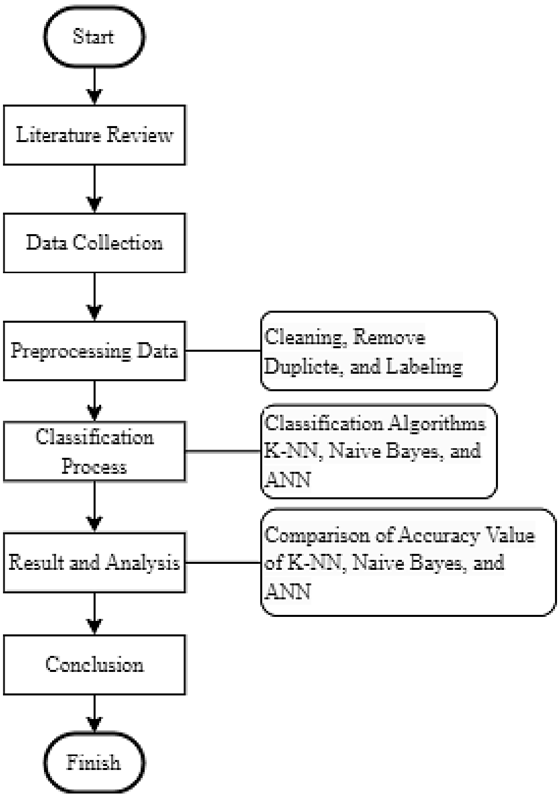

Figure 1. Research Methodology

The stages of this research start from conducting a literature review, activities carried out with data collection techniques from various reference sources to get an overview of scientific publications, websites, and other pertinent information about the subject under investigation are the sources used in this research. In the data collection stage, data is obtained from the Kaggle web in the form of a Customer Shopping Trends dataset, which is then preprocessed with the initial data amount of 3900 to 3808. After that, the application of classification algorithms is carried out, namely K-NN, Naive Bayes, and ANN. At the analysis and results stage, it aims to determine the accuracy of the classification results that have been carried out. The last stage will conclude which algorithm produces the highest accuracy value for the customer shopping trend dataset.

2.1. Data Mining

The technique of extracting meaningful patterns from massive volumes of data is called data mining. Areas of expertise include the areas of text analysis, classification, clustering, classification, visualization, data mining, and machine learning. The target of word mining, which has been widely studied, is web applications and the World Wide Web [18]. In general, many text documents can be processed using text mining algorithms. Some areas of web applications that are studied are news websites, social networks, e-commerce, etc [19]. It cannot be denied that physical stores are still in high demand by various audiences.

2.2. Classification

The classification step is the step in which the learning results are classified based on the defined subclasses. Word classification is the process of determining whether a text belongs to a text class. As the number of documents increases, you will need tools that perform automatic classification. Automatic sorting is a sorting process done by a computer [22].

According to the number of classes, there are two types of classification: two-class classification and multi-class classification. Binary classification is the classification of an object into one of two classes. Multiclass classification is the classification of objects into more than one class. Training data and test data are the two categories into which data items in text categorization are separated [23]. Text that has been manually classified is utilized as training data, and text that has not been classified is used as test data. Finding the class characteristics based on the training data and using them directly to test the data is the aim of classifying this data [18].

2.3. K-Nearest Neighbor

K-NN is the simplest and most widely used algorithm for solving classification problems. K-NN is used to collect data and vocabulary. Object classification method in very close data [24].

1. Suppose there are j training units (C1, C2,...,Cj) and the sum of all training samples is N. The two units produce an m-dimensional vector, a vector field that represents the resulting classification. Suppose there are j training units (C1, C2,...,Cj) in the region where the sum of all training samples is N. The two units result in an m-dimensional vector, a field vector, and an output classification. displayed in the field.
2. Sort the sample X according to the fitter vector in the corresponding binding (X1, X2,... Xm).
3. Determine how similar every training sample is to The cosine similarity formula can be used to determine the similarity.
4. Utilize the following formula to determine X's tendency for each category [18]..

$$P(X, Cj) = \sum Cosine(x, di\ ).\ y(yid\ , Cj) \quad (1)$$

Where: y(yi, Cj) = attribute category function

5. When P(X, Cj) is calculated, the largest number determines which component is X.

2.4. Naive Bayes

When compared to other classification models, the Naive Bayes Algorithm is a classifier algorithm that can perform exceptionally well. Nurhachita and Negara presented in a 2020 journal article titled "A Comparison Between Naive Bayes and K-Means Clustering Algorithms for the Application of Data Mining for New Student Recognition." This demonstrates that compared to other classification models, the Naive Bayes Classifier has higher accuracy [25].

Thomas Bayes developed the probabilistic and statistical techniques used in the Naive Bayes algorithm to calculate a future course based on past performance. It is advantageous to use the Naive Bayes algorithm. This indicates that the method is more efficient because it uses less training data to calculate the parameters or metrics in the classification process. This is because the independent variable uses the variance of the class variable used in the classification decision, not the complete matrix of covariance [14]. Naive Bayes Method Equation The equation 2 of Bayes' theorem is [26].

$$P(H|X)^1 = \frac{P(H|X).\ P(H)}{P(X)} \quad (2)$$

Where:

X : Data with unknown class.
H : Hypothesis that the data belongs to a specific class.

P(H|X) : Probability of hypothesis H based on condition X (a posteriori probability).
P(H) : Probability of hypothesis H (prior probability).
P(X|H) : Probability of X based on the condition in hypothesis H
P(X) : Probability of X

To explain the Naive Bayes method, it is important to understand that there are many indicators for the classification process to determine which class belongs to the analyzed sample [26]. Therefore, the naive Bayes method above is formulated as equation 3.

$$P(C|F1 \ldots Fn)^1 = \frac{P(C)P(F1 \ldots Fn|C)}{P(F1 \ldots Fn)} \tag{3}$$

2.5. Artificial Neural Network (ANN)

Everything written in the reference must be referenced in the text. Using the R package and “Neural Networks”, we built a neural network-based model (ANN) to analyze the top five genes identified based on radio frequency [27]. The expression levels of the top five genes can be converted into “tagged genes”, and the expression levels of a particular gene are compared with the average expression across all gene samples. Points assigned to the mean were assigned a value of 1 if they were above the median and assigned a value of 0. If the expression levels of down-regulated genes were higher than the median, they were given a value of 1, and otherwise, they were given a value of 0. Next, to obtain the computed genetic weights of the "genetic score," a genetic score table was made and the hidden layer of the ANN was set to 5. Ultimately, a model for neural network analysis was created and assessed using the collection [28]. Figure 2., represents an ANN framework with connections known as synapses. This synapse has weights that can be adjusted during learning, its function is to regulate the extent to which the contribution of each input affects the final result.

2.6. Preprocessing

The steps of data mining start with pre-processing. Preprocessing prepares text data as words that can be processed later [20]. Text preprocessing affects the success of the text mining algorithm used. The above steps can be seen in Figure 2.

Figure 2. Preprocessing Steps

3. RESULTS AND DISCUSSION

3.1. Data Collection

The study utilizes a customer experience database sourced from the Kaggle website. This dataset encompasses a range of factors associated with consumer buying preferences, comprising 19 attributes such as customer ID, age, gender, purchased item details, category, transaction amount (in USD), location, size, color, time of purchase, subscription status, payment method, promo code usage, and previous purchase history. Noteworthy aspects highlighted in the dataset include considerations for obtaining the best price, making frequent purchases, and garnering high reviews. Table 1 is dedicated to presenting insights into consumer buying habits.

Table 1. Data Collection Customer Shopping Trends

Customer ID	Age	Gender	Item Purchased	...	Review Rating
1	55	Male	Blouse	...	3,1
2	19	Male	Sweater	...	3,1
3	30	Male	Jeans	...	3,1
4	21	Male	Sandals	...	3,5
5	45	Male	Blouse	...	2,7
...	...	...	...	...	...
3900	52	Female	Handbag	...	4,0

3.2. Preprocessing Data

This process involves cleaning up unnecessary data such as empty data and removing duplicate data. The original data of data 3900 becomes data 3808 after data cleaning. This pre-processing provides labels for the evaluation decision classes, which can be seen in Table 2. The results of the data pre-processing can be seen in Table 3.

Table 2. Labeling

Customer ID	Age	Gender	Item Purchased	…	Review Rating	Class
1	55	1	2	…	3,1	Satisfied
2	19	1	23	…	3,1	Satisfied
3	30	1	11	…	3,1	Satisfied
4	21	1	14	…	3,5	Satisfied
5	45	1	2	…	2,7	Unsatisfied
…	…	…	…	…	…	…
3808	38	0	2	…	4,5	Very Satisfied

Table 3. Preprocessing Data

Customer ID	Age	Gender	Item Purchased	…	Class
1	55	1	2	…	0
2	19	1	23	…	0
3	30	1	11	…	0
4	21	1	14	…	0
5	45	1	2	…	2
…	…	…	…	…	…
3808	38	0	2	…	1

3.3. Implementation of K-NN, Naive Bayes, and ANN Classification Algorithms

The application of the K-NN classification algorithm was performed using the Rapidminer tool and the dataset described above. The class attributes of this data set are "Unsatisfied", "Satisfied" and "Very Satisfied". It indicates customer satisfaction with offline purchases. The results of this step are described in terms of precision, accuracy, and recall values. The first application of the classification algorithm using the K-NN algorithm can be seen in Table 4.

Table 4. Confusion Matrix K-NN

Accuracy: 71.88%

	True Satisfied	True Unsatisfied	True Very Satisfied	Class Precision
Pred. Satisfied	428	60	117	70.74%
Pred. Unsatisfied	38	286	59	74.67%
Pred. Very Satisfied	474	323	2023	71.74%
Class Recall	45.53%	42.75%	92.00%	

The table presented above depicts the confusion matrix generated by the K-NN algorithm, providing insights into the model's efficacy in predicting classes based on true positive, true negative, false positive, and false negative values. The precision metric indicates that "Pred. Unsatisfied" achieves the highest value at 74.67%. In terms of class recall, which gauges the model's ability to identify all positive instances, the class "True Very Satisfied" stands out with the highest recall value of 92.00% compared to other classes. Evaluating the overall performance of the K-NN algorithm, the accuracy metric registers at 71.88%, offering a comprehensive view of the model's effectiveness. Additionally, the Naive Bayes classification algorithm is applied and its outcomes are presented in Table 5.

Table 5. Confusion Matrix Naive Bayes

Accuracy: 57.39%

	True Satisfied	True Unsatisfied	True Very Satisfied	Class Precision
Pred. Satisfied	7	3	11	33.33%
Pred. Unsatisfied	0	1	1	50.00%
Pred. Very Satisfied	275	197	648	57.86%
Class Recall	2.48%	0.50%	98.18%	

The highest precision value is "Pred. Satisfied Once" at 57.86%. "True Satisfied Once" got the highest value for class recall of 98.18%. The accuracy value which is an important parameter for the entire model formed in the application of the Naive Bayes algorithm gets 57.39%. The application of the third algorithm, ANN, is seen in Table 6.

The highest precision values on the ANN confusion matrix are "Pred. Satisfied" and "Pred. Unsatisfied" with a value of 100.00%. The highest recall value of 100.00% is obtained by "True Satisfied" and "True Satisfied". The accuracy value which is an important parameter for the entire model formed in the application of the ANN algorithm is 96.38%. In addition, the ANN application process gets an improved neural network output which can be seen in Figure 4.

Table 6. Confusion Matrix ANN

Accuracy: 96.38%

	True Satisfied	True Unsatisfied	True Very Satisfied	Class Precision
Pred. Satisfied	940	0	0	100.00%
Pred. Unsatisfied	0	531	0	100.00%
Pred. Very Satisfied	0	138	2199	94.09%
Class Recall	100.00%	79.37%	100.00%	

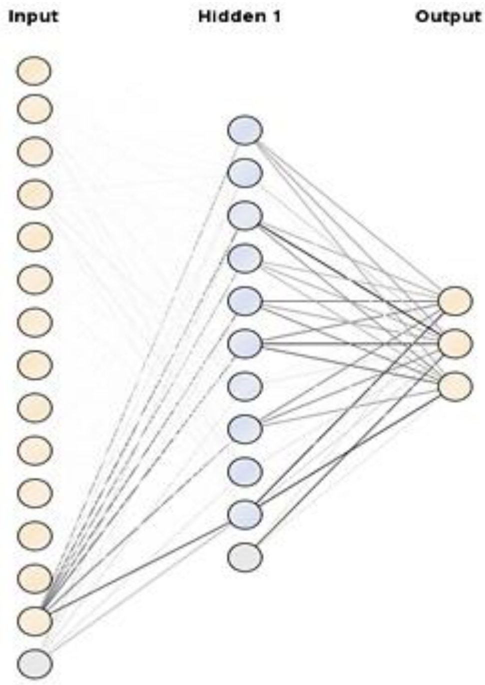

Figure 4. Output Improve ANN

The evaluation of three data classification algorithms K-NN, Naive Bayes, and ANN based on the confusion matrix results reveals insights into customer perceptions in offline shopping across three classes. The analysis involves assessing accuracy, precision, and recall outcomes. Figure 5 presents a comparative diagram illustrating the values within the confusion matrix for the K-NN, Naive Bayes, and ANN algorithms.

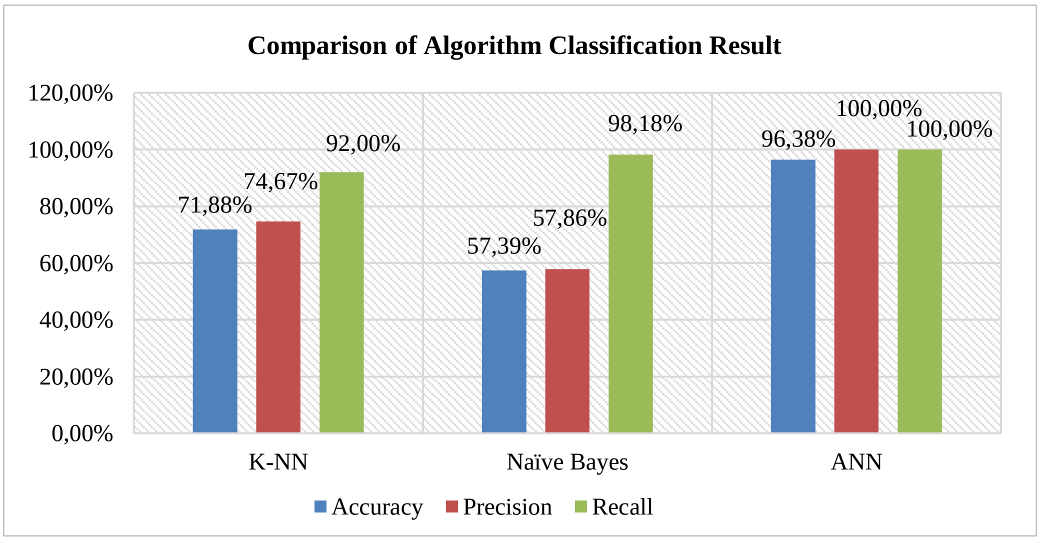

Figure 5. Comparison of Algorithm Classification Results

The depicted graph illustrates the classification outcomes of three algorithms K-NN, Naive Bayes, and ANN based on their respective performance metrics. The K-NN algorithm achieved an Accuracy value of 71.88%, Precision of 74.67%, and Recall of 92.00%. The Naive Bayes algorithm showed an Accuracy value of 57.39%, Precision of 57.86%, and Recall of 98.18%. Meanwhile, the ANN algorithm exhibited an Accuracy value of 96.38%, Precision of 100.00%, and Recall of 100.00%. Comparing these three algorithms, the ANN classification algorithm stands out with the highest values in terms of accuracy, precision, and recall. This enables a comparative assessment of offline marketing methods based on customer satisfaction. Notably, the Naive Bayes algorithm, while having the lowest accuracy and precision values, boasts a Recall value 6.18% higher than the Recall value of K-NN.

4. CONCLUSION

Based on the analysis conducted on a dataset consisting of 3900 offline shopping trends, the remaining 3808 data points were processed using the RapidMiner tool, implementing K-NN, Naive Bayes, and ANN algorithms. The goal was to determine the most favorable confusion matrix results in terms of Accuracy, Precision, and Recall. The findings reveal that the ANN algorithm outperformed the others, yielding the highest confusion matrix results with an accuracy value of 96.38%, a precision value of 100.00%, and a recall value of 100.00%. On the contrary, the Naive Bayes algorithm exhibited the lowest accuracy at 57.39%, precision at 57.86%, and the K-NN algorithm showed the lowest recall at 92.00%. These results indicate that the ANN algorithm excels in classifying offline shopping trend data, demonstrating superior effectiveness compared to K-NN and Naive Bayes algorithms.

REFERENCES

[1] J. Ram and S. Sun, "Business benefits of online-to-offline ecommerce: A theory driven perspective," J. Innov. Econ. Manag., vol. 33, no. 3, pp. 135–162, 2020, doi: 10.3917/jie.033.0135.

[2] F. Rahadi, "Aprindo Sebut 60 Persen Masyarakat Kembali Belanja Luring, Rambla Diluncurkan," REPUBLIKA.CO.ID, 2023. https://rejogja.republika.co.id/berita/rvjunc291/aprindo-sebut-60-persen-masyarakat-kembali-belanja-luring-rambla-diluncurkan (accessed Dec. 04, 2023).

[3] "European: consumers shopping online vs offline Weekly 2023," intellesvizz.net, 2023. https://intellesvizz.net/?_=%2Fstatistics%2F1257230%2Feuropean-consumers-that-shop-online-and-offline-each-week%2F%23KJWqMdlUlBn8PPpbQwnhk4LmbIAuGFCs (accessed Dec. 04, 2023).

[4] D. Tighe, "Percentage of consumers that shop online and offline on a weekly basis in the U.S., UK, and Australia in 2023," statista.com, 2023. https://www.statista.com/statistics/1257243/consumers-that-shop-online-and-offline-each-week/ (accessed Dec. 04, 2023).

[5] V. Khangembam, "Consumers choice of small independent specialty stores in shopping centers during weekday extended trading hours: A qualitative study," Cogent Bus. Manag., vol. 10, no. 1, 2023, doi: 10.1080/23311975.2023.2185071.

[6] Y. Chen, C. Q. L. Xue, and C. Sun, "American shopping malls in China: a mosaic analysis of databases," J. Asian Archit. Build. Eng., vol. 22, no. 6, pp. 3224–3243, 2023, doi: 10.1080/13467581.2023.2182639.

[7] H. Sachdev and M. H. Sauber, "Employee–customer identification: Effect on Chinese online shopping experience, trust, and loyalty," Cogent Bus. Manag., vol. 10, no. 3, 2023, doi: 10.1080/23311975.2023.2275369.

[8] B. N. Swar and R. Panda, "Online Retail Service Quality: Scale Development and Validation," Vis. J. Bus. Perspect., vol. 27, no. 3, pp. 376–385, Jun. 2023, doi: 10.1177/09722629211011282.

[9] M. D. Vo and S. Van Nguyen, "Enhancing store brand equity through relationship quality in the retailing industry: evidence from Vietnam," Cogent Bus. Manag., vol. 9, no. 1, 2022, doi: 10.1080/23311975.2022.2149150.

[10] N. Suarna, Y. A. Wijaya, Mulyawan, T. Hartati, and T. Suprapti, "Comparison K-Medoids Algorithm and K-Means Algorithm for Clustering Fish Cooking Menu from Fish Dataset," IOP Conf. Ser. Mater. Sci. Eng., vol. 1088, no. 1, p. 012034, 2021, doi: 10.1088/1757-899x/1088/1/012034.

[11] and R. B. W. A. Z. Leon, V. A. I. Huvenne, N. M. A. Benoist, M. Ferguson, B. J. Bett, "Assessing the repeatability of automated seafloor classification algorithms, with application in marine protected area monitoring," Remote Sens, vol. 12, no. 10, 2020, doi: 10.3390/rs12101572.

[12] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, "Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors," J. Mater. Chem. C, vol. 3, pp. 10715–10722, 2015, doi: 10.1039/b000000x.

[13] G. Alfian et al., "Customer Shopping Behavior Analysis Using RFID and Machine Learning Models," Inf., vol. 14, no. 10, 2023, doi: 10.3390/info14100551.

[14] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," Int. J. Comput. Appl., vol. 17, no. 8, pp. 43–48, 2011, doi: 10.5120/2237-2860.

[15] M. Dasoomi, A. Naderan, and T. Allahviranloo, "Predict the Shopping Trip (Online and Offline) Using a Combination of a Gray Wolf Optimization Algorithm (GWO) and a Deep Convolutional Neural Network: A Case Study of Tehran, Iran," 2023, doi: 10.20944/preprints202309.0696.v1.

[16] A. M. Collins and G. Maglaras, “Smart shopper feelings in the case of store brands: the role of human capital as a key antecedent and the implications for store loyalty,” Int. Rev. Retail. Distrib. Consum. Res., vol. 00, no. 00, pp. 1–21, 2023, doi: 10.1080/09593969.2023.2200965.

[17] A. A. Salih and A. M. Abdulazeez, “Evaluation of Classification Algorithms for Intrusion Detection System: A Review,” J. Soft Comput. Data Min., vol. 02, no. 01, pp. 31–40, 2021, doi: 10.30880/jscdm.2021.02.01.004.

[18] D. Sebastian, “Implementasi Algoritma K-Nearest Neighbor untuk Melakukan Klasifikasi Produk dari beberapa E-marketplace,” J. Tek. Inform. dan Sist. Inf., vol. 5, no. 1, pp. 51–61, 2019, doi: 10.28932/jutisi.v5i1.1581.

[19] W. T. Wu et al., “Data mining in clinical big data: the frequently used databases, steps, and methodological models,” Mil. Med. Res., vol. 8, no. 1, pp. 1–12, 2021, doi: 10.1186/s40779-021-00338-z.

[20] P. F. Pratama, D. Rahmadani, R. S. Nahampun, D. Harmutika, A. Rahmadeyan, and M. F. Evizal, “Random Forest Optimization Using Particle Swarm Optimization for Diabetes Classification,” Public Res. J. Eng. Data Technol. Comput. Sci., vol. 1, no. 1, pp. 41–46, 2023, doi: 10.57152/predatecs.v1i1.809.

[21] M. Müller, L. Longard, and J. Metternich, “Comparison of preprocessing approaches for text data in digital shop floor management systems,” Procedia CIRP, vol. 107, no. May, pp. 179–184, 2022, doi: 10.1016/j.procir.2022.04.030.

[22] B. Çiğşar and D. Ünal, “Comparison of Data Mining Classification Algorithms Determining the Default Risk,” Sci. Program., vol. 2019, 2019, doi: 10.1155/2019/8706505.

[23] W. Putri, D. Hastari, K. U. Faizah, S. Rohimah, and D. Safira, “Implementation of Naïve Bayes Classifier for Classifying Alzheimer’s Disease Using the K-Means Clustering Data Sharing Technique,” Public Res. J. Eng. Data Technol. Comput. Sci., vol. 1, no. 1, pp. 47–54, 2023, doi: 10.57152/predatecs.v1i1.803.

[24] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” J. Teknoinfo, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

[25] H. Chen, S. Hu, R. Hua, and X. Zhao, “Improved naive Bayes classification algorithm for traffic risk management,” EURASIP J. Adv. Signal Process., vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00742-6.

[26] Rayuwati, Husna Gemasih, and Irma Nizar, “IMPLEMENTASI AlGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID,” Jural Ris. Rumpun Ilmu Tek., vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.

[27] R. Jader and S. Aminifar, “Rasool Jader/ Fast and Accurate Artificial Neural Network Model for Diabetes Recognition,” no. August, 2022, doi: 10.14704/nq.2022.20.10.NQ55189.

[28] Y. F. Yumi Wu, QiWei Xiao, ShouDong Wang, Huanfang Xu, “Establishment and Analysis of an Artificial Neural Network Model for Early Detection of Polycystic Ovary Syndrome Using Machine Learning Techniques,” J. Inflamm. Res., pp. 5667–5676, 2024, doi: https://doi.org/10.2147/JIR.S438838.