

Online Shopper Intention Analysis Using Conventional Machine Learning and Deep Neural Network Classification Algorithm

by Haris Muhammad

Submission date: 26-Jul-2022 08:26PM (UTC+0700)

Submission ID: 1875434302

File name: Jurnal_Penelitian_Pos_dan_Informatika-Paper.pdf (614.82K)

Word count: 5268

Character count: 27947



JPPI Vol 11 No 1 (2021) 89 - 100

10

Jurnal Penelitian Pos dan Informatika

32a/E/KPT/2017

e-ISSN 2476-9266

p-ISSN: 2088-9402



[Doi:10.17933/jppi.2021.110106](https://doi.org/10.17933/jppi.2021.110106)

Online Shopper Intention Analysis Using Conventional Machine Learning and Deep Neural Network Classification Algorithm

Cucu Ika Agustyaningrum¹, Muhammad Haris², Riska Aryanti³, Titik Misriati⁴

¹²Program Studi Ilmu Komputer, Fakultas Teknologi Informasi, ¹²Universitas Nusa Mandiri;

¹²Jl. Kramat Raya No.18, RT.5/RW.7, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450, Indonesia

³⁴Fakultas Teknik dan Informatika, ³⁵Universitas Bina Sarana Informatika;

³⁴Jl. Kramat Raya No.98, RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450, Indonesia

¹14002365@nusamandiri.ac.id, ²muhammad.uhs@nusamandiri.ac.id

Received: 9 October 2021 ; Revision: 23 November 2021 ; Accepted: 23 November 2021

Abstract

The use of e-commerce throughout the world in recent years is very rapid. The continuous increase in sales shows that e-commerce has huge market potential. Store profits are derived from the process of assessing data to identify and classify online shopper intentions. The process of assessing the data uses conventional machine learning algorithms and deep neural networks. Comparison of algorithms in this study using the python programming language by knowing the value of Accuracy, F1-Score, Precision, Recall, and ROCAUC. The test results show that the accuracy of the deep neural network algorithm is 98.48%, the F1 score is 95.06%, precision is 97.36%, recall is 96.81% and AUC is 96.81%. So, based on this research, deep neural network data mining techniques can be an effective algorithm for online shopper intention data sets with cross-validation folds of 10, six hidden layer decoder-encoder variations, relu-sigmoid activation function, adagrad optimizer, and learning rate of 0.01 and no dropout. The value of this deep neural network algorithm is quite dominant compared to conventional machine learning algorithms and related research.

Keywords: Algorithms, Conventional Machine Learning, Deep Neural Networks, E-Commerce, Shopping Intentions.

INTRODUCTION

All over the world, the use of e-commerce or online stores has increased rapidly in recent years. E-commerce is another word for online store, which is a way of shopping using social networking media that is used to process buying and selling transactions, where sellers and buyers don't have to worry about going to the store to see, buy and sell what they are looking for and want. It's just that they can view the goods online, place the desired order, then transfer the money, and then the goods will be sent by the store. online to the buyer's house through a courier service, without the need to go out to the store (Taj & Kumaravel, 2020).

The third party process of buying and selling products online is bridged by e-commerce. Now, online shopping is considered more convenient for users. Over time, online shopping has become an alternative means of shopping, because the internet can certainly be seen as something new in the form of shopping (Sonya Clausis Dea, 2019). The internet has now grown in terms of service, efficiency, security, and ubiquity, so that online purchasing is the third most popular activity involving the internet after the birth of e-commerce (Al-Gasawneh et al., 2020).

E-commerce technology is one of the basic needs of all business organizations. E-commerce is a method for consumers to purchase desired goods through the use of internet technology. Utilization of e-commerce technology can be seen by consumers (business to consumer) and business people (business to business). The use of e-commerce technology by businesses can bring both positive and negative values (Mumtahana et al., 2017). In 2018, retail e-commerce grew to \$1506 billion per year.

The continuous increase in sales shows that e-commerce has huge market potential (Lim et al., 2016).

On the internet, online shops are very diverse, with small or large scales. Even ordinary stores that were originally open for hours turned to online stores. Therefore, consumers can look for goods non-stop for a full 24 hours. They can even look for stores in the country or abroad. The same is true for sellers, who can benefit from consumers around the world (Agustyaningrum et al., 2020).

Effective use of time can affect the level of sales figures, so this experience has a very important influence. Many e-commerce companies invest in early detection (Sakar et al., 2019). Even though e-commerce is emerging and popular today, not everyone is interested. Many online stores fail and fail to make a decent profit because consumers do not perceive online stores as an attractive buying/selling tool (Christian, 2019).

It can be concluded that consumer trust can be determined by the convenience of online store services (Khan & Khan, 2018). Verbal recommendations from family, relatives, close friends or even store star ratings and product prices are also factors that determine consumer trust (Liao et al., 2016). In online shopping, the intention to buy an item also depends on the quality of effective service. So that when ordering, the buyer will trust and believe the store (Han et al., 2018).

It is therefore very important for service providers and manufacturers to understand how these consumers behave in the future virtual market when using technology. Experienced online shoppers have more trust and a better feel for online services (Koththagoda & Herath, 2018). Online stores must be able to assess data that can identify

and classify shopper intentions, so that data can be obtained to predict sales levels. This capability can be applied to online stores using conventional machine learning algorithms. Using two modules, the type of algorithm used and the best prediction is the multilayer perceptron algorithm (Sakar et al., 2019). The random forest algorithm is most suitable for predicting the intention of online shoppers, especially by using gradient boosting in using this algorithm, it will produce high accuracy (Kabir et al., 2019).

In related research, no one has used the deep neural network method, only using conventional machine learning methods such as random forest, multilayer perceptron, and so on. So in this study, the focus is on comparing conventional machine learning algorithms and deep neural networks. Both of these algorithms are the best ways to assess the potential intentions of online shoppers using the Python programming language, so comparing the two algorithms gets the best and most accurate results for online stores receiving predictive results of online shopper intentions for appropriate income.

Based on the analysis of the background description, it can be stated that the identification of the problem in this research is how to prepare data that is ready to be used as training data in machine learning, how to determine the attributes that are very influential in online shopper intention data, and whether the algorithm model used can help online stores analyze the intentions of online shoppers as a source of information.

Based on the identification of the problems described, this study aims to find out how to prepare data that is ready to be used as training data in machine learning, find out how to determine attributes that are very influential in the processing

of online shopper intention data, and find out empirically the effect on online shopper intentions in e-commerce business.

METDHOLOGY

A. Research methods



Figure 1. Research Stages

In online shopper intention research, a model is built based on the research process, namely the dataset, preprocessing, feature selection, modeling and evaluation model stages.

This research begins with collecting data on the UCI Machine Learning Repository website. Then the data is transformed into initial data for preprocessing, then features are selected using the Python programming language, after which it is tested with conventional machine learning algorithms using data cross validation.

At this modeling stage, there are two algorithms that will be compared using a programming language, namely conventional machine learning algorithms and deep neural network algorithms. Before entering the modeling

algorithm, the data to be used is taken using a cross-validation method, where cross-validation is a model validation technique to assess the accuracy of the analysis results. Data that has been preprocessed is carried out by cross-validation by dividing the data into training data and testing data for the classification process (Panggabean et al., 2020). After the data is processed through preprocessing, feature selection, modeling, and testing, the next step is to analyze the data generated by the Python programming language using conventional machine learning algorithms and deep neural networks to get the best results from the algorithms applied to the data.

B. Application of Research Methods

The application of research methods on the Online Shopper's Purchasing Intention Data Set Dataset. In the implementation of this case, using five stages, namely:

1. Dataset

The data used is secondary data, where secondary data is data taken from researchers who have conducted similar research before and general data and the total number of data is 12,330, based on data from UCI Machine Learning (Dataset, 2018), with 18 attributes and 1 class that aims to determine the intention of online shoppers. The use of classification techniques with a high level of prediction and accuracy can help overcome these problems, making the results obtained easier, faster and more accurate. In this study, to be able to provide results with a high level of prediction and accuracy, a comparison was implemented between Conventional Machine Learning and Deep Neural Network algorithms.

2. Preprocessing

At this stage of the research, the amount of data obtained is 12,330, which will be processed and then used to generate predictions of online shopper intentions, based on both the information the shopper buys and does not buy. The data preparation stage begins with conducting data selection, which is reviewing the attributes whose data type will be changed. After the data selection process is carried out, then the data cleaning process is carried out. What is done in this process is to try to check for missing values.

3. Feature engineering and selection

At the feature engineering and selection stage, there are several stages, namely feature binning, feature scaling, and feature selection. In feature binning, it is used to group special days according to their probability values. Feature Scaling is done to standardize numerical features so that there are more outliers. Feature Selection is used to determine which features are the most influential in the data. After that, the data is divided using cross validation, whether it is used when modeling conventional machine learning algorithms or deep neural network algorithms.

4. Modeling

At the modeling or modeling stage, the prediction process is carried out with the proposed algorithm. The proposed algorithm is the Conventional Machine Learning algorithm which will be compared with the Deep Neural Network using the Python programming language to determine the level of accuracy, f1 score, precision, recall and AUC of online shopper intentions.

a. Machine Learning

The term machine learning refers to the automatic detection of meaningful patterns

in data. Machine learning is a way for humans to teach something to computers. Without doing any explicit programming, a computer can learn to process the data given to it. Machine learning algorithms play a role in teaching computers to process data. (Christian, 2019).

b. Decision Tree

A decision tree is the process of finding a set of models or functions that describe and separate data classes from one another, which is used to predict data that does not yet have a certain data class (Shiddiq et al., 2018).

c. Random Forest

Random forest is a classification method that is achieved by developing a Decision Tree method based on a random selection of attributes at each node to determine the classification. During classification, it is based on the highest number of votes in the returned decision tree (Ratnawati & Sulistyaningrum, 2019). Random forest is the union of mutual independent classifiers (CARTs) from the same distribution through a voting process (highest count) to obtain classification predictions. Random forest has features that can reduce correlation, which can reduce the results of random forest prediction errors (As Sarofi et al., 2020). The Random Forest formula (Leonardo et al., 2020):

$$Entropy(Y) = - \sum_i P(Y) \log^2 p(Y) \quad (1)$$

d. Support Vector Machine

Support Vector Machine (SVM) is an integrated classification method (supervised) because at the time of training

it requires certain learning objectives (Nurachim, 2019). The Support Vector Machine formula is as follows (Zulfikar & Lukman, 2016) :

$$similarity = \frac{\sum_{i=1}^n f(T_i, S_i)}{W_i} \quad (2)$$

Information:

T: New case, S: storage cases

n denotes the number of attributes, I: an individual attribute with a value between 1 and n, f: Attributioni have a similarity function between case T and case S, W: the weight assigned to the i-th attribute

e. Adaptive Boosting (Adaboost)

Adaboost is a synthetic learning algorithm that is commonly used in boosting algorithms. These improvements can be combined with other classification algorithms to improve classification performance. The calculation of the Adaboost algorithm is shown in the equation (Gultom, 2020).

$$F(x) = \text{sign} \sum_{t=1}^T \alpha_t h_t(x) \quad (3)$$

Information:

$h_i(x)$ = weak of basic classifier, α_i = learning rate, $F(x)$ = final classifier

f. eXtreme Gradient Boosting (XGBoost)

XGBoost stands for eXtreme Growth Boost. Gradient Boosting is an open source project to implement an efficient, fast and scalable machine learning system called Gradient Tree Boosting for various learning problems based on the Greedy Function Approximation paper. XGBoost is used for supervised learning problems where it uses training data with multiple features x_i to

predict the target variable y_i (Agarwal et al., 1994).

- g. *CatBoost* adalah algoritma *gradient boosting decision tree* (GBDT) baru yang dapat menangani fitur kategoris dengan baik. Algoritma ini berbeda dari algoritma GBDT tradisional (Huang et al., 2019).

h. **Deep Neural Network**

A Deep Neural Network (DNN) is an artificial neural network with multiple layers. A deep neural network typically has more than three layers (input layer, N hidden layers, output layer), making it a Multilayer Perceptron (MLP) with more layers. Since there are relatively many layers, it is deep. The learning process in DNN is called deep learning (Watanabe & Nishimori, 2016).

A Deep Neural Network (DNN) is a type of neural network. This DNN consists of several hidden units with connections between layers but no connections between units in each layer. This method has an architecture similar to an Artificial Neural Network (ANN), with supervised training. By defining the input and matching it to an existing sample. The advantages of the Deep Learning approach for speech recognition are namely better network architecture, the ability to optimize many parameters, DNN is good enough for speech recognition, and DNN is faster to understand various languages/dialects (Fathurrahman et al., 2018).

i. **Confusion Matrix**

The confusion matrix is a very useful tool for analyzing bias when recognizing tuples of different classes (Utami, 2017). At the

evaluation stage using the confusion matrix, the values of accuracy, precision, recall, and error rate will be obtained. Accuracy is the ratio between the number of correctly identified cases and the total number of cases.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision is the proportion of cases with true positive results.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall is the proportion of positive cases that are correctly identified.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

5. **Evaluation**

At the evaluation stage, the prediction process is carried out with several algorithms, namely the Conventional Machine Learning algorithm which will be compared with the Deep Neural Network algorithm to see the results of accuracy, f1 score, precision, recall and auc success and error rates in Python.

C. **The method of collecting data**

Data collection methods can be divided into two sources of data, namely primary data and secondary data. Primary data is data taken from a direct source, while secondary data is data taken from researchers who have conducted similar research before.

In this study, the researcher uses secondary data. The research data was taken from the results of online shoppers' intentions obtained from UCI Machine Learning, with a total of 12,330 records consisting of 18 attributes and 1 class attribute consisting of Table 1.

Table 1. Numerical Attributes and Categories of User Behavior Analysis

No	Nama Atribut	Min. Value	Max. Value	SD	Number of categorical values
1	Administrative	0	27	3.32	
2	Administrative_Duration	0	3398	176.78	
3	Informational	0	24	140.75	
4	Informational_Duration	0	2549	140.75	
5	ProductRelated	0	705	44.48	
6	ProductRelated_Duration	0	63973	1913.67	
7	BounceRates	0	0.2	0.05	
8	ExitRates	0	0.2	0.05	
9	PageValues	0	361	18.57	
10	SpecialDay	0	1.0	0.2	
11	Month		-	-	12
12	OperatingSystems		8.0	0.91	8
13	Browser		13	1.72	13
14	Region		9.0	2.4	9
15	TrafficType		20	4.03	20
16	VisitorType		-	-	3
17	Weekend		1.0	0.42	2
18	Revenue		1.0	0.36	2

RESULTS AND DISCUSSION

In online learner intention research using conventional machine learning algorithms and deep neural networks, it produces several variations in the value of the applied method.

A. Validation of the Preprocessing Step.

This research was conducted using unprocessed data and previously processed data, where unprocessed data is data that does not use feature engineering and selection, while processed data uses feature engineering and selection such as feature scaling and feature binning. The results are shown in Table 2.

From the results of testing the intention of online shoppers without preprocessing, and using preprocessing on conventional machine learning algorithms, the XGBoost preprocessing algorithm has an accuracy value of 90.45%, F1 score 66.16%,

precision 83.11%, recall 78.14% and AUC 78.14% higher than other Random Forest, Catboost, Adaboost, SVM and Decision Tree algorithms. Although the difference in value is only slightly behind the comma. And by using data preprocessing the process of running on the data is faster than without preprocessing.

B. Conventional Machine learning Algorithm Model.

In this case, there are several conventional machine learning algorithms used in this research, namely ¹² Random Forest, Support Vector Machine, ¹² Decision Tree, AdaBoost, XGBoost, CatBoost.

To analyze the intention of online shoppers, this study uses conventional ¹⁵ machine learning algorithms to produce accuracy, F1 Score, Precision, Recall and AUC (Area Under Curve) values.

Table 2. Value Results of Conventional Machine Learning Algorithms using Preprocessing and without Preprocessing

Model	Preprocessing	Accuracy	F1 Score	Precision	Recall	AUC
XGBoost		90,44%	66,12%	83,10%	78,12%	78,12%
XGBoost	√	90,45%	66,16%	83,11%	78,14%	78,14%
Random Forest		90,12%	64,26%	82,74%	76,75%	76,75%
Random Forest	√	90,18%	64,28%	83,01%	76,66%	76,66%
CatBoost		90,22%	65,06%	82,76%	77,39%	77,39%
CatBoost	√	90,07%	64,45%	82,43%	77,04%	77,04%
AdaBoost		89,01%	62,07%	79,55%	76,39%	76,39%
AdaBoost	√	88,95%	61,82%	79,45%	76,23%	76,23%
Decision Tree		85,32%	53,34%	72,03%	72,61%	72,61%
Decision Tree	√	85,41%	52,78%	72,11%	72,04%	72,04%
Support Vector Machine	√	88,71%	52,86%	82,39%	69,17%	69,17%

The test results of several conventional machine learning algorithms that have been carried out, resulting in a comparison of the values of accuracy, F1 Score, precision, recall, and AUC (Area Under Curve), can be seen in the table 3:

than the algorithm. Random Forest, Catboost, Adaboost, SVM, and Decision Tree are some of the algorithms used.

C. Deep Neural Network Algorithm.

In this study, there are several algorithms used after

Table 3. Results of Comparison of Conventional Machine Learning Algorithms

<i>Model</i>	<i>Accuracy</i>	<i>F1 Score</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>
<i>XGBoost</i>	90,45%	66,16%	83,11%	78,14%	78,14%
<i>Random Forest</i>	90,18%	64,28%	83,01%	76,66%	76,66%
<i>CatBoost</i>	90,07%	64,45%	82,43%	77,04%	77,04%
<i>AdaBoost</i>	88,95%	61,82%	79,45%	76,23%	76,23%
<i>Decision Tree</i>	85,41%	52,78%	72,11%	72,04%	72,04%
<i>Support Vector Machine</i>	88,71%	52,86%	82,39%	69,17%	69,17%

From the results of testing the intention of online shoppers without preprocessing, the XGBoost preprocessing algorithm has an accuracy value of 90.45%, F1 score of 66.16%, precision 83.11%, recall 78.14% and AUC of 78.14%, which is higher

the conventional machine learning algorithm, namely there is a deep neural network algorithm that is varied from 3 hidden layers to 8 hidden layers.

The test results from several variations of the deep neural network algorithm that have been carried

out have resulted in a comparison of the values of accuracy, F1 Score, Precision, Recall and AUC (Area Under Curve) which can be seen in table 4.

From the test results above, by evaluating both the confusion matrix and the ROC curve, it is shown that the tests carried out are: optimization of

Table 4. The results of the Accuracy, Precision, Recall, F1-score, and AUC scores on the Deep Neural Network Algorithm

Layers	Activation Function	Optimizer	Learning Rate	Accuracy	F1 Score	Precision	Recall	AUC
3 Hidden Encod-Decod	Sigmoid	Adagrad	0,01	96,82%	89,35%	95,11%	92,48%	92,48%
4 Hidden Encod-Decod	Sigmoid	Adagrad	0,01	97,83%	93,07%	95,59%	96,29%	96,29%
5 Hidden Decod-Encod	Sigmoid	Adagrad	0,01	98,08%	93,72%	96,85%	95,76%	95,76%
6 Hidden Decod-Encod	Sigmoid	Adagrad	0,01	98,84%	95,06%	97,36%	96,81%	96,81%
7 Hidden Decod-Encod	Sigmoid	Adagrad	0,01	98,25%	94,32%	96,90%	96,39%	96,39%
8 Hidden Decod-Encod	Sigmoid	Adagrad	0,01	97,30%	91,01%	95,96%	93,56%	93,56%

From the test results above, by evaluating both the confusion matrix and the ROC curve, it can be ascertained that the tests carried out are optimization of the deep neural network algorithm with variations of 6 hidden layer decoders-encoder, activation function parameter sigmoid, optimizer adagrad, with a learning rate of 0.01 and without dropout, has a higher value than the deep neural network algorithm with other variation.

D. Model Comparison

The test results conclude that the conventional machine learning algorithm with a deep neural network can be seen in table 5.

the deep neural network algorithm with variations of 6 hidden layer decoders-encoder, activation function parameter sigmoid, optimizer adagrad, with a learning rate of 0,01, and without dropout, has a higher value than using conventional machine learning algorithms and related research. The accuracy value of the resulting deep neural network algorithm is 98.48%, F1 score 95.06%, precision 97.36%, recall 96.81% and AUC 96.81% for the conventional machine learning model, xgboost, while for research related to the accuracy of the multi-layer perceptron algorithm, it produces a value of 87.94% and F1-Score 87%.

Table 5. Testing the Conventional Machine Learning Deep Neural Network Algorithm with Related Research

Layers	Activation Function	Optimizer	Accuracy	F1 Score	Precision	Recall	AUC
DNN 6 Hidden Decod-Encod	Sigmoid	Adagrad	98,84%	95,06%	97,36%	96,81%	96,81%
XGBoost			90,45%	66,16%	83,11%	78,14%	78,14%
Multi Layer Perceptron (Sakar et al., 2019)			87,94%	87%	-	-	-

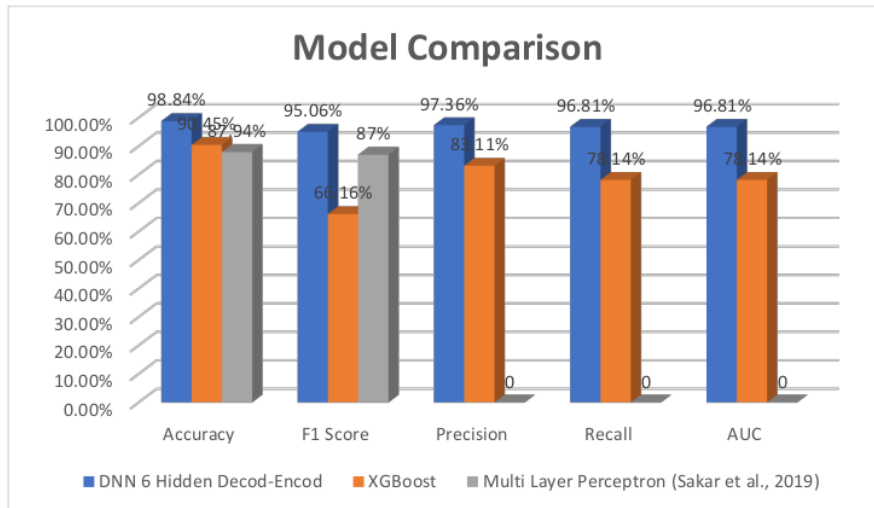


Figure 2. Conventional Machine Learning Algorithms, Deep Neural Networks and Related Research

Based on this value, it can be seen that the average difference in accuracy is 6.705%, F1 score is 14.5%, precision is 14.25%, recall is 18.67% and AUC is 18.67%.

CONCLUSION

From the results of testing online shopper intention data without preprocessing, and using preprocessing on conventional machine learning algorithms, the XGBoost preprocessing algorithm has an accuracy value of 90.45%, F1 score 66.16%, precision 83.11%, recall 78.14% and AUC 78.14% higher than other Random Forest, Catboost, Adaboost, SVM and Decision Tree algorithms. Although the difference in value is only slightly behind the comma. And by using the data preprocessing process that runs on the data faster than without preprocessing.

The preprocessing process uses data selection (data selection) and data repair (data cleaning), as well as engineering and feature

selection, where the stages are feature binning, feature scaling, and feature selection.

The deep neural network data mining technique can be an effective yahoo for online shopper intention data sets with cross validation folds 10 data. Deep neural networks are able to achieve accuracy of 98.48%, precision of 97.36%, recall 96.81%, f1 score 95.06%, and AUC of 96.81% with variations of the deep neural network six hidden layer decoder-encoder, relu and sigmoid activation function, adagrad optimizer, and 0.01 learning rate with no dropout. This value is quite dominant compared to conventional machine learning algorithms and related research.

Based on this explanation, the application of the deep neural network algorithm is able to produce better levels of accuracy, precision, recall, f1 score and auc than conventional machine learning algorithms and related research.

There are no definite rules regarding the learning rate, node and dropout, but the smaller the learning rate, node and dropout, the greater the network accuracy and vice versa. If the learning rate,

nodes, and dropouts are getting bigger, the network accuracy will decrease, with the consequence that the process will take longer.

REFERENCES

- Agarwal, A. K., Wadhwa, S., & Chandra, S. (1994). *Diagnosis of tuberculosis--newer tests. The Journal of the Association of Physicians of India*, 42(8), 665.
- Agustyaningrum, C. I., Gata, W., Nurfalah, R., & Radiyah, U. (2020). *Komparasi Algoritma Naive Bayes, Random Forest Dan Svm Untuk Memprediksi Niat. Jurnal Informatika*, 20(2).
- Al-Gasawneh, J. A., Al-Wadi, M. H., Wadi, B. M. Al, Alown, B. E., & Nuseirat, N. M. (2020). *The Interaction Effect of Comprehensiveness Between Social Media and Online Purchasing Intention in Jordanian Pharmacies The Interaction Effect of Comprehensiveness Between Social Media and Online Purchasing Intention in Jordanian Pharmacies. September*. <https://doi.org/10.3991/ijim.v14i15.15501>
- As Sarofi, M. A., Irhamah, I., & Mukarromah, A. (2020). *Identifikasi Genre Musik dengan Menggunakan Metode Random Forest. Jurnal Sains Dan Seni ITS*, 9(1), 79–86. <https://doi.org/10.12962/j23373520.v9i1.51311>
- Christian, Y. (2019). *JITE (Journal of Informatics and Telecommunication Engineering)*. 3(1), 58–66.
- Dataset, O. S. P. I. (2018). *Online Shoppers Purchasing Intention Dataset*. <https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset>
- Fathurrahman, D. N., Osmond, A. B., & Saputra, R. E. (2018). *Deep Neural Network Untuk Pengenalan Ucapan Pada Bahasa Sunda Dialek Tengah Timur (Majalengka) Deep Neural Network for Speech Recognition on Sundanese Language of the Middle East Dialect*. 5(3), 6073–6080.
- Gultom, S. I. (2020). Implementasi Data Mining Menentukan Pola Hidup Sehat Bagi Pengguna KB Menggunakan Algoritma Adaboost (Studi Kasus : Dinas Serdang Bedagai). *Jurnal Informasi Dan Teknologi Ilmiah (INTI)*, 7(3), 298–304.
- Han, B., Kim, M., & Lee, J. (2018). Exploring consumer attitudes and purchasing intentions of cross-border online shopping in Korea. *Journal of Korea Trade*, 22(2), 86–104. <https://doi.org/10.1108/JKT-10-2017-0093>
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., & Zhou, H. (2019). *Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. Journal of Hydrology*, 574(April), 1029–1041. <https://doi.org/10.1016/j.jhydrol.2019.04.085>
- Kabir, M. R., Ashraf, F. Bin, & Ajwad, R. (2019). *Analysis of different predicting model for online shoppers' purchase intention from empirical data. 2019 22nd International Conference on Computer and Information Technology, ICCIT 2019, March 2020*. <https://doi.org/10.1109/ICCIT48885.2019.9038521>
- Khan, M. A., & Khan, S. (2018). *Service Convenience and Post-Purchase Behaviour of Online Buyers: An Empirical Study. Journal of Service Science Research*, 10(2), 167–188. <https://doi.org/10.1007/s12927-018-0006-x>
- Koththagoda, K. C., & Herath, H. M. R. . (2018). *Factors Influencing Online Purchasing*

- Intention: The Mediation Role of Consumer Attitude. *Journal of Marketing and Consumer Research*, 42(2003), 66–74. www.ij5Ciste.org
- Leonardo, R., Pratama, J., & Chrisnatalis, C. (2020). Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing. *Jurnal Teknologi Dan Ilmu Komputer Prima (Jutikomp)*, 3(2), 1–5.
- Liao, C., Lin, H., Luo, M. M., & Chea, S. (2016). Factors influencing Online Shoppers' Repurchase Intentions: The Roles of Satisfaction and Regret. *Information & Management*. <https://doi.org/10.1016/j.im.2016.12.005>
- Lim, Y. J., Osman, A., Salahuddin, S. N., Romle, A. R., & Abdullah, S. (2016). Factors Influencing Online Shopping Behavior: The Mediating Role of Purchase Intention. *Procedia Economics and Finance*, 35(October 2015), 401–410. [https://doi.org/10.1016/s2212-5671\(16\)00050-2](https://doi.org/10.1016/s2212-5671(16)00050-2)
- Mumtahana, H. A., Nita, S., & Tito, A. W. (2017). Pemanfaatan Web E-Commerce untuk Meningkatkan Strategi Pemasaran. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, 3(1), 6. <https://doi.org/10.23917/khif.v3i1.3309>
- Nurachim, R. I. (2019). *Pemilihan Model Prediksi Indeks Harga Saham Yang Dikembangkan Berdasarkan Algoritma Support Vector Machine (Svm) Atau Multilayer Perceptron (Mlp) Studi Kasus : Saham Pt Telekomunikasi Indonesia Tbk*. 5(1), 29–35.
- Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 56. <https://doi.org/10.30865/jurikom.v7i1.1947>
- Ratnawati, L., & Sulistyaningrum, D. R. (2019). Penerapan Random Forest untuk Mengukur Tingkat Keparahannya Penyakit. *Jurnal Sains Dan Seni Its*, 8(2), A71–A77.
- Sakar, C. O., Polat, S. O., Katircioglu, M., & Kastro, Y. (2019). Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. *Neural Computing and Applications*, 31(10), 6893–6908. <https://doi.org/10.1007/s00521-018-3523-0>
- Shiddiq, A., Niswatin, R. K., & Farida, I. N. (2018). Analisa Kepuasan Konsumen Menggunakan Klasifikasi Decision Tree Di Restoran Dapur Solo (Cabang Kediri). 2(1), 9–18.
- Sonya Clausis Dea, M. S. P. (2019). *FAKTOR-FAKTOR YANG MEMPENGARUHI KEPUTUSAN PEMILIHAN JASA EKSPEDISI JNE (Studi Pada Konsumen Pengguna Jasa JNE di Semarang)*. XVIII(3), 245–258.
- Taj, S. M., & Kumaravel, A. (2020). Intentions of online shoppers prediction by fuzzy petri nets construction. *International Journal of Scientific and Technology Research*, 9(2), 1761–1768.
- Utami, L. A. (2017). Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization. 13(1), 103–112.
- Watanabe, S., & Nishimori, H. (2016). Fall lecture note on statistical learning theory. *Lecture Note for Tokyo Institute of Technology*.
- Zulfikar, W. B., & Lukman, N. (2016). Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata. *Jurnal Online Informatika*, 1(2), 82–86. <https://doi.org/10.15575/join.v1i2.33>

Online Shopper Intention Analysis Using Conventional Machine Learning and Deep Neural Network Classification Algorithm

ORIGINALITY REPORT

11%

SIMILARITY INDEX

7%

INTERNET SOURCES

5%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|---|--|-----|
| 1 | M Khairul Anam, Muhammad Ihza Mahendra, Wirta Agustin, Rahmadden Rahmadden, Nurjayadi Nurjayadi. "Framework for Analyzing Netizen Opinions on BPJS Using Sentiment Analysis and Social Network Analysis (SNA)", INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi, 2022
Publication | <1% |
| 2 | Submitted to Thomas Edison State College
Student Paper | <1% |
| 3 | Submitted to University of Wales, Lampeter
Student Paper | <1% |
| 4 | Submitted to University of Bedfordshire
Student Paper | <1% |
| 5 | ejournal.unisba.ac.id
Internet Source | <1% |
| 6 | ejournal.undip.ac.id
Internet Source | <1% |

<1 %

7

onlinelibrary.wiley.com

Internet Source

<1 %

8

Submitted to Royal Holloway and Bedford
New College

Student Paper

<1 %

9

imsear.searo.who.int

Internet Source

<1 %

10

iztc.ilgabbianoilpino.it

Internet Source

<1 %

11

SY. Yuliani, Shahrin Sahib, Mohd Faizal
Abdollah, Mohammed Nasser Al-Mhiqani,
Aldy Rialdy Atmadja. "Review Study of Hoax
Email Characteristic", International Journal of
Engineering & Technology, 2018

Publication

<1 %

12

docs.google.com

Internet Source

<1 %

13

jurnal.untan.ac.id

Internet Source

<1 %

14

openlibrary.telkomuniversity.ac.id

Internet Source

<1 %

15

www.warse.org

Internet Source

<1 %

16

Submitted to Furness College

Student Paper

<1 %

17

John Hancock, Taghi M. Khoshgoftaar.
"Medicare Fraud Detection using CatBoost",
2020 IEEE 21st International Conference on
Information Reuse and Integration for Data
Science (IRI), 2020

Publication

<1 %

18

datascience.cmu.ac.th

Internet Source

<1 %

19

docshare.tips

Internet Source

<1 %

20

ujcontent.uj.ac.za

Internet Source

<1 %

21

Afina Ramadhani, Edi Sutoyo, Vandha
Pradwiyasma Widartha. "LSTM-based Deep
Learning Architecture of Tourist Review in
Tripadvisor", 2021 Sixth International
Conference on Informatics and Computing
(ICIC), 2021

Publication

<1 %

22

Fani Nurona Cahya, Rangga Pebrianto, Tika
Adilah M. "Klasifikasi Buah Segar dan Busuk
Menggunakan Ekstraksi Fitur Hu-Moment ,
Haralick dan Histogram", IJCIT (Indonesian

<1 %

Journal on Computer and Information Technology), 2021

Publication

23

Submitted to University of Lincoln

Student Paper

<1 %

24

patents.google.com

Internet Source

<1 %

25

Submitted to Charles Sturt University

Student Paper

<1 %

26

ojs.unpkediri.ac.id

Internet Source

<1 %

27

www.ejurnal.its.ac.id

Internet Source

<1 %

28

Ardhana Januar Mahardhani, Iqbal Akbar Imamudin, Fendy Eko Hardiawan. "Upaya Mitigasi Bencana Melalui Aplikasi Dayakan Mitigation Center (DMC)", Jurnal Pemberdayaan Masyarakat Universitas Al Azhar Indonesia, 2021

Publication

<1 %

29

Duy Dang-Pham, Siddhi Pittayachawan, Vince Bruno, Karlheinz Kautz. "Investigating the diffusion of IT consumerization in the workplace: A case study using social network analysis", Information Systems Frontiers, 2017

Publication

<1 %

- | | | |
|----|---|------|
| 30 | Evta Indra, Kevin Ho, Arlinanda, Ridho Hakim, Delima Sitanggang, Oloan Sihombing. "Application of C4.5 Algorithm for Cattle Disease Classification", Journal of Physics: Conference Series, 2019
Publication | <1 % |
| 31 | Umu Hanifah, Purba Daru Kusuma, Casi Setianingsih. "Detection of Pterygium Disease Using Forward Chaining and Viola Jones Algorithm", 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), 2018
Publication | <1 % |
| 32 | dspace.daffodilvarsity.edu.bd:8080
Internet Source | <1 % |
| 33 | 1library.net
Internet Source | <1 % |
| 34 | pdfs.semanticscholar.org
Internet Source | <1 % |
| 35 | www.lamudi.co.id
Internet Source | <1 % |
| 36 | Nikolai Fahrmeier, Simon Frank, Nadine Goeppert, Nico Goldscheider. "Caracterização multiescala de um sistema aquífero cárstico e aluvial complexo no sul da Alemanha usando | <1 % |

37

www.its.ac.id

Internet Source

<1 %

38

Made Adi Widyananda, Irma Palupi. "Implementation of the Spiral Optimization Algorithm in the Support Vector Machine (SVM) Classification Method (Case Study: Diabetes Prediction)", 2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS), 2021

Publication

<1 %

39

Yennimar Yennimar, Rohni Endetta Manihuruk, Etis Landya Br Hotang. "Prediction Models with Machine Learning Against Student Success in Online Learning", SinkrOn, 2021

Publication

<1 %

40

archive.ics.uci.edu

Internet Source

<1 %

41

cognitivescience.ceu.edu

Internet Source

<1 %

42

ejournal.undiksha.ac.id

Internet Source

<1 %

44

Rina Novita, Supratman Zakir, Agus Nur Khomarudin, Efmi Maiyana, Hamimah Hasyim. "Use of the C4.5 Algorithm in Determining Scholarship Recipients", Journal of Physics: Conference Series, 2021

Publication

<1 %

45

Dipanjan Sarkar, Raghav Bali, Tushar Sharma. "Chapter 4 Feature Engineering and Selection", Springer Science and Business Media LLC, 2018

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off