Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 7, No. 1, March 2024, pp. 193 – 199
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Coffee Type Classification Using Backpropagation Artificial Neural Network

[1]Pitrasacha Adytia, [2]Wahyuni, [3]Dimas Asmaramany, [4]Kelik Sussolaikah
[1]Department of Information System, STMIK Widya Cipta Dharma, Indonesia
[2,3]Department of Informatics Engineering, STMIK Widya Cipta Dharma, Indonesia
[4]Department of Informatics Engineering, Universitas PGRI Madiun, Indonesia
Email: [1]pitra@wicida.ac.id, [2]wahyuni@wicida.ac.id,
[3]dimasasmaramany@gmail.com, [4]kelik@unipma.ac.id

Article Info

Article history:
Received Jan 13th, 2024
Revised Feb 28th, 2024
Accepted Mar 26th, 2024

Keyword:
Artificial Neural Network
Backpropagation
Classification
Coffee Type
E-Nose

ABSTRACT

Coffee has several types including robusta coffee, arabica coffee and luwak coffee. Each coffee has certain characteristics of color, texture, aroma and also the quality of the taste. Coffee counterfeiting is also common. This coffee counterfeiting usually uses materials such as corn, wheat, soybeans, husks, sticks and robusta coffee beans. So that a model is needed to be able to classify the type of coffee. This research uses artificial neural network machine learning algorithms to identify and classify coffee. Quality training and testing data is needed in this method because it will affect the final results. Initial data is collected via e-nose, with this equipment data on changes in electrical voltage will be obtained from 4 sensors, namely MQ-2, MQ-3, MQ-7 and MQ-135. These 4 features will be used in the classification process. With 900 sets of training data, the test results show that the neural network is able to provide correct classification 99% of the 3 sets of testing data. The results of training and testing show that the neural network formed can identify and distinguish coffee types with good results.

Corresponding Author:
Pitrasacha Adytia,
Departement of Information System,
STMIK Widya Cipta Dharma,
Jl. M.Yamin, No. 25 Samarinda, Kalimantan Timur, Indonesia
Email: pitra@wicida.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v7i1.28853

1. INTRODUCTION

Coffee is one of the most important plantation commodities in world trade. This commodity has long accompanied and adorned the History of human civilization, across time and national borders [1]. This compound in normal body conditions has several properties, one of which is an analgetic drug that can reduce pain and reduce fever [2]. There are four types of coffee groups known, namely arabica coffee, robusta coffee, liberica coffee, and exelsa coffee. The known coffee groups that have economic value and are commercially traded are arabica coffee and robusta coffee. Meanwhile, the liberica and exelsa coffee groups are less economical and less commercial [3]. Arabica coffee has various varieties based on the region of origin, known as single origin [4]. Arabica coffee and Robusta coffee account for the majority of the world's coffee trade. Arabica coffee has high flavor quality and lower caffeine content compared to robusta coffee, making it more expensive. The gas content of ammonia, hydrogen sulfide, and carbon monoxide in arabica coffee is higher than robusta coffee [5]. The flavor quality of robusta coffee is below that of arabica coffee, but robusta coffee is resistant to leaf rust disease [3]. Kopi luwak is coffee produced from the fermentation process through the stomach of a civet or civet animal that eats ripe (red) and fresh coffee cherries and then excretes them in the form of feces [6]. This coffee adulteration usually uses materials such as corn, wheat, soybeans, husks, sticks and robusta coffee beans [7]. To ensure good coffee quality, it is necessary to pay attention in every process. Based on these problems, technology is needed that can be used to quickly and accurately determine the type

of coffee to be traded so as to reduce the chance of fraud and make it easier to identify the coffee. One of them is by building a coffee type classification model using the backpropagation artificial neural network method.

Artificial Neural Network is an information processing paradigm inspired by biological neural systems, such as the performance of the brain, which processes information [8]. The artificial neural network is a representation that tries to simulate the learning process of the human brain. Some of the JST methods that can be used to identify images or patterns are backpropagation, Learning Vectoration Quantification (LVQ), and perceptron [8]. Backpropagation is a common algorithm for artificial neural networks. In the learning process, the Backpropagation algorithm falls into the category of supervised learning methods [9].

The classification of coffee types using a backpropagation artificial neural network model has been done before, but the classified types of coffee were different [10]. This research classifies 3 types of coffee, namely arabica, robusta, and luwak. For the dataset, it was collected using an Electronic Nose (E-Nose). E-Nose is an instrument that can detect aromas similar to the human sense of smell, consisting of a series of Metal Oxide Semiconductor (MOS) sensors [11]. The electronic nose, also known as an e-nose, is a device consisting of sensors used to detect aromas or odors. The e-nose is made up of an array of gas sensors and operates similarly to the human sense of smell, hence it is often referred to as a human olfactory system [12]. In this study, E-Nose is used to collect datasets from coffee aromas. After obtaining the dataset from E-Nose, the classification of coffee types is continued using the backpropagation artificial neural network method.

2. RESEARCH METHOD

The research methodology used is Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM is an industry-independent process model for data mining, consisting of six iterative phases from business understanding to deployment [13]. CRISP-DM, which is the most popular framework teams use to execute data science projects, provides an easy to understand description of the data science project workflow (i.e., the data science life cycle) [14]. The method widely applied in data mining is CRISP-DM. CRISP-DM elucidates the data mining process in six stages: Business Understanding; Data Understanding; Data Preparation; Modeling; Evaluation; Deployment. [15].

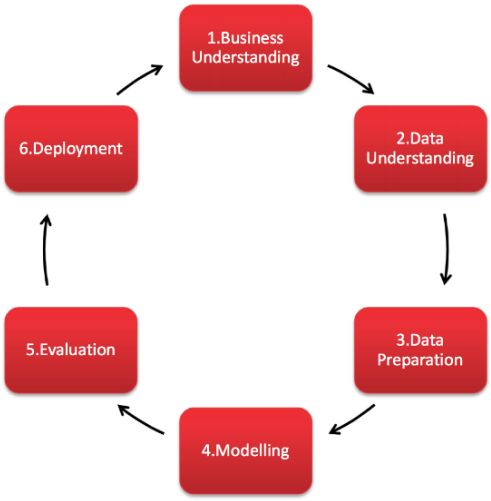

Figure 1. Crisp-DM Methodology

Here are the stages of the CRISP-DM process as shown in Figure 1:

1. Business Understanding
   This is the process of deeply understanding the characteristics of the business to be analyzed [16]. Business Understanding is the process of determining business objectives, understanding the situation and conditions at the time of the research, and establishing a research goal that is translated into problems to be solved using data mining [17]. The aim of this research is to classify types of coffee to accurately determine the types of coffee such as Luwak, Arabica, and Robusta.

2. Data Understanding
   Data understanding involves activities to prepare, assess data needs, and includes data collection [18]. During the data understanding stage, several activities are carried out, including collecting initial data, describing data, exploring data, and verifying data quality. In this research, the dataset used is generated by E-Nose.

3. Data Preparation
   Data preparation may be one of the most difficult steps in any machine learning project. The reason is that each dataset is different and highly specific to the project [19]. The data preparation phase generally includes three steps: Data cleaning, i.e., handling missing data and outliers, data reduction, i.e., reducing the data size by aggregation, elimination redundant feature, etc. and data normalization [20]. In this stage, the final dataset is constructed from raw data. Several tasks will be carried out, including data cleaning, data selection, records and attributes, as well as data transformation to be used as input in the modeling stage.

4. Modelling
   The model building stage involves selecting the appropriate modeling techniques and applying them to the prepared data to meet specific business objectives [21]. The selection of data mining techniques, algorithms, and determining parameters with optimal values is crucial. The chosen model for this research will be a classification model using the backpropagation artificial neural network.

5. Evalution
   This stage involves evaluating the performance of the data model and reviewing the entire process that has been executed to ensure that no data or stages have been missed [22]. The evaluation is conducted thoroughly to ensure that the results in the modeling stage align with the goals set in the business understanding stage. In the evaluation stage, several tasks are carried out, including evaluating the results, reviewing the process, and determining the next steps. This stage is carried out after all models are created, and the program can run, where all software, additional programs, and all programs involved in building the system are tested to ensure that the system can run according to the design or not. The evaluation method used is the confusion matrix, which is one of the methods that can be used to measure the performance of a classification method [23].

		Observed	
		TRUE	FALSE
Predicted Class	TRUE	TRUE POSITIVE (TP)	FALSE POSITIVE (FP)
	FALSE	FALSE NEGATIVE (FN)	TRUE NEGATIVE (TN)

Figure 2. Confussion Matrix.

   True Positive (TP) represents the number of positive data correctly classified by the system. True Negative (TN) represents the number of negative data correctly classified by the system. False Negative (FN) represents the number of negative data incorrectly classified by the system. False Positive (FP) represents the number of positive data incorrectly classified by the system.

6. Deployment
   In the last stage of CRISP-DM is deployment, where the model that has been evaluated is embedded in a user interface for easy use. This classification model will be embedded on a website, so that the results will be visible on the website.

3. RESULTS AND ANALYSIS

In accordance with the methodology used, it will be explained in detail the stages carried out in this research. Which starts from Business Understanding to the evaluation stage. The research carried out only reaches the evaluation stage, where at that stage the final results of the classification of coffee types are obtained.

3.1. Business Understanding

At this stage, identification of problems that occur, what factors influence. To be able to recognize the type of coffee, it is necessary to know the differences in each type of coffee that you want to know such as color, texture, aroma and also taste quality. This coffee counterfeiting usually uses ingredients such as corn, wheat, soybeans, husks, sticks and robusta coffee beans. To ensure good coffee quality, it is necessary to pay attention in every process. Based on these problems, technology is needed that can be used to quickly and accurately determine the type of coffee to be traded so as to reduce the chance of fraud and make it easier to identify the coffee. One of them is by building a coffee type classification model using the backpropagation artificial neural network method.

3.2. Data Understanding

The purpose of data collection is to create a data set about the aroma of coffee types, as well as to understand and analyze the aroma characteristics associated with that coffee. The process of taking a copy dataset using E-Nose can be seen in figure 3.

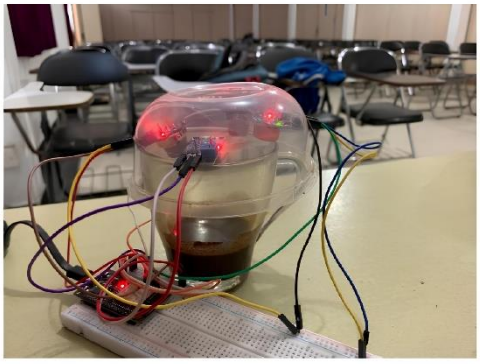

Figure 3. Arabica Coffee Dataset Retrieval Process

The Arabica Coffee data collection process carried out by the STMIK WICIDA campus in room 4/6 takes 15-20 minutes to get 1000 Arabica Coffee datasets.

Table 1. Dataset sample of Arabica Coffee

Id	Mq2	Mq3	Mq7	Mq135	Coffee Type
253	650	2307	576	429	Arabica Coffee
254	654	2300	580	432	Arabica Coffee
255	651	2304	582	429	Arabica Coffee
256	646	2277	582	426	Arabica Coffee
257	644	2273	573	415	Arabica Coffee
258	651	2280	576	421	Arabica Coffee

Followed by taking robusta coffee datasets. The process of collecting Robusta Coffee data carried out by the STMIK WICIDA campus in room 4/6 takes 15-20 minutes to get 1000 Robusta Coffee datasets.

Table 2. Dataset sample of Robusta Coffee

Id	Mq2	Mq3	Mq7	Mq135	Coffee Type
774	503	2241	474	459	Robusta Coffee
775	503	2241	471	459	Robusta Coffee
776	496	2243	474	459	Robusta Coffee
777	494	2249	475	459	Robusta Coffee

Followed by taking the civet coffee dataset. The process of collecting Kopi Luwak data carried out by the STMIK WICIDA campus in room 4/6 takes 15-20 minutes to get 1000 Kopi Luwak datasets.

Table 3. Dataset sample of Luwak Coffee

Id	Mq2	Mq3	Mq7	Mq135	Coffee Type
573	556	2213	539	455	Luwak Coffee
574	564	2219	552	455	Luwak Coffee
575	563	2213	547	453	Luwak Coffee
576	559	2207	537	455	Luwak Coffee
577	568	2224	540	446	Luwak Coffee
578	559	2211	540	447	Luwak Coffee

3.3. Data Preparation

After the data collection stage is carried out, the data preparation stage is carried out. When the data retrieval process is complete, it will be filtered to eliminate incomplete, duplicate, or invalid data. This will ensure better data quality for subsequent analysis.

3.4. Modelling

AI model training will be connected to the sensor data base and take 80% data_set_kopi table for training. When the Arabica Coffee, Robusta Coffee and Civet Coffee dataset retrieval has been completed,

conduct training to teach the model to perform certain tasks automatically by learning patterns, rules, or characteristics hidden in the training data. When you are finished, repeat the training 50 times the data training (epochs = 50) will display a Loss of 0.01288701593875885 Training Accuracy: 100% (100/100) = 1.0.

3.5. Evaluation

For each coffee category (Arabica, Luwak, Robusta), TP, TN, FP, FN, and accuracy can be calculated using the following formula:

1. Arabica:
   TP Arabica = Number of correct predictions for Arabica category
   Total Arabica = Total amount of data with Arabica category
   Arabica Accuracy = TP Arabica / Total Arabica

2. Luwak:
   TP luwak = Number of correct predictions for luwak category
   Total luwak = Total amount of data with luwak category
   luwak Accuracy = TP Arabica / Total Arabica

3. Robusta:
   TP Robusta = Number of correct predictions for the Robusta category
   Total Robusta = Total amount of data with Robusta category
   Robusta Accuracy = TP Robusta / Total Robusta

The calculation of the total 429 predicted to be correct is as follows:

TP Arabica = 118 (correct prediction for Arabica)
Total Arabica = 118
Arabica accuracy = 118 / 118 = 1.0

TP Luwak = 249 (correct prediction for Luwak)
Total Luwak = 249
Luwak accuracy = 249/249 = 1.0

TP Robusta = 62 (correct prediction for Robusta)
Total Robusta = 62
Robusta accuracy = 62/62 = 1.0

Overall accuracy can be calculated using the following formula: Overall Accuracy = (TP Arabica + TP Luwak + TP Robusta) / Total Overall Accuracy Data = (118 + 249 + 62) / 429 = 1.0/100% It can be seen in figure 4 below.

	Kopi Arabika	Kopi Luwak	Kopi Robusta
Kopi Arabika	118	0	0
Kopi Luwak	0	249	0
Kopi Robusta	0	0	62

Figure 4. Confusion Testing

3.5. Deployment

From the results of classification data and sensor values will be displayed through the website in realtime. It can be seen in figure 5.

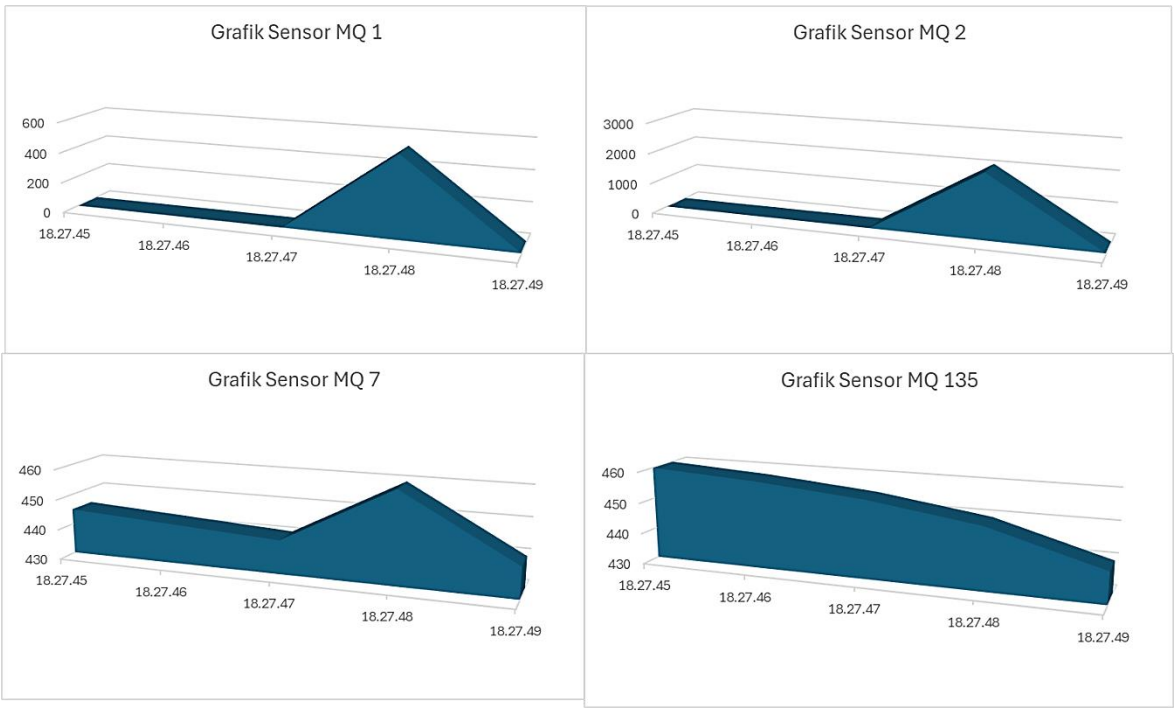

Figure 5. Website Dashboard of Experiment Results

4. CONCLUSION

E-Nose sends analog data to a database and then a Machine Learning Model that classifies coffee types. There are 3 types of coffee, namely, Arabica Coffee, Robusta Coffee and Luwak Coffee. Data collection in different rooms can affect the aroma of coffee. If the room has a lot of cigarette smoke or a strong food smell, then the aroma of coffee will be easier to cover. It takes 15 – 20 minutes to be able to use the sensor from E-Nose so that the analog value is stable. Use a water temperature of about 75-95 degrees Celsius. If the water used to brew coffee is cold, then these substances will not dissolve perfectly, so the coffee aroma of coffee will decrease. And it can also be concluded that the classification of coffee types can run well as evidenced by a very high level of accuracy.

REFERENCES

[1] Asiah N, Epriyani C, Kurnia A, Ramadhan K, Hidayat SG, Apriyantono A. Profil Kopi Arabika Kintamani Bali. Malang: aepublishing; 2022.
[2] Arwangga AF, Asih IARA, Sudiarta IW. Analisis Kandungan Kafein pada Kopi Di Desa Sesaot Narmada Menggunakan Spektrofotometri UV-VIS. JURNAL KIMIA. 2016 Jan;10(1):110–4.
[3] Rhardjo P. Kopi. Jakarta: Penebar Swadaya; 2012.
[4] Nugroho MA, Sebatubun MM. Klasifikasi Varietas Kopi Berdasarkan Green Bean Coffee Menggunakan Metode Machine Learning. Journal of Information System Management (JOISM). 2020 Jan 31;1(2):1–5.
[5] Rabersyah D, . F, . D. Identifikasi Jenis Bubuk Kopi menggunakan Electronic Nose dengan Metode Pembelajaran Backpropagation. Jurnal Nasional Teknik Elektro. 2016 Oct 27;5(3):332.
[6] Pahlevi R, Zakaria WA, Kalsum U. Analisis Kelayakan Usaha Agroindustri Kopi Luwak Di Kecamatan Balik Bukit Kabupaten Lampung Barat. Jurnal Ilmu Ilmu Agribisnis. 2014 Jan;2(1):48–55.
[7] Winkler-Moser JK, Singh M, Rennick KA, Bakota EL, Jham G, Liu SX, et al. Detection of Corn Adulteration in Brazilian Coffee (*Coffea arabica*) by Tocopherol Profiling and Near-Infrared (NIR) Spectroscopy. J Agric Food Chem. 2015 Dec 16;63(49):10662–8.
[8] Edi Ismanto EPC. Jaringan Syaraf Tiruan Algoritma Backpropagation Dalam Memprediksi Ketersediaan Komoditi Pangan Provinsi Riau. Rabit : Jurnal Teknologi dan Sistem Informasi Univrab. 2017 Aug 7;2(2):196–209.
[9] Octariadi BC. Pengenalan Pola Tanda Tangan Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. Jurnal Teknoinfo. 2020 Jan 15;14(1):15.
[10] Novita DD, Sesunan AB, Telaumbanua M, Triyono S, Saputra TW. Identifikasi Jenis Kopi Menggunakan Sensor E-Nose Dengan Metode Pembelajaran Jaringan Syaraf Tiruan Backpropagation. Jurnal Ilmiah Rekayasa Pertanian dan Biosistem. 2021 Sep 29;9(2):205–17.
[11] Sumanto B, Java DR, Wijaya W, Hendry J. Seleksi Fitur Terhadap Performa Kinerja Sistem E-Nose untuk Klasifikasi Aroma Kopi Gayo. MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer. 2022 Mar 31;21(2):429–38.

[12] Kusairi K, Muthmainnah M, Imam Tazi, Moh. Fajrul Falah. Klasifikasi Pola Aroma Teh Hijau Menggunakan Hidung Elektronik (E-Nose) Berbasis Linear Diskriminan Analisis (LDA). JURNAL PENDIDIKAN MIPA. 2022 Sep 21;12(3):868–74.
[13] Schröer C, Kruse F, Gómez JM. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Comput Sci. 2021;181:526–34.
[14] Saltz JS. CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps. In: 2021 IEEE International Conference on Big Data (Big Data). IEEE; 2021. p. 2337–44.
[15] Suhanda Y, Kurniati I, Norma S. Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. Jurnal Teknologi Informatika dan Komputer. 2020 Sep 30;6(2):12–20.
[16] Singgalen YA. Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW. Journal of Information System Research (JOSH). 2023 Jul 31;4(4):1343–53.
[17] Dhewayani FN, Amelia D, Alifah DN, Sari BN, Jajuli M. Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. Jurnal Teknologi dan Informasi. 2022 Mar 31;12(1):64–77.
[18] Pambudi A. Penerapan Crisp-Dm Menggunakan MLR K-Fold pada Data Saham PT. Telkom Indonesia (PERSERO) TBK (TLKM) (STUDI KASUS: Bursa Efek Indonesia Tahun 2015-2022). Jurnal Data Mining dan Sistem Informasi. 2023 Mar 1;4(1):1.
[19] Brownlee J. Data Preparation For Machine Learning. v1.2. 2020.
[20] Masmoudi O, Jaoua M, Jaoua A, Yacout S. Data Preparation in Machine Learning for Condition-based Maintenance. Journal of Computer Science. 2021 Jun 1;17(6):525–38.
[21] Pratama R, Herdiansyah MI, Syamsuar D, Syazili A. Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning. Jurnal Sisfokom (Sistem Informasi dan Komputer). 2023 Mar 17;12(1):96–104.
[22] Christian Y, Qi KOYR. Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM. JURIKOM (Jurnal Riset Komputer). 2022 Aug 30;9(4):966.
[23] Hadianto N, Novitasari HB, Rahmawati A. Klasifikasi peminjaman nasabah bank menggunakan metode neural network. Jurnal Pilar Nusa Mandiri. 2019 Sep 5;15(2):163–70.

BIBLIOGRAPHY OF AUTHORS

Pitrasacha Adytia, S.T., M.T. is a lecturer in the Information Systems study program at STMIK Widya Cipta Dhrama. The author is very interested in researching and writing in the fields of Data Science, Intelligent Systems and Machine Learning.

Wahyuni, S.Kom., M.Kom. is a lecturer at the Informatics Engineering study program at STMIK Widya Cipta Dharma. the author is interested in machine learning and data science.

Dimas Asmaramany is a final year student of the informatics engineering study program at stmik widya cipta dharma. the author is interested in researching machine learning.

Kelik Sussolaikah, S.Kom., M.Kom. is a lecturer in the Informatics Engineering study program at Universitas PGRI Madiun. The author is very interested in researching and writing in the fields of Data Science.