Identification of Signature Authenticity Using Binary Extraction and K-nearest Neighbor Feature Methods

Angela Citra Vidyanti [1]*, Itin Riati[2], Agung Ramadhanu[3]
Department Magister Informatics of Engineering
Universitas Putra Indonesia "YPTK" Padang
Padang, Indonesia
angelacitra86@gmail.com [1] , itinriyantibts@gmail.com[2], , agung_ramadhanu@upiyptk.ac.id [3]

***Abstract*— This research focuses on identifying the authenticity of signatures, which is an important part of the field of biometrics. Identification of signature authenticity has wide applications, including in document security, financial transactions, and identity verification in general. The problem to be resolved is the lack of an effective and efficient method for identifying signature authenticity. The method used is the binary extraction method and the K-nearest Neighbor feature. The main contribution of this research is to propose a new approach in identifying signature authenticity by combining binary extraction methods and K-nearest Neighbor features. This approach is expected to increase the accuracy and efficiency of the signature authenticity identification process. The results of this research are the development of a new model or algorithm for identifying the authenticity of signatures. After testing and validation, the accuracy level of the results of identifying the authenticity of this signature is 75%.**

***Keywords*— Signature, Binary Extraction, K Nearest Neighbor**

I. INTRODUCTION

Signatures have long been used as an authentication method in various contexts, but they emerged when it was necessary to automatically identify their authenticity. A signature is something common to show identification or a sign of someone's identity. A signature is also a sign of validation of a document file[1]. Some signatures can be read, but many signatures are unreadable. However, a signature can be handled as an image so that it can be recognized using pattern recognition applications in image processing, A signature is used to provide proof that the document being signed is genuine and valid[2].

Along with the rapid development of technology and the increasing use of electronic transactions, the need for digital security becomes increasingly crucial. Signatures, which were previously synonymous with physical documents, are now often used digitally. Therefore, sophisticated methods are needed to identify and verify the authenticity of digital signatures, A digital signature is an encryption value that depends on the contents of a digital file and the key of the digital file owner. This signature can be embedded in a digital file or saved in a separate file. The verification process is carried out to verify the authenticity of the electronic signature. If the digital signature is genuine, it means the digital file is still genuine and owned by the same person[3].Signature forgery and identity fraud are serious threats to the security of digital transactions[4]

Electronic signature creation data is a biometric code, cryptographic code, and/or code resulting from converting a manual signature into an electronic signature, including other codes resulting from developments in information technology[5]. Conventional techniques for identifying the authenticity of signatures may not be effective enough in the face of increasingly sophisticated forgery techniques. Therefore, research has focused on developing more sophisticated and reliable methods. The use of the binary extraction feature allows the representation of the signature in a numeric format that can be further processed.

By representing signatures as a series of binaries, specific and in-depth information about the unique patterns of each signature can be obtained.

KNN is a popular classification algorithm because it is simple but effective. The use of KNN in the context of identifying signature authenticity utilizes its ability to compare feature patterns between the signature being tested and a set of signatures that have been used as a reference. In the world of digital transactions, the level of accuracy in identifying the authenticity of a signature is very important.

Errors in the identification process can have serious impacts on the security and integrity of transactions. Therefore, this research aims to provide a high level of accuracy so that it can be relied on in daily practice[6].

Signatures have long been an important form of personal identification in various aspects of life, from business transactions to document security. Its presence in the digital world is also increasingly significant, giving rise to the need for effective authenticity identification methods. One of the main challenges in this context is how to develop a system that can accurately differentiate between legitimate and fake signatures[7].

This research focuses on identifying signature authenticity by tilizing two main components: the binary feature extraction method and the K Nearest Neighbor (KNN) algorithm. Binary extraction features are used to describe signatures in binary format, extracting important information that characterizes each signature uniquely[8].

Meanwhile, the KNN algorithm, as a proximity-based classification algorithm, is implemented to classify signatures

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2063, Copyright ©2024
Submitted : January 17, 2024, Revised : March 30, 2024, Accepted : April 9, 2024, Published : June 15, 2024

based on the extracted binary features.

Document security, transaction authenticity, and recognition of personal identity are some of the applicable aspects that can benefit from the results of this research[9]. By combining the power of binary extraction features and the flexibility of KNN, it is hoped that the resulting system can make a significant contribution to increasing the level of accuracy and reliability of identifying signature authenticity.

In today's digital world, signatures are still one of the main methods for verifying the authenticity and validity of documents and transactions. However, manually disclosing signatures is often a problem. This process is time consuming, sometimes inconsistent, and prone to human error. Moreover, with advances in technology, it has also become easier to create fake signatures, threatening the security of documents and transactions.

The problem faced specifically is the lack of effective and efficient signature identification methods in distinguishing genuine signatures from fakes. Methods often rely on conventional human judgment or simple algorithms that may not be accurate enough or be too time consuming.

Therefore, research is needed to develop a more sophisticated and reliable approach to identifying signature authenticity. This approach must be able to overcome the challenges of dealing with complex variations in signatures, as well as provide a high level of accuracy and efficient processing. This will make a significant contribution to increasing the security of documents and transactions involving signatures, as well as reducing the risk of circumstances and forgery.

It is hoped that this journal can make a real contribution to the development of digital security by presenting innovative and effective methods for identifying the authenticity of signatures. Thus, this research can become a basis for developing better security systems in the future.

II. RESEARCH METHODS

A. System Design

The presence system through identifying the authenticity of signatures designed in this assignment, uses Matlab as system input, then captures the images produced from each image will be processed via the Binary Extraction Feature Method and K Nearest Neighbor (KNN).

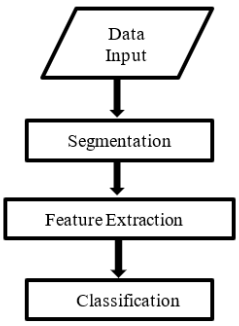

Fig 1. System Design

1. Data Input

Input data in the context of identifying signature uthenticity using the binary feature extraction method and K Nearest Neighbor (KNN) refers to information or recordings used as a basis for identification. In this research, input data can include several components, namely:

a. Signature

The signature is the main data that will be identified. This is in the form of a signature image produced by an individual. This data needs to be processed and analyzed to extract relevant binary features.

b. Training Database

To train the K Nearest Neighbor (KNN) model, there needs to be a training database containing examples of signatures whose authenticity is known. This training data is used to build models and teach the algorithm how to classify new data.

c. Binary Extraction Feature

Binary features resulting from signature extraction. This may include information such as line length, pixel intensity distribution, or certain patterns that can be converted into a binary representation (0 and 1).

d. Label Or Category

Each signature in the training database needs to have a label or category that indicates its authenticity. For example, a real signature can have a label of 1, while a fake signature can have a label of 0. The KNN model uses this information to classify new signatures.

e. Test Data

Once the model is trained, it needs to be tested with never-before-seen signature data. This test data is used to evaluate the extent to which the model can identify the authenticity of signatures that are not present in the training database.

2. Segmentation

Segmentation refers to the process stages for separating and extracting important parts of signature data. Segmentation is an essential first step in processing images or visual data to focus on specific relevant areas or features. Following are the segmentation steps: Signature Segmentation

This process may involve separating individual elements from the signature image, such as the contours or lines of the hand that make up the signature itself. Segmentation can help focus analysis on the most relevant parts of the image.

a. Binary Feature Segmentation

Specifically, in binary feature extraction methods, segmentation can refer to the separation and suppression of certain desired features. For example, if there are features such as hand lines or special shapes in a signature that is the focus of binary extraction, then the segmentation process will help separate and highlight these features.

3. Feature Extraction

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2063, Copyright ©2024
Submitted : January 17, 2024, Revised : March 30, 2024, Accepted : April 9, 2024, Published : June 15, 2024

Feature xtraction results in numerical representations or specific features obtained from signature data after going through the extraction process. These features have an important role in distinguishing real and fake signatures when using the K Nearest Neighbor (KNN) algorithm. The following are some feature extraction results that can be explained:

a. Binary Representation

The feature extraction feature can be a binary representation of the signature. For example, each pixel in a signature image can be converted into a binary value, resulting in a binary representation that reflects the pattern and structure of the signature.

b. Geometric Or Morphological

Features such as line length, angle, or general shape of the signature can be calculated and considered feature extraction. This can help in distinguishing signatures based on certain geometric characteristics.

c. Pixel Intensity Distribution

Feature extraction may involve analyzing the pixel intensity distribution in the signature image. This can help identify distinctive patterns or particular changes in intensity that may be associated with the original signature.

d. Unique Characteristics

Unique features typical of a signature, such as changes in pen pressure or rapid changes in the ink path, can be extracted and used as unique traits for authenticity identification.

4. Classification

Classification is the process of grouping or determining the authenticity category (genuine or fake) of a signature based on the features extracted using this method. The following is an explanation of the classification stages:

a. Binary Feature Extraction

The first stage in the process is binary feature extraction from the signature data. This involves converting the essential information of the signature into a binary representation, such as 0 and 1. These features can include various characteristics, such as pixel patterns, intensity distribution, or geometric shapes.

b. Formation Of Training Data and Test Data

A training database whose authenticity is known is used to train the KNN model. This training data includes real and fake signatures along with their authenticity labels. A subset of data that has never been seen before may be stored as test data to test the performance of the model after it has gone through the training phase.

c. KNN Model

The KNN model is used for the classification of signature authenticity. In this context, KNN compares the binary feature vectors of the signature to be tested with the feature vectors of the signatures in the training database. This model determines the authenticity category based on the majority of labels from the k most similar nearest neighbors.

d. Classification Of New Tag Signs

After going through the training stage, the KNN model can be used to classify new signatures that have never been seen before. The binary feature vector of the signature will be compared with the feature vectors of the training data, and the authenticity label will be determined.

B. System Requirements Analysis

Several components used to support the design of the attendance system in this journal consist of the Matlab application.

Matlab is an abbreviation of Matrix Laboratory. Matlab was first introduced by the University of New Mexico and the University of Stanford in 1970. Matlab is usually used for numerical analysis and computing needs because Matlab is a mathematical programming language that is based on the properties and forms of matrices.(Fatwa et al., 2022).

C. Research Data

The research data in this journal was obtained from the results of signatures, namely 32 signatures for training data and 16 signatures for test data with the same size for all signature images of 626 x 626 pixels. Ensure that the dataset includes enough variation to train and test the model, including a variety of writing styles and conditions.

Fig 2. Training Data

Fig 3. Testing Data

D. Identification Of Signature Authenticity Using The Binary Extraction And K Nearest Neighbor (KNN) Feature Method

The research data in this journal was obtained from the results of signatures, namely 32 signatures for training data and 16 signatures for test data with the same size for all signature images of 626 x 626 pixels. Ensure that the dataset includes enough variation to train and test the model, including a variety of writing styles and conditions.

1. Identify the authenticity of the signature

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2063, Copyright ©2024
Submitted : January 17, 2024, Revised : March 30, 2024, Accepted : April 9, 2024, Published : June 15, 2024

The main focus of the research is to distinguish legitimate signatures from fake or forged ones. Develop a method or system that can automatically or semi-automatically recognize whether a signature is authentic or not.

2. Binary feature extraction method

Feature extraction is a process that identifies and extracts important information from data, in this case, signature images. Using a binary feature extraction method means that the characteristics of the signature will be represented in binary form (0 or 1), possibly involving contour patterns, textures, or other binary features. K Nearest Neighbor (KNN)

E. Binary Extraction and K Nearest Neighbor (KNN) Feature Method

1. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor is a classification method for a set of data by comparing data. This data is training data and testing data. This method uses a distance function to carry out classification. The input image will be tested based on the distance of its features to other image features in the database to obtain an image with the minimum feature distance value.

The K-Nearest Neighbor classification method carries out a matching/recognition process based on the number of nearest neighbors to determine the class. The closeness pattern applied to the K-Nearest Neighbor is defined in the form of a distance matrix such as Euclidean Distance. The data is in the form of X1 (Nur Cahyo et al., 2023).

$$dist(x1, x2) = \sqrt{\sum} (x1 - x2) \text{ n } 2 \text{ i } = 1 \quad (1)$$

Where :

x1 : training data value
x2 : test data value
dist(x1, x2) : Euclidean distance

The following are the steps to apply classification using KNN as follows:

a. Determine the parameter k (number of nearest neighbors)
b. Calculating the distance between the data to be calculated, namely the value that comes from feature extraction using local binary patterns. Based on these values calculated between the extraction training data and the test data, the distance method applied is Euclidean distance.
c. Sort the objects that have been counted from smallest to largest
d. Collect the label results based on the number of k

III. RESULT AND DISCUSSION

From this research, results were obtained with an accuracy level of 75%. Of the 32 test data, the authenticity of 24 signature image data can be detected.

In this study, a signature image object was used with the amount of data used being 32 training data and 16 test data from 8 signers. This image will later become a dataset, then the data will be extracted to take its characteristics after the dataset has been extracted, and then a pattern recognition process is carried out using the K- Nearest Neighbor Algorithm.

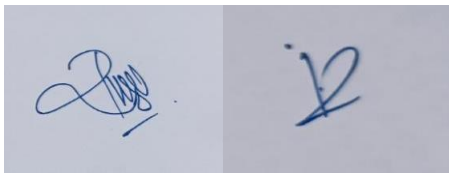

Fig 4. Image Sample

The following are the results of research that has been carried out:

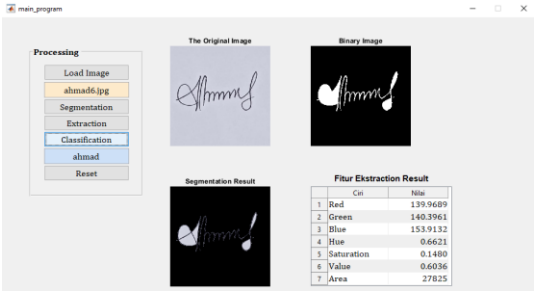

Fig 5. Image Extraction Results of "Ahmad" Signature

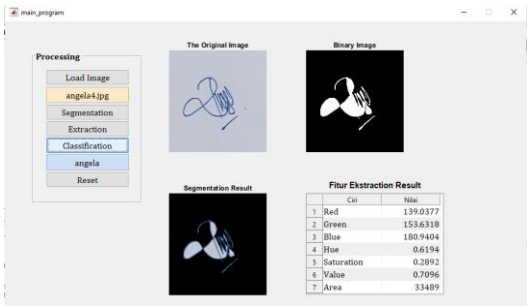

Fig 6. Image Extraction Results of "Angela" Signature

The results of extracting 32 training data images canbe seen in the following database:

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2063, Copyright ©2024
Submitted : January 17, 2024, Revised : March 30, 2024, Accepted : April 9, 2024, Published : June 15, 2024

No	Red	Green	Blue	Hue	Saturation	Value	Area	Label
1	115.5356	132.7176	163.6920	0.6124	0.4035	0.6419	25039	Ahmad
2	142.5202	146.7068	160.2563	0.6286	0.1581	0.6285	17434	Ahmad
3	139.4313	139.5827	152.4720	0.6652	0.1717	0.5979	23581	Ahmad
4	139.9689	140.3961	153.9132	0.6621	0.1480	0.6036	27825	Ahmad
5	144.4337	147.6380	161.7877	0.6360	0.1370	0.6345	29140	Angela
6	152.0171	153.5569	166.6868	0.6499	0.1179	0.6537	25862	Angela
7	149.7128	149.7079	163.4412	0.6669	0.1116	0.6409	28568	Angela
8	139.0377	153.6318	180.9404	0.6194	0.2892	0.7096	33489	Angela
9	55.3534	83.8286	137.6893	0.6110	0.6333	0.5400	16455	Camila
10	25.2001	54.1142	114.5004	0.6135	0.8167	0.4490	22358	Camila
11	89.4784	88.8118	99.6557	0.6767	0.1342	0.3908	13892	Camila
12	46.3756	78.7602	133.4735	0.6055	0.6796	0.5234	20511	Camila
13	162.3224	168.5962	181.8872	0.6122	0.1267	0.7133	14688	Datuak
14	163.9320	169.4681	182.7558	0.6174	0.1262	0.7167	17251	Datuak
15	132.4366	145.5798	169.1548	0.6079	0.3096	0.6634	14119	Datuak
16	159.9200	167.0336	180.9197	0.6088	0.1504	0.7095	14047	Datuak
17	144.7394	146.7812	159.4606	0.6459	0.1086	0.6253	16818	Fernando
18	164.7979	166.8391	179.9435	0.6450	0.1053	0.7057	36829	Fernando
19	154.1957	153.7069	165.2381	0.6737	0.1051	0.6480	19842	Fernando
20	139.2616	138.7437	149.3559	0.6748	0.1194	0.5857	12214	Fernando
21	80.8863	79.0471	90.0758	0.6957	0.1396	0.3532	1530	Itin
22	185.6880	186.1373	196.5957	0.6616	0.0622	0.7710	25624	Itin
23	163.4332	164.0475	176.0227	0.6605	0.0865	0.6903	11078	Itin
24	98.4028	98.9951	109.9706	0.6596	0.1212	0.4313	6097	Itin
25	137.3730	141.8683	162.1467	0.6517	0.1773	0.6359	143563	Putut
26	149.4674	152.6912	169.1196	0.6394	0.1204	0.6632	154931	Putut
27	152.0243	153.0220	168.9114	0.6575	0.1055	0.6624	181513	Putut
28	150.4427	150.1485	164.8722	0.6704	0.0964	0.6466	147376	Putut
29	47.4030	55.5305	77.9274	0.6239	0.4769	0.3056	7255	Zulfikar
30	78.5761	79.0584	92.5359	0.6620	0.1786	0.3629	6096	Zulfikar
31	50.6878	60.5665	83.5346	0.6178	0.4715	0.3276	7389	Zulfikar
32	72.9402	71.8378	84.4126	0.6805	0.1794	0.3310	7193	Zulfikar

Fig 7. Database of Image Data Extraction Results

Next, the extraction results will be carried out in the process of creating the K Nearest Neighbor model.

In this research, a 100:50 data division was carried out. This data division was used in the research to divide the dataset into two parts, namely 100% training data and 50% test data. This approach aims to optimize the use of available data and test model performance.

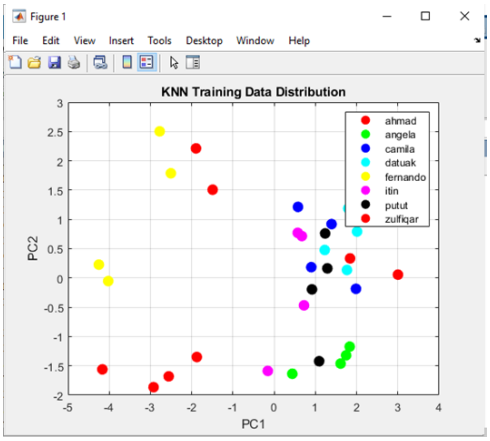

Fig 8. Distribution of KNN Training Data

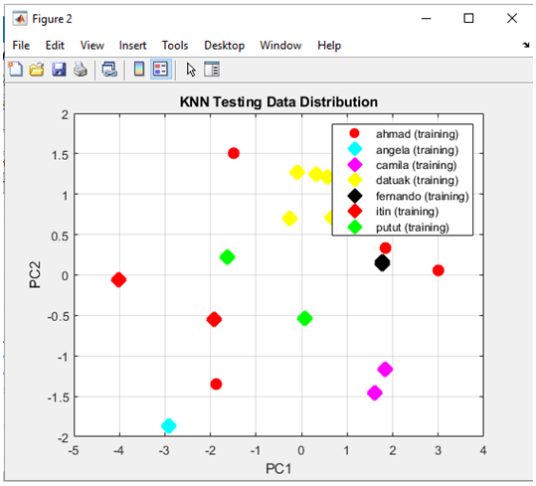

Fig 9. Distribution of KNN Testing Data

IV. CONCLUSION

This research succeeded in developing a signature authenticity identification system that uses the binary feature extraction method and K Nearest Neighbor (KNN). The binary feature extraction method is used to convert the signature into a binary representation, while KNN is used as a classification algorithm to compare and identify the authenticity of the signature based on the detected patterns.

Signature authenticity analysis, a binary feature extraction method helps reveal significant structural information in signatures, while KNN utilizes this information to compare the tested signature with signature samples whose authenticity is known. The experimental results show that the combination of these methods provides a sufficient level of accuracy in identifying the authenticity of signatures.

Practical implications of this research include the ability to efficiently verify signature authenticity in a variety of contexts, including legal document verification or transaction security. The system developed can make a positive contribution to increasing security and authentication in the process of identifying signature authenticity.

The contribution of this research lies in the application of the binary feature extraction method and the use of KNN as an effective approach in the domain of identifying signature authenticity. The success of this system shows the potential for using similar techniques in the development of broader document authentication solutions.

ACKNOWLEDGMENT

Firstly, we would like to thank Allah because Allah has given us health so that we can complete our journal without any problems we also thank the course lecturers who have provided excellent material so that we can make this journal and we also say thank you. to friends who have supported the process of working on this journal and have provided a lot of constructive input.

REFERENCES

[1] Aristantya, R., Santoso, I., & Zahra, A. (n.d.). Identifikasi Tanda Tangan Menggunakan Metode Zoning dan SVM (Support Vector Machine)

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2063, Copyright ©2024
Submitted : January 17, 2024, Revised : March 30, 2024, Accepted : April 9, 2024, Published : June 15, 2024

[2] Damanik, A. R., Hartama, D., & Sumarno, I. G. (2023). Sistem Presensi Pegawai Berbasis Digital Signatures Dan GPS Location. Ejournal.Cvrobema.Com, 1, 30–36. https://ejournal.cvrobema.com/index.php/dike/article/view/11%0Ahttps://ejournal.cvrobema.com/index.php/dike/article/download/11/5

[3] Roji, F. F., Setiawan, R., Gusdiana, R., Cahyadiputra, M. R., & Hamdi, W. H. (2023). Implementasi Tanda Tangan Digital pada Pembuatan Surat Keterangan dengan Metodologi Scrum. Jurnal Algoritma, 20(1), 199–210. https://doi.org/10.33364/algoritma/v.20-1.1301

[4] Simarmata, S. (2023). Deteksi Akurasi Tanda Tangan Menggunakan Metode Support Vector Machine Dan Zernike Moment: Studi Kasus Di SMA Harapan Bangsa. Journal Of Social Science Research, 3(3), 9880–9888. Journal Of Social Science Research.

[5] Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan metode K-Nearest Neighbor. Building of Informatics, Technology and Science (BITS), 3(4), 639–648. https://doi.org/10.47065/bits.v3i4.1408

[6] Distance, E. (2016). Pengenalan Citra Tanda Tangan Menggunakan Metode 2D-LDA dan Euclidean Distance. 3(4).

[7] Fatwa, M., Rizki, R., Sriwinarty, P., & Supriyadi, E. (2022). Pengaplikasian Matlab pada Perhitungan Matriks.Papanda Journal of Mathematics and Science Research, 1(2), 81–93. https://doi.org/10.56916/pjmsr.v1i2.260

[8] Helilintar, R. (2023). Implementasi Region of Interest (ROI) Untuk Segmentasi Citra Tanda Tangan. 7, 1248–1255.

[9] Kamila, C. (2022). Penerapan Metode Scrum pada Pembuatan Aplikasi Sistem Tanda Tangan Digital dengan QR Code Berbasis Website. Intech, 3(1), 36–41. https://doi.org/10.54895/intech.v3i1.1175

[10] Kanugroho, M. T., Rahman, M. A., & Wihandika, R. C. (2022). Klasifikasi Batik dengan Ekstraksi Fitur Tekstur Local Binary Pattern dan Metode K-Nearest Neighbor. 6(10), 4788–4794. http://j-ptiik.ub.ac.id

[11] Lodong, A. T., Widodo, A. W., & Rahman, M. A. (2023).Penentuan Mutu pada Citra Buah Jeruk Keprok menggunakan Metode Local Binary Pattern (LBP).7(4), 1616–1622.

[12] Nur Cahyo, D., Zulfia Zahro', H., & Vendyansyah, N. (2023). Pengenalan Ekspresi Mikro Wajah Dengan Ekstraksi Fitur Pada Komponen Wajah Menggunakan Metode Local Binary Pattern Histogram. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1), 822–829. https://doi.org/10.36040/jati.v7i1.6167

[13] Pengenalan, S., Dokumen, C., & Tangan, T. (2022). Journal Energy. 12(2), 54–61.

[14] Putriana, A. D., Canta, D. S., Hadisaputro, E. L., & Wahyuni, N. (2022). Implementasi Backpropagation untuk Identifikasi Tanda Tangan Digital. Jurnal Informatika Dan Rekayasa Perangkat Lunak, 4(1), 11.

[15] Sunarya, P. A. (2022). Penerapan Sertifikat pada Sistem Keamanan menggunakan Teknologi Blockchain. Journal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi, 1(1), 58–67. https://doi.org/10.34306/mentari.v1i1.139

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2063, Copyright ©2024
Submitted : January 17, 2024, Revised : March 30, 2024, Accepted : April 9, 2024, Published : June 15, 2024