Personality Prediction System Based on Signatures Using Machine Learning

I Maliki*, M A Sidik**

1,2Department of Informatics Engineering, Faculty of Engineering and Computer Science, Universitas Komputer Indonesia, Indonesia.

Email : *irfan.maliki@email.unikom.ac.id, **maulanaamsors@gmail.com

Abstract. Many methods are used to assess a person's character and personality. Some are through the face, body movements, body language, handwriting and signatures. Assessing a person's character and personality by looking at the type of handwriting and signature can be learned with the science of graphology. Detecting a person's personality system based on a signature pattern automatically is still difficult. This study aims to build a system of predicting personality based on signature patterns using machine learning. A person's personality based on a signature has many features. In this study the analysed features consist of four features, namely curve start, end strakes, middle stroke, and underline. The steps taken in developing a prediction system are model training and model testing. The method used to extract features is Principle Component Analysis (PCA) and the method for classifying is Support Vector Machine (SVM). Based on the test results using confusion matrix produces an accuracy value of 71%. It can be concluded that machine learning can be implemented to predict personality based on signatures with good accuracy.

1. Introduction

A person's character can be assessed through many things, one of it is handwriting. Reading one's character through handwriting can be learned through a science called graphology. Graphology is the art of assessing a person's character and personality by looking at the type of handwriting and signature. Information about one's personality can be used for various purposes, one of them is employee recruitment, and graphology is used as a selection device [1]. The existing detection system to detect personalities based on signature patterns automatically is still difficult.

Several studies related to personality detection have been carried out by [2], [3], [4], [5], [6] with accuracy results that vary from 50% -90%. Based on this research, several deficiencies can be developed, including the segmentation and feature extraction processes. This research will develop a focus on the feature extraction process using the Principal Component Analysis (PCA) method. The use of the PCA feature extraction method based on research [7], [8] with a good accuracy of more than 75%. The aim of this research is to find out the accuracy value of the personality detection system based on the signature pattern using the Support Vector Machine (SVM) and PCA methods.

Signatures represent a person's public character, self-ownership, insensitivity, inflection, diversity, and imagination [9]. This study uses 4 signature features based on research [10] namely curve start, end strakes, middle strakes, and underline. Consists of 10 classes of backward curved, curved sharp, curved smoothly, straight ascending, straight descending, no end strakes, middle strakes, no middle strakes, underline, and no underline. Features that represent these personalities can be seen in table 1.
Table 1. Signature Features

| NO | Features    | Image Type       | Personality                                      |
|----|-------------|------------------|--------------------------------------------------|
| 1  | Curve start | Curved backwards | Comfortable going past                           |
|    |             | Curved Sharply   | To formulate a sharp mind                        |
|    |             | Curved smoothly  | Be careful, friendly, diplomatic                 |
| 2  | End streak  | Increase         | Open, foresight, desires ahead, confident       |
|    |             | Down             | Lacking spirit, realistic thinking, lack of confidence, easily discouraged |
| 3  | Middle streaks | Middle streaka | possessive                                       |
| 4  | Underline   | Underline        | Have unique idea and thinking, need support to make decisions, and have reliability in the lead |

2. Method

2.1. Data
In this research, the input data used is digital image. Where the media used is white paper for image photo processing using a scanner. After scanning, process the data collected is transformed into a digital image with a .jpg file type. The input image used has a size of 600x800 pixels. Data that will be used are 530 data signatures with 53 respondents who will write 10 signatures.

2.2. Personality prediction system
The following are the stages of the process undertaken for the personality prediction system can be show in Figure 1. The following are the stages of the personality prediction system process:

1. **Image Training and Image Testing**
   In the training and testing phase, images are entered according to a predetermined size of 600x800 pixels.

2. **Image Processing**
   At this stage several processes are carried out as follows, namely:
   a. **Grayscale**, converts digital images into gray images.
   b. **Edge detection**, determine the edge of the object (signature) to be analyzed. At this stage, it used the canny method [11]. Canny edge detection is used to extract edge values in grayscale images.
   c. **Object segmentation**, the segmentation process aims to divide the image into base elements according to specified criteria. Therefore, the image taken is only important objects. The segmentation process consists of marking objects, where the object here is a character that is in the image of the canny edge detection results. The method used at this object segmentation stage is the Connected Component Labeling method [12].
   d. **Resize**, In the process of segmenting the object pixel size is different for each signature image. The image that has been cut in the object segmentation process produces a variety of sizes in the example used this image produces a size of 288 x 211 pixels. Therefore, to uniform the pixel size the resize process is used by using a pixel size of 36 x 69 pixels.
   e. **Binaryization**, in this process the matrix results of the process that has been changed in value to produce values 0 and 1 in the matrix so that it can be processed for the feature extraction and classification processes.
3. Vertical and Horizontal Segmentation
Vertical and horizontal segmentation are the stages of segmentation that are used to distinguish between image features. Vertical segmentation is processed by dividing three parts of the image area vertically, left, center and right. The area division is done by cutting the size of the 36x69-pixel binaryization matrix to a matrix size of 36x23 pixels. Whereas horizontal segmentation is processed by dividing three parts of the region horizontally, namely the top, middle, and bottom. The area division is done by cutting the size of the 36x69 pixel binaryization matrix to a 12x69 pixel matrix size.

4. Feature Extraction (PCA)
PCA can be used to reduce the dimensions of the input data to be smaller without losing significant information in describing the overall data. The stages in the PCA process are based on [13]. The following is an example of one of the features produced from the PCA process.

\[
0.186650391 -0.042661 -0.01461 -0.01536 -0.021 0.030896 \cdots 0 0
\]

5. SVM Classification
In the classification stage, labeling -1 (negative) and +1 (positive) will be made to determine and detect personalities based on signature patterns. After the feature extraction process is performed, a value is obtained from the PCA feature extraction; a positive label indicates the feature on the signature. Negative labels do not indicate the features of the hand. In this study, the SVM method is based on research [14] used, namely one-against-all.

\[\text{Figure 1. Flowchart For Personality Prediction System}\]

3. Results and Discussion
Accuracy testing in this study was conducted to determine the value of accuracy. This test aims to test the stability of accuracy if tested with training data and different test data. This test data used for training data of 900 signature images for the whole class, by taking 9 images from 10 subsets of each class, with a total of 90 signature images for each class. Distribution of training data and test data, 60% for training data and 40% for test data, so that the data used for training data are 54 and test data are 36 signature images for each class. The following are the results of the classification using the Confusion Matrix method.
Table 2. Initial Curve Start Test Results

| Actual       | Curved backwards | Prediction | Curved Sharply | Curved smoothly |
|--------------|------------------|------------|----------------|-----------------|
| Curved backwards | 23               | 3          | 10             |
| Curved Sharply         | 4                | 24         | 8              |
| Curved smoothly         | 3                | 0          | 33             |

based on the table 2, the accuracy can be determined as follows:

\[
\text{Accuracy} = \frac{23+24+33}{108} = 74.07 \%
\]

Table 3. End strakes Feature Test Results

| Actual           | Increase | Down | No end strakes |
|------------------|----------|------|----------------|
| Increase         | 31       | 1    | 4              |
| Down             | 10       | 19   | 7              |
| No end strakes   | 6        | 5    | 25             |

based on the table 3, the accuracy can be determined as follows:

\[
\text{Accuracy} = \frac{31+19+25}{108} = 69.44 \%
\]

Table 4. Middle Streak Feature Test Results

| Actual            | Prediction | Middle streak | No middle streak |
|-------------------|------------|---------------|------------------|
| Middle streak     | 21         | 15            |
| No middle streak  | 6          | 30            |

based on the table 4, the accuracy can be determined as follows:

\[
\text{Accuracy} = \frac{21+30}{72} = 70.83 \%
\]

Table 5. Underline Feature Test Results

| Actual            | Prediction | underline | No underline |
|-------------------|------------|-----------|--------------|
| Underline         | 24         | 12        |
| No underline      | 10         | 26        |

based on the table 5, the accuracy can be determined as follows:

\[
\text{Accuracy} = \frac{24+26}{72} = 69.44 \%
\]

There are several findings when testing as follows:
1. The use of object segmentation using the connected component labeling method is still lacking, when there is noise that is far from the object signature, it affects the process of vertical and horizontal segmentation.
2. Cutting the object set pixel size 36x23 for vertical segmentation and 12x69 for horizontal segmentation, if there is noise that is far from the signature object, will have an effect when reading the signature pattern to recognize the beginning of the curve, the final stroke, the center stroke, and the bottom line.
Based on these results indicate that this research needs to be continued to achieve better results as in previous studies [3,6].
4. Conclusion
Based on the results of tests that have been done, the accuracy value obtained from the feature extraction of Principal Component Analysis and classification using Support Vector Machine the average accuracy of the method performance is 71%. It can be concluded that machine learning can be implemented to predict personality based on signatures with good accuracy. Signature data is abstract and each signature has its own pattern by adding signature training data and varying, it is likely to be able to recognize the personality of each signature accurately.

Acknowledgment
We would like to thank UNIKOM for supporting this research.

References
[1] Prasetyono, D. S. 2010. Bedah Lengkap Grafologi.
[2] Fadhilla, M., Saf, M. R. I. A., & Sahid, D. S. S. 2017. Pengenalan kepribadian seseorang berdasarkan pola tulisan tangan menggunakan jaringan saraf tiruan. Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), 6(3), pp.365-373.
[3] Prasad, S., Singh, V. K., & Sapre, A. (2010). Handwriting analysis based on segmentation method for prediction of human personality using support vector machine. International Journal of Computer Applications, 8(12), pp.25-29.
[4] Djamal, E. C., & Ramdlan, S. N. 2013. Pengenalan Pola Tanda Tangan Menggunakan Multilayer Perceptron Dalam Identifikasi Kepribadian. SESINDO 2013, 2013.
[5] Hidayat, T. 2015. Pengenalan Pola Tanda Tangan Menggunakan Learning Vector Quantization Dalam IDNentifikasi Kepribadian (Doctoral dissertation, Universitas Komputer Indonesia).
[6] Prasetya Utama, B. 2018. Support Vector Machine Dalam Sistem Pendeteksi Kepribadian Berdasarkan Pola Tanda Tangan (Doctoral dissertation, Universitas Komputer Indonesia).
[7] AM, A. S., & Augustine, P. 2017. Efficient Brain Tumor Classification using PCA and SVM. International Journal of Research in Engineering, IT and Social Sciences, 7, pp.1-7.
[8] I Djahari, K. 2017. Implementasi Metode Principal Component Analysis Dan Support Vector Machines Dalam Mendeteksi Plasmodium Malaria Pada Citra Sampel Darah (Doctoral dissertation, Universitas Komputer Indonesia).
[9] Kedar, S., Nair, V., & Kulkarni, S. 2015. Personality identification through handwriting analysis: a review. International Journal of Advanced Research in Computer Science and Software Engineering, 5(1).
[10] Djamal, E. C., Darmawati, R., & Ramdlan, S. N. 2013. Application image processing to predict personality based on structure of handwriting and signature. In 2013 International Conference on Computer, Control, Informatics and Its Applications (IC3INA), pp. 163-168.
[11] Kabade, A. L., & Sangam, V. G. 2016. Canny edge detection algorithm. International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), 5(5), pp. 1292-1295.
[12] Yudhistiro, K. 2017, September. Menghitung Obyek 2d Menggunakan Connected Component Labeling. In Seminar Nasional Sistem Informasi (SENASIP). 1(1)
[13] Dunteman, G. H. 1989. Principal components analysis, pp. 69.
[14] Maliki, I., & Jarockohir, F. S. 2019. Facial Expressions Recognition Using Markov Stationary Feature-Vector Quantization and Support Vector Machine Method.In IOP Conference Series: Materials Science and Engineering, 662(2), pp. 022023.