HANDWRITING BASELINE PATTERN RECOGNITION ON SENTENCES USING ALGORITHM BACKPROPAGATION

Rasyida Nurlailani

Informatics Engineering Program
Engineering and Computer Science Faculty Indonesia Computer University
Jl. Dipati Ukur No. 112-116 Bandung
E-mail : rasyida_nurlailani@yahoo.com

ABSTRACT

Graphology is the study of a person's character by analyzing a person's handwriting. Analyzing handwriting is very helpful in many fields today, for example in the field of education, crime and can be used as counseling. One of the ways used in graphology to know a person's character, is by analyzing the pattern of a base line of writing sentences tangan.Pengenalan handwriting can use neural networks by classifying input vectors into classes that have been determined, one of the methods artificial neural network is a backpropagation

Image data used for the analysis through the stages of the process of image processing such as the first cropping, grayscale, edge detection, thresholding, normalization and image segmentation. The pattern of the base line used in this study are four basic grid pattern, namely, line patterns dasaar rose pattern down the base line, a straight base line pattern and line pattern random basis.

After making the process of learning and recognition that has been done by using the back propagation algorithm can be used to recognize patterns and handwriting base line values obtained accuracy of 54.33%. Percentage introduction of each - each pattern: on the basis of straight-line pattern 82.35%, on the pattern of a base line rose 55%, to a pattern base line fell by 50% and on the basis of a random pattern of lines 30%.

Keywords: Graphology, Neural Networks, Backpropagation, Image Processing, Baseline, Handwriting.

1. INTRODUCTION

1.1 Background

The work of humans is determined by the nature and character of one's self. In France and Switzerland and more companies are looking for employees in accordance with the character of the company's criteria. Much can be done to be able to judge a person's character, there are through the face, body language and handwriting. To read a person's character through handwriting can be learned through a science called "Graphology". Graphology is the science used to determine the character and personality of a person through writing or graffiti inscribed on paper or other media. Other terms referred to handwriting analysis. In graphology there are several ways to determine a

person's character, by looking at the pattern of handwriting as a basic grid pattern, the font size, the slope of the letters and words, the emphasis pen, spacing between letters and spacing between words [1].

Handwriting pattern recognition is done with the help of software using Artificial Neural Network (ANN). JST is defined as an information processing system that has characteristics resembling human nerve tissue. ANN created as a generalization of mathematical models of biological neural networks [2] One method of ANN, which has an excellent learning capability is Backpropagation algorithm developed by Rumelhart, Hinton and Williams, in 1986.

In view of the results of previous studies, Signature Identification Using Neural network propagation Balik (Backpropagation) denagn t Degree of success of 88% [3]. While the research Face dection Using Backpropagation Neural Networks produces an accuracy value of 90% [4]. And the results of research on fingerprint pattern recognition based on wavelet transform and artificial neural network Backpropagation at yield levels of accuracy is 81.67% [5].

Based excess Backpropagation method in previous studies, in this study the researchers will use backpropagation algorithm is used to find the value of the accuracy of the algorithm back propagation algorithm on the introduction of basic sentence patterns of lines.

1.2 Purpose and Objective

Purpose of this study is to analyze the propagation algorithms in applications that can to know the characteristics of human beings on the basis of handwriting line pattern.

Objectives to be achieved in the research was to determine the accuracy of the algorithm Backpropagation value on the bottom line of sentence patterns.

1.3 Graphology

Graphology is science of handwriting resulting by headwork, based on that thing many graphology expert called handwriting is "writing of brain". The expert say that graphology is na empirical Science which can be proven by phenomena in population and there quantification result or result from statistika test can be accounted.

Graphology from graphos which meaning stroke and logos is science. Based on that mean can be said graphology is a branch of science studies a hand stroke. Handwriting here is not always the letter can be analyzed, but a hand stroke can also have its own meaning. Graphology is also said as art for reading handwriting, because a writing each people has special characteristics.[1]

1.4 Determining basic Handwriting Analysis

There are some basic determination of handwriting analysis, here are the basics in the determination of handwriting analysis: [1]

1. *Margin Posts*

   Margin writing in the science of graphology shows the person and social.
2. *Spacing or spacing between letters*

   In the science of Grammar, space is one type of word separators. Spaces are empty areas that are deliberately held as a giver of space and separation, both separators letters, words, numbers or punctuation. There are no official numbers or spaces that can be used as a benchmark for sure or authorized in this space. Broadly speaking, the space is a reflection comfortable or not a person (the author) to others. Spacing or distance in the text there are three parts, namely:
3. *The distance between words*

For spacing or distance between the written word is one letter. If the spacing between words is less than 1 letter "m" then that space is small, which means the person has the character of those who have social natures. Meanwhile, if the spacing between the letters more than one letter "m" then said width which means the person is a person who is individualistic.

4. *Spacing between letters*

   For spacing or distance between the letters on paper usually 1-2mm. So if spaced posts between letters someone less than 1mm will be said to be small, which means the person has the character of people in need of attention and encouragement from others, while more than 2 mm will say big meaning the person has a full character would caution against someone ,

5. *The distance between rows.*

   For spacing between lines spaces is almost the same as the spacing between words and between letters. For a small line spacing indicate that the person is a happy person directly involved with something, while the wide spaces indicate that the person is a person who likes to work behind the scenes.

6. *Basic line posts*

   Base line or baseline is an imaginary line from each post. In the science grofologi, diperumpamakan base line as the base line for achieving the goal. Basic line has several variations including writing

   a. *The Bottom Line Straight*

      The type of people with this paper are those who have a high enough concentration, it has energy large, tend to try to arrange, usually prepare themselves when dealing with other people, this is very controlled, follow the rules and diplomatically, his orderly, filled with rules and schedule

   b. *The Bottom Line Up*

      The type of people with this paper is an optimist, active, positive thinking, always trying mecapai ideals.

   c. *The Bottom Line Down*

      The type of people with ulisan i ni t often see everything in terms of bad, negative thinking, often feel cynical about the surrounding environment, often feel has always been a victim of circumstances, life is always complicated, difficult and led to bizarre.

   d. *The Bottom Line of Uncertainty*

      The type of people with this paper has the feeling unstable, sometimes upbeat, laughing, but then arrived - arrived lackluster even cry, have a poor self-control, feeling lost sense of security

   e. *The Bottom Line Convex*

      The type of people with this paper has a strong sense of enthusiasm, but gradually melempen / fade because the energy tekuras when reaching the top. Additionally, this article identifies the direction of the fast emerging temperament, but also quickly meghilang

   f. *Concave Bottom Line*

      The type of people with this paper identifies a trait that is full of doubts in terms of acting or making a decision, but this feeling eventually daoat soon resolved and spirits moving up showed their tenacity in him.

7. *Font Size*

   Font size further elucidate the extent to which she sees herself and how she should be seen by their social environment.

8. *Pressure Writing*

Writing pressure on a person related to a person's emotions to write.

9. *Zona writing*
10. *The slope of the article*
11. *The slope of the person writing related to the nature and tendency of their actions.*
12. *This type of literature*
13. *Type someone writing there are two types of printing and writing is continued, and in graphology will relate to social interaction.*
14. *Writing Speed*

    Speed someone in writing to provide an overview of the potential of a person's intelligence and speed of thought. Someone writing speed is also closely linked with the speed of one's acts and spontaneity.
15. *Typeface beginning and end*

    Early and late forms of letter writing can be used as illustration for us to see the character and nature of the author.

1.5 Backpropagation

This learning method is included network architecture multilayer feedforward neural networks. Generally, the network consists of a number of units of neurons as the input layer, one or more layers of vertices neuron hidden layer and a layer of output neurons knots. Incoming input signal propagated from beginning to end layer towards the output layer. This type of network is the result of a single layer perceptron architecture generalization, so-called by the multilayer perceptron (MLPs). Error back propagation is an algorithm that uses the principle of MLPs supervised learning. Propagation occurs (towards the input layer) after the network produces output that contains the error. In this phase, the entire weight of the synapses that do not have a zero activation on the network will be adjusted to correct / minimize error occurred (error correction rule). For network learning, the couple propagation phase ahead and behind is repeated for a set of training data, then repeated a number of epoch (one session throughput for the entire training data in a learning process network) until the error occurs reach small limit certain tolerance or zero , In Figure 1 below outlines the workings of neural networks using backpropagation method:This is step-by-step learning algorithm from learning vector quantization method:

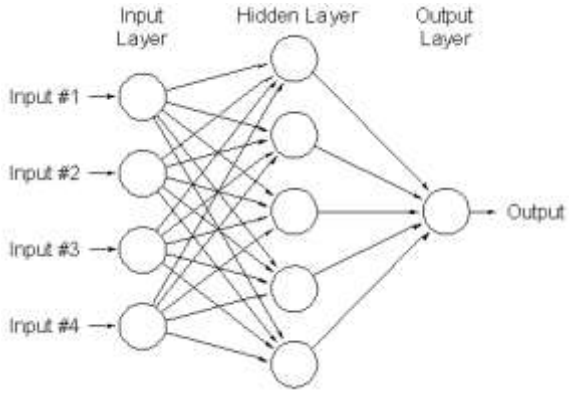


Figure 1. Method Backpropagation

Propagasibalik algorithm can be divided into two parts [12] [13]:

training algorithm

Training algorithm consists of three stages: the stage of advanced propagation, propagation phase reverse, and phase changes in weight.

algorithm application

Algorithm application only uses feed-forward stage only.

Training algorithm to network with one hidden layer (with a binary sigmoid activation function) is as follows:

Step 0: Initialize all weights with small random numbers.

Step 1: If the termination condition is not met, do steps 2-9.

Step 2: For each pair of training data, perform steps 3-8.

Phase I: Propagation forward

Step 3: Each unit receives input signals and forward them to the hidden units on it.

Step 4: Calculate all the hidden unit output in

$$z_in_j = v_{0j} + \sum_{i=1}^{n} x_i \, v_{ji}$$

Where:

v_{0j} *= weight bias*

x_i *= binary inputs*

v_{ji} *= weight initialization v*

$$z_j = f(z_in_j) = \frac{1}{1 + e^{-z_in_j}}$$

where:

z_in_j *= input signal*

e = exponent

Step 5: *Calculate all the network output in y_k unit* $y_k(k = 1,2, \dots, m)$.

$$y_in_k = w_{k0} + \sum_{j=1}^{n} z_j \, w_{kj}$$

where :

w_{k0} *=weight bias* w_{kj}

z_j *= unit of output*

w_{kj}*=weight initialization w*

$$y_k = f(y_in_k) = \frac{1}{1 + e^{-y_in_k}}$$

where

y_in_j *= output signal*

e = exponen

Determine the minimum error and maximum error (k=1, ..., m).Error can be reckoned as the

$$err = \frac{1}{2}\sum_k (T_k - y_k)^2$$

where

T_k *= Target*

y_k *= outputvalue*

etermine the number of epoch minimum and maximum epoch. As in

$$maks_epoch = \frac{sum\ weight}{maks_error}$$

Fase II : Propgasi retreat

Step 6 *: Calculate the factor δ unit of input and output based on the error in each unit of output* $y_k(k = 1,2, \dots, m)$

$$\delta_k = (t_k - y_k) f'(y_{in_k})$$

$$\delta_k = (t_k - y_k) y_k (1 - y_k)$$

wherea

y_in_j *= output signal*

t_k *= Target*

y_k *= output value*

and δ_k *k an error unit unit which will be used to change the weight of the layer below (step7) Calculate the weight* w_{kj} *rate changes (which will be used later to change the weights* w_{kj} *) with the accelerated pace of* α.

$\Delta w_{kj} = \alpha \delta_k z_j; k = (1,2, \dots., m; j = 0,1, \dots, p)$

Where:

a *= learning rate*

δ_k *= factor error unit*

z_j *= input signal*

Calculate the refractivecorrection (Δw_{0k}) *to correct the value* Δw_{0k}

$$\Delta w_{0k} = a\delta_k$$

Where:

a *= learning rate*

δ_k *= factor error unit*

Step 7 : *Calculate the factor δ unit hidden by an error in each unit of output*

$$\delta_in_j = \sum_{k=1}^{m} \delta_k w_{kj}$$

Where:

δ_k *= factor error unit*

w_{kj} *= weight initialization w*

Faktr hidden δ units :

$$\delta_j = \delta_in_j f'(z_in_j) = \delta_{in_j} z_j(1 - z_j)$$

Where:

δ_in_j = *error correction*

z_j = *unit of output*

Calculate the weight v_{ji} *rate changes (yang which will be used later to change the weights* v_{ji}*)*

$$\Delta v_{ij} = \alpha \delta_j x_i$$

Where:

δ_j = *factor hidden units*

x_i = *input value*

a = *learning rate*

Calculate the weight change rate v_{0j}

$$\Delta v_{0j} = \alpha \delta_j$$

Where:

δ_j = *factor hidden units*

a = *learning rate*

Fase III : changes in weight

Step 8 : *count all the weight change*

Change the line that shows the weight of the unit output:

$w_{kj}(new) = w_{kj}(old) + \Delta w_{kj}$

Where:

Δw_{kj} = *weight change* w_{kj}

Changes to the weight of the line that shows hidden units:

$$v_{ji}(new) = v_{ji}(old) + \Delta v_{ji}$$

Δv_{ij} = *weight change* Δv_{ij}

Step 9 :

Test conditions of termination (end of iteration) dismissal testing is done by calculating the error value. If you are looking for the error value that does not have a target value then the formula to find the value of error is as follows :

$$1 - (y)^2$$

If you already have a target value then the formula to find the error value are as follows :

$$\frac{1}{2}\sum_k (T_k - y_k)^2$$

Where

T_k = *Target*

y_k = *output value*

1.6 Digital Image Processing

Literally, the image (image) is a two-dimensional image in the plane (two dimensions). Judging from the mathematical point of view, the image is a continuous function (continue) on the two-dimensional light intensity on the field. The light source illuminates an object, the object reflects back a portion of the light beam. This light reflection is captured by optical devices, such as the human eye, a camera, and a scanner (scanner), so the shadow of the object called the image recorded [14].

1. *Increasing the brightness and contrast*

   Increased brightness and contrast is an improvement in image change the image that has a value of brightness and contrast value becomes a value less brightness and better contrast values.

2. *Busting Noise*

   Removal of noise is the process of removing noise from an image with the distortion state or has noise.

3. *Search object shape*

   Some types of image processing operations is as follows: [6]

a. *Ashing (Grayscale)*

An image conversion process with the actual color (true color) is the image of gray (grayscale). Changing the image into grayscale form does this by taking the pixel values of the

input image is then calculated using the following equation:

*Grayscale = ((0.2989 * R) + (0.5870 * G) + (0.1141 * B))*

b. *The floating (thresholding)*

This operation used to change the image of the gray-scale format that has a value of more than two binary image that has only two values: 0 and 1.

c. *Cutting (Cropping)*

Cut off unwanted parts of an image

d. *Pengskalaan (Scaling)*

Changing the size of the image to be larger or smaller

1.7 Testing BlackBoks

Testing that ignores the internal mechanism of the system or its components and focus solely on the output produced which responds to selected inputs and execution conditions. Tests were conducted to evaluate compliance with the system or components with specific functional requirements. [11]

1. *Interest Black Box is a find:*
2. *Functions that are incorrect or missing*
3. *errors interface*
4. *Error in data structures or external database access*
5. *Error on performance*
6. *Error during initialization and termination*
7. *The system's sensitivity to certain input values*
8. *Limitations of the data*

1.8 Testing Accuracy

Accuracy is how close a number of measurement results against the actual figure (true value or reference value). This level of accuracy is obtained by calculation according to the following equation: [12]

$$Accuracy = \frac{\sum\ TestingData\ True}{\sum Data\ Test} \times 100\%$$

2. RESEARCH CONTENT

2.1 Analysis and System Design

Here is analysis and design from a system built:

2.1.1 Problem Analysis

2.1.2 Input Data Analysis

Input data that need is an image. Image can be a result of scanning. Extension image used is '.jpg' or '.png'. Image used as input data will divided into several class based on analysis result. Here are available classes for input image:[1]

Table 1 Class Input Data



2.1.3 Process Analysis

In general, system have some process in doing analyze against input data image. Each process have a each role for recognition input image. Here figure process analysis:

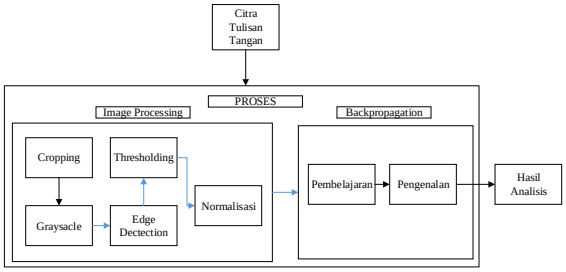


Figure 1 Process Analysis

2.1.4 Input Image

The procees of taking iamge that used as input is a handwriting people was scanned or photo with format '.jpg' or '.png'. Here a figure input image that used for input process:



Figure 2 Input Image

2.1.5 Image Preprocessing

In this process there several process image processing that need in simulation will built begin from scaling, grayscaling, edge detection, thresholding, segmentation and normalize. Here is a figure steps performed in image processing process:

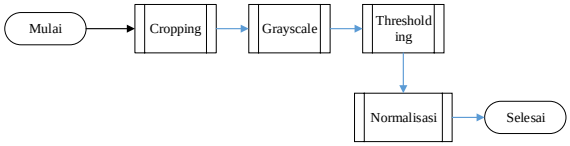


Figure 3 Image Preprocessing Process

2.1.5.1 Cropping

Cropping or cutting is done to get the image of the image that is used as input data. From the initial image measuring 77 x 1390 pixels and after doing the process of cropping the image size of 66 x 1216 is the cropped image of the input image:



Figure 4 Result of Cropping Process

2.1.5.2 Grayscale

Grayscaling is a image processing was change normal image (RGB) to grayscale image. Change image to grayscale do with take a pixel value from input data then calculate with this equation:

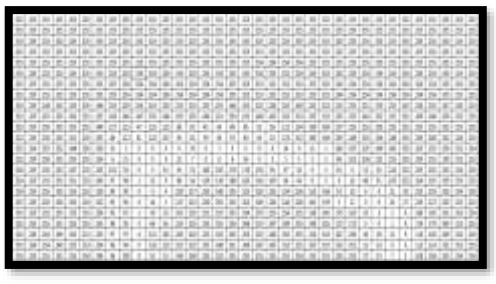


Figure 5 Result of Grayscale Process

2.1.5.3 Thresholding

Thresholding is a process to representation digital image from edge detection process to binary image with 0 and 1. 0 and 1 get from change threshold intensity value which has been determined. If pixel value in upper value intensity then change 0 which mean that pixel is background image, meanwhile if pixel value lower than value intensity then change 1 which mean that handwriting character.

Here is figure result of thresholding process:

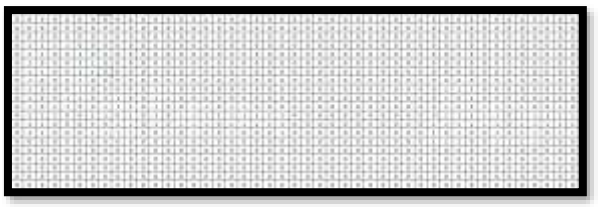


Figure 6 Result of Thresholding Process

2.1.5.4 Normalize

Normalize is similarity size of image to be same size. Because resulting data can different size. Therefore to get same similar data, accurate and consistent from each sample, image data will normalize with pixel size 20x2500 pixel. Here figure result of normalize:



Figure 7 Result of Normalize Process.

2.1.6 Analysis Requirements System Functionality

Here is a context diagram from system:

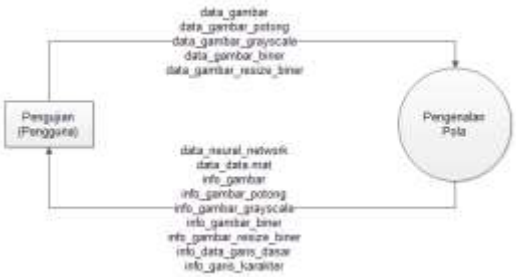


Figure 8 Context Diagram

Here is a data flow diagram level 1:

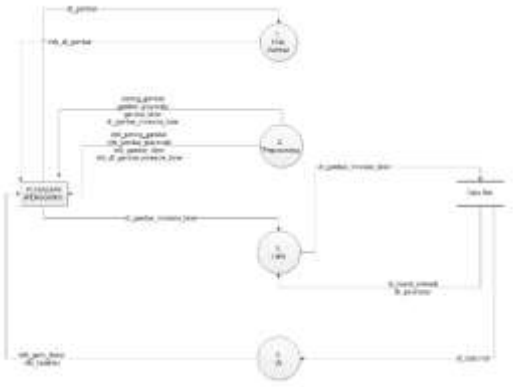


Figure 9 Data Flow Diagram Level 1

2.2 Implementation and System Testing

The following will describe about implementation and testing of system.

2.2.1 Implementation Environment

Implementation environment is a specification of hardware and software which the system will be used. Here is the implementation environment:

1. *Hardware*

   Here is hardware spesification in developed this simulator:

   a. *Processor : Intel Core i5-4210u 2,7 GHz*
   b. *Memory : RAM DDR 3 4 GB*
   c. *Harddisk : 1 TB*
   d. *VGA : NVIDIA GeForce 840m*
   e. *Monitor 14" with resolution 1366 x 768*
   f. *Keyboard*
   g. *Mouse*
   h. *Scanner*

2. *Software*

   Here is software used in developer simulator:

   a. Operating System : Microsoft Windows7
   b. Matlab R2012b

2.2.2 Interface Implementation

Here is interface implementation from simulator:

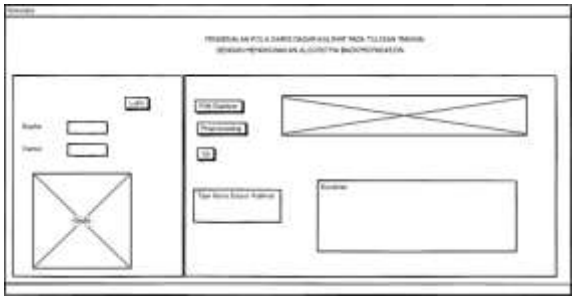


Figure 10 Interface Beranda

2.2.3 Testing

Here is testing result from developer simulator:

1. *Evaluation of functional testing*

Based on functional testing result done do, can be concluded that simulator was built has been running as expected.

2. *Evaluation of accuracy testing*

From 5 accuracy testing performed, the highest accuracy obtained from 5th accuracy testing. Accuracy testing 5 do with using combination input parameter as below:

a. Learning rate : 0.05
b. Minimum error : 0.05
c. Maximum Epoch : 100

Here is accuracy testing result :

Here is result accuracy from accuracy testing using this equation:

$$Baseline\ accuracy\ Straight = \frac{14}{17} x100\% = 82.35\%$$

$$Accuracy\ Baseline\ Rise = \frac{11}{20} x100\% = 55\%$$

$$Basic\ Line\ Accuracy\ Down = \frac{10}{20} x100\% = 50\%$$

$$Accuracy\ Baseline\ Random = \frac{6}{20} x100\% = 30\%$$

Accuracy value of the average of the base line pattern is equal to 54.33%

3. CLOSING

3.1 Conclusion and Suggestion

Based on the results of research, analysis, design and manufacturing system simulator base line pattern recognition of handwriting using backpropagation algorithm then kesimpulanya is back propagation method can be used to recognize the pattern of a base line of handwriting, with accuracy values obtained at 54.33%. Percentage introduction of each - each pattern: on the basis of straight-line pattern 82.35%, on the pattern of a base line rose 55%, to a pattern base line fell by 50% and on the basis of a random pattern of lines 30%.

To further improve the performance of the system are made, it is proposed some suggestions as follows:

1. *The introduction of the characters in handwriting not only see the bottom line of sentence patterns, but from the tilt of the writing, spaces between words, writing pressure, size and margi literary writings.*
2. *Search the specific value of the image of the handwriting can be done by other methods.*

REFERENCES

[1]https://www.karohs.com/books/GraphologyBooks ,Handwriting Analsysis.

[2] Hermawan, Arief. Jaringan saraf tiruan, Teori dan Aplikasi. Yogyakarta : Andi.2006

[3] Achmad Hidayanto, R.Rizal Isnanto , Identifikasi Tanda Tangan Menggunakan Jaringan Saraf Tiruan Perambatan Balik (Backpropagation).

[4] Anifuddin Aziz, Muhamad Haikal, Face Detection Using Backpropagation Neural Network,Universitas Gadjah Mada., Yogyakarta

[5] Sayful Hakam, Adiwijaya , Andrian Rakhmatsyah, Pengenalan Pola Sidik Jari berbasis Transformasi Wavelet dan Jaringan Syaraf Tiruan Backpropagation., Teknik Informatika. Universitas Telkom.

[6] tutorialspoint.com, Software Devlopment Life Cycle, tutorialspoint.com

[7] Ludvianto, Bayu. Analisis Tulisan Tangan. Gramedia, Jakarta. 2011

[8] M. H. Purnomo dan A. Kurniawan, Supervised Neural Networks dan Aplikasinya, Yogyakarta: Graha Ilmu, 2006

[9] Hermawan, Arief. 2006. Jaringan Saraf Tiruan Teori dan Aplikasi. Yogyakarta: Andi.

[10] Puspitaningrum, Diyah. 2006. Pengantar Jaringan Saraf Tiruan. Yogyakarta: Andi.

[11] S. Kusumadewi, Artificial Intelligence (Teknik dan Aplikasinya), Yogyakarta: Graha Ilmu, 2003

[12] Siang, Jong Jek. 2009 Jaringan Saraf Tiruan & Pemrogramannya Menggunakan Matlab. Yogyakarta: Andi.

[13] F. L, Fundamentals of Neural Network: Architectures, Algorithms, and Applications, New Jersey: Prentice-Hall, 2002

[14]http://informatika.stei.itb.ac.id/~rinaldi.munir/Buku/Pengolahan%20Citra Digital/ diakses 1 Maret 2015

.