Apllication of Naïve Bayes Algorithm in Education Games Learn to Write Aksara Bali

I Made Satya Vyasa[1)], I Gede Arta Wibawa [2)], I Gusti Ngurah Anom Cahyadi Putra [3)], I Gusti Agung Gede

Arya Kadyanan [4),] Ngurah Agus Sanjaya ER[5),] I Putu Gede Hendra Suputra [6)]

[a)]Informatics Department, Faculty of Mathematics and Natural Sciences, Udayana University
South Kuta, Badung, Bali, Indonesia
[1] satyavyasa77@gmail.com
[2] gede.arta@unud.ac.id
[3] anomcp@unud.ac.id
[4] gungde@unud.ac.id
[5] agus_sanjaya@unud.ac.id
[6] hendra.suputra[@]unud.ac.id

Abstract

Balinese script is a script from the Balinese area that was commonly used by people in ancient times to describe a word, there are many ways to keep this script sustainable and not extinct. One of them is by making a Balinese character recognition game. Which in this application the user will provide input in the form of Balinese characters written on the application, which then the input will go through a preprocessing process and then proceed with diagonal feature extraction, then the results of the feature extraction will go through a classification process using the Naïve Bayes method. The result is a web-based application that can recognize Balinese script writing using the Naïve Bayes classification method with an accuracy rate of 70.3% and get a good response from every respondent who has tested the application.

Keyword : *Balinese Scription, Diagonal Feature Extraction, Naïve Bayes.*

1. Introduction

Balinese script itself is divided into 4 types when viewed in terms of its function, namely, Wreastra script which is a Balinese script commonly used by Balinese people, both for writing place names, object names, and the like. The next is the Swalalita script. This script is used by the Balinese people in writing Kekawin (poems in Balinese culture) and also seloka derived from Kawi and Sanskrit languages. After that there is the Wijaksana script, this script is a script that is sacred by Hindus in Bali because this script is used to write religious mantras.

By utilizing the media technology that is currently developing, one of which is a smartphone. Almost everyone around the world must have this device, including the people on the island of Bali. By utilizing this technology, the learning process will run more pleasantly, Walat. W also emphasized that the use of media in the learning process is one of the efforts to create more meaningful and quality learning [7]. One of the features of smartphones that can be applied as a learning tool is games. Using games as learning media can give a relaxed and fun impression to attract students' interest in learning.

In the character recognition process, feature extraction and methods are needed to classify the data so that the data held can be separated into certain data classes. One of the classification methods is the Naive Bayes method. Based on research conducted by Arif Muhamad, this nave Bayes classification has a reasonably high accuracy rate of 85.10%.

2. Literature
2.1 Aksara Bali

Aksara comes from Sanskrit, which by type is classified as a noun neutral or sissy which means letters, syllables, or words [1]. One of the scripts used by people in Indonesia to date is the Balinese

Vyasa, dkk.
Application of Naïve Bayes Algorithm in Educational
Game Learn to write Aksara Bali

script. Balinese script is a sign or symbol used by Balinese people to write words or sentences using Balinese, but in its development apart from being written in Balinese script, Balinese is also often written in Latin script.

Based on its function, the Balinese script can be divided into four, namely the Wreastra script, Swalelita script, Wijaksana script, and Modre script. The most common script used by Balinese people is the Wreastra script, where this script is used to write street names, building names, and the name of an agency. This Wreastra script consists of 18 consonant scripts, namely ha, na, ca, ra, ka, da, ta, sa, wa, la, ma, ga, ba, nga, pa, ja, yes, nya [1].

2.2   Color Model in Image

In applications that use the concept of digital image processing, the digital image is known into several types, namely the RGB color model which generally in this color image each pixel has a certain color, namely red (red), green (green), and blue (blue). This RGB color space can be visualized as a cube, which on the 3 axes will represent the red (red) R, green (green) G, and blue (blue) B color components. Then the grayscale or gray color model is an image with black and white color gradations that produce a gray color effect. Each pixel value of the image represents the degree of gray or white intensity. and the last is the binary color model which is a color space that only has 2 possible values for each pixel, namely black (0) and white (1). Binary images are also often referred to as B & W images (black and white) or monochrome images [5]. The process of converting RGB color to binary is needed to make it easier at the preprocessing stage, feature extraction, and data classification process.

2.3   Diagonal Feature Extraction
Feature extraction is a process that is needed to get the characteristics of an image which will later describe how the character of the object. There are many algorithms in feature extraction including diagonal. A diagonal algorithm in feature extraction is an algorithm used to get character traits from handwriting. This diagonal algorithm is done by dividing an image into several equal zones. In each zone, the point characteristics are calculated on each diagonal [3].

2.4   Naïve Bayes
The Naïve Bayes classification method is an information retrieval method that uses a probabilistic approach in its inference process, which is based on Bayes' theorem in general. The application that uses this method the most is text classification. This algorithm assumes that an attribute in an object is independent. This method assumes that the absence of a certain feature will not affect the presence of other features.

3.   Research Method
3.1   Research Data

In this study, the data used is primary data which is data obtained from interviews or data collection through sources or respondents. The primary data in this study was in the form of Balinese script characters obtained through collecting data from several volunteers who wrote Balinese script characters manually or by hand. In the data collection process itself, one respondent will be asked to write 10 times per character [2].

3.1.1 Training Data

Training data is data used to train the architecture of the Naïve Bayes method. The training data used in this study were primary. Data were obtained from 10 respondents who wrote each character on the paper provided 10 times so that the total training data collected was 100 pieces of writing on each Balinese script character . The details of the characters collected are the characters in the Wreastra script, totaling 18 characters, and the voice character cast, totaling 5 characters so that the total characters are 23 characters so that the total of all training data obtained is 2300 training data. An example of the training data can be seen in figure 1.

Figure 1 Training Data

3.1.2 Testing Data

The test data is the same as the training data, namely in the form of images of Balinese script characters. Taking the image of the Balinese script character by writing the Balinese script character on the application that was built directly.

3.2   Design Method
3.2.1 Preprocessing

The initial data processing process, is divided into 2 processes, namely the process of processing test data and training data, where these two processes have the same stages but use different data. This process consists of binarization, normalization, and thining .

3.2.2 Diagonal Feature Extraction

The preprocessing results are broken down into a size of 10x10 pixels so that the image has 54 areas or features. Next, the value for each feature will be searched. For each of these features, a diagonal line will be drawn according to the size of the feature, where with a size of each feature of 10x10 pixels it will produce a total of 19 diagonal lines, of which 19 lines are sub-features. Next, the total value of all foreground pixels in the feature will be calculated. These values will be averaged to form a single value on the feature it occupies [4]. This process will be repeated until all features are scored. The entire value of this feature will be the weighted value that will be used in the classification process. The diagonal feature extraction process can be seen in figure 2.
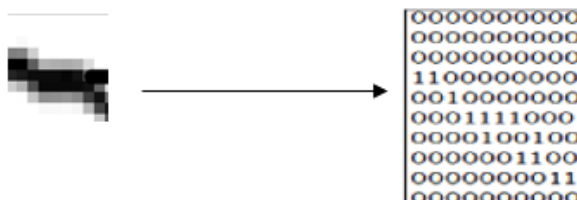

Image 2 Diagonal Feature Extraction

For example, in one of the features x with a size, the value will be searched based on the binary color it has. Number 1 shows the foreground part of the image which has black color, as can be seen in figure 3.
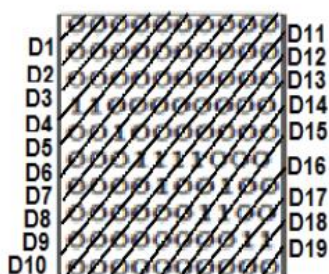

Image 3 Feature Diagonal Line

The search for the value of this feature is done by moving diagonally. Where this movement produces 19 diagonal lines or equal to 19 sub-features. The values of these sub-features will be averaged to form the value for the x feature. For example, in Figure 4 it has a feature value of 0.6842. This is

Vyasa, dkk.
Application of Naïve Bayes Algorithm in Educational
Game Learn to write Aksara Bali

obtained from all the foreground values in the image which is worth 13, then the average feature will be searched so that.

$$Feature\ Value\ x = \frac{13}{19} = 0,6842$$

After this feature x value is found, it will proceed to the next feature until all feature values are obtained. After all 54 features are obtained, as many as 15 additional features are obtained from the 54 feature areas obtained previously [8]. The number of additional features is obtained through 9 features from the row average and 6 features from the column average. The number of features that are obtained after the entire process is complete is 69 feature areas.

3.2.2 Data Classification

The first step in this process is to input data from a character dataset where the data set will be used as training data in this application. Furthermore, after inputting the data, the process of calculating the amount of data in each class will be carried out, whereas in this study there were 23 classes. With each class having 90 data each, then the prior probability calculation process will be carried out.

After getting the prior probability in each class, the next process is to calculate the number of cases per class or the probability of each feature where the probability X is based on the conditions in the hypothesis H or P(X|H). After the prior probabilities and cases per class are obtained, then the probability value is multiplied by all class variables. The last step is to calculate the value and compare the results per class using equation (2.3) (posterior probability).

3.3 Application Needs Analysis

This section will explain the functional and non-functional requirements of the application to be made. Functional Requirements are application requirements that are seen from their functions. This requirement usually describes the functionality that the author expects to have in the application and to run properly. Meanwhile, non-functional requirements are application requirements for software (software) and hardware (hardware), both in making applications and in implementing the application itself. The two needs can be seen in table 1 and tables 2 and 3.

Table 1 Application Functional Requirements

No.	Functional Needs
1.	The application can display information about how to use the application.
2.	Applications can display information about the application developer.
3.	The application can be a canvas on which the user writes Balinese script.
4.	The application can apply the Bayesian naive classification algorithm when performing the classification process for the Balinese script entered by the user.

Table 2 Developer Non-functional Requirements

Type	Non-Functional Needs
Hardware	Prosessor : ryzen 5
	Kartu grafik : Intel Iris Plus
	RAM : 8GB
	Storage : 128GB
	OS : windows 10
Software	IDE :Visual Studio Code
	Image Processing Software : Adobe Illustrator
	Programming Language : Python 3.9.6, HTML, CSS, JS
	Framework : Flask, Boostrap
	Library : OpenCV, numpy, os, glob, sklearn, pickle

Table 3 User Non-functional Requirements

Type	Non-Functional Needs
Hardware	A laptop that can open a browser
Software	Web Browser : Google Chrome, Safari, Opera, Mozilla Firefox

4.   Result and Discussion
4.1   App Workflow

The application flow design is used as a description of the application application from function to function. With regard to the game request, the user is faced with 3 choices, namely, how to start, start and get started. If you choose to play, the game application offers information about the playback playback. If you want the user to be confronted with information about the developer, the user will be returned to the main menu using the function by using the function, and the last selection starts the user presenting the game in the game. This start menu contains 2 more options, namely learning and training. If the user wants to learn, the user will enter the learning mode of the written Balinese script, and when it is finished, it returns to the main menu. If you, if you select the practice menu, the user is faced with the challenge mode of writing Balinese script, and when finished, returns to the main menu.

4.2   Interface Implementation

Interface ImplementationImplementation of the interface in this system will be divided into 8 pages. Can be seen in figures 4 and 5, namely the main menu page (1), how to play page (2), application page (3), game start page (4), study page (5), exercise page (6) , true page (7), and false page (8).

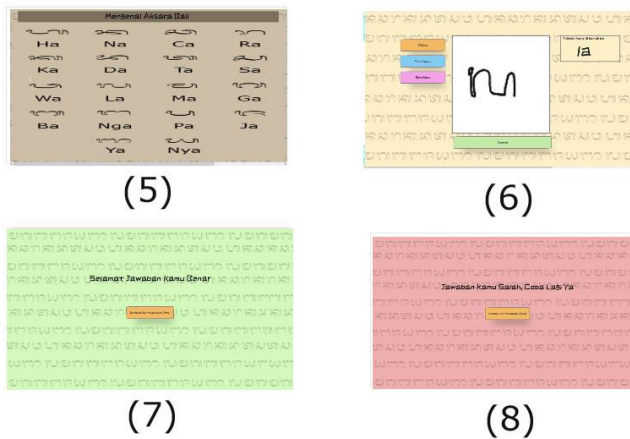**Image 4 Interface Implementation 1**

Vyasa, dkk.
Application of Naïve Bayes Algorithm in Educational
Game Learn to write Aksara Bali

Image 5 Interface Implementation2

4.3 Testing

4.3.1 Black Box Testing

In this test, the functionality of the application will be tested whether it is following the expected expectations or not. The results of this test are as follows:

Table 3 Black Box Test Results

No	Testing	Result	Description
1.	Starting the app	Success	Can be opened and display the main page.
2.	Pressing the "Start" button	Success	The button can direct the application to the game mode select page.
3.	Pressing the "How to Play" button"	Success	The button can direct the application to the how-to page.
4.	Pressing the "About" button	Success	The button can direct the app to the about page.
5.	Pressing the "Learn" button	Success	The button can direct the application to the study page.
6.	Pressing the "Practice" button	Success	The button can direct the app to the workout page.
7.	Writing on canvas	Success	User can write on canvas.
8.	Pressing the "Eraser" button	Success	Users can change from pencil mode to eraser mode, and can erase doodles on the canvas well.
9.	Pressing the "Pencil" button	Success	Users can change the mode from eraser back to pencil mode just fine.
10.	Hit the "Clean" button	Success	Users can clean the entire canvas content automatically.
11.	Pressing the "Gather" button	Success	This button will stop writing activities and then send data to enter into the classification process.
12.	Question	Success	Questions can appear and display questions on the practice menu.
14.	Pressing the "Back to Start Page" button	Success	This button will direct the user to the start page.

4.3.2 UAT Testing (User Acceptnance Test)

This test is carried out to measure the user's perception of the application that is built. In this test, 20 respondents will assess the application built. This test is carried out by asking several questions to users who have tried this application based on a predetermined rating scale. This application is run locally and is not included in the URL link. Here are the questions and the percentage of answers from respondents who have tried this application:

Table 4 UAT Test Results (User Acceptance Test)

(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)					(User Acceptnance Test)				
		(User Acceptnance Test)	(User Acceptnance Test)	C	D	E	A	B	C	D	E
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	1	0	0	70%	25%	5%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	0	0	0	50%	50%	0%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	3	2	0	35%	40%	15%	10%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	0	0	0	60%	40%	0%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	3	0	0	75%	10%	15%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	9	1	0	30%	20%	45%	5%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	2	0	0	65%	25%	10%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	1	0	0	60%	35%	5%	0%	0%

From the table above, it can be found that, as many as 70% of respondents strongly agree that the appearance of the application is attractive. Then as many as 50% of respondents strongly agree that the menus of the application are easy to understand. As many as 35% of respondents agree that the application is easy to use. As many as 60% of respondents strongly agree that the application provides examples of Balinese script and can help respondents to know the shape of the Balinese script. Then as many as 75% of respondents strongly agree that the exercise menu can help respondents to know their ability to write Balinese script. As many as 30% of respondents strongly agree that the application can improve writing skills in Balinese script. As many as 65% of respondents strongly agree that with the application that is built the experience of learning to write Balinese script becomes fun. Finally, 60% of respondents strongly agree that the application built can be a medium for learning Balinese script well.

4.3.3 Accuracy Test

This accuracy test is done by calculating the average of the results of the tests carried out. In this accuracy test, the Balinese script characters are written 20 times per character so the tests carried out are 460 times. This test is carried out directly on the application that has been built. accuracy test results can be seen in Figure 6.
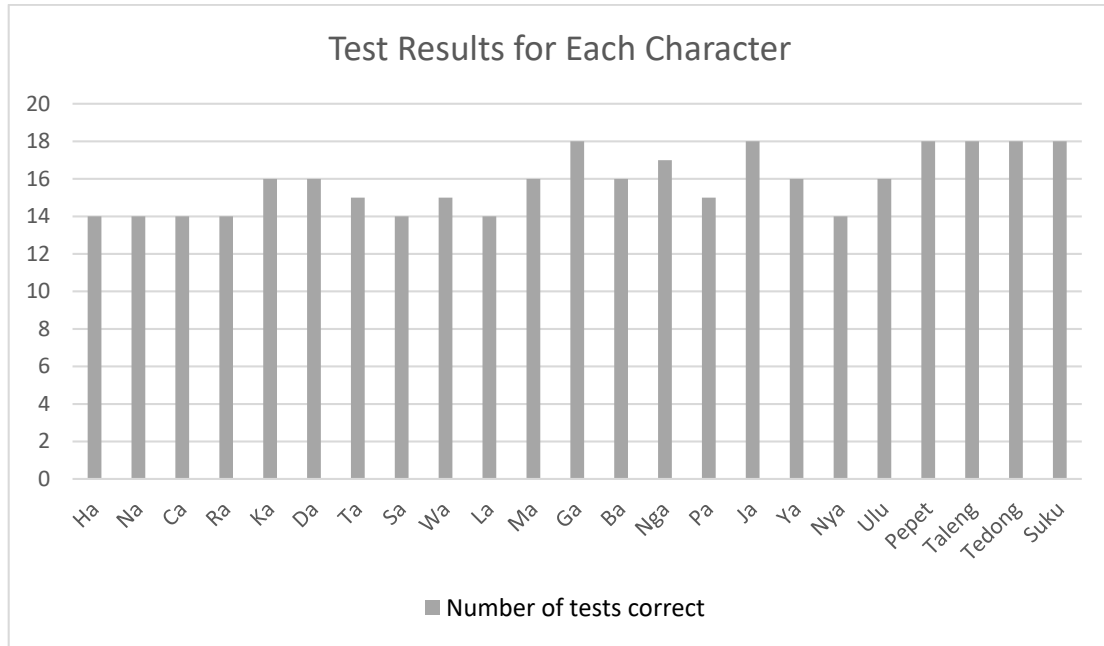
Vyasa, dkk.
Application of Naïve Bayes Algorithm in Educational
Game Learn to write Aksara Bali

Figure 6 Accuracy Test Results

From the tests carried out, where each character was tested 20 times with a total of 364 trials which were true. It was found that the accuracy of the classification of Balinese characters using the Naïve Bayes method with diagonal feature extraction was 79.1%.

5.   Conclusion

obtained are in the form of a Balinese script recognition game application that is intended for the community, both students and the general public. In testing, this research applies 3 types of testing, namely black box testing, where the results of black box testing in this study indicate that all the functions and features in this application can run well as expected. Then in the UAT test (user acceptance test), this application received a positive response from 20 respondents who had tested this application based on the questions and assessment criteria that had been determined. Furthermore, this study also tested the accuracy of the classification method used, in this study the classification method used was the Naïve Bayes method with diagonal feature extraction in classifying Balinese characters written by the user. The result of this accuracy test is the percentage of success of the Naïve Bayes algorithm in classifying Balinese characters, which is 79.1% in a total of 460 trials.

References

[1] Astra, I. S. (1982). *Prasasti Sibang Kaja di Kabupaten Badung.* Badung: Fakultas Sastra, Universitas Udayana.

[2] Cahyadhi, I. P., Sunarya, I. M., & Wirawan, I. M. (2016). Pengembangan Game Edukasi "Aksara Bali" Berbasis Android. *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*.

[3] Firmansyah, M. A., Ramadhani, K. N., & Arifianto, A. (2018). *Pengenalan Angka Tulis Tangan Menggunakan Diagonal Feature Extraction dan Klasifikasi Artificial Neural Network Multiayer Perceptron*, 10.

[4] Parker, J. (2010). *Algorithms for Image Processing and Computer Vision, Second Edition.* Alberta: Wiley Publishing, Inc.

[5] Putra, D. (2010). *Pengolahan Citra Digital.* Yogyakarta: C.V ANDI OFFSET.

[6] Zang, T. Y., & Suen, C. Y. (1984). A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the ACM*, 239.

[7] Walat, W. (2010). Conception of Media Education. *Journal of Technology and Information Education vol. 2*, 10.

[8] Win , T., Dr. , E. H., & Dr. , S. Y. (2019). *License Plate Detection and Recognition using OCR based on Morphological Operation*, 5.

Vyasa, dkk.
Application of Naïve Bayes Algorithm in Educational
Game Learn to write Aksara Bali

This page is intentionally left blank.