NAIVE BAYES CLASSIFICATION FOR DETECTION OF HARAM STATUS ON JAPANESE PACKAGING FOOD

Teguh Syam Mardiana[1], Ken Kinanti Purnamasari. S.Kom., M.T.[2]

[1,2]Teknik Informatika - Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : teguhsyamm@gmail.com [1], kinantiken@gmail.com[2]

ABSTRAK

Composition is a label in which there is written, drawing, or a combination of both included on the packaging food by way of pasted or printed. Composition is a very important part to provide a information of the ingredients contained in a packaging food so that consumers can obtain the expected product, healthy, and safe for consumption. The language on the composition in the packaging food is different, one of which is contained in Japanese packaging food that uses Japanese script as information of ingredients contained in it. In this research is made an OCR system by utilizing image composition as the media identification. The method of classification used is Naïve Bayes method combined with the Zoning (ICZ and ZCZ) feature extraction method to perform character recognition on the image of Japanese packaging food composition so then can do detection of its haram status. Japanese characters are classified only characters ze, ra, chi, and n that form the haram ingredients of zerachin. Training data used in this research as the 75 characters in which there are 5 ze characters, 11 ra characters, 18 chi characters, and 41 n characters. While the test data used as much as 20 images of composition are 5 images of halal composition, 8 image of haram composition containing zerachin, and 7 images of haram composition containing other haram ingredients. Based on some test results obtained the best accuracy of 33,33% on the test using the parameters with the division of 6 zones on the Zoning feature extraction (ICZ and ZCZ). It is concluded that the Naive Bayes method can classify the image of Japanese packaging food composition, it is just the result of accuracy is still not good. The result of accuracy is influenced from the testing process at the classification which is also low accuracy.

keywords : Atrificial Intelligence, Image Processing, OCR, Zoning, Naive Bayes.

1. INTRODUCTION

Previously, there were several apps already made to detect illegal status on Japanese packaging food. Some applications are still using barcodes as a medium of identification. Lack of barcode usage that must be done periodically renewal along with the emergence of new packaging food. Updates will take up a lot of capacity in memory so it is still less effective for now. Different things if there is an Optical Character Recognition (OCR) system that can detect the illicit status of Japanese packaging food by using the image of the composition as its identification. The system will conduct training on the characters forming illicit materials, so that in the testing process, the system can automatically perform character recognition on the image of the composition, which then the system can detect the illegal status of the food packaging.

Naive Bayes is one classical classification method that uses probabilistic concepts. Previously Naive Bayes method ever implemented in research related to OCR, one of which is research for the identification of handwritten alphabet letters[1]. The results of the study get an average accuracy of 80%. In addition, the Naive Bayes method has also been used to classify several other cases such as: Detection of human tongue disease with an accuracy of almost 85%[2], the classification of nutritional status with an accuracy of 93.2%[3] and the highest accuracy of 96% in pattern recognition For the identification of five grains[4]. From some facts found that the method of Naive Bayes is still very optimal in its use because it can classify some cases well.

Zoning is one of the most popular feature extraction methods for OCR systems. This feature extraction method has already been implemented in research for the introduction of the printed letters with an accuracy of more than 90%[5]. In this study using Zoning ICZ and ZCZ combined feature extraction methods with high accuracy results compared only using ICZ or ZCZ method only[5]. Taken together with a high degree of accuracy in each of these studies, the Naive Bayes classification method and Zoning feature extraction methods (ICZ and ZCZ) will be combined and implemented for detection systems of disguised status in Japanese packaging foods in this study. The use of Naive Bayes classification method and Zoning feature extraction method (ICZ and ZCZ) have not known accuracy results if combined so it is expected to provide high accuracy in the case in this study.

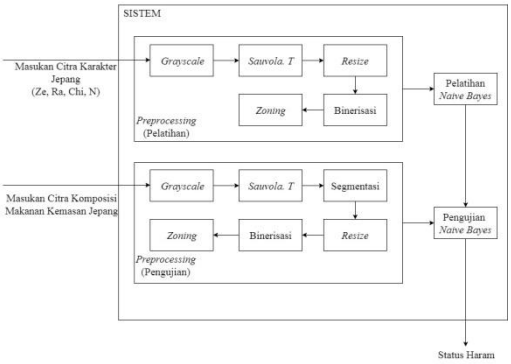
2. CONTENT

2.1 Problem Analysis

Based on background and problem identification, it will be explained about problem analysis found in this research. The problem in this research is not yet known the accuracy of Naïve Bayes classification method combined with Zoning combined feature extraction method (ICZ and ZCZ) which is implemented on an OCR system detecting illicit status on Japanese packaging food.

2.2 Process Analysis

Analysis of the solution is a stage where the process undertaken in this system is divided into several parts, where each process has their respective roles to detect the status of haram to eat this Japanese packaging.



Picture 1 System Architecture

In the block diagram above, the initial stage of the training process where the first time to do is enter the image of Japanese characters ze, ra, chi, and n into the system. The next step is to do the preprocessing process including grayscale, sauvola threshold, resize, binary and feature extraction using Zoning ICZ & ZCZ so as to get a decimal array value. In addition, in the testing stage is to include the image of Japanese packaging food composition. Next is to do the preprocessing process as in the training process which also later obtained a decimal array value. The value of the decimal array will then be processed again in the process of classification by the method of Naive Bayes in both training and testing. Results of the classification will be used to determine the illicit status of the Japanese packaging food.

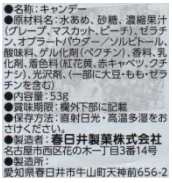
2.2.1 Input Data Analysis

Examples of input data used in the training process are Japanese script image including the characters ze, ra, chi, and n as in figure 2. While as the input data used in the testing process is the image of Japanese food packaging composition as in figure 3. Both input data use the .jpg image format, The image is obtained from the capture directly using the camera phone with 8 megapixel resolution with a distance of 15cm. Image retrieval is not too dark for the application can be more leverage in preprocessing.



ゼラチン

Gambar 2 Citra Masukan Data Latih



Gambar 3 Citra Masukan Data Uji

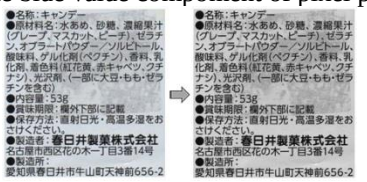
2.2.1 Grayscale

Grayscale is a process to convert colored image (RGB) into a gray image that serves to shrink the color range to 0 to 255. Here is the color image conversion formula (RGB) to grayscale.

$$I = (0.2989* R)+(0.5870*G)+(0.1141*B) \quad (1)$$

Information :

R : the red value component of pixel point.
G : the green value component of pixel point.
B : the blue value compoment of pixel point.



Picture 4 Grayscale Image Result

2.2.3 Sauvola Threshold

Sauvola threshold is a threshold method capable of providing enhancement by calculating the threshold using the dynamic value range of the standard devaluation value of grayscale image. Sauvola threshold has advantages over other thresholding methods such as bernsen, eikvil, parker, niblack and otsu[20][21]. Here is the sauvola threshold formula.

$$m(x,y) = \frac{\sum_{i=min}^{i=max}\sum_{j=min}^{j=max} img(i,j)}{(max\ i - min\ i+1)*(max\ i - min\ i+1)} \quad (2)$$

$$s(x,y) = \sqrt{\frac{\sum_{i=min}^{i=max}\sum_{j=min}^{j=max}(img(i,j)-m(x,y))^2}{((max\ i - min\ i+1)*(max\ i - min\ i+1))-1}} \quad (3)$$

$$T(x,y) = m(x,y) * \left(1 + k * \left(\frac{s(x,y)}{R} - 1\right)\right) \quad (4)$$

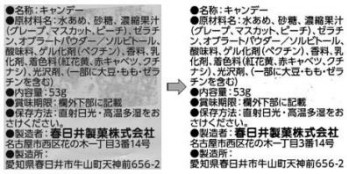
Information :

R : Maximum value of standard deviation (128 for grayscale image)
k : kernel with value between 0,2 – 0,5.
m: Fungsi yang menghasilkan nilai rata-rata dari sejumlah *pixel* citra.
s : Function that produces the average value of a number of image pixels.
T : Function that produces threshold value (threshold).
x : Coordinate value of image width.

y : High image coordinate value..

After the threshold value T (x, y) has been obtained, then enter into the following equation.

$$f(x,y) = \begin{cases} 0, & \text{img}(x,y) < \text{T}(x,y) \\ 255, & \text{img}(x,y) \geq \text{T}(x,y) \end{cases} \quad (5)$$



Gambar 5 Threshold Image Result

2.2.4 Segmentation

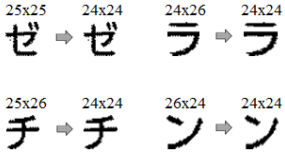
Segmentation is a technique to divide an image into several regions (regions) where each region has similarity attributes. In this research segmentation is done by cutting row and column image. The first cut is done for the entire line of handwritten image, then continued by cutting the column for each row cutting image.

Table 1 Segmentation Process

Pemotongan Baris	1	(グレープ、マスカット、ピーチ)、ゼラチ
	2	ン、オブラートパウダー／ソルビトール、
Pemotongan Kolom	1	(グレープ、マスカット、ピーチ)、ゼラチ
	2	ン、オブラートパウダー／ソルビトール、

2.2.5 Resize

Resizing the image means changing the size of the digital image in pixels. In this process, any existing characters will be forced to be in a predetermined size so that the shape is not proportional. But if this is done consistently and thoroughly on every character then it will not damage the information needed for the next process.

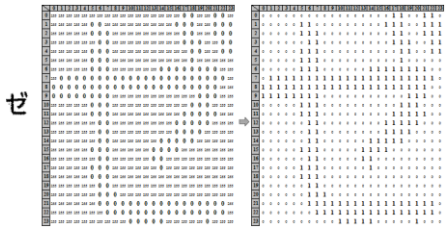


Picture 6 Image Resize Results

2.2.6 Binery

Binary imagery is an image with each pixel only represented by a value of two possibilities ie values 0 and 1. The image converted to binary is usually from an image in black and white. In this study binary image is used as input data at the classification stage. Here is the image binaryization equation.

$$f(x) = \begin{cases} 0, & x = 0 \\ 1, & x = 255 \end{cases} \quad (6)$$



Picture 7 Character image Binary result

2.2.7 Zoning Feature Extraction

Zoning feature extraction is a feature extraction with the main concept of dividing the image into the same number of zones in order to recognize the characteristics of each letter character and number [12], which will result in a value to be processed at the classification stage. Steps in Zoning ICZ algorithm (Image Centroid Zone).

1. Calculate the centroid value of the input image.

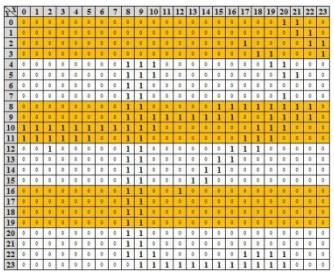
$$xc = \frac{\sum_{y=0}^{y=max} \sum_{x=0}^{x=max} img(x,y) * Bx}{t}$$

$$yc = \frac{\sum_{x=0}^{x=max} \sum_{y=0}^{y=max} img(x,y) * By}{t}$$



Gambar 8 Examples of Characters to Extract

2. Divide into the same number of zones. In this study used several zone divisions, ie 6, 8, and 12 zones. As an illustration the following example of the 6 zone division and its calculations.



Picture 9 Zoning Matrix 24x24 Character Ze

3. Calculate the pixel black distance from the centroid in each zone.

$$d(p,c) = \sqrt{(px - xc)^2 + (py - yc)^2}$$

atau bisa diturunkan sebagai

$$d((px,py),(xc,yc)) = \sqrt{(px - xc)^2 + (py - yc)^2}$$

Calculation of pixel distance against the centroid in zone 1 can be seen below. Do it on all zones.

Perhitungan pada Zona 1

$$d((px,py),(xc,yc)) = \sqrt{(px - xc)^2 + (py - yc)^2}$$

$$\text{Jarak}_{1\,1}((20,0),(12,12)) = \sqrt{(20-12)^2 + (0-12)^2} = 14{,}4222051$$

$$\text{Jarak}_{2\,1}((21,0),(12,12)) = \sqrt{(21-12)^2 + (0-12)^2} = 15$$

…

$$\text{Jarak}_{13\,1}((23,23),(12,12)) = \sqrt{(23-12)^2 + (3-12)^2} = 14{,}2126704$$

4. Calculate the average distance between points

$$\sum_{m=1}^{m=6} RJ_m = \frac{Jarak_{1,m} + Jarak_{2,m} + Jarak_{3,m} + \cdots Jarak_{n,m}}{t_m}$$

Calculation of the average distance in zone 1 can be seen below. Do it on all zones.

Perhitungan pada Zona 1

$$\sum_{m=1}^{m=6} RJ_m = \frac{Jarak_{1.1} + Jarak_{2.1} + Jarak_{3.1} + \cdots Jarak_{n.1}}{t_m}$$

$$\text{Rataan Jarak}_1 = \frac{Jarak_{1.1} + Jarak_{2.1} + Jarak_{3.1} + \cdots Jarak_{12.1}}{t_1}$$

$$= \frac{164{,}7518879}{12} = 12{,}67322214734602$$

5. Obtained 6 values as follows.

Ze1|ICZZona 1 = 12,67322214734602
Ze1|ICZZona 2 = 12,012497207679937
Ze1|ICZZona 3 = 12,447487267849914
Ze1|ICZZona 4 = 11,409830874191181
Ze1|ICZZona 5 = 11,185689842746477
Ze1|ICZZona 6 = 10,709418276009616

Steps for Zoning ZCZ (Zone Centroid Zone) feature extraction.

1. The steps taken are the same as on Zoning ICZ, It's just the value of the features that must be done is to divide the zone first before the calculation of centroid. So in Zoning ZCZ obtained some centroid value in each zone.
2. The resulting value is as follows.

Ze1|ZCZZona 1 = 4,721785887569821
Ze1|ZCZZona 2 = 5,2598512621445055
Ze1|ZCZZona 3 = 6,266306046689112
Ze1|ZCZZona 4 = 4,025892101809964
Ze1|ZCZZona 5 = 1,46964141804604
Ze1|ZCZZona 6 = 4,095808930800819

Table 2 Value of Extraction Result Feature

Karakter	Nilai Fitur ICZ	Nilai Fitur ZCZ
ゼ	Zona 1 = 12,67322214734602 Zona 2 = 12,012497207679937 Zona 3 = 12,447487267849914 Zona 4 = 11,409830874191181 Zona 5 = 11,185689842746477 Zona 6 = 10,709418276009616	Zona 1 = 4,721785887569821 Zona 2 = 5,2598512621445055 Zona 3 = 6,266306046689112 Zona 4 = 4,025892101809964 Zona 5 = 1,46964141804604 Zona 6 = 4,095808930800819
ゼ	Zona 1 = 12,152092034699395 Zona 2 = 11,517637496096641 Zona 3 = 12,918093375696683 Zona 4 = 11,774304565748105 Zona 5 = 11,717404647201132 Zona 6 = 11,204265885334451	Zona 1 = 6,585186117885765 Zona 2 = 5,343669893271543 Zona 3 = 6,028100403797702 Zona 4 = 5,386656689061503 Zona 5 = 2,033548153373633 Zona 6 = 3,789256623257417 6
ラ	Zona 1 = 9,815704917563945 Zona 2 = 9,29752113327062 1 Zona 3 = 11,375360590190143 Zona 4 = 9,147618075919464 Zona 5 = 7,456429991169783 Zona 6 = 9,39874519725485	Zona 1 = 5,270038649525 98 Zona 2 = 6,403924215678749 Zona 3 = 7,673956413475698 Zona 4 = 1,412899020449196 Zona 5 = 1,729126786466033 8 Zona 6 = 2,145973745760809
ラ	Zona 1 = 10,685100190679 76 Zona 2 = 0 Zona 3 = 12,344423773295137 Zona 4 = 12,218913863708 77 Zona 5 = 10,562231499771158 Zona 6 = 10,456940163557647	Zona 1 = 4,338280731159644 Zona 2 = 0 Zona 3 = 6,258517038284154 Zona 4 = 1,330056307974577 Zona 5 = 1,627277986846677 6 Zona 6 = 3,133835219296107
チ	Zona 1 = 8,012749103327335 Zona 2 = 9,418413025752 18 Zona 3 = 10,158762579985678 Zona 4 = 7,592153617140983 Zona 5 = 7,656575506990917 Zona 6 = 8,965168581122711	Zona 1 = 2,616887652671708 Zona 2 = 4,465719115888621 Zona 3 = 5,759863980887081 Zona 4 = 1,258068188778074 75 Zona 5 = 1,412335645338953 7 Zona 6 = 1,984631787785905 4
チ	Zona 1 = 9,141663910276941 Zona 2 = 8,798054549379625 Zona 3 = 10,089916519544389 Zona 4 = 8,740215188717475 Zona 5 = 8,623828443755032 Zona 6 = 10,363979509874575	Zona 1 = 4,171640436000164 Zona 2 = 1,275412073846563 7 Zona 3 = 4,920798851643807 Zona 4 = 1,412696783230135 Zona 5 = 1,530421515165979 Zona 6 = 2,098183374044004 3
ン	Zona 1 = 13,439618456275456 Zona 2 = 14,967811074186626 Zona 3 = 14,609756273435398 Zona 4 = 12,389560516003 07 Zona 5 = 11,948957391616256 Zona 6 = 13,439973808660172	Zona 1 = 3,395414262140956 Zona 2 = 8,593759175857413 Zona 3 = 1,516123775561495 Zona 4 = 1,725456785586119 8 Zona 5 = 2,754679910652 39 Zona 6 = 1,868587419059603 5
ン	Zona 1 = 12,979078303553663 Zona 2 = 13,685679595518 46 Zona 3 = 13,321328228344745 Zona 4 = 11,950069732803595 Zona 5 = 10,614544829711189 Zona 6 = 12,604500897567021	Zona 1 = 4,801682613375976 Zona 2 = 7,104436990221722 Zona 3 = 1,521482441874218 8 Zona 4 = 1,662769960302215 Zona 5 = 2,085941674157796 Zona 6 = 2,987641033811849

2.2.8 Naive Bayes Classification

Naive Bayes is the simplest method of classification using existing opportunities, where it is assumed that each variable x is independent (independence). The data used can be categorical (discrete) or continuous. However, in this study will be used continuous data, because the extraction of image features is a continuous data in the form of calculation results. So for continuous data can be completed by using the steps as follows.

2.2.8.1 Training

1. Calculate the mean value

$$\mu = \frac{\sum x_i}{\text{n}} \quad (7)$$

Here is an example calculation. Do it for all feature values in each existing class.

$$\mu = \frac{\sum \text{Nilai Fitur Ze}|\text{ICZ}_{\text{Zona 1}}}{\text{n}}$$

$$\mu = \frac{\text{Ze}_1|\text{ICZ Zona 1} + \text{Ze}_2|\text{ICZ Zona 1}}{\text{n}}$$

$$\mu = \frac{12{,}67322214734602 + 12{,}152092034699395}{2}$$

$$\mu = \frac{24{,}82531418}{2} = 12{,}41265709$$

2. Counting variance

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 \quad (8)$$

Here is an example calculation. Do it for all feature values in each existing class.

$$\sigma^2 = \frac{1}{2-1}\sum_{i=1}^{n}(\text{Nilai Fitur Ze}|\text{ICZ}_{\text{Zona 1}} - \mu)^2$$

$$\sigma^2 = \frac{1}{2-1}.(12{,}67322214734602 - 12{,}41265709)^2 + (12{,}152092034699395 - 12{,}41265709)^2$$

$$= 0{,}135788297$$

So we get the recapitulation of mean and variance as follows.

Table 3 Mean & Variance Value

Kelas Karakter	Fitur	μ	σ^2	Kelas Karakter	Fitur	μ	σ^2
Ze	$ICZ_{Zona\ 1}$	12,41265709	0,135788297	Ra	$ICZ_{Zona\ 1}$	10,25040255	0,37792407
	$ICZ_{Zona\ 2}$	11,76506735	0,122443067		$ICZ_{Zona\ 2}$	4,648760567	43,22194961
	$ICZ_{Zona\ 3}$	12,68279032	0,110735054		$ICZ_{Zona\ 3}$	11,85989218	0,469541726
	$ICZ_{Zona\ 4}$	11,59206772	0,066420536		$ICZ_{Zona\ 4}$	10,68326597	4,716428908
	$ICZ_{Zona\ 5}$	11,45154724	0,141360317		$ICZ_{Zona\ 5}$	9,009330745	4,823001505
	$ICZ_{Zona\ 6}$	10,95684208	0,122437078		$ICZ_{Zona\ 6}$	9,92784268	0,559888293
	$ZCZ_{Zona\ 1}$	5,653486003	1,736130211		$ZCZ_{Zona\ 1}$	4,80415969	0,434086409
	$ZCZ_{Zona\ 2}$	5,301760578	0,003512781		$ZCZ_{Zona\ 2}$	3,201962108	20,50512268
	$ZCZ_{Zona\ 3}$	6,147203225	0,028370964		$ZCZ_{Zona\ 3}$	6,966236726	1,001734312
	$ZCZ_{Zona\ 4}$	4,706274395	0,925840131		$ZCZ_{Zona\ 4}$	1,371477664	0,003431458
	$ZCZ_{Zona\ 5}$	1,751594786	0,158995403		$ZCZ_{Zona\ 5}$	1,678202387	0,005186589
	$ZCZ_{Zona\ 6}$	3,942532777	0,046987159		$ZCZ_{Zona\ 6}$	2,639904483	0,487935145
Chi	$ICZ_{Zona\ 1}$	8,577206507	0,637224321	N	$ICZ_{Zona\ 1}$	13,20934838	0,106048616
	$ICZ_{Zona\ 2}$	9,108233788	0,19242232		$ICZ_{Zona\ 2}$	14,32674452	0,821932662
	$ICZ_{Zona\ 3}$	10,12433955	0,00236989		$ICZ_{Zona\ 3}$	13,96554225	0,830023414
	$ICZ_{Zona\ 4}$	8,166184403	0,659022686		$ICZ_{Zona\ 4}$	12,16981512	0,096576074
	$ICZ_{Zona\ 5}$	8,140201975	0,467789122		$ICZ_{Zona\ 5}$	11,28175111	0,890328443
	$ICZ_{Zona\ 6}$	9,664574045	0,978336007		$ICZ_{Zona\ 6}$	13,02223735	0,349007493
	$ZCZ_{Zona\ 1}$	3,394264044	1,208628109		$ZCZ_{Zona\ 1}$	4,098548438	0,988795338
	$ZCZ_{Zona\ 2}$	2,870565595	5,089029511		$ZCZ_{Zona\ 2}$	7,849098083	1,109040286
	$ZCZ_{Zona\ 3}$	5,340331416	0,352015146		$ZCZ_{Zona\ 3}$	1,518803109	1,43577E
	$ZCZ_{Zona\ 4}$	1,335379336	0,011955975		$ZCZ_{Zona\ 4}$	1,694113373	0,001964819
	$ZCZ_{Zona\ 5}$	1,47137858	0,006972136		$ZCZ_{Zona\ 5}$	2,420310792	0,223605414
	$ZCZ_{Zona\ 6}$	2,041407581	0,006446981		$ZCZ_{Zona\ 6}$	2,428114226	0,626140496

2.2.8.2 Testing

Here is a snippet of Japanese packaged food composition image.

Table 4 Examples of Characters To Be Tested

No	Karakter	Nilai Fitur ICZ	Nilai Fitur ZCZ
1.	ゼ	Zona 1 = 13,025233469603139 Zona 2 = 11,410754257617823 Zona 3 = 13,791873159443282 Zona 4 = 11,705892525941424 Zona 5 = 12,284985913444595 Zona 6 = 11,432536236765731	Zona 1 = 4,559097203215948 Zona 2 = 5,897394017768255 Zona 3 = 4,994586374032392 Zona 4 = 5,094496037888883 Zona 5 = 1,082842712474619 Zona 6 = 3,8859577983839477
2.	ラ	Zona 1 = 11,482920100648125 Zona 2 = 9,739332964963 53 Zona 3 = 11,966315386442417 Zona 4 = 10,812547792676513 Zona 5 = 9,019370143936545 Zona 6 = 10,420798224556306	Zona 1 = 5,3054106155503415 Zona 2 = 5,952380952380 9526 Zona 3 = 6,245963343061068 Zona 4 = 1,3377069674653514 Zona 5 = 2,024604141377068 Zona 6 = 3,28261919020708

To know which class includes the character of the test image. Then step testing in the Naive Bayes algorithm should be done, the following stages are as follows.

1. Look for probability (prior) for each class on the training data. The trick is to calculate the amount of data per class divided by the total amount of data as a whole

   Prior karakter Ze = 5/75=0,0666
   Prior karakter Ra = 11/75=0,1466
   Prior karakter Chi = 18/75=0,24
   Prior karakter N = 41/75=0,5466

2. Calculates the probability density

$$\varphi_{\mu\sigma}(x) = \frac{1}{\sqrt{2\pi(\sigma^2)}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (9)$$

   Here is an example calculation. Do it for all feature values in each existing class.

Perhitungan pada kelas karakter Ze

$$= \frac{1}{\sqrt{2*3,14*(0,135788297)}} e^{\frac{(13,025233469603139-12,41265709)^2}{2*0,135788297}}$$

$$= \frac{1}{\sqrt{6,28*0,135788297}} e^{\frac{0,612576389^2}{0,271576594}}$$

$$= \frac{1}{\sqrt{0,852750505}} e^{\frac{0,375249821}{0,271576594}}$$

$$= \frac{1}{0,923444912} e^{\frac{0,375249821}{0,271576594}}$$

$$= 1,08290163 \ e^{-1,38174581}$$

$$= 1,08290163 * 0,251139728$$

$$= 0,271959621$$

So we get the recapitulation of the density value of each class feature.

Table 5 Each Class Density Value

Kelas Karakter	Fitur	φ	Kelas Karakter	Fitur	φ
Ze	$ICZ_{Zona\ 1}$	0,271959621	Ra	$ICZ_{Zona\ 1}$	2,44478E-5
	$ICZ_{Zona\ 2}$	0,682995003		$ICZ_{Zona\ 2}$	0,035764128
	$ICZ_{Zona\ 3}$	0,004642665		$ICZ_{Zona\ 3}$	0,010939668
	$ICZ_{Zona\ 4}$	0,404468274		$ICZ_{Zona\ 4}$	0,164462117
	$ICZ_{Zona\ 5}$	0,090956268		$ICZ_{Zona\ 5}$	0,059740062
	$ICZ_{Zona\ 6}$	0,452624419		$ICZ_{Zona\ 6}$	0,070608864
	$ZCZ_{Zona\ 1}$	0,214500263		$ZCZ_{Zona\ 1}$	0,565184027
	$ZCZ_{Zona\ 2}$	7,88943E-22		$ZCZ_{Zona\ 2}$	0,073815848
	$ZCZ_{Zona\ 3}$	1,60783E-10		$ZCZ_{Zona\ 3}$	0,057275129
	$ZCZ_{Zona\ 4}$	0,382298966		$ZCZ_{Zona\ 4}$	0
	$ZCZ_{Zona\ 5}$	0,245203398		$ZCZ_{Zona\ 5}$	8,01065E-15
	$ZCZ_{Zona\ 6}$	0,779257011		$ZCZ_{Zona\ 6}$	0,116374853
Chi	$ICZ_{Zona\ 1}$	9,05209E-8	N	$ICZ_{Zona\ 1}$	0,044375912
	$ICZ_{Zona\ 2}$	9,46397E-7		$ICZ_{Zona\ 2}$	0,002495663
	$ICZ_{Zona\ 3}$	0		$ICZ_{Zona\ 3}$	0,430114837
	$ICZ_{Zona\ 4}$	3,65681E-5		$ICZ_{Zona\ 4}$	0,421369311
	$ICZ_{Zona\ 5}$	6,18603E-9		$ICZ_{Zona\ 5}$	0,240308391
	$ICZ_{Zona\ 6}$	0,081660384		$ICZ_{Zona\ 6}$	0,018082068
	$ZCZ_{Zona\ 1}$	0,207061275		$ZCZ_{Zona\ 1}$	0,360484843
	$ZCZ_{Zona\ 2}$	0,071907855		$ZCZ_{Zona\ 2}$	0,068033838
	$ZCZ_{Zona\ 3}$	0,567543567		$ZCZ_{Zona\ 3}$	0
	$ZCZ_{Zona\ 4}$	8,1738E-257		$ZCZ_{Zona\ 4}$	0
	$ZCZ_{Zona\ 5}$	9,49911E-5		$ZCZ_{Zona\ 5}$	0,015456918
	$ZCZ_{Zona\ 6}$	1,2527E-114		$ZCZ_{Zona\ 6}$	0,092389301

3. After getting the density value of each character, then look for posterior values for each class.

$$Posterior = \frac{Prior * Likehood}{Evidence} \quad (10)$$

   Previously searching the value of evidence first as below.

Evidence =
(*Prior* Ze) * φ(Ze|ICZ$_{Zona\ 1}$) * φ(Ze|ICZ$_{Zona\ 2}$) * φ(Ze|ICZ$_{Zona\ 3}$) * φ(Ze|ICZ$_{Zona\ 4}$) * φ(Ze|ICZ$_{Zona\ 5}$) * φ(Ze|ICZ$_{Zona\ 6}$) * φ(Ze|ZCZ$_{Zona\ 1}$) * φ(Ze|ZCZ$_{Zona\ 2}$) * φ(Ze|ZCZ$_{Zona\ 3}$) * φ(Ze|ZCZ$_{Zona\ 4}$) * φ(Ze|ZCZ$_{Zona\ 5}$) * φ(Ze|ZCZ$_{Zona\ 6}$) +
(*Prior* Ra) * φ(Ra|ICZ$_{Zona\ 1}$) * φ(Ra|ICZ$_{Zona\ 2}$) * φ(Ra|ICZ$_{Zona\ 3}$) * φ(Ra|ICZ$_{Zona\ 4}$) * φ(Ra|ICZ$_{Zona\ 5}$) * φ(Ra|ICZ$_{Zona\ 6}$) * φ(Ra|ZCZ$_{Zona\ 1}$) * φ(Ra|ZCZ$_{Zona\ 2}$) * φ(Ra|ZCZ$_{Zona\ 3}$) * φ(Ra|ZCZ$_{Zona\ 4}$) * φ(Ra|ZCZ$_{Zona\ 5}$) * φ(Ra|ZCZ$_{Zona\ 6}$) +
(*Prior* Chi) * φ(Chi|ICZ$_{Zona\ 1}$) * φ(Chi|ICZ$_{Zona\ 2}$) * φ(Chi|ICZ$_{Zona\ 3}$) * φ(Chi|ICZ$_{Zona\ 4}$) * φ(Chi|ICZ$_{Zona\ 5}$) * φ(Chi|ICZ$_{Zona\ 6}$) * φ(Chi|ZCZ$_{Zona\ 1}$) * φ(Chi|ZCZ$_{Zona\ 2}$) * φ(Chi|ZCZ$_{Zona\ 3}$) * φ(Chi|ZCZ$_{Zona\ 4}$) * φ(Chi|ZCZ$_{Zona\ 5}$) * φ(Chi|ZCZ$_{Zona\ 6}$) +
(*Prior* N) * φ(N|ICZ$_{Zona\ 1}$) * φ(N|ICZ$_{Zona\ 2}$) * φ(N|ICZ$_{Zona\ 3}$) * φ(N|ICZ$_{Zona\ 4}$) * φ(N|ICZ$_{Zona\ 5}$) * φ(N|ICZ$_{Zona\ 6}$) * φ(N|ZCZ$_{Zona\ 1}$) * φ(N|ZCZ$_{Zona\ 2}$) * φ(N|ZCZ$_{Zona\ 3}$) * φ(N|ZCZ$_{Zona\ 4}$) * φ(N|ZCZ$_{Zona\ 5}$) * φ(N|ZCZ$_{Zona\ 6}$) =

   Get results.

   1,90083E-39 + 0 + 0 + 0 = 1,90083E-39

   The value will be used in subsequent posterior counting. Here's an example of a show on ze characters.

Posterior **Karakter Ze =**

$$\frac{(Prior\ Ze) * \varphi(Ze|ICZZona\ 1) * \varphi(Ze|ICZZona\ 2) * \varphi(Ze|ICZZona\ 3) * \varphi(Ze|ICZZona\ 4) * \varphi(Ze|ICZZona\ 5) * \varphi(Ze|ICZZona\ 6) * \varphi(Ze|ZCZZona\ 1) * \varphi(Ze|ZCZZona\ 2) * \varphi(Ze|ZCZZona\ 3) * \varphi(Ze|ZCZZona\ 4) * \varphi(Ze|ZCZZona\ 5) * \varphi(Ze|ZCZZona\ 6)}{Evidence}$$

$$\frac{0,25 * 0,271959621 * 0,682995003 * 0,004642665 * 1,404468274 * 0,090956268 * 0,452624419 * 0,214500263 * 7,88943E-22 * 1,60783E-10 * 0,382298966 * 0,245203398 * 1,779257011}{5,6571E-36}$$

$$\frac{1,90083E-39}{1,90083E-39} = 1$$

   Perform posterior calculations for all possible characters.

Table 6 Posterior Calculation Results

Posterior Karakter	Hasil
Ze	1
Ra	0
Chi	0
N	0

From the posterior calculations in each class above, find the largest value, where the largest posterior value shows the class recognition results for the test data that is the character of ze. Perform for all test characters to do classification testing, so obtained the results of recognition on bebeberapa test character on the image of the food packaging composition.

Tabel 7 Results Introduction

No	Citra Karakter	Kelas	Hasil Pengenalan	Keterangan	No	Citra Karakter	Kelas	Hasil Pengenalan	Keterangan
1.		Ze	Ze	Benar	11.		X	N	Salah
2.		Ra	Ra	Benar	12.		X	N	Salah
3.		Chi	Chi	Benar	13.		X	Chi	Salah
4.		N	N	Benar	14.		X	Chi	Salah
5.		X	Ze	Salah	15.		X	N	Salah
6.		X	N	Salah	16.		X	N	Salah
7.		X	N	Salah	17.		X	Ra	Salah
8.		Ra	Ra	Benar	18.		X	N	Salah
9.		X	N	Salah	19.		X	N	Salah
10.		X	Chi	Salah	20.		X	N	Salah

From the test plan analysis table above, the following conclusions are obtained.

a. The number of characters is correct = 5
b. The number of characters is wrong = 15
c. The total number of characters = 20

Then calculated the accuracy of the introduction so obtained the following results.

$$\text{Akurasi} = \frac{\text{Jumlah Karakter Benar}}{\text{Jumlah Keseluruhan Karakter}} * 100\%$$

$$\text{Akurasi} = \frac{5}{20} * 100\% = 25\%$$

2.2.9 Detection of Haram Status

In this process, the system will perform the search and matching of illicit materials that come on listHaram on the results of character recognition on the image of the composition of pre-packaged Japanese food. There are some conditions where the system can detect the status of the haram, as shown in the following table.

Table 9 Status Status Detection

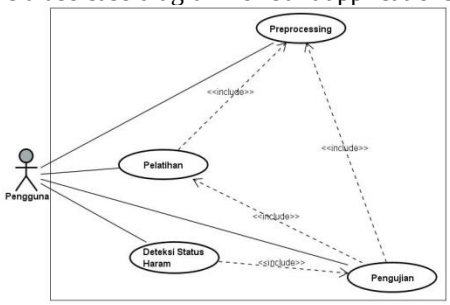
No	Hasil Pengenalan	Status
1.	Ze-Ra-Chi-N	Haram
2.	Ze-Ra-Chi-X	Halal
3.	Ze-Ra-X-N	Halal
4.	Ze-X-Chi-N	Halal
5.	X-Ra-Chi-N	Halal
6.	X-X-X-X	Halal

Table 3.23 above explains that if the result of character recognition is appropriate and sequential as in condition no. 1 ie Ze-Ra-Chi-N, then the status of the Japanese packaging food is haram. In addition to these conditions, then the status obtained is lawful.

From the results of the introduction in cases that have been done before, then the status obtained is forbidden on the input image of the packaging food composition.

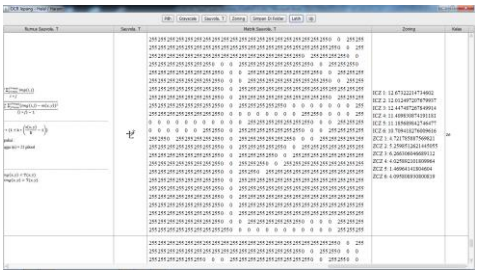
2.3 Use Case Diagram

Use case diagram is a model that shows the interaction between the system with the user (actor). In use cases are created using five use cases, where there are use cases that are related to other use cases. Here is a use case diagram for built applications.

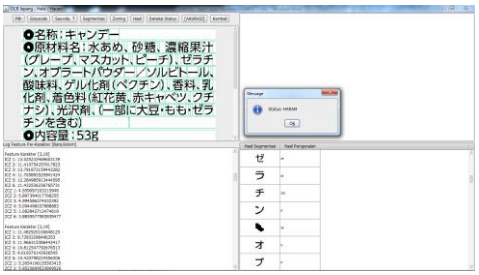


Picture 8 Use Case Diagram

2.4 Interface Implementation



Picture 9 Course Interface Menu



Picture 10 Test Menu Interface

2.5 Testing Accuracy

In this research, the accuracy testing is to conduct the testing process with the number of Japanese food packaging composition image that is recognized as zerachin in the system to the number of composite whole haram image. In this research, accuracy testing used several parameters in Zoning feature extraction process, which is division with several zones such as division of 6 zones, 8 zones and 12 zones. Table below are some of the parameters used.

Table 8 Test result

Pengujian ke-	Akurasi
1. Pembagian 6 zona	33,33%
2. Pembagian 8 zona	20%
3. Pembagian 12 zona	0%

From the table above test results, obtained The best test on the 1st parameter, which obtained an accuracy of 33.33%.

3. CLOSING

3.1 Conclusion

Based on the test results of the detection system of the disposable status of Japanese packaging food using Zoning feature extraction method (ICZ and ZCZ) combined with Naive Bayes classification method obtained the highest accuracy reached 33.33% on the test results using the 6 zone division on the extraction of its features. It is concluded that the Naive Bayes method can classify the image of Japanese packaging food composition, only the accuracy. The result of accuracy is influenced from the testing process at the classification stage which is also low accuracy.

3.2 Suggestion

In order for subsequent research on detection of the illegitimate status of Japanese packaging food to

obtain higher accuracy, the following are suggestions that can be taken into account :

1. Using other segmentation methods such as the modern Canny edge detection method or other segmentation methods that can better separate all characters on the test image.
2. Can use other classification methods such as Support Vector Machine (SVM) or Deep Learning as a comparison.

BIBLIOGRAPHY

[1] Akhmad R. M. Rizky B. S. Akhmad F. H. Ach Dwi A. Nuri N. A. U, "Optical Character Recognition Dengan Metode Naive Bayes".

[2] Fahmi Riekal, "Analisis Gray Level Difference Method Dan Metode Naive Bayes Mengidentifikasi Penyakit Lidah Manusia", Universitas Komputer Indonesia, Bandung, 2016.

[3] Sri K, "Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification" Universitas Islam Indonesia, Yogyakarta.

[4] Kusmian A, "Implementasi Algoritma Naive Bayes Untuk Klasifikasi Citra Berdasarkan Ekstraksi Ciri Tekstur Dengan Metode Matriks Kookurensi", Universitas Komputer Indonesia, Bandung, 2015.

[5] I. Mulia, "Pengenalan Aksara Sunda Menggunakan Ekstraksi Ciri Zoning Dan Klasifikasi Support Vector Machine", Institut Pertanian Bogor, Bogor, 2012.

[6] Sugiyono, "Metode Penelitian Kuantitatif, Kualitatif, dan Kombinasi (Mixed Methods), Bandung : Alfabeta, 2014.

[7] Sutoyo T. dkk, "Teori Pengolahan Citra Digital", Yogyakarta : Andi.

[8] U. Ahmad, "Pengolahan Citra Digital Dan Teknik Pemrogramannya", Yogyakarta : Graha Ilmu.

[9] Purnama. Agus, "Operasi Pengolahan Citra Digital", Tersedia di http://elektronika-dasar.web.id/operasi-pengolahan-citra-digital/, [Di akses 28 April 2017].

[10] A. Kadir. A. Susanto, "Teori Dan Aplikasi Pengolahan Citra", Yogyakarta : ANDI, 2013.

[11] C. Solomon. T. Breckon, "Fundamental Of Digital Image Processing", West Sussex : John Wiley & Sons. Ltd, 2011.

[12] K. Khurshid. I. Siddiqi. C. Faure. dan N. Vincent, "Comparison Of Niblack Inspired Binarization Methods For Ancient Document", Document Recognition And Retrieval XVI, San Jose, 2009.

[13] S. V. Rajashekararadhya. P. Vanaja R, "Zone Based Feature Extraction Algorithm For Handwritten Numeral Recognition Of South Indian Scripts", Digital Technology Journal, vol. 2, pp.41-51, 2009.

[14] Prasetyo. E, "Data Mining Konsep Dan Aplikasi Menggunakan Matlab", Yogyakarta : ANDI.

[15] B. Hariyanto, "Rekayasa Sistem Berorientasi Objek", Bandung : Informatika, 2004.

[16] S. R. A. Shalahudin. M, "Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek", Bandung : Informatika, 2013.

[17] R. Sianipar, "Teori Dan Implementasi Java", Bandung : Informatika, 2013.

[18] B. Hariyanto, "Rekayasa Sistem Berorientasi Objek", Bandung : Informatika, 2004.

[19] Kim Hamilton. Russel Miles, "Learning UML 2.0", O'Reilly : Sebastopol, 2006.

[20] J. Sauvola. T. Seppanen. S. Haapakoski. M. Pietikainen, "Adaptive Document Binarization", 4th Int. Conf. On Document Analysis And Recognition, Ulm, Germany, pp.147-152, 1997.

[21] Z. A. Agus. Y. W. Arya. Laili, "Algoritma Thresholding Adaptif Berdasarkan Deteksi Blok Terhadap Citra Dokumen Terdegradasi", Institut Teknologi Sepuluh Nopember, Surabaya.