

## CHAPTER 4

### ANALYSIS AND DESIGN

#### 4.1 Analysis

Classification of freshness level of ikan bandeng using Naive Bayes is a program to help determine whether the ikan bandeng is fresh or not

Table 4.1: Sample training data table

	Eye	Gill	Skin
Fresh			
Not Fresh			

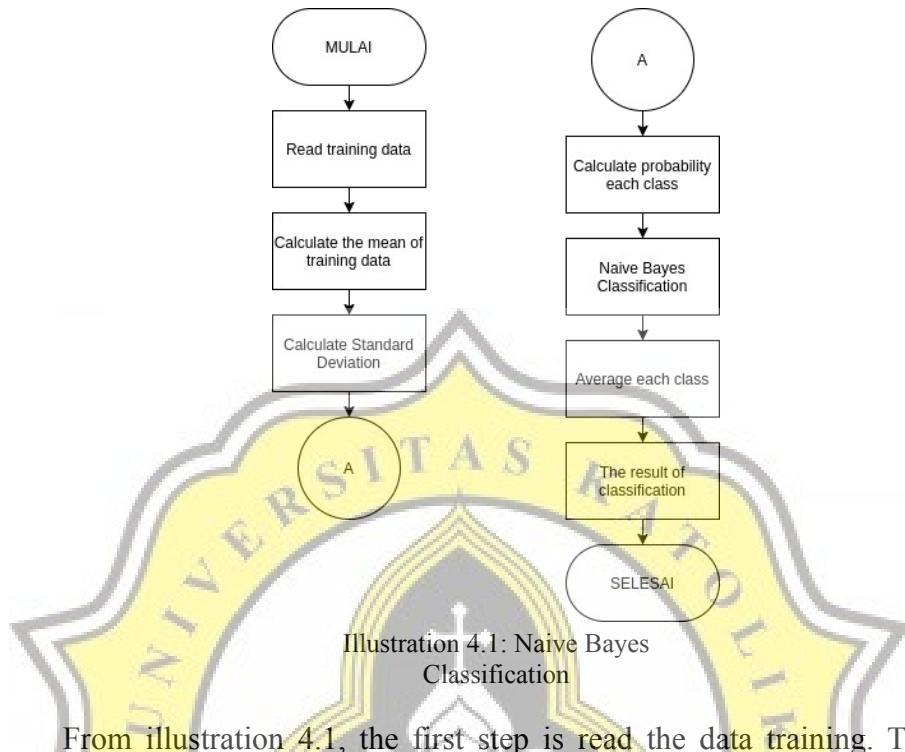
Then to find out whether the program can run well or not, it can be seen from the data provided in table 4.1 which is an example of training data for training algorithms. From the example image, the training data processing is carried out. First convert image to grayscale then convert grayscale into binary, because the results from convert to binary makes it easier to calculate the mean value. Then the next step is calculate mean R,G,B of each image. Thereafter the mean R,G,B values of each image in training data used for data processing on the Naive Bayes classification.

Table 4.2: Sampe testing data table

Mata	Insang	Daging	Kelas
			?

After training data is obtained, the next step is enter testing data that will be used to determine performance algorithm that have been previously trained when finding new data that have never been seen before, for example testing data can be seen in table 4.2. In testing data there is the same stage as training data that is, convert image into grayscale and convert grayscale into binary. Then looking for the mean value of R,G,B from the image. The mean value of R,G,B in testing data will be used to serach probability in Naive Bayes algorithm.

After doing the training data process and inputting data testing, then the next step is classification process using the Naive Bayes algorithm. Naive Bayes algorithm is a statistical approach to calculating the probability or probability of a classification problem. In this method all attributes contribute to decision making. This method is used to find the probabiltiy value in each factor. This is the Naive Bayes flowchart classification used in this program:



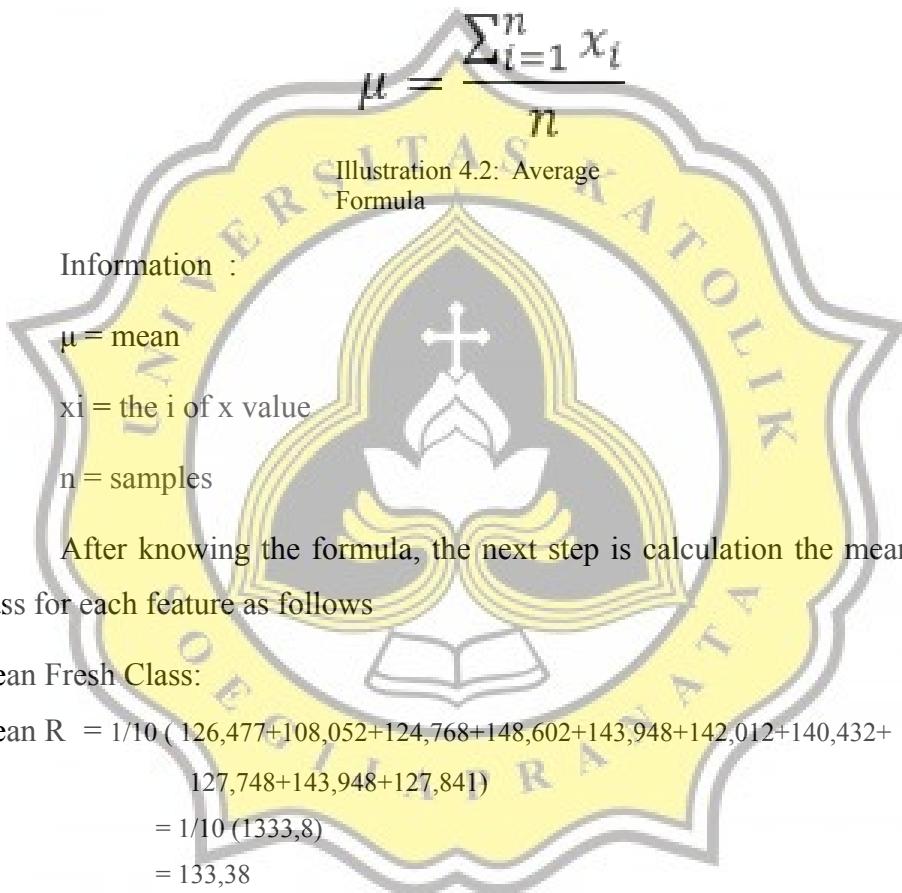
From illustration 4.1, the first step is read the data training. This is an example of the training data used in this program:

Table 4.3: Training data table

No	Mean Red	Mean Green	Mean Blue	Class
1	126,477	125,257	125,575	Segar
2	108,052	108,523	102,915	Segar
3	124,768	133,232	136,733	Segar
4	148,602	152,766	151,547	Segar
5	143,948	133,538	126,11	Segar
6	142,012	138,77	130,258	Segar
7	140,432	136,147	129,074	Segar
8	127,748	128,142	129,119	Segar
9	143,948	133,537	126,109	Segar
10	127,841	128,158	129,065	Segar
11	140,946	129,952	122,737	Tidak Segar
12	123,641	111,94	103,089	Tidak Segar
13	127,921	108,496	112,804	Tidak Segar
14	104,122	107,035	94,529	Tidak Segar
15	122,374	106,174	105,568	Tidak Segar
16	134,212	137,565	132,252	Tidak Segar

17	104,837	101,369	97,251	Tidak Segar
18	118,981	109,956	99,032	Tidak Segar
19	98,978	55,255	65,939	Tidak Segar
20	77,513	32,422	71,278	Tidak Segar

Next step is calculate the average value from each class which will be used as a parameter on the gaussian formula



After knowing the formula, the next step is calculation the mean of each class for each feature as follows

Mean Fresh Class:

$$\begin{aligned}\text{Mean R} &= 1/10 (126,477+108,052+124,768+148,602+143,948+142,012+140,432+ \\ &\quad 127,748+143,948+127,841) \\ &= 1/10 (1333,8) \\ &= 133,38\end{aligned}$$

$$\begin{aligned}\text{Mean G} &= 1/10 (125,257+108,523+133,232+152,766+133,538+138,77+136,147+ \\ &\quad 128,142+133,537+128,158) \\ &= 1/10 (1318,07) \\ &= 131,81\end{aligned}$$

$$\begin{aligned}\text{Mean B} &= 1/10 (125,575+102,915+136,733+151,547+126,11+130,258+129,074+ \\ &\quad 129,119+126,109+129,065) \\ &= 1/10 (1286,5) \\ &= 128,65\end{aligned}$$

Mean Not Fresh Class:

$$\begin{aligned}
 \text{Mean R} &= 1/10 (140,946+123,641+127,921+104,122+122,374+134,212+104,837+ \\
 &\quad 118,981+98,978+77,513) \\
 &= 1/10 (1153,5) \\
 &= 115,35
 \end{aligned}$$

$$\begin{aligned}
 \text{Mean G} &= 1/10 (129,952+111,94+108,496+107,035+106,174+137,565+101,369+ \\
 &\quad 109,956+55,255+32,422) \\
 &= 1/10 (100,02) \\
 &= 100,02
 \end{aligned}$$

$$\begin{aligned}
 \text{Mean B} &= 1/10 (122,737+103,089+112,804+94,529+105,568+132,252+97,251+ \\
 &\quad 99,032+65,939+71,278) \\
 &= 1/10 (1004,5) \\
 &= 100,45
 \end{aligned}$$

This is a table mean result of each class on every feature:

Table 4.4: Average result table

No	Red	Green	Blue	Class
1	133,38	131,81	128,65	Fresh
2	115,35	100,02	100,45	Not Fresh

Then, after obtained mean value from each class. The next step is look for the Standard Deviation value that will be used as the 2nd parameter in the gaussian formula. This is the formula:

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

Illustration 4.3: Standard Deviation Formula

Information :

$\sigma$  = standard deviation

$x_i$  = the i of x value

$\mu$  = mean

n = samples

After know the formula, this is how to calculate the standard deviation from each class for each feature:

Fresh class standard deviation: :

$$\text{Standard Deviation R} = (((126,477-133,38)^2) + ((108,052-133,38)^2) + \sum_{i=1}^{10} (X_i - \mu)^2 + ((124,768-133,38)^2) + ((148,602-133,38)^2) + ((143,948-133,38)^2) + ((142,012-133,38)^2) + ((140,432-133,38)^2) + ((127,748-133,38)^2) + ((143,948-133,38)^2) + ((127,841-133,38)^2))$$

$$= 1405,04$$

$$s^2 = \frac{\sum_{i=1}^{10} (X_i - \mu)^2}{n-1} = \frac{1405,04}{9} = 156,11$$

$$s = \sqrt{156,11} = 12,49.$$

$$\text{Standard Deviation G} = (((125,257-131,81)^2) + ((108,523-131,81)^2) + \sum_{i=1}^{10} (X_i - \mu)^2 + ((133,232-131,81)^2) + ((152,766-131,81)^2) + ((133,538-131,81)^2) + ((138,77-131,81)^2) + ((136,147-131,81)^2) + ((128,142-131,81)^2) + ((133,537-131,81)^2) + ((128,158-131,81)^2))$$

$$= 1151,24$$

$$s^2 = \frac{\sum_{i=1}^{10} (X_i - \mu)^2}{n-1} = \frac{1151,24}{9} = 125,15$$

$$s = \sqrt{156,11} = 11,19.$$

$$\begin{aligned}
 \text{Standard Deviation B} &= (((125,575-128,65)^2) + ((102,915-128,65)^2) + \\
 &\quad \sum_{i=1}^{10} (X_i - \mu)^2 \\
 &\quad (((136,733-128,65)^2) + ((151,547-128,65)^2) + \\
 &\quad (((126,11-128,65)^2) + ((130,258-128,65)^2) + \\
 &\quad (((129,074-128,65)^2) + ((129,119-128,65)^2) + \\
 &\quad (((126,109-128,65)^2) + ((129,065-128,65)^2)) \\
 &= 1277,41
 \end{aligned}$$

$$\begin{aligned}
 s^2 &= \frac{\sum_{i=1}^{10} (X_i - \mu)^2}{n-1} \\
 &= \frac{1277,41}{9} = 141,93 \\
 s &= \sqrt{141,93} = 11,91
 \end{aligned}$$

Not Fresh class standard deviation:

$$\begin{aligned}
 \text{Standard Deviation R} &= (((140,946-115,35)^2) + ((123,641-115,35)^2) + \\
 &\quad \sum_{i=1}^{10} (X_i - \mu)^2 \\
 &\quad (((127,921-115,35)^2) + ((104,122-115,35)^2) + \\
 &\quad (((122,374-115,35)^2) + ((134,212-115,35)^2) + \\
 &\quad (((104,837-115,35)^2) + ((118,981-115,35)^2) + \\
 &\quad (((98,978-115,35)^2) + ((77,513-115,35)^2)) \\
 &= 3236,49
 \end{aligned}$$

$$\begin{aligned}
 s^2 &= \frac{\sum_{i=1}^{10} (X_i - \mu)^2}{n-1} \\
 &= \frac{3236,49}{9} = 359,61 \\
 s &= \sqrt{359,61} = 18,96
 \end{aligned}$$

$$\begin{aligned}
 \text{Standard Deviation G} &= (((129,952-100,02)^2) + ((111,94-100,02)^2) + \\
 &\quad \sum_{i=1}^{10} (X_i - \mu)^2 \\
 &\quad (((108,496-100,02)^2) + ((107,035-100,02)^2) + \\
 &\quad (((106,174-100,02)^2) + ((137,565-100,02)^2) + \\
 &\quad (((101,369-100,02)^2) + ((109,956-100,02)^2) + \\
 &\quad (((55,255-100,02)^2) + ((32,422-100,02)^2)) \\
 &= 9280,50
 \end{aligned}$$

$$s^2 = \frac{\sum_{i=1}^{10} (X_i - \mu)^2}{n-1} = \frac{9280,50}{9} = 1031,16$$

$$s = \sqrt{1031,16} = 32,11.$$

$$\begin{aligned} \text{Standard Deviation B} &= (((122,737-100,45)^2) + ((103,089-100,45)^2) + \\ &\quad ((112,804-100,45)^2) + ((94,529-100,45)^2) + \\ &\quad ((105,568-100,45)^2) + ((132,252-100,45)^2) + \\ &\quad ((97,251-100,45)^2) + ((99,032-100,45)^2) + \\ &\quad ((65,939-100,45)^2) + ((71,278-100,45)^2)) \\ &= 3783,17 \end{aligned}$$

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^{10} (X_i - \mu)^2}{n-1} = \frac{3783,17}{9} = 420,35 \\ s &= \sqrt{420,35} = 20,50. \end{aligned}$$

This is standard deviation results table from each class on every parameter:

Table 4.5: Standard Deviation value table

No	Red	Green	Blue	Class
1	12,49	11,19	11,91	Fresh
2	18,96	32,11	20,50	Not Fresh

After obtaining the result of mean and standard deviation from each class on every feature, the next step is looking for probability each class which will be used for parameter in finding the final value using the following methods:

$$\begin{aligned} P(\text{Fresh}) &= \text{The amount of "Fresh" data / total data} \\ &= 10 / 20 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} P(\text{Not Fresh}) &= \text{The amount of "Not Fresh" data / total data} \\ &= 10 / 20 \\ &= 0.5 \end{aligned}$$

This is followings probability table from each class:

Table 4.6: Probability table each class

No	Class	The amount of data	Probability
1	Fresh	10	0,5
2	Not Fresh	10	0,5
		20	

After obtaining two parameter (mean and standard deviation) from each class and probability each class. The next step is looking for probability by using a gaussian function. This is followings example testing data one image:

Table 4.7: Testing data one image table.

No	Mean R	Mean G	Mean B	Class
1	122,583	117,258	110,337	?

After knowing the data being tested. Then used Gaussian function to find out the probability value, probability value in this function will be used for specifies the parameter to be used in the final value, this is followings function will be used:

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Illustration 4.4: Rumus Gaussian

Information :

P : Chance

X<sub>i</sub> : Attribute i

x<sub>i</sub> : Attribute Value i

Y : Class

y<sub>i</sub> : Sub class

μ : Mean,

σ : Standard deviation

From this formula, then the next step is to calculate the probability value of each class for each parameter as follows:

Fresh Class Probability :

$$P_{\text{Red}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{12,49\sqrt{2*3,14}} e^{-\frac{(122,583-133,38)^2}{12,49^2}}$$

$$= \frac{1}{0,03194903460639} e^{-\frac{116,57}{156}}$$

$$= 0,021981831388316$$

$$P_{\text{Green}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{11,19\sqrt{2*3,14}} e^{-\frac{(117,258-131,81)^2}{11,19^2}}$$

$$= \frac{1}{0,035660718698285} e^{-\frac{211,76}{125,21}}$$

$$= 0,015312268342002$$

$$P_{\text{Blue}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{11,91\sqrt{2*3,14}} e^{-\frac{(110,337-128,65)^2}{11,91^2}}$$

$$= \frac{1}{0,033494592942028} e^{-\frac{335,36}{141,84}}$$

$$= 0,010277084064843$$

Not Fresh Class Probability:

$$\begin{aligned}
 P \text{ Red} &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
 &= \frac{1}{18,96\sqrt{2*3,14}} e^{-\frac{(122,583-115,35)^2}{18,96^2}} \\
 &= \frac{1}{0,021042825107724} e^{-\frac{52,31}{359,48}} \\
 &= 0,019567493548337
 \end{aligned}$$

$$\begin{aligned}
 P \text{ Green} &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
 &= \frac{1}{32,11\sqrt{2*3,14}} e^{-\frac{(117,258-100,02)^2}{32,11^2}} \\
 &= \frac{1}{0,012426697258864} e^{-\frac{297,07}{1031,05}} \\
 &= 0,010758576824818
 \end{aligned}$$

$$\begin{aligned}
 P \text{ Blue} &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
 &= \frac{1}{20,50\sqrt{2*3,14}} e^{-\frac{(110,337-100,45)^2}{20,50^2}} \\
 &= \frac{1}{0,01946315556922} e^{-\frac{97,75}{420,25}} \\
 &= 0,017325841623032
 \end{aligned}$$

This is followings probability value table from each class for each parameter:

Table 4.8: Gaussian value table

No	Prob. Red	Prob. Green	Prob. Blue	Class
1	0,021981831388316	0,015312268342002	0,010277084064843	Fresh
2	0,019567493548337	0,010758576824818	0,017325841623032	Not Fresh

Then, the final step to find the final value in the classification by multiplying all the probabilities R,G,B along with the probability of each class.

Fresh Class

Final Value = P Blue x P Green x P Red x Probability per class  

$$0,021981831388316 \times 0,015312268342002 \times 0,010277084064843 \times 0,5$$
  

$$= 1,72959060266694E-06$$

Not Fresh Class

Final Value = P Blue x P Green x P Red x Probability per class  

$$0,019567493548337 \times 0,010758576824818 \times 0,017325841623032 \times 0,5$$
  

$$= 1,82370407790941E-06$$

From the final results of the 2 classes, it can determined that image is in the fresh or not fresh category by looking at the largest number. From the result of classification, the image is classified as not fresh.

Table 4.9: Table of classification results of one image

No	Mean R	Mean G	Mean B	Class
1	122,583	117,258	110.337	Tidak Segar

Example of testing data by using three image:

Table 4.10: Testing data three image table.

No	Mean R	Mean G	Mean B	Class
1	122,583	117,258	110.337	?
2	91,34	82,43	74,96	
3	57,45	61,60	61,21	

To do testing with three image, the methods is same as one image. Looking for final probability each class, then added and averaged. Then the biggest final value will be used for the final result, as in the following table:

Table 4.11: Average result table

No	Image 1	Image 2	Image 3	Average	Class
1	1,7295906026669 4E-06	1.5272549598441 044E-16	5.6821229730677 05E-29	5.7609891573040 14E-7	Fresh
2	1,8237040779094 1E-06	4.5371789253648 186E-7	1.8868662382035 646E-9	7.5931332033037 E-7	Not Fresh

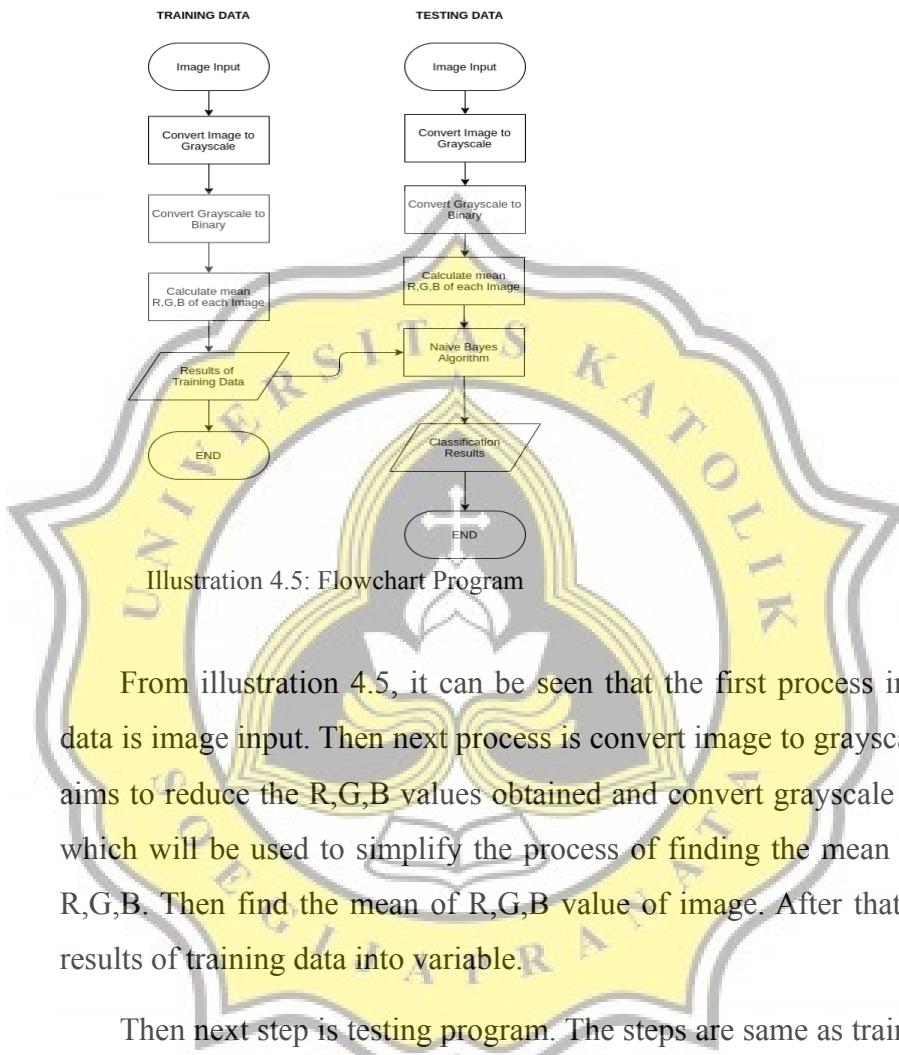
So, the final result is that the fish is not fresh. Because in the calculation average shows the highest value in the Not Fresh class.

Table 4.12: Tabel Hasil Testing 3 Gambar

Image Data	Mean Red	Mean Green	Mean Blue	Output
Image 1	122,58	117,25	110,336	Not Fresh
Image 2	91,34	82,43	74,96	
Image 3	57,45	61,60	61,21	

## 4.2 Design

### a. Flowchart Program



From illustration 4.5, it can be seen that the first process in training data is image input. Then next process is convert image to grayscale which aims to reduce the R,G,B values obtained and convert grayscale to binary which will be used to simplify the process of finding the mean values of R,G,B. Then find the mean of R,G,B value of image. After that save the results of training data into variable.

Then next step is testing program. The steps are same as training data, that is, Image input to be convert in to grayscale which aims to reduce the R,G,B values obtained and convert grayscale to binary which will be used to simplify the process of finding the mean values of R,G,B. Then looks for the mean R,G,B value of image. The next step is using Naive Bayes Classifier for doing classification image, whether it is categorized as fresh fish or not fresh fish. After get the results process of testing data is complete.

b. Use Case Diagram

