

## Analysis of the Naïve Bayes Method in Classifying Formalized Fish Images Using GLCM Feature Extraction


Ayu Pariyandani<sup>1</sup>, Eka Pirdia Wanti<sup>2</sup>, Muhathir<sup>3</sup>

<sup>1,2,3</sup>Departement of Information Technology, University of Medan Area, Indonesia

### ABSTRACT

Fish is one of the foods that are high in protein so that many Indonesians consume fish as protein intake for health. Fish can be found in any waters including Indonesian marine waters, so that some of the Indonesian people work as fishermen. This causes the number of fish catches to increase and the fishermen have to sell the fish quickly in at least one day because the fish will rot easily if not consumed immediately. This has led some traders to cheat by mixing formaldehyde with fish that are not sold out. This action is very detrimental to consumers, so they must be more vigilant in choosing or buying fish on the market. One way for consumers to recognize formaldehyde fish is a technology that can distinguish fresh fish or formalin fish based on the image of the fish, Naive Bayes and GLCM (Gray Level Co-Occurrence Matrix) by using this method the accuracy of this system can reach up to 70%.

**Keyword : Fish, Formalin, Classification, Naive bayes, GLCM (Gray Level Co-Occurrence Matrix).**

 This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

#### **Corresponding Author:**

muhathir,  
Department of Information Technology  
University of Medan Area,  
Jl. Kolam Nomor 1 Medan Estate/Jalan Gedung PBSI, Medan 20223, Indonesia.  
Email : [muhathir@staff.uma.ac.id](mailto:muhathir@staff.uma.ac.id)

#### **Article history:**

Received Aug 27, 2020  
Revised Sep 03, 2020  
Accepted Sep 06, 2020

### 1. INTRODUCTION

The vast area of Indonesian waters is one of the marine and marine sectors that has great potential and is superior in the main movement of the economic development of the Indonesian people. The availability of abundant natural resources in the territory of Indonesia creates tremendous economic potential, so that it will produce products and services with high competitiveness, as long as they can be managed properly and appropriately (Rizka, 2017). In contrast to the existing land areas in Indonesia, the availability of natural resources is increasingly limited, especially in land acquisition, this is the impact of the implementation of development. Along with the increasing population of Indonesia, making Indonesia a developing country with a very large population.

Catching fresh fish on a fisherman ship is the most important and very decisive first step of the next fresh fish customer, when the fish is removed from the water, the quality of the fresh fish can be determined. From the abundance of fish catches of fishermen, fish stocks in the market are increasing. This invites fish sellers to spend their fish sales (Rizal, 2017). This is so that the fish does not rot. From this occurrence, fish traders have to turn their minds so that they do not experience losses due to the fish they sell not being sold completely. So that the traders mix fresh fish with formaldehyde so that the fish can last a long time so that there is no loss as a result of the fish they sell are not sold as a whole.

With this, consumers experience a huge loss because the fish they buy is no longer fresh fish but fish that has been mixed with formalin. Even though formaldehyde is not good for health. If consumed in the long term and in a large capacity, it will make health worse, especially causing disease in humans.

Therefore, to be able to assist the public in recognizing fresh fish or fish that have been mixed with formalin, technology is created that can classify fish with formalin based on the image of the fish. This technology was created using the Naive Bayes method and GLCM, while the Naive Bayes method

was used as a classification of fish images. Naive Bayes is a machine learning method that uses probability calculations. This algorithm utilizes probability and simple statistical methods with the assumption that one class and another are independent (Lowd, 2005). The feature extraction used is GLCM (Gray Level Co-Occurrence Matrix) which is a method using statistical calculations in image texture extraction which considers the spatial relationship of pixels in the image (Sidubyo et al., 2018). Pattern recognition using GLCM feature extraction has been carried out, such as research by Pariyandani et al (2019) classifying formalized fish images using the KNN (K-Nearest Neighbor) and GLCM (Gray Level Co-Occurrence Matrix) methods with a dataset of 500 images consisting of from 250 fresh fish images and 250 formalin fish images with an accuracy rate of 92%.

So this study uses the naïve bayes method and feature extraction of the GLCM (Gray Level Co-Occurrence Matrix) which is expected to help the wider community in classifying formalized fish images with fresh fish images. So that it helps the community, especially mothers, to help choose fish that is good for consumption so that it does not cause disease if consumed for a long time and to awaken naughty fish traders, the existence of a formalized fish image classification system will reduce fraud between fish traders and consumers.

## 2. LITERATURE REVIEW

### A. Fish

Fish are organisms that are classified as vertebrates and have a very wide variety of species. Fish are animals that live in large water pools. Fish can be found anywhere in the waters such as brackish waters, freshwater and marine waters (Dewi & Harsindhi, 2018).

The availability of abundant natural resources, especially in water areas, makes the majority of Indonesians work as fishermen. Sea waters are one of the potential areas with a large number of commodity types, including fish, shellfish, sea cucumber, crab, squid, shrimp, seaweed and so on. Fish is a sea catch that is well known by all people. This is because fish is mostly known and consumed by upper economic communities to lower economic communities. This is because the price of fish is quite affordable for all customers compared to other protein sources, such as beef, chicken, milk and eggs. Fish and other fishery products are relatively affordable animal source foods.

Fish is a food with a high protein source that makes many consumers need this food. Fish are animals that are very much found in Indonesian waters. This is because Indonesia is a marine country that produces many fish species living in Indonesian waters (Sarwanto et al., 2014). This makes Indonesian fish catch quite a lot so that it can meet the protein needs of the Indonesian nation.

### B. Formalin

Formalin is a chemical compound whose chemical formula is  $H_2CO$ , where this substance is a highly reactive aldehyde gas formed by incomplete oxidation or combustion of hydrocarbons (Eftika & Passarella, 2018). Formalin has uses, among others, as the manufacture of resins and textiles, as a disinfectant and as a fixative or preservative for objects in the laboratory. Formalin is a poisonous gas that is colorless, formalin is also readily dissolved in water and is generally distributed as a 37% solution in water, while a solution of formalin with 10% in water is used as a desinfectant and species preservative in the laboratory (Tatuh, Rorong & Sudewi, 2016; Handayani & Mutiara, 2020).

Formalin is a clear, colorless solution, but has an offensive odor. Formalin can be used to preserve small animals to humans. However, some people use formaldehyde to preserve food ingredients so that they last longer. Whereas the use of formaldehyde to preserve food is clearly prohibited because of the effect that will be generated if humans consume this substance (Trisnawati & Setiawan, 2019). The United States Environmental Protection Agency (EPA) and the International Institute for Cancer Research (IARC) state that formalin is carcinogenic, that is, it can lead to cancer. In the environment, formaldehyde can be found in cigarette smoke, car exhaust, smoke from fires and in the atmosphere [33]. If the formaldehyde in the air exceeds the permissible level of 0.1 ppm (parts per million), half of the individuals are at risk of experiencing symptoms including burning sensations in the eyes, nose and throat, not only that the effect of this formalin can cause nausea, dizziness and experiencing skin irritation if exposed to this formalin directly (Puspawiningtyas, Pamungkas & Hamad, 2017; Singgihm 2017).

### C. GLCM (Gray Level Co- occurrence Matrix)

GLCM (Gray Level Co-occurrence Matrix) is a method that functions to analyze the texture or feature extraction (Haba & Pelangi, 2020). In 1973 Haralick proposed the Gray Level Co-Occurrence Matrix (GLCM) as texture extraction from an image. To explain the spatial pattern the GLCM uses 28 features (Kumar, 2020). The GLCM (Gray Level Co-occurrence Matrix) technique is the oldest technique for analyzing texture or feature extraction. The gray level of the GLCM has 2 parameters, namely direction and distance (Lusiana et al., 2019). The feature extraction of the GLCM is carried out in 4 angular directions which include the respective angle intervals of 45° namely 0°, 45°, 90°, 135°. So that GLCM is a technique for obtaining statistical values from the second order, this is the initial process of creating a matrix work area and determining the value. angles and the closest distance (Rizal et al., 2019) GLCM was developed by Robert Haralick, where this method is very often used, so the following are the steps in performing feature extraction using GLCM (Nabella, Sari, Wihandika, 2019; Ahsani, Sari & Adikara, 2019).

- a. Convert from RGB image to grayscale
- b. Determine the value of the maximum intensity on a grayscale image
- c. Forming a co-occurrence (CM) matrix in the directions of 0°, 45°, 90°, and 135° with a maximum intensity value size of 1
- d. Calculates the element value of each matrix where the value of each matrix element states the number of neighbors, the frequency of occurrence of the sequence of values  $i, j$
- e. Perform normalization by dividing the value of each matrix element by the number of values from the matrix. Normalization is done to eliminate dependence on image measurement values
- f. Counting the GLCM features on each co-occurrence (CM)

As for the GLCM (Gray Level Co-occurrence Matrix) method, there are 4 calculated features, namely (Sudibyoto et al., 2018; Pristanti, 2019):

- a. Contrast  
The contrast value shows that as the intensity variation between images, the higher the contrast value, the more varied the intensity value in the image.
- b. Energy  
This feature is used to see the level of texture uniformity, where the higher the energy value, the higher the level of texture homogeneity.
- c. Entropy  
From this entropy value, it can provide information on coarse and fine texture features from the image extraction that has been carried out. The higher the entropy value, the coarser the resulting texture on the image
- d. IDM (Inverse Different Moment)  
IDM (Inverse Difference Moment) is a feature used for measurement that defines the intrinsic level variation of pixel pairs in an image to be tested.

### 4. Naive Bayes

Naive Bayes is one of the most efficient and effective inductive learning algorithms for data mining and machine learning. This is due to the competitive performance of Naive Bayes in the classification process even though it uses the assumption of independence. The assumption of independence (there is no link between attributes) in the actual data rarely occurs, but even though the assumption of independence of these attributes is violated, the performance of the naive bayes classification is quite high, this is proven in various empirical studies.

The naive bayes method was proposed by a scientist in the 18th century named Thomas Bayes (Annur, 2018; Salvador & Perez-Pellitero, 2015). This method is an approach method for an uncertain thing which is measured by probability, so that the Bayes method will produce the category label with the highest probability value (McCann & Love, 2011). Naïve Bayes is also a very popular classification method and is included in the ten best algorithms in data mining (Zaidin, 2013).

Naive Bayes is a simple probabilistic classification method. In this method calculates a set of probabilities by adding the values of the frequencies and the combination of values from the existing dataset (Narayana, Arora & Bhatia, 2012). Naïve Bayes is a method that uses a branch of mathematics with probability theory to seek from the greatest possible probability of classification, by looking at the frequency of each classification in the training data (Saritas & Yasar, 2019). In this method, each

attribute is considered to have no relation with one another or is called independent, but this method requires training data before classification (Siregar, 2018; Guntur et al., 2018; Devita, 2018). The advantage of the Naive Bayes method is that it only requires a small amount of training data to determine the mean and variance parameters of a variable required for the classification process (Calders & Verwer, 2010). Here is a step up from the naïve Bayes method.

The steps in the naïve Bayes method are as follows (Fadlan, Ningsij & Windarto, 2018):

- a. Read training data
- b. Calculate from sums and probabilities
  - If there is numeric data, then determine the mean and deviation value of each parameter indicating the numerical value
  - If the data is not numeric, then calculate the probability value for each of the same categories, by means of the appropriate amount of data from the same category then divided by the number of data so that it can determine the probability value.
- c. Obtain values from the mean, standard deviation and probability tables
- d. Calculating the test data, but if there is numeric data look for the value of the gaussian distribution of each feature. If there is no numeric data then go to the next step
- e. The final probability of each class, enter and calculate all probability data into the same class
- f. Final probability, multiplying the probability of each class by the probability of the end of each class

### 3. METHOD

#### A. DataSet

The research data used includes images of fresh fish and fish that have been given a mixture of formaldehyde for their preservation. Where the image size used is 480x640 which is taken using a cellphone camera.

#### B. Research Steps

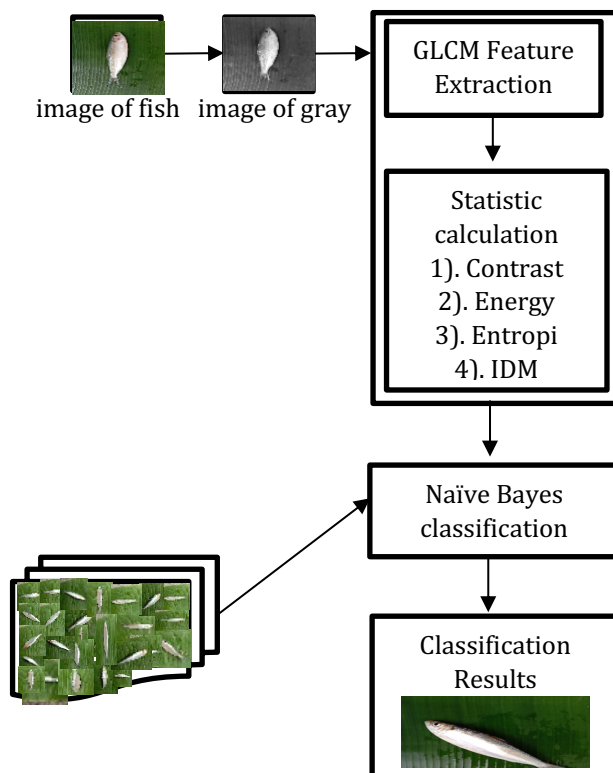


Figure 1. Research Steps

#### 4. RESULTS AND DISCUSSION

##### A. Sample

The sample in this study used images of fresh fish and images of fish that had been mixed with formalin. Where the number of each fish image (fresh fish image and fish image mixed with formaldehyde) consists of 250, so the total number of samples used is 500 fish images. Where of the total sample, 60% is used for training samples and 40% is used for testing samples. Where in Figures 2 and 3 are part of the sample used in this study.



Figure 2 Fresh Fish



Figure 3 Fish with Formalin

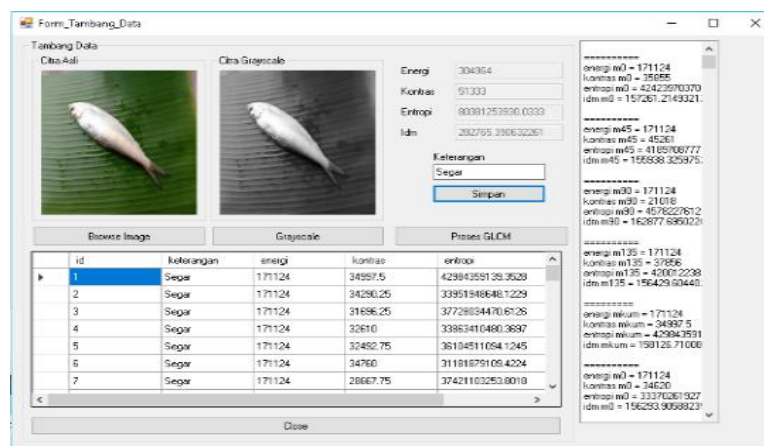


Figure 4. Fish Image Pattern Training Process

##### B. Fish Image Classification Process



Figure 5. Fish Image Pattern Classification Process

**Table 1 :** Results of the Fish Image Pattern Classification Test

	Fresh	Formalin
Fresh	246	4
Formalin	146	104
Accuracy	175	

**Table 2 :** Results of the Accuracy Test for the Classification of Fish Image Patterns

	Fresh	Formalin
Fresh	98.4	1.6
Formalin	58.4	41.6
Accuracy	70	

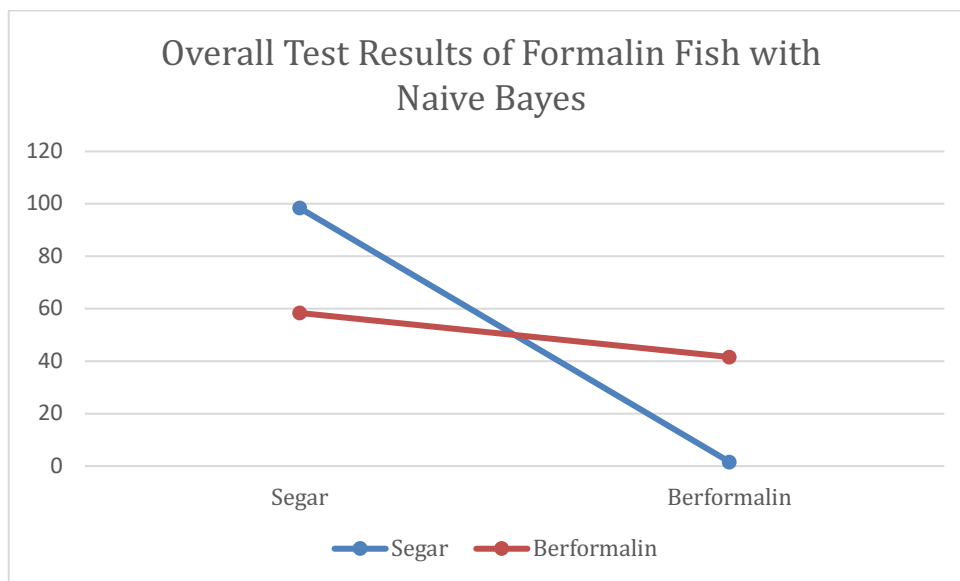


Figure 6. Test Results of Formalin Fish Using the Naïve Bayes Method

**C. Discussion result**

Figure 3.5 shows the results of a formalized fish test using the Naïve Bayes method and the extraction of the GLCM (Gray Level Co-occurrence Matrix) feature. Where the image of fresh fish was identified as the image of fresh fish by 98.4%, while the image of fresh fish was identified as the image of fish with formaldehyde by 1.6% and the image of fish mixed with formaldehyde was identified as the image of fresh fish by 58.4%, Meanwhile, the image of fish mixed with formaldehyde was identified as the image of fish mixed with formalin by 41.6%. With the total accuracy of the test the fish taste pattern is 70%.

**5. CONCLUSION**

From the results of the above research, it can be concluded that the use of the Naïve Bayes method and the extraction of the GLCM (Gray Level Co-occurrence Matrix) feature in classifying formalin-shaped

fish images is good enough in recognizing the image of the fish that is the object of this study. This is evident from the results of research which shows the accuracy rate of testing fish image patterns by 70%.

## REFERENCES

- A. F. Ahsani, Y. A. Sari and P. P. Adikara, "Temu Kembali Citra Makanan Menggunakan Ekstrasi Fitur Gray-Level-Co-occurrence Matrix dan CIE  $L^*a^*b$  Color Moments Untuk Pencarian Resep Makanan," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. Vol.3 No.3, pp. 2261-2268, Maret 2019.
- A. K. Haba and K. C. Pelangi, "Sistem Cerdas Dalam Klasifikasi Kematangan Buah Jeruk Berdasarkan Fitur Ekstrasi GLCM dengan Metode Naive Bayes," *Jurnal Teknologi & Manajemen Informatika*, vol. Vol.6 No.1, pp. 1-7, 2020.
- A. Pariyandani, D. A. Larasati, E. P. Wanti and M. , "Klasifikasi Citra Ikan Berformalin Menggunakan Metode K-NN dan GLCM," *Prosiding Seminar Nasional Teknologi Informatika*, vol. Vol.2 No.1, pp. 42-47, November 2019.
- A. Trisnawati and M. A. Setiawan, "Pelatihan Identifikasi Boraks dan Formalin Paada Makanan di Desa Bareng,Babadan, Ponorogo," *Jurnal Widya Laksana*, vol. Vol.8 No.1, pp. 69-77, Januari 2019.
- C. Fadlan, S. Ningsih and A. P. Windarto, "Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rasta," *JUTIM*, vol. Vol.3 No.1, pp. 1-8, Juni 2018.
- C. S. U. Dewi, C. J.Harsindhi and S. , Karang dan Ikan Terumbu Pulau Bawean, Malang, Maret 2018.
- C. Sarwanto, E. S. Wiyono, T. W. Nurani and J. Haluan, "Kajian Sistem Pemasaran Ikan Hasil Tangkapan Nelayan Di Kabupaten Gunung Kidul, Provinsi DIY," *Jurnal Sosek KP*, vol. Vol.9 No.2, pp. 207-217, 2014.
- D. Lowd, "Naive Bayes Models for Probabilyty Estination," *University of Washington*, pp. 1-18, 17 January 2005.
- D. T. Satya, N. Hidayat and S. , "Sistem Pakar Diagnosis Penyakit Sapi Ternak Potong Menggunakan Metode Naive Bayes-Certainty Factor," *Jurnal Pengembangan Teknologi Informmasi dan Ilmu Komputer*, vol. Vol2 No.10, pp. 3406-3410, Oktober 2018.
- E. Puspawiningtyas, R. B. Pamungkas and A. Hamad, "Upaya Peningkatan Pengetahuan Bahan Tambahan Pangan Melalui Pelatihan Deteksi kandungan Boraks dan Formalin," *Jurnal Pengabdian dan Pemberdayaan Masyarakat*, vol. Vol.1 No.1, pp. 46-55, Maret 2017.
- E. Rizka, "Diplomasi Maritim Indonesia Terhadap Aktivitas Penangkapan Ikan Ilegal (Illegal Fishing) Oleh Nelayan China Di ZEEi Perairan Kepulauan Natuna," *Jurnal Prodi Diploma Pertahanan*, Vols. Vol.3, No.2, pp. 33-47, Juni 2017.
- E. S. Kartamihardja and C. Umar, "Kebijakan Pemacuan Sumber Daya Ikan Di Perairan Umum Daratan Indonesia : Teknologi Alternatif Untuk Meningkatkan Produksi Ikan Dan Pendapatan Nelayan," *Jurnal Kebijak.Perikan.Ind*, vol. Vol.1 No.2, pp. 99-111, November 2017.
- F. Y. Nabella, Y. A. Sari and R. C. Wihandika, "Seleksi Fitur Information Gain Pada Klasifikasi Citra Makanan Menggunakan Hue Saturation Value dan Gray Level Co-occurrence Matrix," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. Vol.3 No.2, pp. 1892-1900, Februari 2019.
- G. and P. Kumar, "Pulmonary Tumor Detection by Virtue of GLCM," *Journal of cientific & Industrial Research*, vol. Vol.79, pp. 132-1334, February 2020.
- H. A. Tatuh, J. Rorong and S. Sudewi, "Analisis Kandungan Formalin Pada Berbagai Jenis Ikan di Kota Manado," *Jurnal Ilmiah Farmasi*, vol. Vol.5 No.4, pp. 162-167, November 2016.
- H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *ILKOM Jurnal Ilmiah*, vol. Vol.10 No.2, pp. 160-165, Agustus 2018.
- H. Singgih, "Uji Kandungan Formalin Pada Ikan Asin Menggunakan Sensor Warna Dengan Bantuan FMR (Formalin Main Reagent)," *Jurnal ELTEK*, vol. Vol.11 No.1, pp. 55-62, April 2019.
- J. Salvador and E. Perez-Pellitero, "Naive Bayes Super-Resulation Forest," *ICCV Paper*, pp. 325-333, 2015.
- M. Guntur, J. Santony and Y. , "Prediksi Harga Emas Dengan Menggunakan Metode Naive Bayes Dalam Investasi Untuk Meminimalisasi Resiko," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol.

Vol.2 No.1, pp. 354-360, 2018.

- M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *International KJournal of Intelligent Systems and Application in Engineering*, Vols. 2147-6977, pp. 88-91, Maret 2019.
- M. Rizal, "Diversifikasi Produk Olahan Ikan Bgi Ibu-Ibu Nelayan Di Dusun Mamua Kabupaten Maluku Tengah," *Jurnal Biology Science & Education 2017*, vol. Vol.6 No.2, pp. 1-12, Jul-Des 2017.
- N. A.Zaidi, J. Cerquides, M. J.Carman and G. I.Webb, "Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting," *Journal of Machine Learning Research 14*, pp. 1947-1988, 2013.
- R. A. Rizal, S. Gulo, O. Della, A. B. M. Napitupulu, A. Y. Gultom and T. J. Siagian, "Analisis Gray Level Co-occurrence Matrix (GLCM) Dalam Mengenali Citra Ekspresi Wajah," *Jurnal Mantik*, vol. Vol.3 No.2, pp. 31-38, Agustus 2019.
- R. N. Devita, H. W. Herwanto and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer(JTIK)*, vol. Vol.5 No.4, pp. 427-434, September 2018.
- R. R. Eftika and R. Passarella, "Implementasi Sensor Gas dan Fuzzy Logic Untuk Mendeteksi Formalin Pada Tahu," *Prosiding Annual Research Seminar (ASR) 2018*, vol. Vol.4 No.1, pp. 156-160, 2018.
- S. McCann and D. G.Love, "Local Naive Bayes Nearest Neighbor for Image Classification," *Journal of Technical Report TR-2011-11*, pp. 1-10, December 2011.
- T. Calders and S. Verwer, "Three Naive Bayes approaches for Discrimination-Free Classification," *Eindhoven University of Technology*, pp. 277-292, 2010.
- T. Handayani and S. Mutiara , "Pemeriksaan Kandungan Zat Kimia Formalin Pada Bakso Ikan dan Tahu," *Jurnal Katalisator*, vol. Vol.5 No.1, pp. 81-87, 2020.
- T. Handayani, "Identifikasi Formalin Pada Ikan yang Dijual di Pasar Lasi Kabupaten Agam Tahun 2019," *Prosiding SENAKES 1.0*, no. 978-623-93603-0-6, pp. 75-79, 2019.
- U. Sudibyo, D. P. Kusumaningrum, E. H. Rachmawanto and C. A. Sari, "Optimasi Algoritma Learning Vector Quantization (LVQ) Dalam Pengklasifikasian Citra Daging Sapi dan Daging Babi Berbasis GLCM dan HSV," *Jurnal SIMETRIS*, vol. Vol.9 No.1, pp. 1-10, April 2018.
- V. Lusiana, I. H. Al Amin, B. Hartono and T. Kristianto, "Ekstrasi Fitur Menggunakan Matriks GLCM Pada Citra Variasi Arah Objek," *Prosiding SENDI\_U 2019*, vol. vol.3 No.2, pp. 272-275, 2019.
- V. M. M. Siregar, "Sistem Pendukung Keputusan Penentuan Intensif Bulanan Pegawai Dengan Menggunakan Metode Naaive Bayes," *Jurnal SISTEMASI*, vol. Vol.7 No.2, pp. 87-94, Mei 2018.
- V. Narayana, I. Arora and A. Bhatia, "Fast and Accurate Sentimrnt Classificaation Using an Enhanced Naive Bayes Model," *Journal of Indian Institute of Technology*, pp. 1-8, 2012.
- Y. D. Pristanti , P. Mudjirahardjo and A. Basuki, "Identifikasi Tanda Tangan Dengan Ekstrasi Ciri GLCM dan LBP," *Jurnal EECCIS*, vol. Vol.13 No.1, pp. 6-10, April 2019.