Journal of Intelligent Decision Support System (IDSS) x (x) (20xx), pp. x-xx

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Classification of fresh chicken meat and tainted chicken meat using naive bayes classifier algorithm

Ahmad Muzakky Zain[1] , Aang Ali Murtopo[2] , Nurul Fadila[3] , Gunawan Gunawan[4]

[1,2,3,4] *STMIK YMI Tegal, Tegal City, Central Java, Indonesia*

Article Info

Article history:

Received Mar 19, 2024
Revised Mar 20, 2024
Accepted Mar 26, 2024

Keywords:

Classification;
Chicken Meat;
GLCM;
Naive Bayes;
RapidMiner.

ABSTRACT

This research discusses the classification of fresh and tainted chicken meat using the Naive Bayes Classifier (NBC) algorithm based on Gray Level Co-occurrence Matrix (GLCM) feature extraction, with the aim of developing an efficient and accurate classification method. This research aims to utilize image processing and machine learning technologies to distinguish fresh chicken meat from tainted ones, which is crucial for the food industry. The research methodology involved the use of GLCM for texture feature extraction from chicken meat images, with the implementation of the NBC model through RapidMiner, offering an intuitive and efficient approach. The results showed the success of the model in achieving 80% accuracy, with an average precision of 81.25%, recall of 80%, and F1-score of 80.62%, confirming its ability in chicken meat classification. The integration of GLCM and RapidMiner in the application of NBC not only improves accuracy and objectivity in chicken meat classification but also provides a foundation for the wider application of machine learning techniques in ensuring food safety and consumer satisfaction.

Corresponding Author:

Ahmad Muzakky Zain,
Informatics Engineering
STMIK YMI TEGAL
#1 Pendidikan Street, Tegal City, Central Java, 52142, INDONESIA
Email: muzakkyzain@gmail.com

Introduction

In modern agriculture and with the increasing food demand, the selection of quality chicken meat species is crucial in the food industry. Chicken meat, as a significant source of protein, influences consumer preferences, nutritional value, and applications in various dishes (Agustina & Ardiansyah, 2020),(Putri & Yodfiatfinda, 2021). The importance of efficient and accurate classification between fresh and tainted chicken meat is not only to safeguard consumer health but also to guarantee the accuracy of product labeling. This is due to the health risks of consuming poorly stored chicken meat.

Traditional classification methods involve time- and resource-consuming physical and chemical analysis (Pirmansyah & Wahyudi, 2023). However, advances in information technology and artificial intelligence have enabled the development of computationally-based methods that are faster, more objective, and cost-effective (Mondejar et al., 2021). In this regard, the Naive Bayes Classifier (NBC), a probabilistic classification approach based on the Bayes theorem, has demonstrated its effectiveness in text analysis and pattern recognition, including in classification (Ali et al., 2022), (Putra & Syafira, 2023). NBC is effective in managing data uncertainty and variability, making it a suitable choice for the needs of this industry (Alizadeh et al., 2021).

This study uses the Gray Level Co-occurrence Matrix (GLCM) for texture feature extraction from chicken meat images, which provides quantitative data for further analysis (Maneno et al., 2023), (Ullu et al., 2022). Through GLCM, texture features such as contrast, homogeneity, entropy, and energy are identified, enabling the analysis of microstructures that cannot be captured through visual inspection (Mengiste et al., 2024). The integration of image data from GLCM with NBC in digital image processing offers an objective method to distinguish between fresh and tainted chicken meat.

In addition, this research utilizes RapidMiner, an advanced platform for data analysis and machine learning, which supports the implementation and testing of NBC models in an intuitive and flexible environment (Taranto-Vera et al., 2021). The use of RapidMiner enhances the ability to design, test, and refine classification models with high efficiency (Alghamdi & Al-Baity, 2022), (Anugrah Pratama et al., 2023). The use of RapidMiner in this study also ensures that the classification process is not only supported by powerful statistical analysis but also by efficient and reproducible data processing.

By integrating GLCM and RapidMiner in the application of NBC, this study not only contributes to the development of innovative methodology in chicken meat classification but also strengthens the application of machine learning techniques in the food industry. Specifically, this contribution signifies an advancement in faster and cost-effective food safety detection methods, reducing public health-related risks and enhancing consumer trust in food products. In the research context, these findings enrich the literature with practical applications of machine learning algorithms in addressing real challenges in the food industry, offering new insights and potential for future research. This research promises broad benefits, encapsulating both scientific and practical contributions to ensuring food quality and safety.

Method

Research Procedure

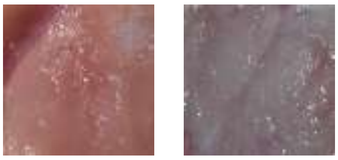
This research will test the effectiveness of classifying fresh chicken meat and tainted chicken meat using the Naive Bayes Classifier (NBC) with several steps, starting from collecting samples and image data, pre-processing data, implementing algorithms, and evaluating results.

Figure 1. Research Procedure

Figure 1. Shows a procedure or research flow that will be carried out. Using the right procedure will provide good data integrity for analysis and model testing.

Data Collection

The data collected is in the form of images or pictures of fresh chicken meat and tainted chicken meat. Data collection is carried out with steps, namely taking samples of fresh and tainted chicken meat taken from chicken slaughterhouses, documenting images using a smartphone camera with as many as 40 images for each meat or a total of 80 images, cropping images to remove irrelevant parts of the image or focus on the meat area and grouping chicken meat image data according to its type into different directories.



(a) (b)

Figure 2. (a) Fresh Chicken Meat, (b) Tainted Chicken Meat

Figure 2 shows two samples of chicken meat in different conditions. In part (a), the chicken meat looks fresher, while in part (b), the chicken meat looks duller and paler, which could be a sign that the meat has decreased in freshness or has been stored for a long time.

Data Pre-processing

Data pre-processing includes image data normalization steps by equalizing image dimensions at 1386 x 1386 pixels, then converting color images to grayscale with image processing on each image using the Gray Level Co-occurrence Matrix (GLCM) method to retrieve homogeneity, contrast, energy, entropy, correlation, and dissimilarity features. Image processing is done using Python, which then exports the results to an Excel file.

```
import numpy as np
import cv2
from skimage.feature import graycomatrix, graycoprops
import os
import pandas as pd # Import pandas

def entropy(glcm):
with np.errstate(divide='ignore', invalid='ignore'):
entropy = -np.sum(np.multiply(glcm, np.log2(glcm + np.finfo(float).eps)))
return entropy
def extract_features(image_path):
img = cv2.imread(image_path)
gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
glcm = graycomatrix(gray, distances=[1], angles=[0], levels=256,
symmetric=True, normed=True)
features = {
'homogeneity': graycoprops(glcm, 'homogeneity')[0, 0],
'contrast': graycoprops(glcm, 'contrast')[0, 0],
'energy': graycoprops(glcm, 'energy')[0, 0],
'entropy': entropy(glcm),
'correlation': graycoprops(glcm, 'correlation')[0, 0],
'dissimilarity': graycoprops(glcm, 'dissimilarity')[0, 0], # Calculate
dissimilarity }
return features
folder_path = 'E:/CHICKEN IMAGE PROCESSING / Tiren Chicken'
images = os.listdir(folder_path)
results = []
for image_name in images:
image_path = os.path.join(folder_path, image_name)
features = extract_features(image_path)
features['image'] = image_name # Add image name to feature
results.append(features)
df = pd.DataFrame(results)
excel_path = 'E:/CHICKEN IMAGE PROCESSING/GLCM_TIREN.xlsx'
df.to_excel(excel_path, index=False)
print(f'The feature extraction results have been saved to {excel_path}')
```

Figure 3. GLCM Source Code

In Figure 3. The image processing process starts with image reading, conversion to grayscale, calculation of GLCM for image feature extraction with the parameters of distance between pixels at 1 and angle of pixel pairs at 0°, and then saving the results in a structured format for further analysis.

After the dataset is obtained, feature selection will be carried out on data that is considered irrelevant or does not contribute significantly to the classification stage to improve the performance of the classification model by focusing on the features or data that are most informative or relevant in predicting the target class (Ulinnuha et al., 2023), (Sen et al., 2021). Then, the data will be divided into two, namely the training dataset and the test dataset, with a composition of 3 1 to be implemented in the NBC model.

Algorithm Implementation

The Naive Bayes Classifier (NBC) algorithm, which is the method for classification in this study, is based on the Bayes theorem with the assumption of independence between predictors (Gunawan et al., 2021). The Bayes theorem equation is generally as follows (Rachman & Moritami, 2020).

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{1}$$

In Equation 1 P (A|B) is the probability that chicken meat is fresh (hypothesis A) conditioned by the characteristics observed in the sample (B), P(B|A) is the probability of the characteristic (B) assuming the meat is fresh (A), P (A) is the a priori probability of chicken meat to be fresh, and P (B) is the a priori probability of observing a particular characteristic, regardless of its freshness status. In this research, the implementation of the NBC algorithm is carried out using the RapidMiner application for model testing and performance.

Evaluation

The evaluation aims to help calculate how well the model can distinguish between fresh chicken meat and tainted chicken meat. Evaluation is carried out using the confusion matrix method by calculating the accuracy, precision, recall, and F1-score value of the classification results with the following equation (Admojo & Sulistya, 2022).

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Amount\ of\ Data} \tag{2}$$

Equation (2). Shows the accuracy equation to measure how often the classification model gives the correct prediction out of all predictions made (Fadli & Saputra, 2023).

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{3}$$

Equation (3). Showing the precision equation to measure the proportion of positive predictions that are actually correct out of all positive predictions made (Fajrin Mustafa et al., 2024).

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{4}$$

Equation (4). Showing the Recall equation to measure the proportion of positive instances successfully identified by the model out of all positive instances that actually exist (Antika et al., 2023).

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

Equation (5) shows the F1 score equation to calculate the harmonic mean of precision and recall. It provides a balance between precision and recall (Giglioni et al., 2021). F1 score becomes low if either precision or recall is low, so it is useful for measuring the overall quality of the classification model (Chicco & Jurman, 2020).

Results and Discussions

From the results of data collection in the form of images of fresh chicken meat and tainted chicken meat, data pre-processing is then carried out in the form of image feature extraction for each type of meat using the Gray Level Co-occurrence Matrix (GLCM) method to retrieve homogeneity, contrast, energy, entropy, correlation and dissimilarity features, using the python language as a dataset to be analyzed in the Naive Bayes Classifier (NBC) model later.

Table 1. Image Dataset of Fresh Chicken Meat and Tainted Chicken Meat

Image	Homogeneity	Contrast	Energy	Entropy	Correlation	Dissimilarity	Type of Meat
Fresh 1.jpg	0,730246889	3,591178416	0,069009645	8,673888943	0,997578725	0,761895385	Fresh
Fresh 2.jpg	0,730174732	4,09836946	0,102805592	7,672138804	0,991595285	0,797268716	Fresh
Fresh 3.jpg	0,734761286	2,7624804	0,058076873	9,029096715	0,998934552	0,716251218	Fresh
Fresh 4.jpg	0,672190353	6,424087705	0,059094729	9,20756393	0,996114451	1,071724465	Fresh
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Tainted 37.jpg	0,549498957	20,61390178	0,052279819	9,896960513	0,987356437	2,111320529	Tainted
Tainted 38.jpg	0,728847421	3,722984356	0,111178754	7,294376686	0,988086838	0,761828705	Tainted
Tainted 39.jpg	0,703251705	3,776746318	0,119810072	7,372981784	0,990308553	0,852685702	Tainted
Tainted 40.jpg	0,437875753	9,095350097	0,046704198	9,596383281	0,990386361	1,838807362	Tainted

Table 1 shows a dataset of 80 numerical image data from images of each type of meat after extracting GLCM features using variable coding in Figure 3.

After pre-processing the data from the obtained dataset, feature selection was performed by removing the "image" column in order to help improve the accuracy and efficiency of the model. After that, the dataset is divided into a training dataset and a test dataset with a composition of 3:1 to be applied to the NBC model.

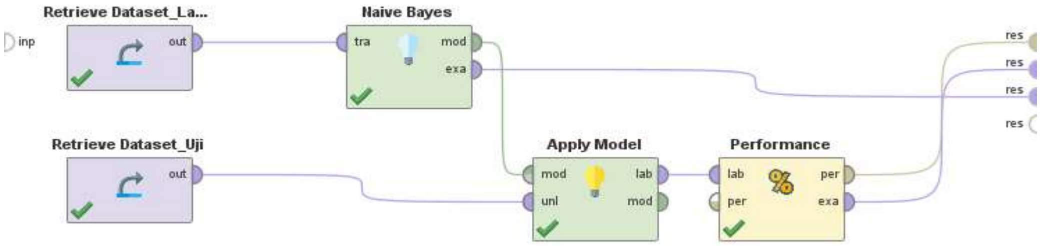


Figure 3. Model Implementation Process in RapidMiner

In Figure 4. Describes the process starting with inputting two datasets, one for training ("Retrieve Dataset_Training") and one for testing ("Retrieve Dataset_Test"). Once the training dataset is retrieved, the "Naive Bayes" algorithm is used to train the model with that dataset. The trained model is then applied to the test dataset using the "Apply Model" process. The results of the model application are evaluated using the "Performance" process to measure how well the model works with the test dataset to provide an assessment of accuracy, precision, recall, or other metrics that show the performance of the trained NBC model.

After all operators are properly connected, then the process is run to see the performance results of the NBC model testing.

accuracy: 80.00%

	true Segar	true Tiren	class precision
pred. Segar	9	3	75.00%
pred. Tiren	1	7	87.50%
class recall	90.00%	70.00%	

Figure 4. NBC Model Performance Results on RapidMiner

In Figure 5. Shows the confusion matrix that illustrates the performance of the Naive Bayes model in RapidMiner with 80% accuracy. This model has 75% precision for the "Fresh" class and 87.5% for the "Tainted" class, as well as 90% recall for the "Fresh" class and 70% for the "Tainted" class. Meanwhile, F1-Score can be calculated using equation (5). As a result, the F1-Score for the "Fresh" class is 81.8%, and for the "Tainted" class is 77.8%.

Table 2. Confusion Matrix Results

	Fresh	Tainted	Average
Precision	75%	87,5%	81,25%
Recall	90%	70%	80%
F1-Score	81,8%	77,8%	79,8%

Table 2. shows the average value for the confusion matrix results of each class in the form of precision, recall, and F1-Score.

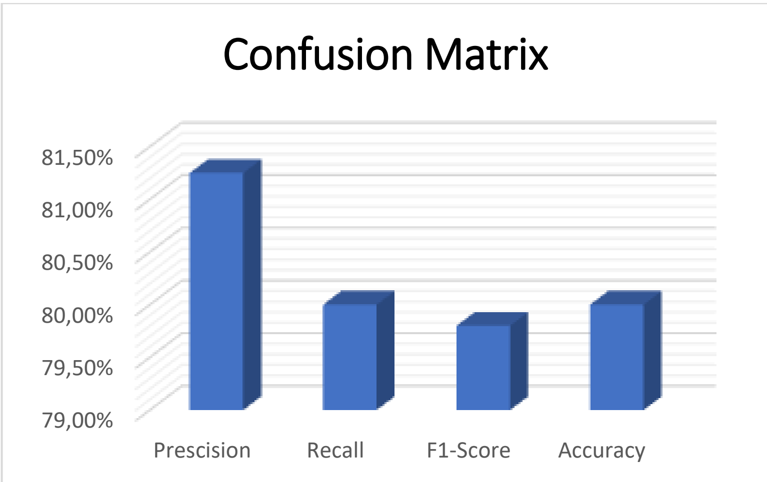


Figure 5. Confusion Matrix Diagram

Figure 6. shows the confusion matrix diagram, where the Naive Bayes model achieved 80% accuracy with an average precision of 81.25% and recall of 80%, and F1-score of 80.62%. This performance indicates the effectiveness of the model in classification, with a balance between prediction accuracy and the ability to detect positive cases.

In the evaluation of our Naive Bayes Classifier (NBC) model, we achieved an overall accuracy of 80%, with a precision of 81.25%, a recall of 80%, and an F1-score of 80.62%. These results indicate a robust performance in distinguishing between fresh and tainted chicken meat, leveraging the Gray Level Co-occurrence Matrix (GLCM) for texture feature extraction and the RapidMiner platform for model implementation. Compared to traditional methods, which often involve time-consuming physical and chemical analyses, our approach offers a more efficient and objective alternative. Previous studies, such as those by (Agustina & Ardiansyah, 2020), have also explored the classification of chicken meat using different methodologies. However, our integration of GLCM and NBC through

RapidMiner enhances the classification process by providing a faster and more cost-effective solution. This not only aligns with the advancements in machine learning applications in food safety, as noted by (Mondejar et al., 2021), but also pushes the boundary by offering a practical tool for industry stakeholders. The potential for future research is vast, with opportunities to explore advanced feature extraction techniques and other classification algorithms to further improve the model's precision, recall, and F1-score

Conclusions

The research results of the classification of fresh chicken meat and tainted chicken meat using the Naive Bayes Classifier (NBC) algorithm show satisfactory performance with an overall accuracy of 80%. The model showed an average precision of 81.25%, recall of 80%, and F1-score of 80.62%, signifying its ability to classify chicken meat effectively. Although the results are satisfactory, there is potential for improvement through optimization to reduce prediction errors, which will increase precision and recall. Furthermore, this research has significant theoretical and practical implications in the field of food safety and machine learning applications. Theoretically, our findings expand the understanding of the application of the Naive Bayes Classifier (NBC) and Gray Level Co-occurrence Matrix (GLCM) in chicken meat classification, offering new insights into image processing and texture analysis. Practically, the method we developed offers a more efficient and objective approach to distinguishing between fresh and spoiled chicken meat, with the potential to enhance food safety and consumer satisfaction. For future research, it is recommended that more advanced feature extraction techniques and other classification algorithms be explored to improve precision, recall, and F1-score, as well as test the model in more diverse datasets to validate the effectiveness of the classification method under various conditions.

References

Admojo, F. T., & Sulistya, Y. I. (2022). Analisis performa algoritma Stochastic Gradient Descent (SGD) dalam mengklasifikasi tahu berformalin. *Indonesian Journal of Data and Science (IJODAS)*, *3*(1). https://doi.org/https://doi.org/10.56705/ijodas.v3i1.42

Agustina, F., & Ardiansyah, Z. A. (2020). Identifikasi Citra Daging Ayam Kampung dan Broiler Menggunakan Metode GLCM dan Klasifikasi-NN. *25 Jurnal Infokam*, *XVI*(1). https://doi.org/https://doi.org/10.53845/infokam.v16i1.196

Alghamdi, N. A., & Al-Baity, H. H. (2022). Augmented Analytics Driven by AI: A Digital Transformation beyond Business Intelligence. *Sensors (Basel, Switzerland)*, *22*(20). https://doi.org/10.3390/s22208071

Ali, A., Samara, W., Alhaddad, D., Ware, A., & Saraereh, O. A. (2022). Human Activity and Motion Pattern Recognition within Indoor Environment Using Convolutional Neural Networks Clustering and Naive Bayes Classification Algorithms. *Sensors*, *22*(3). https://doi.org/10.3390/s22031016

Alizadeh, S. H., Hediehloo, A., & Harzevili, N. S. (2021). Multi independent latent component extension of naive Bayes classifier. *Knowledge-Based Systems*, *213*, 106646. https://doi.org/10.1016/J.KNOSYS.2020.106646

Antika, R., Rifa, A., Dikananda, F., Indriya Efendi, D., & Narasati, R. (2023). PENERAPAN ALGORITMA DECISION TREE BERBASIS POHON KEPUTUSAN DALAM KLASIFIKASI PENYAKIT JANTUNG. *JITA*, *7*(6). https://doi.org/https://doi.org/10.36040/jati.v7i6.8264

Anugrah Pratama, D., Rizal Mutaqin, I., & Rafael Manuela, K. (2023). Analisis Terjadinya Kanker Paru-Paru Pada Pasien Menggunakan Decision Tree: Penerapan Algoritma C4.5 Dan RapidMiner Untuk Menentukan Risiko Kanker Pada Gejala Pasien. *JTMEI*, *2*, 156–170. https://doi.org/10.55606/jtmei.v2i4.3004

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-019-6413-7

Fadli, M., & Saputra, R. A. (2023). Classification And Evaluation Of Performance Models Random Forest For Stroke Prediction. *Jurnal Teknik*, *12*(02). https://doi.org/http://dx.doi.org/10.31000/jt.v12i2.9099

Fajrin Mustafa, W., Hidayat, S., & Hatta Fudholi, D. (2024). Prediksi Retensi Pengguna Baru Shopee Menggunakan Machine Learning. *JURNAL MEDIA INFORMATIKA BUDIDARMA* , *8*(1). https://doi.org/10.30865/mib.v8i1.7074

Giglioni, V., García-Macías, E., Venanzi, I., Ierimonti, L., & Ubertini, F. (2021). The use of receiver operating characteristic curves and precision-versus-recall curves as performance metrics in unsupervised structural damage classification under changing environment. *Engineering Structures*, *246*, 113029. https://doi.org/10.1016/J.ENGSTRUCT.2021.113029

Gunawan, E., Wahyudi, J., & Sari, Y. (2021). PENDEKATAN BERBASIS KECERDASAN BUATAN DENGAN METODE NAÏVE BAYES UNTUK WEBSITE BAZNAS. *JTIULM*, *6*(1). https://doi.org/https://doi.org/10.20527/jtiulm.v6i1.68

Maneno, R., Baso, B., Manek, P. G., & Fallo, K. (2023). Deteksi Tingkat Kematangan Buah Pinang Menggunakan Metode Support Vector Machine Berdasarkan Warna Dan Tekstur. *Journal of Information and Technology*, *3*(2), 60–66. https://doi.org/10.32938/jitu.v3i2.5323

Mengiste, E., Mannem, K. R., Prieto, S. A., & Garcia de Soto, B. (2024). Transfer-Learning and Texture Features for Recognition of the Conditions of Construction Materials with Small Data Sets. *Journal of Computing in Civil Engineering*, *38*(1). https://doi.org/10.1061/jccee5.cpeng-5478

Mondejar, M. E., Avtar, R., Diaz, H. L. B., Dubey, R. K., Esteban, J., Gómez-Morales, A., Hallam, B., Mbungu, N. T., Okolo, C. C., Prasad, K. A., She, Q., & Garcia-Segura, S. (2021). Digitalization to achieve sustainable development goals: Steps towards a Smart Green Planet. *Science of The Total Environment*, *794*, 148539. https://doi.org/10.1016/J.SCITOTENV.2021.148539

Pirmansyah, F., & Wahyudi, T. (2023). IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA C4.5 UNTUK PREDIKSI EVALUASI ANGGOTA SATUAN PENGAMANAN STUDI KASUS PT. YIMM PULOGADUNG. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *4*(3), 1566–1580. https://doi.org/10.35870/jimik.v4i3.370

Putra, A. P., & Syafira, A. F. (2023). Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan Convolutional Neural Networks (Cnn). *Jurnal Ilmiah Wahana Pendidikan, Oktober*, *9*(20), 36–41. https://doi.org/10.5281/zenodo

Putri, O. O., & Yodfiatfinda, Y. (2021). ANALYSIS OF MARKETING MARGIN OF THE BROILER CHICKEN TRADERS IN PD PASAR JAYA, PASAR MINGGU. *Jurnal Bioindustri*, *03*(02). https://doi.org/https://doi.org/10.31326/jbio.v3i2.910.g471

Rachman, R., & Moritami, S. (2020). Sistem Pakar Deteksi Penyakit Refraksi Mata Dengan Metode Teorema Bayes Berbasis Web. *JURNAL INFORMATIKA*, *7*(1). http://ejournal.bsi.ac.id/ejurnal/index.php/ji

Sen, S., Saha, S., Chatterjee, S., Mirjalili, S., & Sarkar, R. (2021). A bi-stage feature selection approach for COVID-19 prediction using chest CT images. *Applied Intelligence*, *51*(12), 8985–9000. https://doi.org/10.1007/s10489-021-02292-8

Taranto-Vera, G., Galindo-Villardón, P., Merchán-Sánchez-Jara, J., Salazar-Pozo, J., Moreno-Salazar, A., & Salazar-Villalva, V. (2021). Algorithms and software for data mining and machine learning: a critical comparative view from a systematic review of the literature. *Journal of Supercomputing*, *77*(10), 11481–11513. https://doi.org/10.1007/s11227-021-03708-5

Ulinnuha, N., Fanani, A., & Sunan Ampel Surabaya, U. (2023). Classification of Drop Out Status for Students Using Naïve Bayes with Information Gain Feature Selection. *Jurnal Teknologi Informasi*, *22*(4), 1014–1025. https://doi.org/https://doi.org/10.33633/tc.v22i4.9004

Ullu, H. H., Baso, B., Risald, R., Manek, P. G., & Chrisinta, D. (2022). Ektraksi Fitur Berbasis Tekstur Pada Citra Tenun Timor Menggunakan Metode Gray Level Co-occurrence Matrix (GLCM). *Journal of Information and Technology*, *2*(2), 70–74. https://doi.org/10.32938/jitu.v2i2.3245