Performance of the Decision Tree Algorithm in the Classification of Edible and Poisonous Mushrooms with Information Gain Optimization

Arif Rifan Rudiyanto[1], Pujiono[2], M. Arief Soeleman[3], Mustagfirin[4*]

[1,2,3]Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
[4]Department of Informatics Engineering, Faculty of Engineering, Universitas Wahid Hasyim, Indonesia

Abstract.
Purpose: This study proposes a new mushroom classification model using a decision tree algorithm to classify edible and poisonous mushrooms by applying machine learning whose algorithm has better performance in terms of accuracy.
Methods: The information gain technique was applied at the data feature selection stage to increase the accuracy of the suggested decision tree model. This study used the same mushroom dataset as that employed in previous studies.
Result: The proposed decision tree model in this study can classify edible and poisonous mushrooms with a good accuracy of 99.61%, outperforming a previous study whose final accuracy was 97.05%.
Novelty: The novelty of this s is the use of information gain as a filter technique at the feature selection stage. This study aims to optimize the previous mushroom classification models with improved accuracy.

Keywords: Machine learning, Decision tree, Information gain, Mushrooms.
Received October 2023 / **Revised** November 2023 / **Accepted** November 2023

INTRODUCTION

Mushroom is one of the garden products with attractive future prospects. It refers to a group of plants with no chlorophyll that depends on other living beings. In general, mushrooms can be defined as eukaryotic organisms that have a nucleus and organelles. They are composed of long unicellular filaments called hyphae. The collection of hyphae is called mycelium. Mushrooms are easily recognized by the color of their mycelium, a sizable mass of filaments formed by a network of hyphae as the mushroom grows [1], [2], [3], [4]. In the microbial world, mushrooms belong to the kingdom Mycota (fungi). Mycota is derived from a Greek word *mykes*, while the word fungus (plural fungi) is borrowed directly from Latin. There are several terms known to refer to fungi: (a) mushrooms, namely fungi that can produce large fruiting bodies, including edible ones; (b) molds, threadlike-shaped fungi; and (c) yeasts, which are unicellular [5], [6].

Some mushrooms are consumable as a food ingredient and can be utilized in various applications in the health sector. In addition, they also act as decomposers in the food chain of organisms. There are more than 1,500,000 species of fungi in the world. However, only about 7,000 of them have been identified. Some of the identified fungi have multiple uses in the fields of nutrition, health, economics, and agriculture, whereas some others can cause infections in humans [7].

Mushroom is a cheap and readily available functional food ingredient and substitute, which is not only protein-rich but also a source of vitamins, carbohydrates, fiber, minerals and essential amino acids, unsaturated fatty acids, and saturated fats. Furthermore, mushrooms also contain active ingredients that are beneficial for health, such as polysaccharides (glucans), triterpenoids, nucleotides, mannitol, and alkaloids [8], [9]. The most commonly consumed mushrooms include champignon, button mushroom (soft mushroom), oyster mushroom, abalone mushroom, and shiitake mushroom [10], [11]. As a source of nutrients, the protein content in mushrooms is relatively high, namely 15-20% of the dry weight, while the digestibility is 34-89%.

The terms species and genus are commonly used to classify living organisms. Mushrooms can be recognized by their size, color, cap, and stem shape. However, mushroom identification remains challenging because there are numerous species of mushrooms. Of the many species of mushrooms, only 50-100 species are edible [12]. In other words, most mushrooms are inedible [13]. Meanwhile, various poisonous mushrooms are identical to edible ones, making it difficult to differentiate them by their physical characteristics. Therefore, it is crucial to classify mushrooms accurately to help identify which mushrooms are edible and which are poisonous. Several studies have explored pharmaceutical data extraction, including mushroom taxonomy. In Indonesia, there are less than 200,000 different species of mushrooms. Of the millions of mushroom species found throughout the world, only 2,000 are edible or can be used as food ingredients [14]. Consuming mushrooms directly without knowing species can be very dangerous as it may lead to various symptoms of poisoning and even death [15].

To avoid these risks, there are several ways to find out which mushrooms are edible and which are poisonous. The utilization of data mining is an important aspect that is used as a benchmark in identifying mushrooms. Data mining is information processing with basic artificial intelligence techniques from data analysis related to large databases [16]. Meanwhile, classification analysis is a multivariate method that deals with separating observations and assigning new observations to data categories with the aim of obtaining a discriminant function that can separate as many observations as possible and gaining rules to assign classes to new observations [17].

Methods in classification analysis include Decision Tree, Support Vector Machine (SVM), Discriminant Analysis, Artificial Neural Network (ANN), Random Forest, Naïve Bayes Classifier (NBC) [18], and many others. This study applied the Decision Tree method, which is one of the most effective data mining methods that has been widely used in several scientific disciplines [19], [20], [21]. Recently, this method has become increasingly popular in medical research [22]. Furthermore, the Classification and Regression Tree (CART) algorithm contained in decision tree has long been employed for the purpose of forming a decision tree to find variable classification patterns [23], [24]. Learning Samples are used to form a decision tree model by using a certain splitting algorithm that separate them into two subgroups (nodes) recursively or nested until they reach the terminal node (leaf). In addition to decision tree, this study also utilized Information Gain to analyze and predict which mushrooms are edible and which are poisonous based on certain characteristics.

Categorization or classification is a method for imposing group membership based on existing knowledge[25][26]. In this study, the classification of poisonous mushrooms according to their types was carried out using the RapidMiner application as a statistical technique to test and model the relationship between one response variable and one or more predictor variable(s) [27].

Classification of mushrooms using machine-learning techniques has been done in several previous studies. A study by Essa and Dhanalakshmi (2023) combined five machine-learning models and achieved accuracy values of 0.83 for Logistic Regression, 0.98 for K-Nearest Neighbor (K-NN), 0.095 for Support Vector Classification (SVC), 0.97 for Decision Tree, 0.99 for Random Forest, and 0.93 for Gradient Boosting Classifier [14]. In their study, feature selection becomes an important step in mushroom classification which helps improve the accuracy of their model.

Tarawneh *et al*. (2022) also combined several ML methods and obtained accuracies of 90% for SVM classification, 72% for Naive Bayes, 92% for Decision Tree, 84% for Logistic Regression, 93% for ANN[13]. The proposed hybrid approach, which combines KNN, ANN, and SVM and comprises 3 phases of raw data processing, description, and combination, is considered the best model with slightly better performance. In another prior study by Wati, Fauzan, and Harliana (2022), the classification using Decision Tree (DT) C5.0 algorithm in scenario 3 had an accuracy of 97.11%, while the utilization of DT C4.5 algorithm in scenario 3 produced an accuracy of 97.05% [28]. This means that DT C5.0 algorithm for identifying poisonous mushrooms is superior in terms of accuracy.

Another previous study conducted by Ortega *et al.* (2020) found the accuracies of the logistic regression and decision tree algorithms of 87.8% and 88.2%, respectively [29]. The procedures and steps in their study are very effective, with the most important steps being preprocessing and modeling; preprocessing involves various techniques to improve the quality of the dataset, while modeling examines the performance of the algorithms and develops predictive and cluster models that are used to profile and predict future cases.

Lastly, Lavanya and Preethi (2017) employed the CART algorithm based on decision tree and reached an accuracy value of 90% [30]. In their study, the Hoeffding Tree algorithm, which is also part of decision tree, can be used to handle unstructured data with the aims of reducing time when processing data.

This present study implemented machine learning through data processing for mushroom classification and accuracy calculation using the decision tree algorithm and information gain optimization at the feature selection stage in the RapidMiner programming language. This system classified mushrooms into two classes, namely edible mushrooms and poisonous mushrooms. The dataset used is from the UCI Machine Learning (ML) repository [31], which is categorical data that is converted into numerical data in Microsoft Excel 2013. This dataset was then divided into training data and testing data to find the best performance. The decision tree algorithm obtained the best performance in each evaluation.

METHODS
The first step in the research planning process is to create a block diagram of the entire system. Such flowchart is essential to provide an overview of the stages in the research methodology. Figure 1 displays a block diagram of the overall research stages in developing the system to work efficiently.
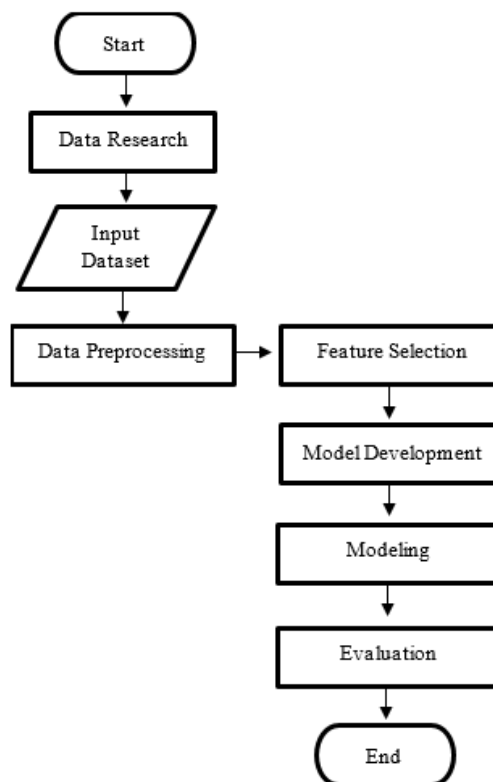

Figure 1. Flowchart of the research methodology

Data Research
In this stage, models used in previous studies were examined to obtain better model estimates. Data exploration employed Kaggle and RapidMiner. More detailed steps and explanations are provided in the next sections, which are scrutinized by presenting accuracy, precision, recall, and F1-score.

Dataset
The public dataset[1] used in this study was retrieved from the Kaggle website on June 2023. This dataset has been previously tested in numerous studies [13], [14], [28], [29], [30], and contains approximately 8,124 records covering both edible and poisonous mushrooms. There are 22 attributes in this dataset, namely class, cap-shape, cap-surface, cap-color, bruises, odor, gill-attachment, gill-spacing, gill-size, gill-color,

---

[1] https://www.kaggle.com/datasets/uciml/mushroom-classification

stalk-shape, stalk-root, stalk-surface-above-ring, stalk-surface-below-ring, stalk-color-above-ring, stalk-color-below-ring, veil-type, veil-color, ring-number, ring-type, spore-print-color, population, and habitat. Each mushroom species is identified as edible or poisonous. Figure 2 shows data on the numbers of mushrooms in the dataset, of which 4,208 are edible and 3,916 are poisonous. Meanwhile, the attribute information of the mushroom dataset is presented in Table 1.

Table 1. Mushroom dataset attribute information

| No	Attribute	Value
1	cap-shape	bell=b, comical=c, convex=x, flat=f, knobbed=k, sumken=s
2	cap-surface	fibrous=f, grooves=g, scaly=y, smooth=s
3	cap-color	brown=b, buff=b, cinnamon=c, gray=g, green=g, pink=p, purple=u, red=e, white=w, yellow=y
4	bruises	bruises=t, no=f
5	odor	almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s
6	gill_attachment	attached=a, free=f
7	gill-spacing	close=c, crowded=w
8	gill-size	broad=b, narrow=n
9	gill-color	black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y
10	stalk-shape	enlarging=e, tapering=t
11	stalk-root	bulbous=b, club=c, equal=e, rooted=r
12	stalk-surface-above-ring	fibrous=f, scaly=y, silky=k, smooth=s
13	stalk-surface-below-ring	fibrous=f, scaly=y, silky=k, smooth=s
14	stalk-color-above-ring	brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
15	stalk-color-below-ring	brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
16	veil-type	partial=p
17	veil-color	brown=n, orange=o, white=w, yellow=y
18	ring-number	none=n, one=o, two=t
19	ring-type	evanescent=e, flaring=f, large=l, none=n, pandant=p
20	spore-print-color	black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y
21	population	abundant=a, chusted=c, numerous=n, scattered=s, several=v, solitary=y
22	habitat	grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d
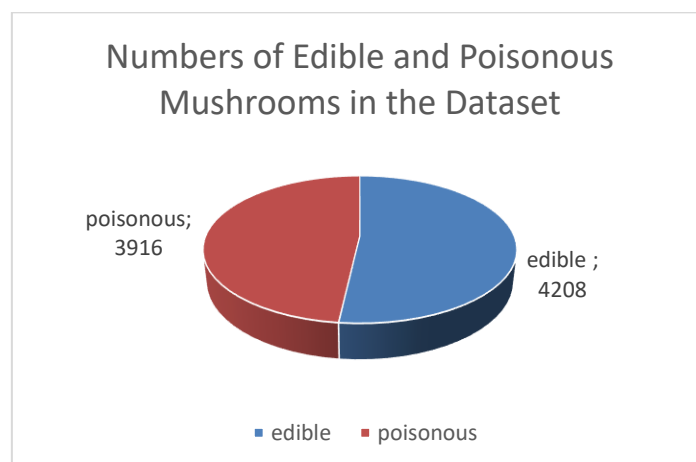
Figure 2. Edible and poisonous mushroom dataset

Data Preprocessing

Data preprocessing techniques were employed to improve the quality of the algorithm applied in this study. At this stage, data manipulation was carried out on the mushroom dataset before being used in the decision tree model. This involved data cleaning and transformation techniques [32]. Data cleaning processed unwanted noises or inconsistent data in the dataset, e.g., removing the veil-type attribute which only has one value, making it easier to use the dataset as input data in the decision tree model. The next steps of data preprocessing stage in this study include normalization, selection, and design of features from the dataset. After that, data integration, which is combining data from several different sources, was performed. Meanwhile, data transformation was used to increase the efficiency or accuracy of the algorithm.

The application of machine learning in mushroom classification includes feature simulation, covering the selection of relationships and desired features in the dataset suitable for distinguishing edible mushrooms from poisonous mushrooms. There are 22 non-equivalent attributes in the mushroom dataset which describe different aspects of the mushroom context. These 22 attributes are used to differentiate edible and poisonous mushrooms based on their characteristics. The main preparation for machine learning classification of the dataset in this study was data preprocessing and cleaning to condition data that has empty values by changing categorical values to numeric values, with the assumption that the values of the attributes have the same distance. Then, the next step was to select recursive features to recognize important features in the classification process.

Feature Selection

The feature selection stage used Information Gain technique to reduce attributes that do not have much influence on the dataset by separating each record by class or label. Information gain is a feature selection method that calculates the reduction in entropy from before to after split. This is used to determine the attributes to be used or discarded [33].

Attribute determination was carried out by employing Information Gain by calculating the values of all attributes. Information Gain value was measured using Equation 1.

$$info(D) = -\sum_{i=1}^{m} pi \, log_2(pi) \tag{1}$$

where *D* is the set of cases, *m* is the number of partitions of *D*, and *pi* is the probability of a tuple *D*.

Meanwhile, the process of determining the entropy values after split uses Equation 2.

$$info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} X \, I\left(D_j\right) \tag{2}$$

Model Development

In this study, a new model was developed for the classification of edible and poisonous mushrooms with greater accuracy than other models proposed in previous studies. This stage can be achieved after carrying out the research on the processed dataset. The model development process in this study used Information Gain optimization.

Modeling

After carrying out data transformation at the data preprocessing stage, the next step was to build the decision tree model. Decision tree is a popular classification algorithm used in various applications and encompasses a variety of models, such as ID3 and C4.5 [34], [35], [36].

The general structure of a decision tree consists of leaves, branches, and roots, with the bottom part being the leaves and the top part being the root. Each feature in the dataset represents a node. The structures between two nodes are called the branches. In building a decision tree, determining which attributes to branch into is the most vital step [37]. The Gini index, information validation, and the rule of attraction are usually used as decision-making conditions [34], [38]. Meanwhile, this study employed information gain technique in the decision tree model. With this technique, the influence of related attributes on the results was calculated with a value based on the entropy of each attribute. If it is assumed that there are n classes

in the decision tree structure and these classes are estimated to be repeated T times, the probability of data belonging to these classes can be calculated using Equation 3.

$$P_i = \frac{c_i}{|D|} \tag{3}$$

where ci represents the value of a certain class.

In this study, the attributes were processed through several stages of information gain optimization, resulting in 21 attributes to be used optimally in the mushroom classification. Table 2 shows the information gain values of the 22 attributes in the mushroom classification.

Table 2. Information gain of dataset attributes

No	Attribute	Information Gain
1	odor	0,906
2	spore-print-color	0,481
3	gill-color	0,417
4	ring-type	0,318
5	stalk-surface-above-ring	0,285
6	stalk-surface-below-ring	0,272
7	stalk-color-above-ring	0,254
8	stalk-color-below-ring	0,241
9	gill-size	0,230
10	population	0,202
11	bruises	0,192
12	habitat	0,157
13	gill-spacing	0,101
14	stalk-root	0,063
15	cap-shape	0,049
16	ring-number	0,038
17	cap-color	0,036
18	cap-surface	0,029
19	veil-color	0,024
20	gill-attachment	0,014
21	stalk-shape	0,080
22	veil-type	0,000

Evaluation

The main objective of system evaluation is to test the system's ability to classify mushrooms. A key step in evaluating the accuracy of the mushroom classification model is the measurement of the extent to which the decision tree model can classify edible and poisonous mushrooms. To do so, this study utilized the confusion matrix validated with a crossover model. The validation results can be seen in Table 3.

Table 3. Confusion matrix

	False Positive (FP)	False Negative (FN)
True Positive (TP)	TP	FN
True Negative (TN)	FP	TN

Simplicity in the classification of mushrooms in this study was evaluated with the confusion matrix. This matrix provides a visual summary of the model's predicted results against the actual predictions made by the model. In the context of mushroom classification, confusion matrix can be used to measure how well the model can differentiate between edible and poisonous mushrooms. The confusion matrix can show how

many poisonous mushrooms were correctly identified (TP), how many edible mushrooms were correctly identified (TN), how many poisonous mushrooms were correctly identified (FP), and how many poisonous mushrooms were correctly identified (FN).

Confusion matrix is an evaluation of performance by means of estimation based on the entries entered. Predictions can be made by measuring the accuracy of the correct proportions. Accuracy, recall, precision, and F1 score are the preferred performance evaluation metrics in numerous applications in the literature. In this present study, these metrics were used to evaluate the performance and success of the model developed. These metrics are measured using Equations 4-7.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Precition = \frac{TP}{FP+TP} \tag{5}$$

$$Recall = \frac{TP}{FN+TP} \tag{6}$$

$$F1\ score = 2x\ \frac{Precision\ x\ Recall}{Precision+Recall} x100\% \tag{7}$$

In the confusion matrix by entry, the accuracy value includes the results of describing the model's good expressions by distinguishing edible mushrooms from poisonous ones. However, it should be noted that precision can provide inaccurate information if the data is not balanced. Meanwhile, the accuracy of the model developed in this study is expressed as a measure of the effectiveness of the decision tree model in identifying poisonous mushrooms from the dataset.

RESULTS AND DISCUSSIONS
After carrying out data preprocessing, feature selection, model development, and modeling, the next stage is the application of K-Fold cross-validation as a calculation on training data and testing data in mushroom classification. K-Fold cross-validation is a technique that minimizes the shortcomings of the hold-out method. K-Fold splits the dataset in a new way that overcomes the "test only once bottleneck".

This ensures that every observation from the original dataset has a chance to appear in both the training and testing sets, resulting in a model that is less likely to be biased and particularly suitable for applications with a limited amount of input data. In this study, there is a scenario where the dataset is split into 60% training data and 40% testing data.

The procedure of K-Fold technique includes:
- Selecting the desired number of k-folds; k = 8 was chosen for this study as the end point in determining the value.
- Splitting the dataset into k equal parts
- Starting from k - 1 folds as the training set, with the remaining folds belonging to the testing set
- Training the model on the training set. On each iteration of cross-validation, a new model should be trained independently of the model trained on the previous iteration.
- Validating on the testing set
- Saving the validation results

The average values obtained in various scenarios in cross validation of the decision tree algorithm are shown in Table 4 and Figure 3.

Table 4. Results of the average value in k-fold cross-validation

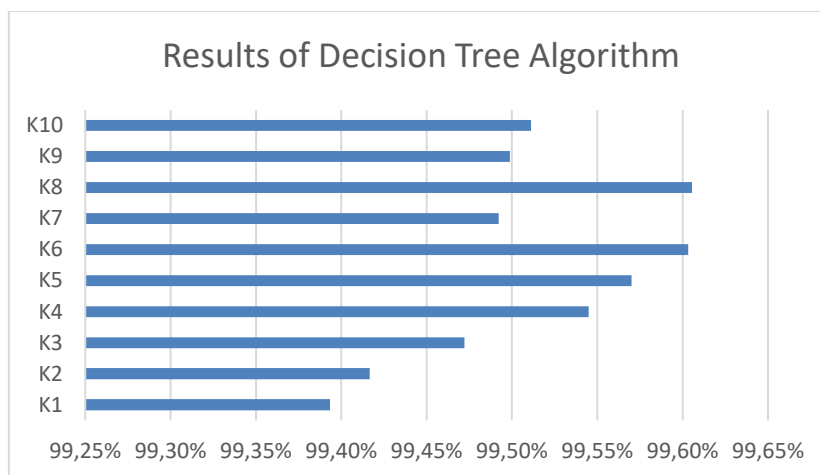| K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
99.39%	99.42%	99.47%	99.55%	99.57%	99.60%	99.49%	**99.61**%	99.50%	99.51%

Figure 3. Results of the decision tree algorithm with k-fold cross-validation

CONCLUSION

From this study, it can be concluded that the machine learning approach can be successfully applied to mushroom classification. In identifying and classifying mushroom species, the decision tree has a good performance, with high accuracy, recall, precision, and F1 score. In addition, this study highlights the importance of feature selection and deep dataset to improve the accuracy of machine learning-based mushroom classification. To further improve the performance of mushroom classification systems, future studies are strongly recommended to use better feature selection and extraction methods as well as larger and more diverse dataset.

REFERENCES

[1]   I. N. P. A. Indah Juwita Sari, "Screening and Identification of Mushrooms Growth Promoting Bacteria on Straw Mushrooms (Volvariella volvacea)," *J. Trop. Biodivers. Biotechnol.*, vol. 6, no. 1, 2021, doi: 10.22146/jtbb.60618.

[2]   H. Rathore, S. Prasad, M. Kapri, A. Tiwari, and S. Sharma, "Medicinal importance of mushroom mycelium: Mechanisms and applications," *J. Funct. Foods*, vol. 56, pp. 182–193, May 2019, doi: 10.1016/j.jff.2019.03.016.

[3]   S. C. Khoo *et al.*, "Development of formaldehyde-free bio-board produced from mushroom mycelium and substrate waste," *J. Hazard. Mater.*, vol. 400, p. 123296, Dec. 2020, doi: 10.1016/j.jhazmat.2020.123296.

[4]   R. W. deVere White, R. M. Hackman, S. E. Soares, L. A. Beckett, and B. Sun, "Effects of a mushroom mycelium extract on the treatment of prostate cancer," *Urology*, vol. 60, no. 4, pp. 640–644, Oct. 2002, doi: 10.1016/S0090-4295(02)01856-3.

[5]   H. A. B. W. Nienke Braat, Margot C. Koster, "Beneficial interactions between bacteria and edible mushrooms," *Fungal Biol. Rev.*, vol. 39, pp. 60–72, 2022, doi: https://doi.org/10.1016/j.fbr.2021.12.001.

[6]   H. L. Jiang, C. A. Mbajiorgu, J. W. Ngambi, and C. P. Tlabela, "Economic Importance of Lentinulas edodes as an Alternative to Antibiotics Growth Promoters to Small Scale Rural Pig Farmers of Limpopo Province, South Africa: A Review," vol. 4, no. 2, pp. 486–493, 2016.

[7]   N. A. R. G. Alistair J.P. Brown, Gordon D. Brown, Mihai G. Netea, "Metabolism impacts upon candida immunogenicity and pathogenicity at multiple levels," *Trends Microbiol.*, vol. 22, no. 11, pp. 614–622, 2014, doi: 10.1016/j.tim.2014.07.001.

[8]   D. Tjokrokusumo, "Diversitas jamur pangan terhadap kandungan beta-glukan dan manfaatnya terhadap kesehatan," vol. 1, no. September, pp. 1520–1523, 2015, doi: 10.13057/psnmbi/m010646.

[9]   J. Preechasuk, O. Chaowalit, F. Pensiri, and P. Visutsak, "Image Analysis of Mushroom Types Classification by Convolution Neural Networks," in *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*, Dec. 2019, pp. 82–88. doi: 10.1145/3375959.3375982.

[10]   N. S. Iwan Saskiawan, Nur Hasanah, "Cultivation of Pleurotus ostreatus using sorghum-supplemented spawn on various substrates," vol. 23, no. 4, pp. 179–182, 2016.

[11] K. M. Phillips, D. M. Ruggio, and D. B. Haytowitz, "Folate composition of 10 types of mushrooms determined by liquid chromatography–mass spectrometry," *Food Chem.*, vol. 129, no. 2, pp. 630–636, Nov. 2011, doi: 10.1016/j.foodchem.2011.04.087.

[12] M. Nagulwar, DR More, and L. Mandhare, "Nutritional properties and value addition of mushroom: A review," *Pharma Innov.*, vol. 9, no. 10, pp. 395–398, 2020, doi: 10.22271/tpi.2020.v9.i10f.5266.

[13] O. Tarawneh, M. Tarawneh, and Y. Sharrab, "Mushroom classification using machine-learning techniques," *researchgate*, vol. 9, no. July, pp. 395–398, 2022.

[14] I. A. Essa and R. Dhanalakshmi, "Machine Learning-based Classification of Edible and Poisonous Mushrooms : A Performance," *Ijraset*, vol. 11, no. May, pp. 1364–1370, 2023.

[15] K. Tank, "A Comparative Study on Mushroom Classification using Supervised Machine Learning Algorithms," *Int. J. Trend Sci. Res. Dev.*, vol. 5, no. 5, pp. 716–723, 2021.

[16] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2016.

[17] E. N. O. C.E. Onwukwe, "Fisher's Linear Discriminant Classifier And Rank Transformation Approach To Discriminant Analysis," *J. Nat. Sci. Res.*, vol. 4, no. 8, pp. 16–30, 2014.

[18] A. Nurdina and A. B. I. Puspita, "Naive Bayes and KNN for Airline Passenger Satisfaction Classification: Comparative Analysis," *J. Inf. Syst. Explor. Res.*, vol. 1, no. 2, pp. 83–92, 2023, doi: 10.52465/joiser.v1i2.167.

[19] P. Bielza, C., Robles, V., dan Larrañaga, "Regularized Logistic Regression Without a Penalty Term: An Application to Cancer Classification With Microarray Data," *Expert Syst. Appl.*, vol. 38, no. 5, p. 5110—5118, 2011.

[20] K. L. Kohsasih and B. H. Hayadi, "Classification SARS-CoV-2 Disease based on CT-Scan Image Using Convolutional Neural Network," *Sci. J. Informatics*, vol. 9, no. 2, pp. 197–204, 2022, doi: 10.15294/sji.v9i2.36583.

[21] N. D. Gökhan Aksu, "Comparison of decision trees used in data mining," *Pegem J. Educ. Instr.*, vol. 9, no. 4, pp. 1183–1208, 2019, doi: https://doi.org/10.14527/pegegog.2019.039.

[22] N. T. & A. J. Barbara Fallon, Jennifer Ma, Kate Allan, Melanie Pillhofer, "Opportunities for prevention and intervention with young children: lessons from the Canadian incidence study of reported child abuse and neglect," *Child Adolesc. Psychiatry Ment. Health*, vol. 7, no. 4, 2013.

[23] W. S. Sumitra Nuanmeesri, "Development of the Edible and Poisonous Mushrooms Classification Model by using the Feature Selection and the Decision Tree Techniques," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 2, 2019.

[24] A. Nazir, A. Akhyar, Y. Yusra, and E. Budianita, "Toddler Nutritional Status Classification Using C4.5 and Particle Swarm Optimization," *Sci. J. Informatics*, vol. 9, no. 1, pp. 32–41, 2022, doi: 10.15294/sji.v9i1.33158.

[25] D. K. R. Sudrajat, I. Irianingsih, "Analysis of Data Mining Classification by Comparison of C4.5 and ID algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, 2017.

[26] Y. F. Sihombing, A. Septiarini, A. H. Kridalaksana, and N. Puspitasari, "Chili Classification Using Shape and Color Features Based on Image Processing," *Sci. J. Informatics*, vol. 9, no. 1, pp. 42–50, 2022, doi: 10.15294/sji.v9i1.33658.

[27] O. D. Rifan Halomoan Tua Sinaga, "Binary Logistic Regression Analysis Using Stepwise Method on Tuberculosis Events," *J. Math. Educ. Appl.*, vol. 2, no. 1, pp. 01–10, 2023.

[28] H. Citra Mirna Wati, Abd. Charis Fauzan, "Performance Comparison of Mushroom Type Classification Based on Multi-Scenario Dataset Using Decision Tree C4.5 and C5.0," *J. Ris. Inform.*, vol. 4, no. 3, 2022.

[29] J. Heland *et al.*, "Analysis of Performance of Classification Algorithms in Mushroom Poisonous Detection using Confusion Matrix Analysis," vol. 9, no. 1, 2020.

[30] G. R. P. B. Lavanya, "Performance Analysis of Decision Tree Algorithms on Mushroom Dataset," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 5, no. 11, 2017, doi: 10.22214/ijraset.2017.11029.

[31] R. Hamonangan, M. B. Saputro, C. Bagus, S. Dinata, and K. Atmaja, "Accuracy of classification poisonous or edible of mushroom using naïve bayes and k-nearest neighbors," *J. Soft Comput. Explor.*, vol. 2, no. 1, 2021, doi: 10.52465/joscex.v2i1.26.

[32] S. A. A. and W. S. Bhaya, "Review of Data Preprocessing Techniques in Data Mining," *J. Eng. Appl. Sci*, vol. 12, no. 16, pp. 4102–4107, 2017.

[33] S. Tangirala, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020.

[34] M. husni Omar Tarawneh, Monther Tarawneh, Yousef Sharrab, "Mushroom Classification using Machine-Learning Techniques," *Proc. 4th Int. Comput. Sci. Informatics Conf.*, 2022, doi:

doi.org/10.1063/5.0174721.

[35]   Y. Fakir, M. Azalmad, and R. Elaychi, "Study of The ID3 and C4.5 Learning Algorithms," *J. Med. Informatics Decis. Mak.*, vol. 1, no. 2, pp. 29–43, Apr. 2020, doi: 10.14302/issn.2641-5526.jmid-20-3302.

[36]   Ismail Yusuf Panessai, M. M. Lakulu, M. H. Abdul Rahman, N. A. Z. Mohd Noor, N. S. Mat Salleh, and Aldrin Aran Bilong, "PSAP: Improving Accuracy of Students' Final Grade Prediction using ID3 and C4.5," *Int. J. Artif. Intell.*, vol. 6, no. 2, pp. 125–133, 2019, doi: 10.36079/lamintang.ijai-0602.42.

[37]   I. Ç. Taşkin Kavzoğlu, "Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması," *Har. Teknol. Elektron. Derg*, vol. 2, no. 1, pp. 36–45, 2010.

[38]   M. B. Gülen Yildiz Turp, "Tıbbi ve Yenilebilir Mantarlar & Et Ürünlerinde Kullanımı," *YYU J AGR SCI*, vol. 28, no. 1, pp. 144–153, 2018.