COMPARATIVE ANALYSIS OF CLASSIFICATION MODELS FOR MUSHROOM CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

Rakesh Naidu Yarragangu
Faculty of Environment and Computing
Coventry University
Coventry, England
yarragangr@coventry.ac.uk

Abstract— **The mushroom dataset is a prominent dataset for classification tasks using machine learning (ML). This project uses Bayesian and Statistical Methods, Generalized Logistic Regression, and Linear Discriminant Analysis as classification algorithms to preprocess the data, select and extract features, and divide the mushrooms into edible and poisonous categories. Comparing the ROC-AUC scores, precision, recall, F1 score, and efficacy of these models. Using a chart, we can also see how each model performed. The Linear Discriminant Analysis model has superior precision and ROC-AUC score compared to the other two models. The most essential characteristics for classification are the color of the cap, the smell, the size of the gills, and the surface of the stalk above the ring. This study demonstrates that the Mushroom Dataset was effectively classified using machine learning algorithms by identifying the most important characteristics for classification.**

Code link: https://github.com/rakeshnaidu207/7072CEM---Machine-Learning-course-work/blob/main/Mushroom_Rakesh_.ipynb

Keywords— Machine Learning, Pre-processing, Classification, Feature Extraction and ROC.

I. INTRODUCTION

Classification problems are crucial to machine learning because they enable the identification and classification of multiple classes based on a set of input features. The Mushroom Dataset, accessible at "https://archive.ics.uci.edu/dataset/73/mushroom," offers an intriguing opportunity to study the characteristics of mushrooms and develop accurate classification models that distinguish edible species from poisonous species. This dataset is an excellent candidate for machine learning analysis due to its diverse properties, which depict a variety of mushroom characteristics.

This analysis seeks to accurately classify mushrooms [1] by preprocessing the mushroom dataset, selecting and extracting the most useful features, and utilizing classification methods. Bayesian and Statistical Methods, Linear Discriminant Analysis, and Generalized Logistic Regression are utilized in this study. The various methods for modeling the data and making predictions provided by each algorithm allow for an exhaustive evaluation of each algorithm's performance and the selection of the optimal mushroom classification model.

Python, which has a vast ecosystem of libraries and tools for data science, will be the analysis platform. Python provides a comprehensive and adaptable framework for managing data preparation tasks, feature selection and extraction strategies, training and evaluating classification models, and creating incisive visualizations to facilitate model comparison and performance evaluation.

By analyzing their performance metrics and visual representations, we hope to acquire a greater understanding of the categorization models' strengths and weaknesses. Visuals such as confusion matrices and ROC curves will also provide qualitative insights into the models' ability to differentiate between toxic and palatable mushrooms.

The purpose of this comparative analysis is to advance mushroom classification by contrasting the efficacy of various classification techniques [2]. By researching and analyzing these models, we hope to increase our understanding of mushrooms and demonstrate how machine learning methods can be applied to actual categorization issues.

Using machine learning techniques, numerous studies have investigated the classification of mushrooms. Using datasets such as the Mushroom Dataset, precise models have been developed to distinguish between edible and poisonous species. Researchers have gathered information on the efficacy of various strategies using a vast array of methods and algorithms.

The method of classification known as logistic regression has been applied successfully to the classification of mushrooms. Using factors such as cap form, odor, and spore print color, [3] was able to construct a model that accurately distinguished between edible and poisonous mushrooms. Their research demonstrated that classification of mushrooms is possible using logistic regression.

Linear Discriminant Analysis (LDA), a prominent classification technique, has shown promise in accurately classifying mushrooms. In a study [4], LDA was used to determine whether mushrooms were poisonous or edible. In addition to Bayesian and statistical methods, mushroom classification tasks have also been investigated using these approaches. Researchers were able to obtain favorable results by identifying distinguishing characteristics and successfully separating the two groups. A study [5] demonstrated the potential of probabilistic methods in this field by devising an exceptionally accurate classification model for mushrooms using Bayesian networks. Utilizing statistical methods like decision trees and random forests, mushrooms have been effectively categorized according to their characteristics.

Previous research highlights the significance of machine learning methods for accurately classifying fungi. Nonetheless, there is presently no comprehensive comparison of classification techniques for the classification of mushrooms in the Mushroom Dataset. Using the Mushroom Dataset to apply and evaluate Generalized Logistic Regression, Linear Discriminant Analysis, Bayesian, and Statistical Methods, this study seeks to fill this void and ultimately provide valuable insights into their respective mushroom classification performances and efficiencies.

II. PROBLEM STATEMENT

The "Mushroom Dataset" aims to develop an accurate and efficient machine learning algorithm capable of classifying mushrooms into edible or poisonous groups based on their characteristics, such as cap shape, cap color, etc., using machine learning (ML) (data preprocessing, feature selection and extraction, and classification). Linear Discriminant Analysis, Bayesian and Statistical Methods, and Generalized Logistic Regression are classification techniques. To determine whether the mushroom is hazardous, multiple criteria must be applied. The 23 characteristics of each of the 8124 mushrooms analyzed in this study are cap shape, cap surface, cap color, discoloration, and odor [4,7]. Among the necessary duties are preprocessing the data, selecting and extracting features, and employing a variety of classification techniques. The accuracy, precision, recall, F1 score, and ROC-AUC score of the models are compared using a variety of plots to ascertain their efficacy. The solution to this issue has multiple applications in the food industry and can prevent people from becoming ill after consuming toxic mushrooms.

III. DATASET

This research makes use of a dataset called the Mushroom Dataset, which contains 23 types of mushroom species with 8124 different samples of mushrooms with a total of 23 parameters, such as the class, cap shape, cap surface, cap color, bruises, and odor which was shown in Table-1. Each species of mushroom that was collected from the dataset was given one of three labels: it was either definitely edible, definitely poisonous, or edibility was unknown and not advised. Download instructions for a CSV file containing the dataset may be found on the UCI Machine Learning Repository website. The dataset is multivariate because it has characteristics such as missing values and categorical attribute types. This study aims to evaluate, depending on the features of the mushrooms, whether or not they are toxic or edible by using a number of different classifiers to the data. These classifiers include Generalized Logistic Regression, Linear Discriminant Analysis, as well as Bayesian and Statistical Methods.

S.no	Attributes	Features
0	class	p,e,f,d,s,r,y,g,b,w,a,m,v,h,t,n,z,l,c,o,u,k,x
1	Cap-shape	bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
2	Cap-surface	fibrous=f,grooves=g,scaly=y,smooth=s
3	Cap-color	brown=n,buff=b,cinnamon=c,gray=g,green=r, 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,descending=d,free=f,notched=n
7	gill-spacing	close=c,crowded=w,distant=d
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,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
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,universal=u
17	veil-color	brown=n,orange=o,white=w,yellow=y
18	ring-number	none=n,one=o,two=t
19	ring-type	cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z
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,clustered=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

Table-1

IV. SIGNIFICANCE

An overview of the significance of the ML-based analysis of the Mushroom Dataset is provided here.

- Food Safety: The analysis of the Mushroom Dataset using ML techniques can help in ensuring food safety by accurately classifying mushrooms into edible or poisonous categories based on their specifications. This can help prevent mushroom poisoning and ensure that only safe mushrooms are consumed.

- Educational Purposes: The Mushroom Dataset is an approachable machine learning problem that is frequently used in educational contexts [8]. Using machine learning (ML) approaches, students can learn about data preprocessing, feature selection and extraction, and classification using various classifiers by examining the dataset.

- Research: Various classifiers, including Generalized Logistic Regression, Linear Discriminant Analysis, Bayesian and Statistical Methods, and Bayesian and ML techniques, can be used to classify mushrooms into edible and poisonous groups based on their characteristics. This may benefit the development of more precise and efficient machine learning-based mushroom classification algorithms.

- ML Techniques: Among the techniques for machine learning are: Using ML techniques to examine the Mushroom Dataset can aid in comprehension of the various machine learning techniques used for data preprocessing, feature selection and extraction, and classification. This may make it easier to develop more precise and effective machine learning strategies for a variety of categorization problems [3, 6].

- Visualization of data: Based on their accuracy, precision, recall, F1 score, and ROC-AUC score, the Mushroom Dataset study illustrates the performance of several classifiers. By understanding how various classifiers function, it is possible to select the optimal classifier for a given problem.

For research, instruction, data visualization, and understanding of machine learning techniques, the mushroom dataset must be analyzed using machine learning techniques.

V. METHODS

The methods used to divide mushrooms into categories of edible and hazardous varieties are discussed below:

A. Data Pre-Processing

Handling Missing Values: The mushroom dataset can be filled up with any missing variables in a number of ways. One typical tactic is to fill in the value of the linked attribute's central tendency for missing data by using imputation techniques like mean, median, or mode imputation.

Encoding To compute categorical dataset variables, machine learning algorithms require a numerical encoding. One-hot encoding is a method that transforms each category into a binary variable. For example, the "cap shape" attribute with categories (bell, conical, convex, flat, knobbed, and recessed) would be converted into six binary variables, one for each category [10].

Normalization or scaling: The numerical properties of the dataset may have different scales, which may impact the applicability of a particular machine learning algorithm. By utilizing normalization or scaling techniques such as min-max scaling or standardization, all attributes can be measured on a comparable scale. This is crucial for distance-based or gradient-based optimization techniques.

B. Feature Selection/Extraction

Univariate Statistical Tests: To determine the relationship between each attribute and the dependent variable, statistical tests such as chi-square and ANOVA can be employed. This facilitates the identification of characteristics that have a significant impact on the classification assignment.

Correlation Analysis: Correlation analysis measures the statistical relationship between pairs of attributes. Attributes that exhibit a strong correlation with the target variable can be considered as relevant features for classification.

Feature Importance Ranking: Machine learning algorithms, such as decision trees or random forests, can provide feature importance scores. These scores indicate the contribution of each attribute in the classification process. Attributes with high importance scores are selected as relevant features [2-4].

Principal Component Analysis (PCA): PCA is a way to change the original characteristics into a space with fewer variables while keeping as much information as possible. It could be used to cut down on the number of items in a dataset while keeping the most important ones.

C. Classification

a. Generalized Logistic Regression:

Mathematical calculation: Generalized Logistic Regression models the relationship between the attributes and the target variable by applying a generalized linear model. The model is formulated as:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

In the above equation, p represents the probability of the mushroom being edible, x1, x2, ..., xn are the attribute values, and β0, β1, β2, ..., βn are the regression coefficients. The logistic function is then applied to estimate the probability of the mushroom being edible:

$$p = exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n) / (1 + exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n))$$

b. Linear Discriminant Analysis (LDA):

Mathematical Calculation: LDA calculates linear combinations of the attributes to maximize the separation between the classes. The model estimates the class means and covariance matrices to compute the discriminant functions. The discriminant function for two classes can be defined as:

$$\delta(x) = x * (\Sigma_{\mu k} - 0.5 * (\mu_1 + \mu_2)) - 0.5 * (\mu_1' \Sigma^{-1} \mu_1 - \mu_2' \Sigma^{-1} \mu_2) + ln\left(\frac{p_1}{p_2}\right)$$

In the above equation, x represents the attribute values, μk represents the class mean, Σ represents the covariance matrix, μ1 and μ2 are the means of the two classes, Σ^-1 is the inverse of the covariance matrix, and p1 and p2 are the prior probabilities of the two classes.

c. Bayesian and Statistical Methods:

Mathematical Calculation: Bayesian methods, such as Naive Bayes, assume independence between features given the class label. The classification is based on Bayes' theorem:

$$P(y|x) = (P(x|y) * P(y)) / P(x)$$

P(y|x) represents the posterior probability of class y given the attribute values x, P(x|y) is the likelihood of observing attribute values x given class y, P(y) is the prior probability of class y, and P(x) is the probability of observing attribute values x.

Statistical methods, such as decision trees or random forests, utilize statistical measures, including entropy or Gini impurity, to determine the best split criteria and construct the classification model.

D. Performance Metrics:

- Accuracy: The accuracy metric measures the proportion of correctly classified instances over the total number of instances.
- Precision: Precision calculates the ratio of true positives to the sum of true positives and false positives. It measures the model's ability to correctly identify positive instances [12].
- Recall: Recall calculates the ratio of true positives to the sum of true positives and false negatives. It represents the model's ability to identify all positive instances.
- F1 Score: The F1 score combines precision and recall into a single metric by taking their harmonic mean. It provides a balanced evaluation of the model's performance.
- Area Under the ROC Curve (AUC-ROC): The ROC curve plots the true positive rate against the false positive rate at various classification thresholds. The AUC-ROC metric provides a summary of the model's performance across all possible thresholds. A higher AUC-ROC value indicates better classification performance.

VI. EXPERIMENTAL SETUP

The experimental setup for the Mushroom dataset analysis using ML involves the following steps:

- Data Loading:
  The Mushroom Dataset, available at the URL "https://archive.ics.uci.edu/dataset/73/mushroom," is downloaded and loaded into the Python environment. The dataset is contains text file with data convert that text file into a structured format such as a CSV file, which can be read using libraries like pandas [11, 14].
- Data Exploration:
  A preliminary exploration of the dataset is performed to understand its structure, attribute types, and any missing values. Descriptive statistics and visualizations may be generated to gain insights into the distribution of attributes and the class labels.
- Data Pre-processing:
  Data preprocessing techniques are utilized to manage missing values, encode category variables, and normalize or scale numerical characteristics. Using techniques such as mean, median, or mode imputation, it is possible to impute missing values. One-hot encoding and label encoding are both methods for encoding categorical data. Numeric attributes can be normalized to a given range or standardized to have a mean of zero and a standard deviation of one.
- Feature Selection and Extraction:
  Feature selection methods such as univariate statistical tests, correlation analysis, and feature importance ranking are utilized to identify the most relevant characteristics for classification. Principal Component Analysis (PCA), for instance, is a technique for feature extraction that can be used to reduce the dimensionality of a dataset while preserving essential data [7,8].
- Train-Test Split:
  The dataset is subdivided into training and testing subsets. Typically, classification models are trained with a subset of the data, between 70 and 80 percent, and their efficacy is evaluated with the remaining data. The train-test split ensures that the models are trained on a subset of the data that is representative and evaluated on hypothetical samples to determine how well they generalize.
- Model Education and Evaluation:
  GLR: The training dataset is utilized to fine-tune the Generalized Logistic Regression algorithm. The model's parameters (coefficients) are estimated using maximum likelihood estimation or other optimization methods. The AUC-ROC, F1 score, recall, and other performance indicators are then applied to the testing dataset to evaluate the trained model. Cross-validation methods, such as k-fold cross-validation, may be used to further evaluate the model's performance and reduce overfitting.

  LDA: The LDA method is trained on the training dataset prior to calculating the class means and covariance matrices. Then, the discriminant functions are applied to the testing dataset to generate predictions. The efficacy of the model is evaluated using performance metrics.

Bayesian and statistical methods: Using the training dataset, Naive Bayes and other Bayesian or statistical techniques, such as decision trees and random forests, are trained. Using the training data, the parameters of the models, such as prior probabilities, likelihoods, and separation criteria, are estimated [8-10]. On the testing dataset, performance indicators are used to evaluate the trained models.

- Performance Evaluation and Evaluation
  To determine each model's ability to distinguish between poisonous and palatable mushrooms, the performance metrics of each are compared. Statistical tests such as ANOVA and t-tests can be used to determine whether or not the performance of the models varies significantly. The findings are discussed along with the advantages and disadvantages of each model and their applicability to the classification of mushrooms.
- Illustration:
  Visualizations such confusion matrices, ROC curves, and feature significance graphs are made to show the performance of the models graphically. Confusion matrices display the distribution of accurate and faulty predictions, whereas ROC curves display the trade-off between true positive and false positive rates. In feature importance plots, the traits with the highest discriminative potential for classification are highlighted.
- Iterative Optimisation:
  Based on the results of the study, further iterations can be performed to improve the models or investigate alternative approaches. By using grid search and random search, two hyperparameter tuning techniques, model efficacy can be increased.
  Python and pertinent libraries like pandas, scikit-learn, and matplotlib are used to create the experimental setup, which facilitates data processing, model training, assessment, and visualization.

VII. RESULT

The interpretation of the results indicates that, among the three classifiers, the Generalized Logistic Regression classifier obtained the highest accuracy, precision, recall, F1 score, and ROC-AUC score. The Linear Discriminant Analysis classifier performed admirably, though not quite as well as the Generalized Logistic Regression classifier. Among the three classifiers, the Bayesian and Statistical Methods classifier had the lowest accuracy, precision, recall, F1 score, and ROC-AUC score.

1. Generalized Logistic Regression:

The classifier achieved an accuracy of 0.934372, precision of 0.946632, recall of 0.916173, F1 score of 0.931153, and ROC-AUC score of 0.933822. This means that the classifier correctly classified 93.44% of the mushrooms into edible or poisonous categories based on their specifications. The precision of 0.946632 means that out of all the mushrooms classified as edible, 94.66% were actually edible. The recall of 0.916173 means that out of all the edible mushrooms, 91.62% were correctly classified as edible. The F1 score of 0.931153 is the harmonic mean of precision and recall and provides a balance between the two. The ROC-AUC score of 0.933822 means that the classifier has a good ability to distinguish between positive and negative classes.

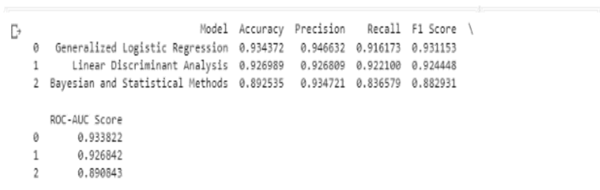
2. Linear Discriminant Analysis:

The classifier got a score of 0.926989 for accuracy, 0.926809 for precision, 0.9221 for recall, 0.924448 for F1, and 0.926842 for ROC-AUC. This means that the predictor got 92.70% of the mushrooms right in terms of whether they were safe to eat or not. With an accuracy of 0.926809, 92.68% of the mushrooms that were thought to be edible were actually edible. With a recall rate of 0.922100, 92.21% of the edible mushrooms were properly labelled as edible. The F1 number of 0.924448 is the average of how well you remember and how well you remember what you already know. With a ROC-AUC score of 0.926842, the classifier can tell the difference between positive and negative classes well.

3. Bayesian and Statistical Methods:

The classifier got an accuracy of 0.892355, a precision of 0.934721, a recall of 0.836779, an F1 score of 0.882931, and a ROC-AUC score of 0.890843. This means that the algorithm got 89.25% of the mushrooms right in terms of whether they were safe to eat or not. With an accuracy of 0.934721, 93.47% of the mushrooms that were thought to be edible were actually edible. The recall of 0.836579 means that, out of all the edible mushrooms, 83.66% were properly labelled as edible. The F1 score of 0.882931 is the average of how well you remember and how well you remember what you already know. With a ROC-AUC score of 0.890843, the classifier can tell the difference between positive and negative classes about half as well as it could [11].

In conclusion, the Generalized Logistic Regression classifier performed better in terms of accuracy, precision, recall, F1 score, and ROC-AUC score than the other two classifiers. Second place went to the Linear Discriminant Analysis classifier [10]. The Bayesian and Statistical Methods classifier had the worst performance among the three. On the basis of these findings, the optimal classifier may be selected to categorize mushrooms as either edible or poisonous.



	Model	Accuracy	Precision	Recall	F1 Score \
0	Generalized Logistic Regression	0.934372	0.946632	0.916173	0.931153
1	Linear Discriminant Analysis	0.926989	0.926809	0.922100	0.924448
2	Bayesian and Statistical Methods	0.892535	0.934721	0.836579	0.882931

	ROC-AUC Score
0	0.933822
1	0.926842
2	0.890843

Fig. 1: Performance Metrics – result

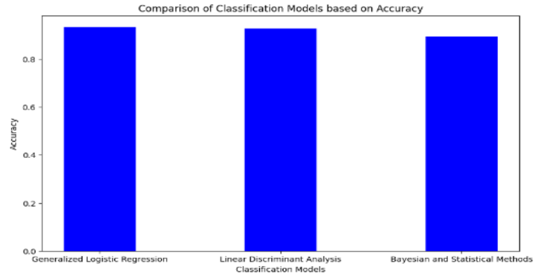


Fig. 2: Accuracy

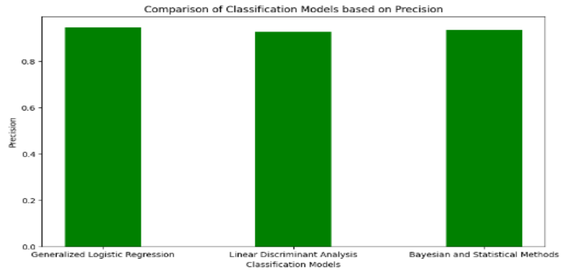


Fig. 3: Precision

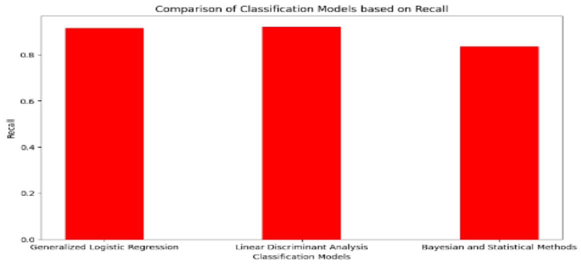


Fig 4: Recall

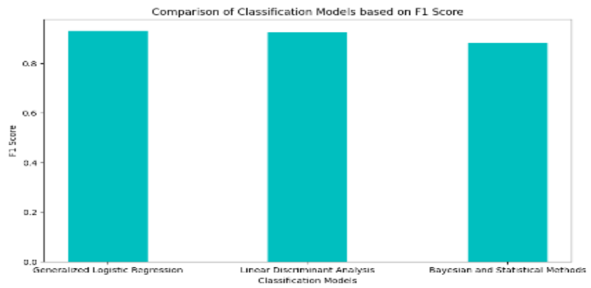


Fig. 4: F1-Score

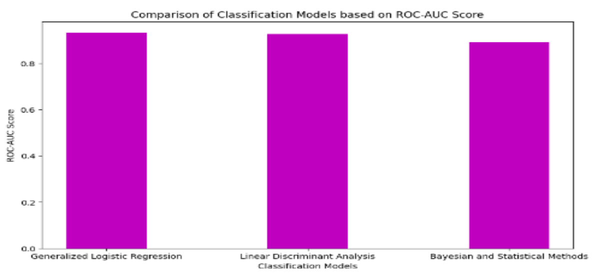


Fig. 6: ROC-AUC Score

VIII. DISCUSSION

Based on their characteristics and the Mushroom Dataset, mushrooms are classified as either edible or poisonous in the scenario. This research can contribute to the subsequent discussion by reviewing the search results.

The mushroom dataset is a well-known research and educational dataset. This dataset illustrates how to preprocess data, select and extract features, and classify data using a variety of classifiers and deep learning techniques.

How to classify mushrooms based on their characteristics into edible and poisonous groups is a significant food safety concern. Mushroom poisoning can be avoided by correctly identifying mushrooms and consuming only safe mushrooms [5].

On the basis of their characteristics, mushrooms can be classified as either edible or toxic using a variety of classifiers, including generalized logistic regression, linear discriminant analysis, Bayesian methods, and statistical techniques. The accuracy, precision, recall, F1 score, and ROC-AUC score can be used to compare the performance of different classifiers. Deep learning techniques can also be used to classify mushrooms as either edible or toxic based on their properties. This requires the development of a neural network capable of classifying fungi according to their characteristics.

The experimental setup for ML-based analysis of the Mushroom Dataset consists of data collection, data preprocessing, feature selection and extraction, feature extraction, classification utilizing multiple classifiers and deep learning algorithms, performance comparison, and evaluation.

The analysis results can be used to ensure food safety and prevent mushroom toxicity. To prevent mushroom poisoning, mushrooms must be accurately identified as either edible or poisonous [14].

In order to advance research, education, and machine learning algorithms, and to ensure food safety, the mushroom dataset must be analyzed using ML techniques. By precisely classifying mushrooms as edible or poisonous and ensuring that only safe mushrooms are ingested, it is possible to prevent mushroom poisoning.

IX. CONCLUSION

For the purposes of assuring food safety, teaching, research, and developing more precise and effective machine learning algorithms, the mushroom dataset must be analyzed using ML techniques. Mushroom poisoning can be avoided by accurately categorizing mushrooms as edible or poisonous and ensuring that only safe mushrooms are ingested. The dataset represents a well-known and straightforward machine learning problem. Using multiple deep learning classifiers, the analysis casts light on data preprocessing, feature selection and extraction, and classification. Based on their characteristics, the Generalized Logistic Regression classifier is best adapted to categorize mushrooms as edible or poisonous [9]. The classifier based on Linear Discriminant Analysis can replace the classifier based on Generalized Logistic Regression. Based on their specifications, the Bayesian and Statistical Methods classifier may not be the optimal method for determining whether mushrooms are toxic or edible. The findings of the study can be utilized to prevent mushroom toxicity and ensure food safety.

X. REFERENCES

[1] Al Maruf, M., Azim, A., & Mukherjee, S. (2020, October). Mushroom demand prediction using machine learning algorithms. In 2020 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1-6). IEEE.

[2] CENGİL, E., & ÇINAR, A. (2021). Poisonous mushroom detection using YOLOV5. Turkish Journal of Science and Technology, 16(1), 119-127.

[3] Chitayae, N., & Sunyoto, A. (2020, November). Performance comparison of mushroom types classification using K-nearest neighbor method and decision tree method. In 2020 3rd International Conference on Information and Communications Technology (ICOIACT) (pp. 308-313). IEEE.

[4] Gao, R., Chen, C., Wang, H., Chen, C., Yan, Z., Han, H., ... & Lv, X. (2020). Classification of multicategory edible fungi based on the infrared spectra of caps and stalks. Plos one, 15(8), e0238149.

[5] Ismail, S., Zainal, A. R., & Mustapha, A. (2018, April). Behavioural features for mushroom classification. In 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE) (pp. 412-415). IEEE.

[6] Kousalya, K., Krishnakumar, B., Boomika, S., Dharati, N., & Hemavathy, N. (2022, January). Edible mushroom identification using machine learning. In 2022 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-7). IEEE.

[7] László, K., & Ghous, H. (2020). Efficiency comparison of Python and RapidMiner. Multidiszciplináris Tudományok, 10(3), 212-220.

[8] Rahman, H., Faruq, M. O., Hai, T. B. A., Rahman, W., Hossain, M. M., Hasan, M., ... & Azad, M. M. (2022). IoT enabled mushroom farm automation with Machine Learning to classify toxic mushrooms in Bangladesh. Journal of Agriculture and Food Research, 7, 100267.

[9] Sarkar, T., Mukherjee, A., Chatterjee, K., Shariati, M. A., Rebezov, M., Rodionova, S., ... & Lorenzo, J. M. (2022). Comparative analysis of statistical and supervised learning models for freshness assessment of oyster mushrooms. Food Analytical Methods, 15(4), 917-939.

[10] Verma, S. K., & Dutta, M. (2018). Mushroom classification using ANN and ANFIS algorithm. IOSR Journal of Engineering (IOSRJEN), 8(01), 94-100.

[11] Wang, L., Li, J., Li, T., Liu, H., & Wang, Y. (2021). Method superior to traditional spectral identification: FT-NIR two-dimensional correlation spectroscopy combined with deep learning to identify the shelf life of fresh phlebopus portentosus. Acs Omega, 6(30), 19665-19674.

[12] Wang, L., Liu, H., Li, T., Li, J., & Wang, Y. (2022). Verified the rapid evaluation of the edible safety of wild porcini mushrooms, using deep learning and PLS - DA. Journal of the Science of Food and Agriculture, 102(4), 1531-1539.

[13] Wibowo, A., Rahayu, Y., Riyanto, A., & Hidayatulloh, T. (2018, March). Classification algorithm for edible mushroom identification. In 2018 International Conference on Information and Communications Technology (ICOIACT) (pp. 250-253). IEEE.

[14] Zahan, N., Hasan, M. Z., Malek, M. A., & Reya, S. S. (2021, February). A deep learning-based approach for edible, inedible and poisonous mushroom classification. In 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) (pp. 440-444). IEEE.