

PAPER NAME

mushrrom classification- 2086.docx

AUTHOR

Suresh kumar Mandala

WORD COUNT

2364 Words

CHARACTER COUNT

13942 Characters

PAGE COUNT

8 Pages

FILE SIZE

1.4MB

SUBMISSION DATE

May 2, 2024 11:21 AM GMT+5:30

REPORT DATE

May 2, 2024 11:21 AM GMT+5:30

### ● 13% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 9% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 4% Submitted Works database

### ● Excluded from Similarity Report

- Bibliographic material
- Cited material

# CLASSIFICATION OF MUSHROOMS

Goundla Sai Hasini

Department of Computer Science and Engineering

SR UNIVERSITY

Warangal, India

Hasinigoundla335@gmail.com

**Abstract**— The classification of fungi is important in many fields, including mycological research on food safety and the environment. In this paper, we investigate the use of machine learning methods for accurate and efficient categorization of mushrooms. We collected a collection of mushroom images classified into edible, inedible, and toxic categories from mushroom photographs. We extracted a number of characteristics such as eigenfeature histograms and parametric features to train classification models. We used various machine learning techniques including neural networks nn support vector machines svm decision trees and k-nearest neighbors knn overall our work shows how machine learning can be used for effective mushroom categorization. The results highlight the importance of considering the context selection algorithm and feature extraction methods to achieve high accuracy these discoveries have practical implications for several applications involving the identification and classification of fungi as advances in the field of fungal classification

**Save Report**

*IndexTerms—Mushroom, fungi, Classification, Machine Learning.*

## I. INTRODUCTION :

ml which is used to classify mushrooms the goal is to reliably differentiate between hazardous and edible varieties based on traits including gill size stalk structure color and shape of the cap for this work a labeled collection of mushroom samples with many attributes and their classification as edible or dangerous is needed to make sure this data compatible with the ml algorithms preprocessing is an essential step it might be included encoding categorical data scaling numerical properties and managing incomplete or missing data a machine learning algorithm is selected trained during the preprocessed data this makes it possible for the algorithm to recognize trends and accurately categorize mushrooms as harmful or edible precision one of the parameters used to assess the models efficacy learning mo classify shrooms the goal is to reliably differentiate between hazardous and edible varieties based on traits including gill size stalk structure color and shape of the cap for this work a labeled collection of mushroom samples with many attributes and their classification as

edible or dangerous is needed to make sure the data is compatible with machine learning algorithms preprocessing is an essential step this may include encoding categorical data scaling numerical properties and managing incomplete or missing data a machine learning algorithm is selected and trained on the preprocessed data this makes it possible for the algorithm to recognize trends and accurately categorize mushrooms as harmful or edible precision one of the parameters used to assess the models efficacy

## LITERATURE REVIEW

[1] Author : AgungWibowo; Yuri Rahayu; AndiRiyanto; TaufikHidayatulloh.(2018) Classification method of toxic mushroom or not are going to be simply conducted by learning machine exploitation mining study mutually of the ways in which to extract pc power-assisted data

[2] Author: Jinhua Dong; LixinZheng.(2019) As there are a few traditional algorithms for enoki mushroom detection, this paper planned a automatic enoki mushroom caps classification rule, and engineered a convolutional neural network model supported LeNet..

[3] Author: Suhaida Ismail; Amy Rosshaida Zainal; Aida Mustapha.(2018) This paper is about to check mushroom activity options like the form, surface and color of the cap, gill and stalk, additionally because the odour, population and environment of the mushrooms.

[4] Author: AlirezaMasoudian and Kenneth A.McIsaac.(2013) Mushrooms should be classified as healthy or unhealthy to make sure correct handling and maximize crop yield.

[5] Author: You-ming Shi; Cui-qiong Yan; Dong-yu Li; Gang Liu.(2011) In order to develop a speedy and correct methodology for discriminating Amanita mushrooms, Fourier rework infrared (FTIR)

[6] Author:He Xin-yi; Liu Jin-fu; Cheng Li-li.(2011) Quality properties of dehydration mushroom dried by freeze drying, explosion puffing drying and hot air drying were studied in this paper. The results showed that dehydration mushroom dried by freeze drying exhibited the simplest color values and rehydration capability.

[7] Author: Y.Munirah; M.Rozlini; Y.Siti Mariam.(2012) In this paper, the main objective is to develop a collection of rules that may be accustomed diagnose oyster agaric

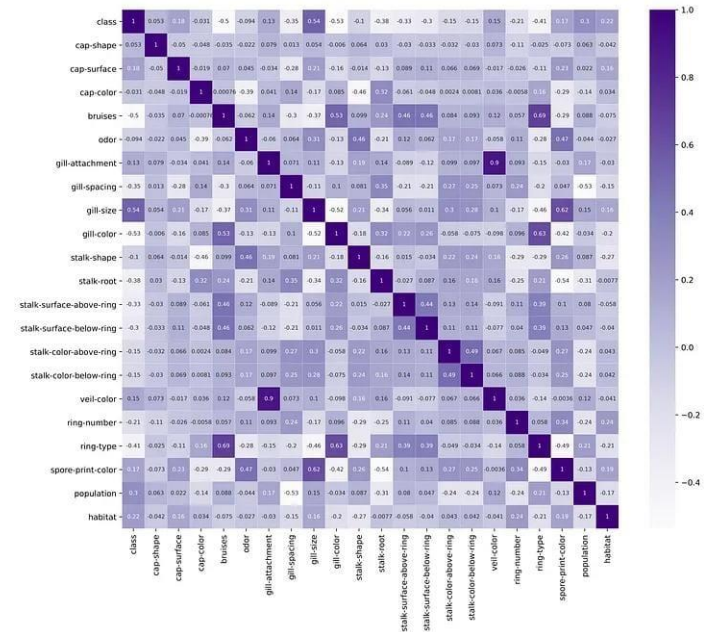
diseases. Within the close to future, we'll develop associate degree skilled System for designation oyster agaric Diseases to help person concerned in mushroom production and cultivation in designation the diseases cause by mildew, bacteria, virus, insects and different pests

[8]. Author: Al-mejibli and Hamad in [1] developed an application can be applied on a mobile phone and web application named Mushroom Diagnosis Assistance System, the purpose of this application is to realize safety when gathering mushroom. They used decision tree and naïve bays classifiers to group themushrooms types.

[9] Author: Chowdhury and S. Ojha in [13] identified a manner to distinguished several mushroom diseases using different datamining classification methods. They used actual dataset gatheredfrom mushroom farm by using data mining like Naïve Bayes, RIDOR and SMO algorithms.

[10]. Author: Onudu in [11] suggested modified K-means techniquebased on the traditional k-mean algorithm to enhance the clustering categorical dataset and solving the inherentproblem in the traditional clustering algorithm.

machine, the training and testing is very important. Training data is used to train the model because of that the model give the accurate answer. The testing data is used to test the data and give the predicted values then it will tell the accuracy of that particular model. In my model I divided the dataset in to two pieces 1st one training and 2nd is testing where it is divided in the ratio of 75percentage and 30percentage using the in-build functions.



### III. PROPOSED APPROCH

#### A. Overview

machine learning is used to accurately identify and categorize mushrooms some common methods include logistic regression lightgbm xgboost naive bayes multi-layer perceptron decision trees and svm logistic regression predicts the probability that a fungus belongs to a certain category svm is a technique that can handle complex data and is often used for mushroom classification these models monitor various mushroom features such as cap shape surface color stem shape color bruising odor and gill attachment to make predictions this information is useful for mycologists researchers and others who need to accurately identify fungi

#### B. Data Pre-Processing

- In the dataset the target or the analysing variable is Grade in that there are 0-7 values repeatedly so i came to known that the dataset is a classification.it is also known as multi classification dataset.With original dataset the accuracy is not proper so i changes to a classification dataset by changing 0-3 are 0 and 3-7 are 1.In dataset there is one string with the column name class i dropped it.later i filled the null values using fillna().
- after completing the process we will do split the data set into training and testing dataset.For the model or the

#### C. Implementation

- logistic regression:

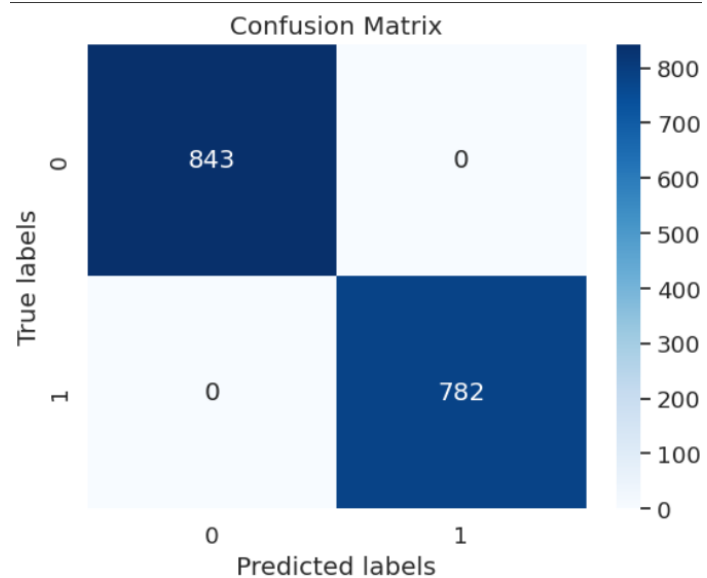
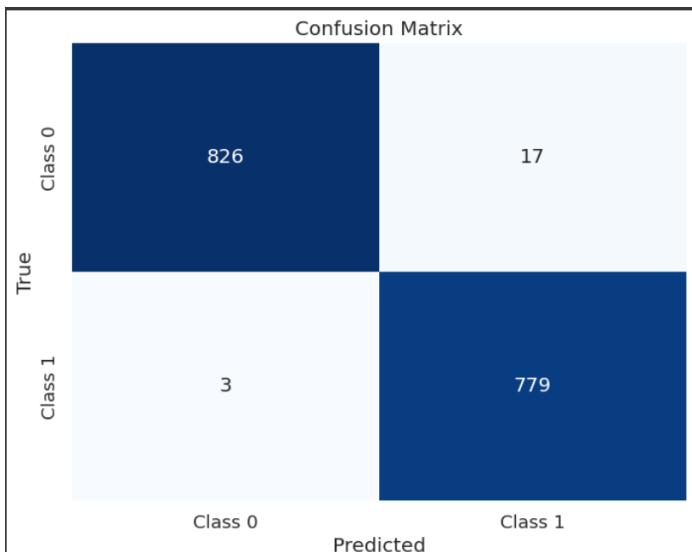
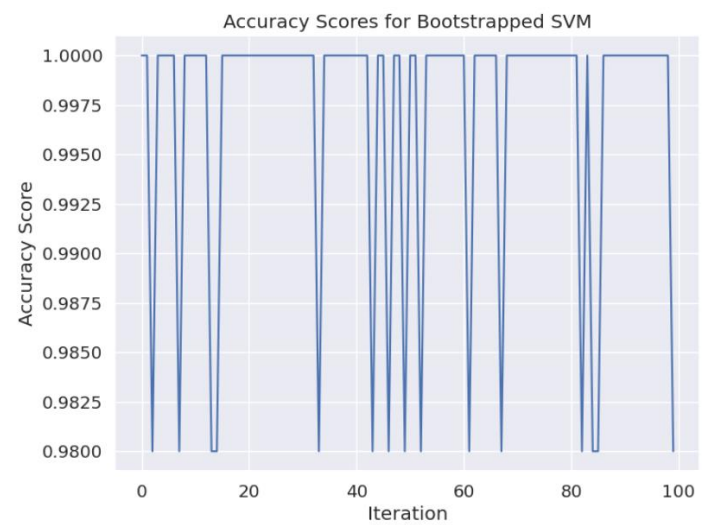
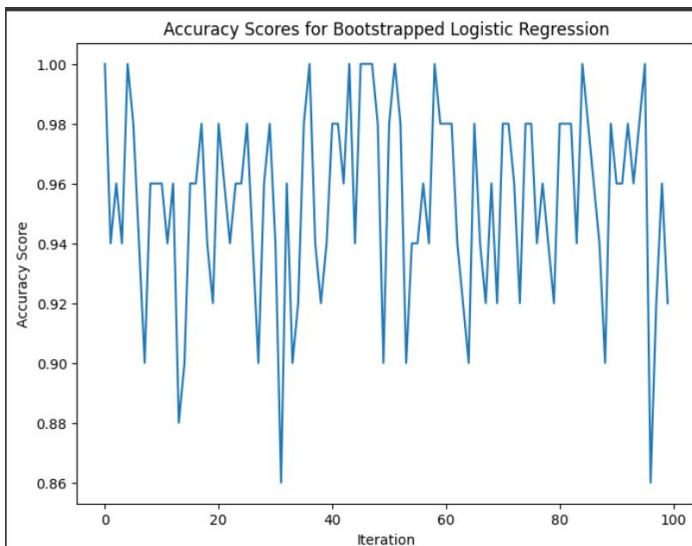
Logistic regression is a mathematical version for binary type duties, wherein the intention is to expect certainly one of two feasible consequences (e.G., 0 or 1, yes or no, real or false) primarily based on one or greater enter traits on. Unlike linear regression, which predicts a continuous outcome, logistic regression predicts the possibility of two consequences. A logistic characteristic (also referred to as a sigmoid characteristic) is used to version this probability. Here is an evidence of logistic regression with its formula. Measurement: A logistic regression model uses the logistic characteristic to estimate the chance (P) of a couple of events (e.G., the probability of a tremendous final results) the usage of the logistic function:

$$z = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$$

$$(1) y_p = 1 / (1 + e^{-z}) \quad (2)$$

$$z = -Y \log(YP) - (1-Y) \log(1-YP)$$

$$\text{INTERPRETATION}(1-YP) \quad (3)$$



#### 6 Support Vector Machine:

-Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, it's far used for Classification in addition to for Regression issues. However, it's miles specially used for classification issues in machine learning.

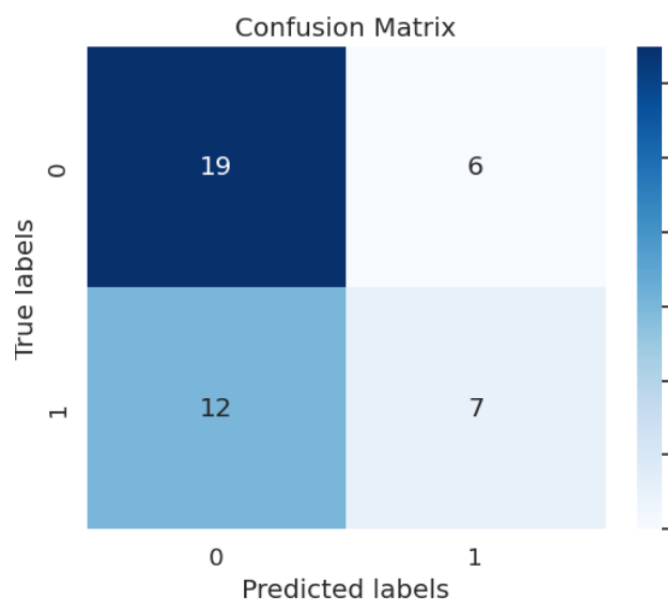
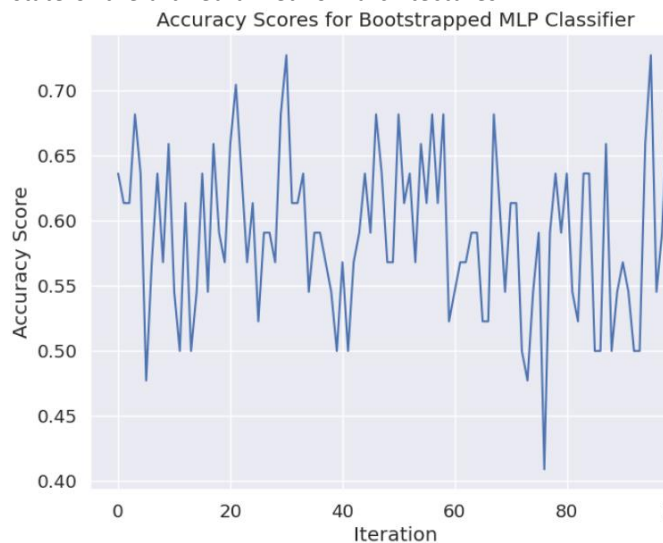
-The intention of the SVM set of rules is to create an most useful line or decision boundary that can separate the n-dimensional area into training so that within the destiny we can easily assign a brand new facts point to the suitable range. This choicest choice constraint is called a hyperplane.

#### 8 Multi-layer Perceptron:

-A Multi-Layer Perceptron (MLP) is a fundamental form of artificial neural network such as more than one layers of interconnected nodes or neurons. It generally accommodates an enter layer, one or greater hidden layers, and an output layer. Each neuron within the community gets enter alerts, strategies them thru a non-linear activation function, and transmits an output to neurons within the next layer.

-MLPs are able to learning complex styles and relationships within statistics, making them widely used in various fields consisting of pattern recognition, class, and regression tasks. Through a manner called backpropagation, MLPs regulate the weights of connections among neurons all through schooling, aiming to decrease the difference between expected and actual outputs. -Despite their effectiveness, MLPs have barriers, together with vulnerability to overfitting and problem in handling excessive-dimensional data.

Nevertheless, they remain a cornerstone within the subject of deep studying, forming the idea for more state-of-the-art neural network architectures.



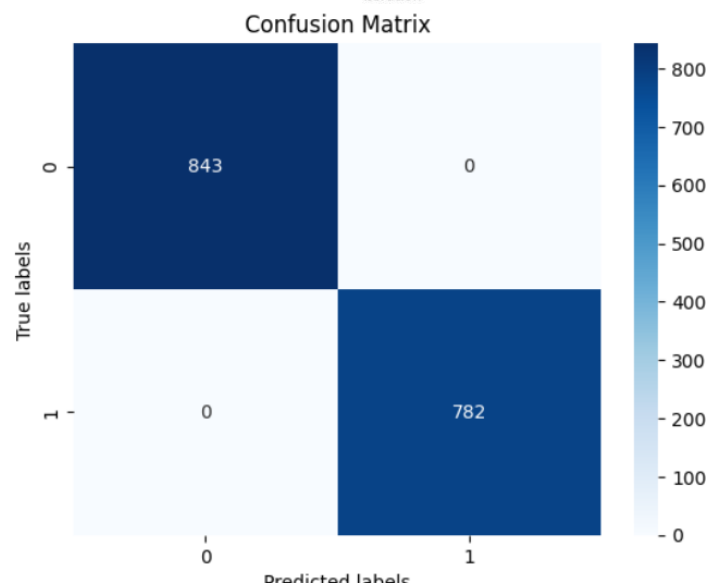
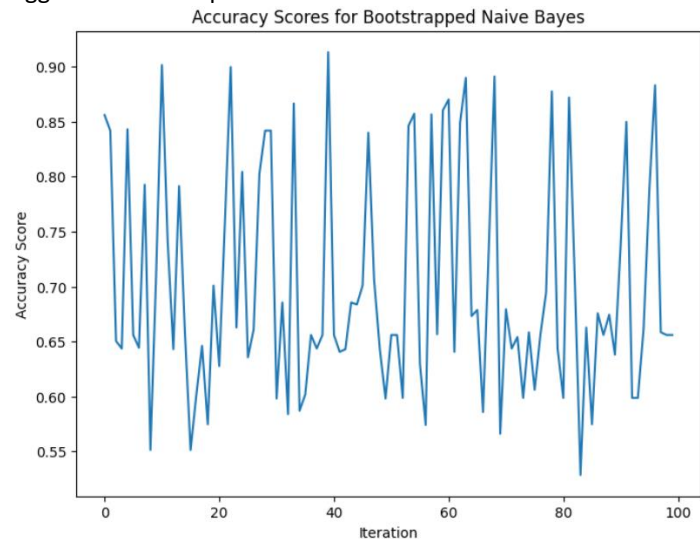
- Naive Bayes:

-Naive Bayes is a famous set of rules in machine learning. It's based totally on Bayes' theorem, which calculates the possibility of a hypothesis given the proof. Despite its simplicity, Naive Bayes often performs remarkably nicely, mainly in text type and spam filtering.

-The "naive" issue of Naive Bayes comes from its assumption of independence among functions. It assumes that each function contributes independently to the chance of a selected final results, which may not hold actual in real-international scenarios. However, this simplifying

assumption allows for immediate schooling and prediction, making Naive Bayes specially beneficial for massive datasets.

-Naive Bayes works with the aid of calculating the possibility of each elegance given the enter features and choosing the class with the very best possibility as the prediction. Despite its "naive" assumptions, Naive Bayes may be extraordinarily effective and is broadly utilized in numerous programs because of its simplicity, efficiency, and frequently aggressive overall performance.

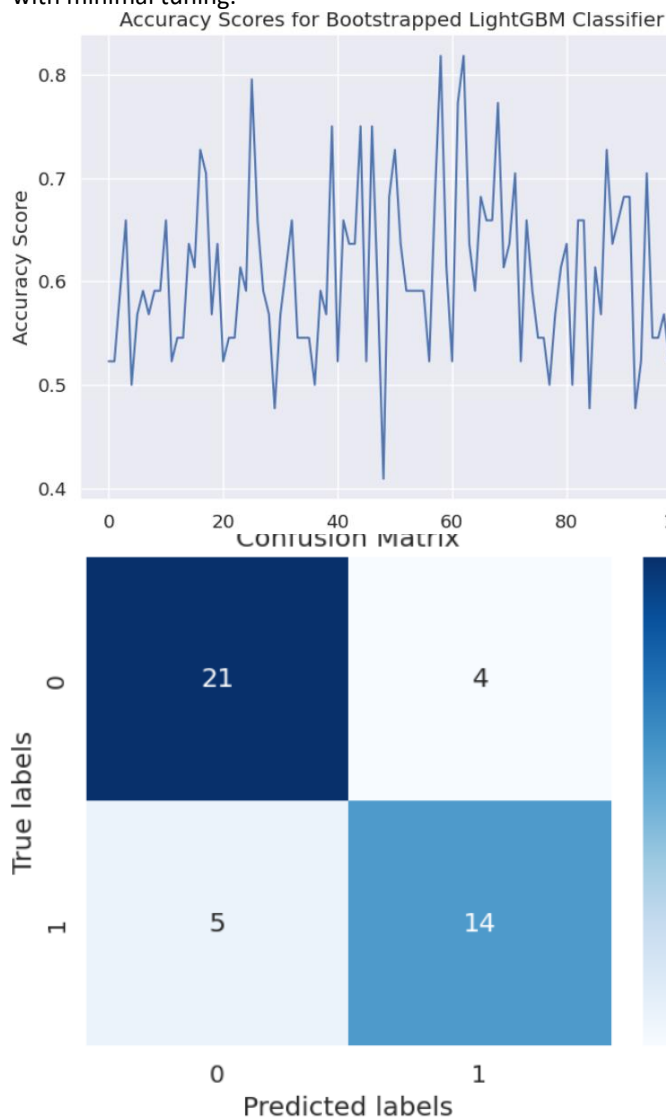


- Lightgbm:

LightGBM, brief for Light Gradient Boosting Machine, is a high-performance gradient boosting framework developed via Microsoft. It's designed for performance and speed in handling large datasets and is widely used in device studying competitions and enterprise programs. LightGBM makes use of a tree-based mastering set of rules, using a singular technique referred to as Gradientbased totally



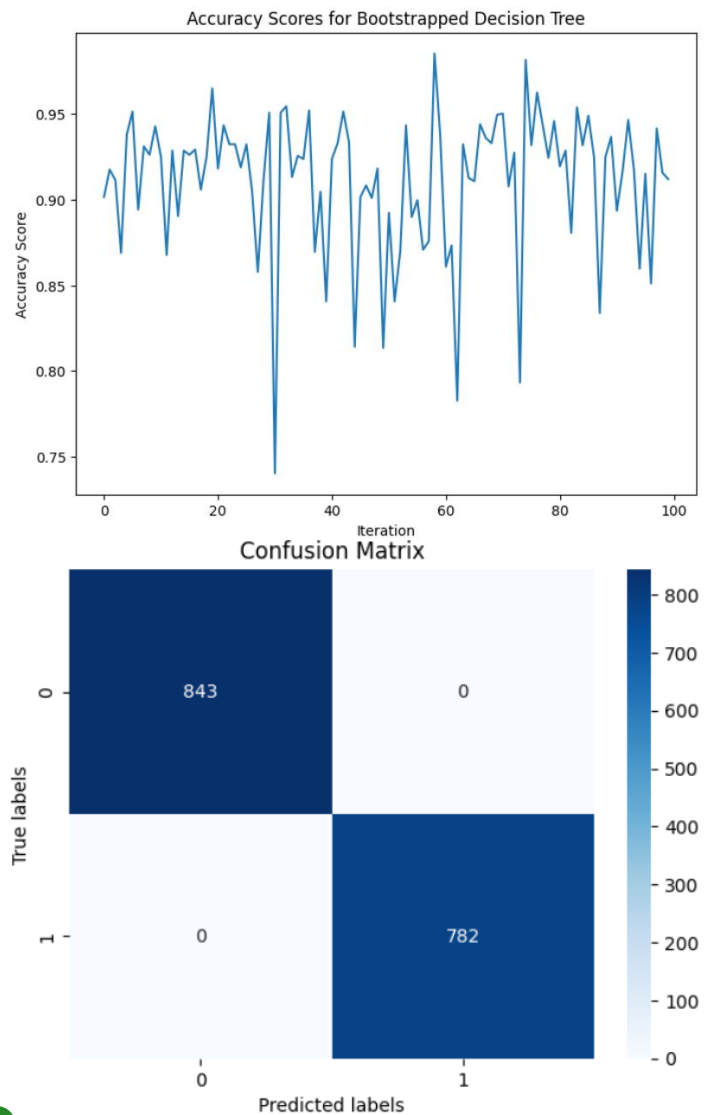
One-Side Sampling (GOSS) to reduce memory usage and enhance schooling velocity. It additionally utilizes Histogram-primarily based algorithms for quicker computation. Its flexibility, scalability, and capacity to handle specific features make it a famous desire for diverse tasks, along with category, regression, and ranking problems, continuously handing over competitive effects with minimal tuning.



- Decision Tree:

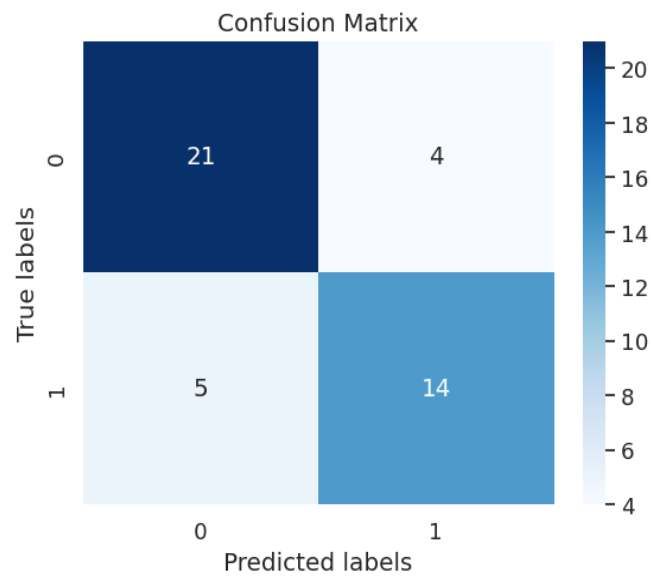
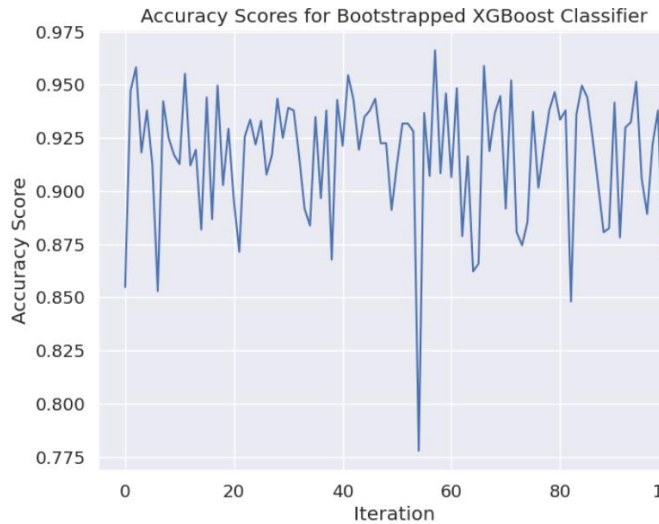
A decision tree is a predictive modeling tool utilized in facts mining and machine getting to know. It organizes records right into a tree-like structure of choices and their viable outcomes. Each internal node represents a "take a look at" on an characteristic, every branch represents the final results of the test, and every leaf node represents a class label or a choice. Decision timber are famous due to their simplicity and interpretability, making them treasured for know-how and explaining the decisionmaking technique. They're employed in various fields, from finance to

healthcare, aiding in class, regression, and solving complicated choice-making troubles easily and transparency.



#### 4 XG Boost:

XGBoost, short for Extreme Gradient Boosting, is a powerful machine learning set of rules famend for its efficiency and performance in classification and regression responsibilities. It employs a gradient boosting framework, iteratively combining vulnerable beginners, normally decision trees, to shape a robust predictive model. XGBoost contains regularization techniques to save you overfitting and makes use of parallel computing to beautify velocity. Its versatility, scalability, and effectiveness in handling huge datasets make it a popular desire in numerous domain names, from finance to healthcare and past. With its capability to deliver excessive accuracy and interpretability, XGBoost stands as a cornerstone in modern-day device studying workflows.



#### D. Results

- logistic regression:

```
Accuracy is: 0.6818181818181818
      precision    recall  f1-score   support

     0       0.70      0.76      0.73         25
     1       0.65      0.58      0.61         19

 accuracy          0.68
 macro avg         0.68      0.67      0.67
 weighted avg      0.68      0.68      0.68
```

- Support Vector Machine:

```
Accuracy: 0.6363636363636364
      precision    recall  f1-score   support

     0       0.67      0.72      0.69         25
     1       0.59      0.53      0.56         19

 accuracy          0.64
 macro avg         0.63      0.62      0.62
 weighted avg      0.63      0.64      0.63
```

- Decision tree:

```
Accuracy: 0.6136363636363636
```

```
Classification Report:
      precision    recall  f1-score   support

     0       0.63      0.76      0.69         25
     1       0.57      0.42      0.48         19

 accuracy          0.61
 macro avg         0.60      0.59      0.59
 weighted avg      0.61      0.61      0.60
```

- multi-layer perceptron

```
Accuracy: 0.5909090909090909
```

```
Classification Report:
      precision    recall  f1-score   support

     0       0.61      0.76      0.68         25
     1       0.54      0.37      0.44         19

 accuracy          0.59
 macro avg         0.58      0.56      0.56
 weighted avg      0.58      0.59      0.57
```

- Light gbm:

```
Accuracy: 0.7954545454545454
```

```
Classification Report:
      precision    recall  f1-score   support

     0       0.81      0.84      0.82         25
     1       0.78      0.74      0.76         19

 accuracy          0.80
 macro avg         0.79      0.79      0.79
 weighted avg      0.79      0.80      0.79
```

#### Naive bayes

```
Accuracy: 0.6136363636363636
```

```
Classification Report:
      precision    recall  f1-score   support

     0       0.65      0.68      0.67         25
     1       0.56      0.53      0.54         19

 accuracy          0.61
 macro avg         0.60      0.60      0.60
 weighted avg      0.61      0.61      0.61
```

- xgboost:

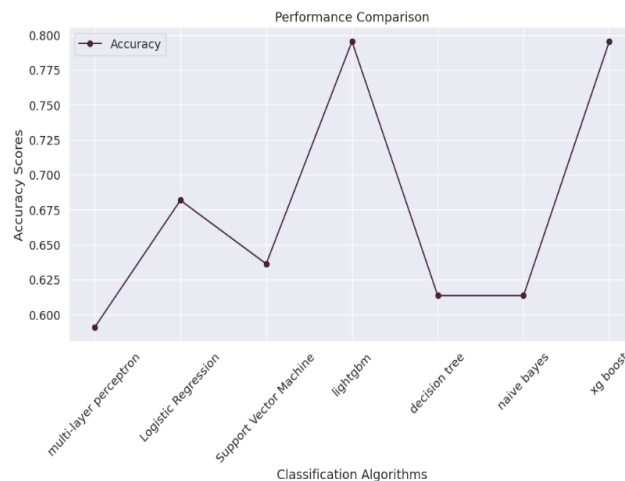
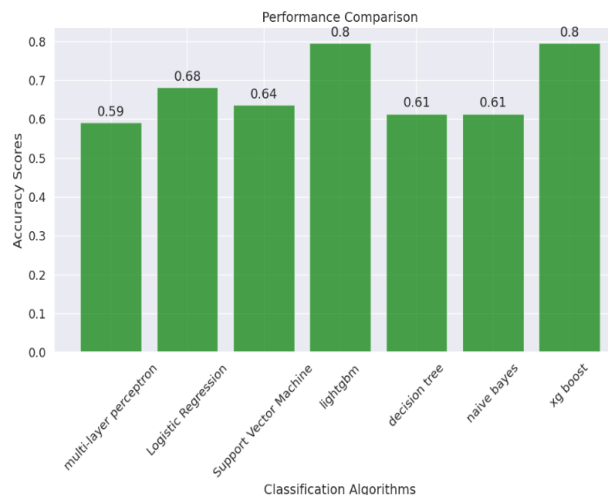
Accuracy:0.7954545454545454

Classification Report:				
	precision	recall	f1-score	support
0	0.81	0.84	0.82	2
1	0.78	0.74	0.76	1
accuracy			0.80	4
macro avg	0.79	0.79	0.79	4
weighted avg	0.79	0.80	0.79	4

Confusion Matrix

True \ Predicted	0	1
0	17	8
1	13	6

- overall graph:



## 1 CONCLUSION:

In conclusion, we tried with many algorithms to get the best mushroom classification outcomes. With regard to both photos with and without backgrounds, we used neural networks (NN), support vector machines (SVM), decision trees, and k-nearest neighbours (KNN) to a variety of scenarios. From mushroom photos, we retrieved several characteristics, such as Eigen features, histogram features, and parametric features. Our experimental findings demonstrate the benefits of using backdrop photos, particularly when using the KNN method with Eigen feature extraction and the actual mushroom dimensions. These results advance our knowledge of the variables that may affect how accurately mushrooms are classified

## REFERENCES:

- [1] M. Alameady, "Classifying Poisonous and Edible Mushrooms in the Agaricus," International Journal of Engineering Sciences & Research Technology, vol. 6, no. 1, pp. 154–164, 2017
- [2] Mohammad Ashraf Ottom1 , Noor Aldeen Alawad2 , Khalid M. O. Nahar2 "Classification of Mushroom Fungi Using Machine Learning Techniques"
- [3] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, "recaptcha: Human-based character recognition via web security measures," Science, vol. 321, no. 5895, pp. 1465–1468, 2008.
- [4] P. Kumar, V. K. Sehgal, D. S. Chauhan, and others, "A benchmark to select data mining based classification algorithms for business intelligence and decision support systems," arXiv preprint arXiv:1210.3139, 2012
- [5] Udrea, A.; Mitra, G.; Costea, D.; Noels, E.; Wakkee, M.; Siegel, D.; de Carvalho, T.; Nijsten, T. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. J. Eur. Acad. Dermatol. Venereol. 2020, 34, 648–655.



[6] Classification Algorithm For Edible Mushroom Identification – AgungWibowo, Yuri Rahayu, AndiRiyanto, TaufikHidayatulloh in 2018.

[7] Quality Classification of Enoki Mushroom Caps Based On CNN - Jinhua Dong, LixinZheng in 2019.

[8] Behavioural Features For Mushroom Classification - Shuhaida Ismail, Amy RosshaidaZainal, Aida Mustapa in 2018.

[9] Application Of Support Vector Machine To Detect Microbial Spoilage Of Mushrooms - AlirezaMasoundian and Kenneth A.McIsaac in 2013.

[10] Discrimination Of Amanita Mushrooms Using Fourier Transform Infrared Difference Spectroscopy And Cluster Analysis – You-ming Shi, Cui-qiong Yan, Dong-yu Li, Gang Liu in 2011.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.

## ● 13% Overall Similarity

Top sources found in the following databases:

- 9% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 4% Submitted Works database

### TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	<b>irjmets.com</b> Internet	7%
2	<b>He Xin-yi, Liu Jin-fu, Cheng Li-li. "Effects of drying methods on the qu..."</b> Crossref	1%
3	<b>Jinhua Dong, Lixin Zheng. "Quality Classification of Enoki Mushroom C..."</b> Crossref	<1%
4	<b>CSU, San Diego State University on 2024-04-26</b> Submitted works	<1%
5	<b>Bhumi Singh, Namrata Dhanda, Rajat Verma. "Applications of Machine ..."</b> Crossref	<1%
6	<b>pearl.plymouth.ac.uk</b> Internet	<1%
7	<b>researchgate.net</b> Internet	<1%
8	<b>University of Western Ontario on 2023-12-11</b> Submitted works	<1%

9	Daffodil International University on 2022-04-28	<1%
	Submitted works	
10	acikbilim.yok.gov.tr	<1%
	Internet	
11	community.uthm.edu.my	<1%
	Internet	