

FINAL PROJECT REPORT
ON
CLASSIFICATION OF MUSHROOM BASED ON PHYSICAL
CHARACTERISTICS USING CLASSIFICATION TECHNIQUES

By

NAME - VISHNU CHOUNDUR
U# - U00869233

Instructor

CHING CHI YANG





PORCINI
edible



LOBSTER
edible



MOREL
edible



FLY AGARIC
non-edible



YELLOW
FOOT
edible



DEATH CAP
non-edible



BLACK
TRUMPET
edible



KING
STROPHARIA
edible



CONOCYBE
FILARIS
non-edible



CHANTERELLE
edible



HONEY
MUSHROOM
edible



DESTROYING
ANGEL
non-edible



BLEWIT
edible



HEN OF
THE WOODS
edible



SHAGGY
MANE
edible

1. INTRODUCTION

Mushroom prediction is a challenging data science project that predicts poisonous or edible. However, the identification of poisonous and non-poisonous mushrooms is a challenging task. Consumption of poisonous mushrooms can lead to serious health consequences. Therefore, it is very important to identify and distinguish between the edible and non-edible mushrooms accurately.

In recent years, Machine Learning algorithms have gained popularity in various fields for their ability to perform accurate and efficient predictions. The main task of the project is to predict the model. It is a classification problem, because the target variable binary values i.e., poisonous, or edible. So, we will perform classification techniques such as Logistic Regression, Naive Bayes, Linear Discriminant Analysis (LDA) and Random Forest. Among this mode, we found the logistic Regression provides best results with accuracy and AUC.

The mushroom data set contains two datasets i.e., primary, and secondary dataset. The primary data set has a wide variety of mushroom family with its physical features. The dataset contains 20 physical features such as cap diameter, stem width, stem color, veil color, ring type, seasons etc. The primary data set contains 173 species from 23 families. Whereas the secondary dataset contains mushroom physical characteristics and poisonous or non-poisonous. The dataset has been collected from the UCI Machine Learning and consists of 61069 records.

The goal of the project is to create a strong classification model that reliably predicts the poisonous or non-poisonous mushroom based on the physical characteristics. This project can have wide applications in real-world scenarios such as Environmental monitoring system and Quality control system for mushroom, Mushroom Inventory Management System.

2. PROBLEM STATEMENT

The identification of poisonous and non-poisonous mushrooms is a critical issue in the agricultural sector. The consumption of poisonous mushrooms can lead to serious health issues. Hence, it is essential to develop an accurate and efficient method to differentiate between the edible and poisonous mushroom.

Currently, there is a lack of reliable methods to detect or identify the poisonous mushroom and existing methods are time consuming and labor intensive. Moreover, these methods rely on the subjective interpretation of human experts, which can lead to errors in identification.

3. MOTIVATION

The primary motivation behind this project is to develop an accurate and efficient method for identifying edible and non-edible mushrooms, thereby reducing the risk of having mushroom poisonous. The development of a machine learning, statistical based classification model can provide an automated and objective method for identification of mushrooms, which can significantly reduce the dependence on human experts.

Moreover, the project aims to compare the performance of different classification techniques and identify the most accurate algorithm for predicting the edibility of mushrooms. The outcome of this project can contribute to the development of improved classification models for predicting the edibility of mushrooms, which can have a significant impact on the safety and health of consumers.

Overall, the project motivation is to contribute to the agricultural industry by providing a reliable and efficient method for identifying edible and poisonous mushrooms, which can improve the safety and health of consumers and reduce the economic losses associated with mushroom poisoning.

4. CHALLENGES

The development of a machine learning-based classification model for predicting the edible mushroom presents several challenges. Some of these challenges include:

- **Limited data:** The dataset contained for this project may not have sufficient samples of certain mushroom varieties, which can affect the accuracy of the classification model. Example: Veil Type.
- **Data Preprocessing:** The dataset contains missing values, outliers or inconsistent data, which require careful preprocessing before building the classification model. Example: gill attachment, cap surface, stem surface, spore-print-color.
- **Feature Selection:** Selecting relevant features that are more significant for predicting the target variable (Class). It is the crucial step in developing an accurate and efficient model. Example: cap shape, cap diameter is most informative for predicting the class of mushroom.
- **Overfitting and Underfitting:** The classification model may overfit or underfit the training data, leading to poor performance on new data.

5. RELATED WORK

This survey of the literature can provide insights on the various approaches and strategies employed by other researchers to predict the classification of mushroom. *Omar Tarawneh*^[1] et al. (2022) states that classification techniques such as SVM, LDA, QDA etc. This paper has produced the best accuracy with Random Forest of 99.6%. But this paper doesn't provide any evidence for multicollinearity. This paper doesn't improve the multicollinearity. *Shuhaida ismail*^[2] et al (2018) states that behavioral features for mushroom classification. Additionally, this study eliminated the missing data and used PCA techniques for feature selection. PCA is a type of multivariate analysis that uses fewer, uncorrelated new variables to replace the original, correlated ones. Applied the Decision Tree as algorithm and got accuracy of 100% for full model. But, this model contains multicollinearity.

Kanchi Tank^[3] states the comparative study on mushroom classification using supervised learning algorithm. In this paper, eliminated all missing data and empty data. Almost 80% of the data has been removed from data. So, the model is developed with approx. 8000 data. Additionally, the model develops with all variables i.e., full model. So, multicollinearity will exist in the model. Because

there are few variable correlated. for example, cap diameter and stem width, veil color and veil type are pair strong correlated. This paper states that Random forest will predict the best model with accuracy of 99.95%. *Sumitra Nuanmee sri^[4]* et al states that classification of mushrooms by using the feature selection and decision tree techniques. In this paper describes that the decision tree model is the best model when compares with other classification methods such as logistic Regression.

In most of the papers, missed the multicollinearity between the variables which might affect the prediction. Almost 75% of the data is removed in the form of missing data/empty data. During this missing data treatment, most of the information was lost.

6. DESIGN OF APPROACH / METHODOLOGY

The dataset used for this project is mushroom secondary data. The dataset is taken from the UCI ML Repository (<http://archive.ics.uci.edu/ml/datasets/Secondary+Mushroom+Dataset>). The dataset contains 61069 data and contains the 21 physical features of mushroom (attribute) such as cap diameter, stem width, stem height, veil type, has ring, gill attachment etc. The data contains the missing values in some attributes such as cap surface, stem root, stem surface, veil type, veil color, gill attachment, gill spacing, ring type, spore print color. And we found that greater than 70% data missing in the stem root, veil type, spore print color. The target variable is class, and the goal is to build a model that can accurately predict the edible or poisonous mushroom based on physical characteristics of mushroom. The data contain the 17 categorical variables, 3 numerical variables and one target variable (categorical variable).

```
> summary(mmm)
class      cap.diameter  cap.shape  cap.surface  cap.color  does.bruise.or.bleed  gill.attachment
e:27181    Min.   : 0.380  b: 5694    :14120  n      :24218  f:50479          a      :12698
p:33888    1st Qu.: 3.480  c: 1815    t      : 8196  y      : 8543  t:10590          d      :10247
           Median : 5.860  f:13404   s      : 7608  w      : 7666             : 9884
           Mean   : 6.734  o: 3460   y      : 6341  g      : 4420          x      : 7413
           3rd Qu.: 8.540  p: 2598   h      : 4974  e      : 4035          p      : 6001
           Max.   :62.340  s: 7164   g      : 4724  o      : 3656          e      : 5648
           x:26934  (Other):15106  (Other): 8531             (Other): 9178

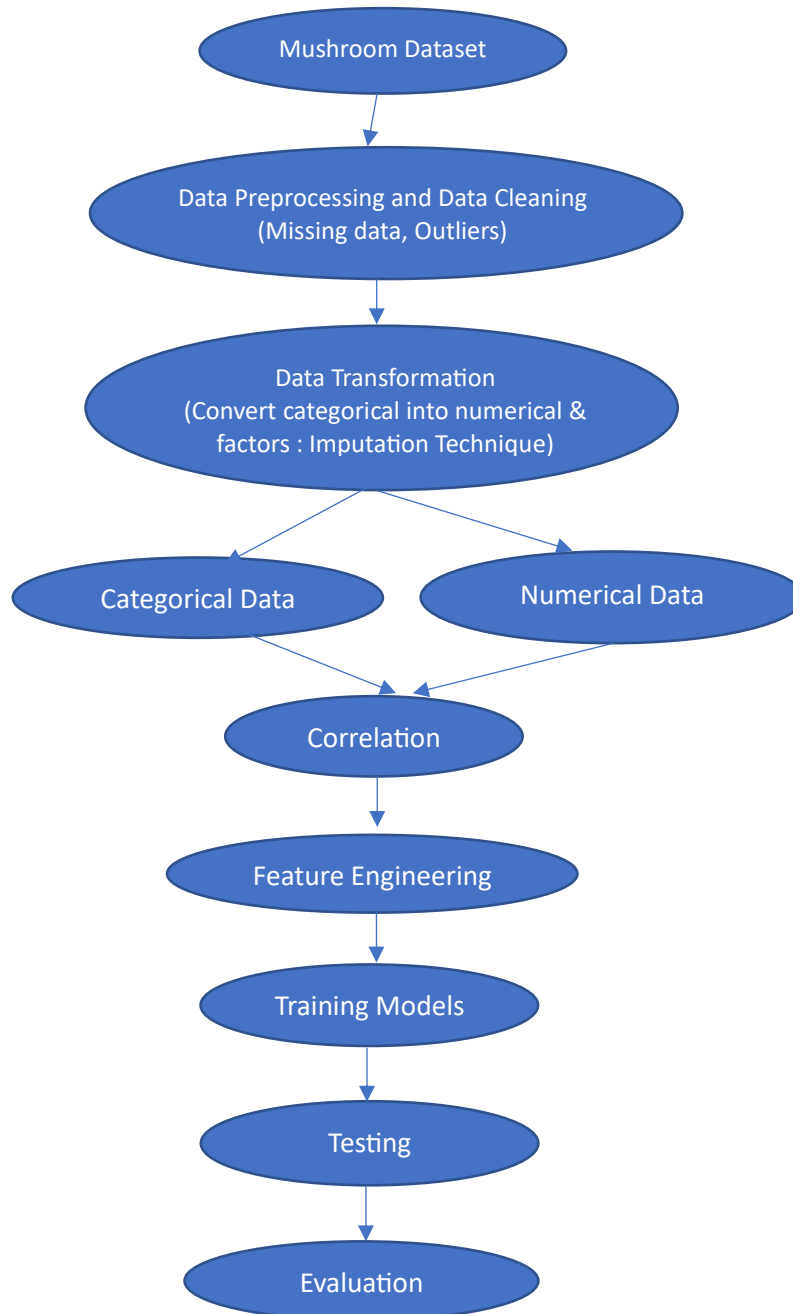
gill.spacing  gill.color  stem.height  stem.width  stem.root  stem.surface  stem.color
:25063        w      :18521  Min.   : 0.000  Min.   : 0.00  b:51538          w      :22926
c:24710        n      : 9645  1st Qu.: 4.640  1st Qu.: 5.21  c: 3177          s      : 6025
d: 7766        y      : 9546  Median : 5.950  Median : 10.19  f: 706          y      :18063
f: 3530        p      : 5983  Mean   : 6.582  Mean   : 12.15  r:1059          t      : 7865
           g      : 4118  3rd Qu.: 7.740  3rd Qu.: 16.57  s:1412          g      : 2626
           f      : 3530  Max.   :33.920  Max.   :103.91  (Other): 1765          o      : 2187
           (Other): 9726             (Other): 3175          e      : 2050
           veil.type  veil.color  has.ring  ring.type  spore.print.color  habitat  season  (Other): 5352
:57892        :53656  f:45890  f      :48361  :54715  d      :44209  a:30177
u: 3177        e: 181    t:15179  e      : 2471  k      : 2118  g      : 7943  s: 2727
           k: 353          : 2435  p      : 1259  l      : 3168  u:22898
           n: 525          : 2118  w      : 1212  m      : 2920  w: 5267
           u: 353          : 1427  n      : 1059  h      : 2001
           w: 5474          r: 1399  g      : 353  p      : 360
           y: 527          (Other): 2858  (Other): 353  (Other): 468
```

Summary of Mushroom dataset

Initially, read the data, converted all categorical into factors and perform the visualization of data. The data set contains 55.49% poisonous and 44.51% are edible. Removed all variables that contain more than 70% of missing data and only one factor in the variable such as veil type. All other variables that contain the missing data perform data manipulation and converted all categorical into numerical using multiple imputation techniques. And performed the correlation between the variables to find the most correlated variable with other variables except class and find aliasing between the variable to void the multicollinearity. Found that gill.spacingf, gill.colorf, stem.surfacef, stem.colorf is aliasing with gill.attachment, gill.spacing, stem.root etc. Removed the variables that are most correlated, missing data and aliased.

Finally, removed the stem.root, gill.attachment, stem.surface, gill spacing, veil type variables because contain most missing values and correlated with other variables. Build the Logistic Regression, Naive Bayes, Linear Discriminant Analysis, Random Forest model for this dataset. Accuracy, AUC is used as scoring metrics for model to decide the best model or not. Based on accuracy and model summary results, I am trying to find out the most suitable model for this mushroom data. Finally plotted the ROC curves and AUC for the models and found the best model. At last, conclude that Random Forest is the best model and predict most accurate output.

FLOW CHART



Data Gathering: The primary dataset describes the type family belong a mushroom belong with 20 physical characteristics, one variables as family belongs to and one classification variable (edible (e) or poisonous (p)). The secondary dataset is important data for classifying a mushroom as edible or poisonous. This dataset contains the 20 physical characteristics of mushroom and one target variable as class (e or p). Based on 173 species, this dataset has 61069 potential mushrooms with caps, with 353 mushrooms per species. Each mushroom is classified as either edible, poisonous, or possibly edible but not advised. The complete dataset description is given below image.

```
One binary class divided in edible=e and poisonous=p (with the latter one also containing mushrooms of unknown edibility).
Twenty remaining variables (n: nominal, m: metrical)
1. cap-diameter (m): float number in cm
2. cap-shape (n): bell=b, conical=c, convex=x, flat=f, sunken=s, spherical=p, others=o
3. cap-surface (n): fibrous=i, grooves=g, scaly=y, smooth=s, shiny=h, leathery=l, silky=k, sticky=t, wrinkled=w, fleshy=f
4. cap-color (n): brown=n, buff=b, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y, blue=l, orange=o, black=k
5. does-bruise-bleed (n): bruises-or-bleeding=t,no=f
6. gill-attachment (n): adnate=a, adnexed=x, decurrent=d, free=e, sinuate=s, pores=p, none=f, unknown=?
7. gill-spacing (n): close=c, distant=d, none=f
8. gill-color (n): see cap-color + none=f
9. stem-height (m): float number in cm
10. stem-width (m): float number in mm
11. stem-root (n): bulbous=b, swollen=s, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r
12. stem-surface (n): see cap-surface + none=f
13. stem-color (n): see cap-color + none=f
14. veil-type (n): partial=p, universal=u
15. veil-color (n): see cap-color + none=f
16. has-ring (n): ring=t, none=f
17. ring-type (n): cobwebby=c, evanescent=e, flaring=r, grooved=g, large=l, pendant=p, sheathing=s, zone=z, scaly=y, movable=m, none=f, unknown=?
18. spore-print-color (n): see cap color
19. habitat (n): grasses=g, leaves=l, meadows=m, paths=p, heaths=h,urban=u, waste=w, woods=d
20. season (n): spring=s, summer=u, autumn=a, winter=w
```

Data Preprocessing and Data Cleaning: Identify the missing values, outliers and remove the columns that have missing data more than 70%. The dimension of the dataset is (61069, 21) i.e., 20 are independent variables and one target variable. There is total 61069 samples data.

```
[1] "class has 0 empty values and 0 null values"
[1] "cap.diameter has 0 empty values and 0 null values"
[1] "cap.shape has 0 empty values and 0 null values"
[1] "cap.surface has 14120 empty values and 0 null values"
[1] "cap.color has 0 empty values and 0 null values"
[1] "does.bruise.or.bleed has 0 empty values and 0 null values"
[1] "gill.attachment has 9884 empty values and 0 null values"
[1] "gill.spacing has 25063 empty values and 0 null values"
[1] "gill.color has 0 empty values and 0 null values"
[1] "stem.height has 0 empty values and 0 null values"
[1] "stem.width has 0 empty values and 0 null values"
[1] "stem.root has 51538 empty values and 0 null values"
[1] "stem.surface has 38124 empty values and 0 null values"
[1] "stem.color has 0 empty values and 0 null values"
[1] "veil.type has 57892 empty values and 0 null values"
[1] "veil.color has 53656 empty values and 0 null values"
[1] "has.ring has 0 empty values and 0 null values"
[1] "ring.type has 2471 empty values and 0 null values"
[1] "spore.print.color has 54715 empty values and 0 null values"
[1] "habitat has 0 empty values and 0 null values"
[1] "season has 0 empty values and 0 null values"
```

Data Transformation: Converted the categorical variables into numerical and factors using one hot encoding and Multiple Imputation techniques. Applied Multiple Imputation technique and mice technique and package to replace the missing data with appropriate data.


```

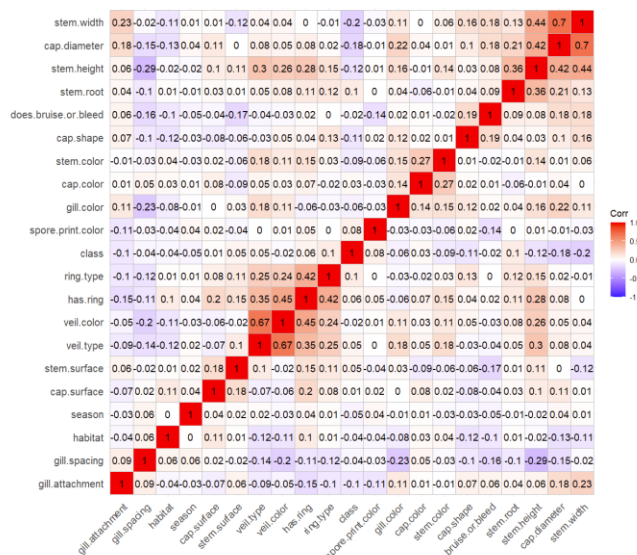
[1] "Column Name : class"
[1] "No. of Categories : 2"
[1] "Types : " "e" "p"
[1] "-----"
[1] "Column Name : cap.diameter"
[1] "No. of Categories : 0"
[1] "Types : " NULL
[1] "-----"
[1] "Column Name : cap.shape"
[1] "No. of Categories : 7"
[1] "Types : " "b" "c" "f" "o" "p" "s" "x"
[1] "-----"
[1] "Column Name : cap.surface"
[1] "No. of Categories : 12"
[1] "Types : " " " "d" "e" "g" "h" "i" "k" "l" "s" "t" "u" "y"
[1] "-----"
[1] "Column Name : cap.color"
[1] "No. of Categories : 12"
[1] "Types : " "b" "e" "g" "k" "l" "n" "o" "p" "r" "u" "w" "y"
[1] "-----"
[1] "Column Name : does.bruise.or.bleed"
[1] "No. of Categories : 2"
[1] "Types : " "f" "t"
[1] "-----"
[1] "Column Name : gill.attachment"
[1] "No. of Categories : 8"
[1] "Types : " " " "a" "d" "e" "f" "p" "s" "x"
[1] "-----"
[1] "Column Name : gill.spacing"
[1] "No. of Categories : 4"
[1] "Types : " " " "c" "d" "f"
[1] "-----"
[1] "Column Name : gill.color"
[1] "No. of Categories : 12"
[1] "Types : " "b" "e" "f" "g" "k" "n" "o" "p" "r" "u" "w" "y"
[1] "-----"
[1] "Column Name : stem.height"
[1] "No. of Categories : 0"
[1] "Types : " NULL
[1] "-----"
[1] "Column Name : stem.width"
[1] "No. of Categories : 0"
[1] "Types : " NULL
[1] "-----"

[1] "-----"
[1] "Column Name : stem.width"
[1] "No. of Categories : 0"
[1] "Types : " NULL
[1] "-----"
[1] "Column Name : stem.root"
[1] "No. of Categories : 6"
[1] "Types : " " " "b" "c" "f" "r" "s"
[1] "-----"
[1] "Column Name : stem.surface"
[1] "No. of Categories : 9"
[1] "Types : " " " "f" "g" "h" "i" "k" "s" "t" "y"
[1] "-----"
[1] "Column Name : stem.color"
[1] "No. of Categories : 13"
[1] "Types : " "b" "e" "f" "g" "k" "l" "n" "o" "p" "r" "u" "w" "y"
[1] "-----"
[1] "Column Name : veil.type"
[1] "No. of Categories : 2"
[1] "Types : " " " "u"
[1] "-----"
[1] "Column Name : veil.color"
[1] "No. of Categories : 7"
[1] "Types : " " " "e" "k" "n" "u" "w" "y"
[1] "-----"
[1] "Column Name : has.ring"
[1] "No. of Categories : 2"
[1] "Types : " "f" "t"
[1] "-----"
[1] "Column Name : ring.type"
[1] "No. of Categories : 9"
[1] "Types : " " " "e" "f" "g" "l" "m" "p" "r" "z"
[1] "-----"
[1] "Column Name : spore.print.color"
[1] "No. of Categories : 8"
[1] "Types : " " " "g" "k" "n" "p" "r" "u" "w"
[1] "-----"
[1] "Column Name : habitat"
[1] "No. of Categories : 8"
[1] "Types : " "d" "g" "h" "l" "m" "p" "u" "w"
[1] "-----"
[1] "Column Name : season"
[1] "No. of Categories : 4"
[1] "Types : " "a" "s" "u" "w"
[1] "-----"

```

Factors in each variables

Feature Engineering and Selection: As categorical variables are converted into numerical variables and find the correlation between all variables and find most significant variables with target and other variables. For example, Cap diameter is most significant with class and stem width is most significant with cap diameter. Similarly, Veil type is most significant with veil color. If two independent variables are most significant then there will exist multi collinearity. So, In order to eliminate multicollinearity, I'm removing the veil type because it has only one factor and empty data. Similarly found correlation in gill spacing, gill color, stem surface, stem color with gill attachment, gill spacing, stem root. So, In reduced model, I have removed the stem surface, veil type, gill attachment, gill spacing, stem root to eliminate multi collinearity.



Correlation between the variables

Training Model: Divided the data into two splits i.e., training data (80%) and testing data (20%). Developed classification models using training data and full data such as logistic Regression, LDA, Naïve Bayes, Random Forest and calculated the accuracy, ROC, AUC as scoring metrics for the model.

Modeling:

Logistic Regression is a popular statistical technique used for modeling binary classification problems. It models the relationship between a binary response variable and one or more predictor variables, estimating the probability that an observation belongs to a particular class. Logistic Regression assumes a linear relationship between the predictors and the logit function of the response variable, and uses maximum likelihood estimation to estimate the parameters of the model. Logistic Regression is a simple and interpretable model that can be used to predict the probability of an event occurring.

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem with the assumption of independence between the features. It is called "naive" because it assumes that all the predictors are independent of each other, which is often not the case in real-world problems. Despite its simplicity, Naive Bayes is often surprisingly accurate, and can be very fast to train and predict.

Linear Discriminant Analysis (LDA) is a classification algorithm that models the distribution of the predictor variables in each class using multivariate Gaussian distributions. It then uses Bayes' theorem to calculate the probability of an observation belonging to each class, and assigns it to the class with the highest probability. LDA assumes that the predictor variables are normally distributed and that the covariance matrices are equal across classes.

Random Forest is a powerful and widely used machine learning algorithm for classification and regression problems. It creates an ensemble of decision trees by randomly selecting a subset of features and a subset of data points for each tree, and then aggregating the predictions of the individual trees. Random Forest is robust to noise and outliers, can handle high-dimensional data, and can capture nonlinear and interactive relationships between the predictors and the response variable.

Testing: Testing is performed with testing data, training data, and whole data. Calculated the Accuracy, ROC, and AUC to evaluate performance of the model.

Evaluation: Evaluated each model based on the testing data and scoring metrics (AUC, ROC, Accuracy). Finally concluded that Random Forest is the best model with accuracy of approx. 99.9% and AUC is approximately 100%.

Accuracy: Accuracy is a commonly used evaluation metric in classification that measures how well a classification model is able to predict the correct class labels for a given set of instances. It is defined as the ratio of the number of correct predictions to the total number of predictions made by the model. In other words, accuracy is the percentage of instances for which the model predicted the correct class label.

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

AUC-ROC: AUC (Area Under the Curve) ROC (Receiver Operating Characteristic) is a popular metric used for evaluating the performance of binary classification models. The ROC curve is a graphical representation of the true positive rate (TPR) versus the false positive rate (FPR) at different classification thresholds. The AUC ROC is the area under this curve, which is a measure of the overall ability of the model to discriminate between the positive and negative classes.

An AUC ROC value of 1 indicates that the model has perfect discrimination, meaning that it can perfectly distinguish between positive and negative instances. An AUC ROC value of 0.5 indicates that the model is no better than random guessing, while a value below 0.5 indicates that the model is performing worse than random guessing. In general, higher AUC ROC values indicate better model performance.

7. RESULTS

This section explores the results of each step-by-step process in the project.

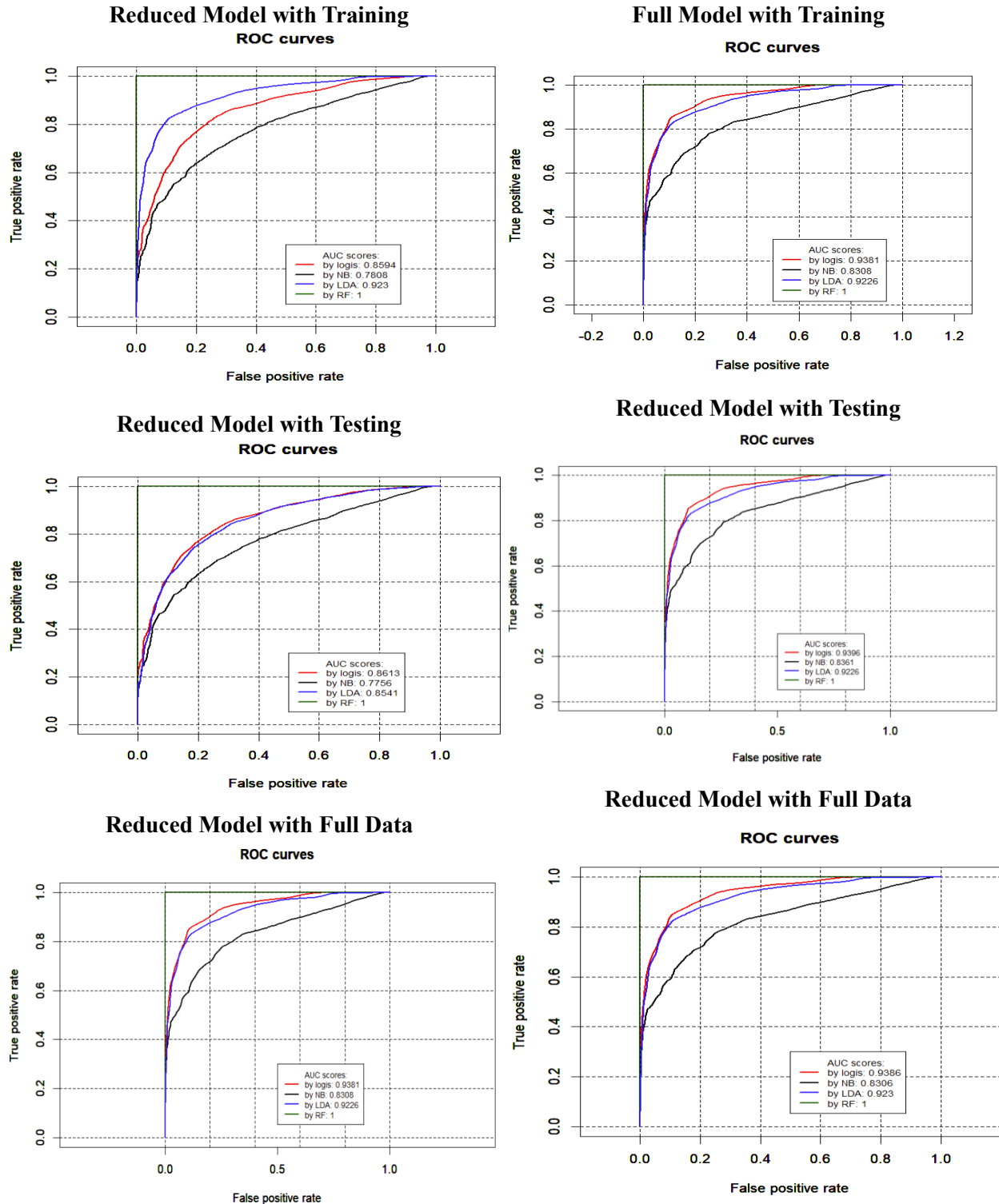
	Reduced Model Train Accuracy	Reduced Model Test Accuracy	Reduced model with full data	Full Model Train Accuracy	Full model Test accuracy	Full Model Accuracy
Logistic Regression	78.44	78.53	78.63	86.48	86.92	86.55
Naïve Bayes	68.95	69.41	68.83	75.24	75.54	75.05
LDA	77.63	77.44	77.79	84.92	85.11	84.95
Random Forest	99.99	99.97	100	100	100	100

If we observe the above result, we can say that full model will have more accuracy when compared to reduced model. But, in full model, multi collinearity arises because of two independent variables are highly correlated such as veil color and veil type are one pair that is highly correlated. Similarly, we have a few more pairs of gill color and gill spacing etc. Removed the highly correlated variable among one and that have more missing data. But, in reduced model, I have selected feature that most significant with target variable and build the model. The accuracy of reduced model is decreased in all models, but we eliminate multicollinearity. I would say that reduced model is better though accuracy is less. But, irrespective of full or reduced, Random Forest has maintained significant accuracy when compared to other classification techniques. Finally conclude that Random Forest is the better model.

If we observe the ROC curves, we can see almost all curves are above 75% except Naïve Bayes. I feel Naïve Bayes doesn't fit this problem. Because accuracy and AUC are very low when compared to other models. The Random Forest AUC curve is 100% for all training, test and full data.

Out of all models, Only Random Forest will predict the output with the most efficient and accurate output.

ROC and AUC of All Model



Out of all combinations and models, I feel Random Forest is the best model with best accuracy score and AUC.

8. CONCLUSION

In this project, we applied four popular classification algorithms, Logistic Regression, Linear Discriminant Analysis (LDA), Naïve Bayes, and Random Forest, to classify mushrooms as either poisonous or edible based on their physical characteristics. The dataset used in this project contained null values which were cleaned using multiple imputation technique.

After cleaning the data, we evaluated the performance of these algorithms using various metrics such as accuracy, AUC-ROC. Our results showed that all four algorithms were able to achieve high accuracy, with Random Forest performing the best with an accuracy of 99.9%. Logistic Regression also performed well, achieving an accuracy of 86.5%. LDA and NB achieved accuracies of 85% and 75%, respectively.

The results of this project demonstrate that the classification algorithms used in this study can be effective in classifying mushrooms as poisonous or edible based on their physical characteristics even when the dataset contains null values. The high accuracy achieved by all four algorithms suggests that they can be useful tools for identifying the toxicity of mushrooms in various applications, such as food safety and medical diagnosis.

However, it is important to note that while multiple imputation can help address the issue of null values, it may not always be the best approach. Further analysis using other techniques for handling missing values, such as deleting the rows or columns with null values or using domain knowledge to impute the missing values, may be necessary to fully evaluate the effectiveness of these algorithms in practice.

References

1. Omar Tarawneh, Monther Tarawneh, Yousef Sharrab, Moath Husni (2022). “*Mushroom classification using machine-learning techniques*”, Publication: Journal of Ambient Intelligence and Humanized Computing, 2022.
2. Shuhaida ismail, Amy Rosshaida Zainal, Aida Mustapha (2018), “*Behavioural Features for Mushroom Classification*”, (2018) DOI:10.1109/ISCAIE.2018.8405508.
3. Kanchi Tank (2021), “*A comparative study on Mushroom Classification using supervised Learning Algorithms*”, International Journal of Trend in scientific Research and Dvelopment (IJTSRD) volume 5 Issue 5, July-August 2021. ISSN: 2456-6470
4. Firdous, DrxHina. “Health Benefits Of Mushroom, Uses And Its Side Effects.” Lybrate, 2020. <https://www.lybrate.com/topic/benefits-ofmushroom-and-its-side-effects> (accessed Sep. 24, 2020).
5. Sumitra Nuanmeesri, Wongkot Sriurai, “*Development of the Edible and Poisonous Mushrooms Classification Model by using the Feature Selection and the Decision Tree Techniques*”, International Journal of Engineering and Advanced Technology (IJEAT). ISSN: 2249-8958