

MUSHROOM CLASSIFICATION USING LOGISTIC REGRESSION ALGORITHM

Submitted in partial fulfillment of the requirements for the award
of Bachelor of Engineering Degree in Electronics and Communication

Engineering

By

JAYAPUNITHA S (37130167)

JEMMY JASMIN S (37130168)



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SCHOOL OF ELECTRICAL AND ELECTRONICS

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119.

APRIL 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of JAYAPUNITHA S (36130167), JEMMY JASMIN S (36130168) who carried out the project entitled **"MUSHROOM CLASSIFICATION USING LOGISTIC REGRESSION ALGORITHM"** under my supervision from **AUGUST 2020** to **MARCH 2021**.

Internal Guide

Dr. S. POORNAPUSHPAKALA Ph.D

Head of the Department

Dr.T.RAVI,M.E.,Ph.D.,

Submitted for Viva voce Examination held on _____

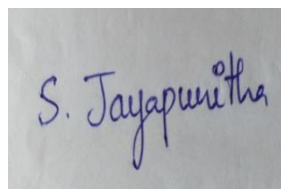
Internal Examiner

External Examiner

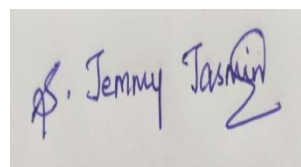
DECLARATION

We, **JAYAPUNITHA S (37130167)**, **JEMMY JASMIN S (37130168)** hereby declare that the Project Report entitled “**MUSHROOM CLASSIFICATION USING LOGISTIC REGRESSION ALGORITHM**” done by us under the guidance **Dr. S. POORNAPUSHPAKALA Ph.D** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Electronics and Communication Engineering**.

1)



2)



DATE:

PLACE:

SIGNATURE OF THE CANDIDATES

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr.N.M.NANDHITHA,M.E.,Ph.D., Dean, School of Electrical and Electronics** and **Dr.T.RAVI,M.E.,Ph.D, Head, Department of Electronics and Communication Engineering** for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our project guide **Dr.S.POORNAPUSHPAKALA Ph.D** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of our project work.

We wish to express our thanks to all teaching and Non-teaching staffmembers of the **Department of Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Mushroom is a fleshly fruiting body of fungi arising from a bunch of plant part buried in substrate. Mushroom is a crucial plant that contains a decent supply of vitamin B and an outsized quantity of macromolecule in comparison to any or all alternative vegetables. It helps to prevent cancer and will increase the immunity power of human. On the opposite hand, some mushrooms are harmful and will prove dangerous once consumed. Therefore, it's essential to spot, the edible and toxic mushrooms. This project focuses on developing a Logistic Regression Algorithmic rule for classification of mushroom exploitation its texture feature. With the Logistic regression algorithmic rule, the accuracy of the projected approach is found to be 95% that is found higher with reference to the other classifiers like KNN.

TABLE OF CONTENTS

CHAPTER NUMBER	TITLE	PAGE NUMBER
1	INTRODUCTION	8
2	LITERATURE SURVEY	11
3	AIM OF THE PROJECT	15
3.1	DESIGN AND METHODOLOGY	16
3.1.1	HARDWARE REQUIREMENTS	16
3.1.2	SOFTWARE REQUIREMENTS	16
4	PROPOSED SYSTEM	17
4.1	DESCRIPTION OF THE PROPOSED WORK	17
4.2	BLOCK DIAGRAM	17
4.3	IMPLEMENTATION ALGORITHM	18
4.4	LOGISTIC REGRESSION ALGORITHM	19
4.5	DATASET	19
5	RESULT AND DISCUSSION	20
6	CONCLUSION	26

LIST OF FIGURES

FIGURE NUMBER	FIGURE NAME	PAGE NUMBER
4.1	BLOCK DIAGRAM	17
5.1	NUMBER OF POISONOUS AND EDIBLE MUSHROOM	20
5.2	FACTOR PLOT FOR VARIOUS VALUES OF GILL COLOR	21
5.3	CORRELATION MATRIX	23
5.4	CONFUSION MATRIX FOR LOGISTIC REGRESSION CLASSIFIER	24
5.5	ACCURACY PREDICTION	25

CHAPTER 1

INTRODUCTION

Mushroom, the conspicuous umbrella-shaped plant organ (sporophore) of basidiomycetes, generally of the Agaricales within the phylum subdivision Basidiomycotina, however additionally of other teams. Popularly, the term mushroom is employed to spot the edible sporophores; the term agaric is usually reserved for poisonous or toxic sporophores. However, no scientific distinction between the 2 names is properly applied to any fleshy fruiting body. In a very restricted sense, mushroom indicates the common edible fruiting body of fields and meadows (*Agaricus campestris*). A awfully closely connected species, *A. bisporus*, is that the mushroom mature commercially and seen in markets. Umbrella-shaped sporophores species found mainly within the agaric family (Agaricaceae), members of that bear skinny, unsubdivided gills on the face of the cap from that the spores are shed. The reproductive structure of agaric consists of a cap (pileus) and a stalk (stipe). The reproductive structure emerges from an intensive underground network of threadlike strands (mycelium). Example of agaric is *Armillariella mellea* (*Armillariella*). Mushroom mycelia might live many years or die in a very few months, looking on the accessible food offer. As long as nourishment is offered and temperature and wetness are appropriate, a plant part can turn out a brand new crop of sporophores annually throughout its mature season. Mature bodies of some mushrooms occur in arcs or rings known as fairy rings. The plant part starts from a reproductive structure falling in a very favourable spot and manufacturing strands (hyphae) that grow up in all directions, eventually forming a circular mat of underground hyphal threads. Mature bodies, created close to the edge of this mat, might widen the ring for many years. Large mushrooms belong to the order Boletales, that bear pores on the underside of the cap. Among these are the hydnums or hedgehog mushrooms, that have teeth, spines, or warts on the face of the cap (e.g., *Dentinum repandum*, *Hydnum imbricatum*) or at the ends of branches (e.g., *H. coralloides*, *Hericium caput-ursi*). The polypores, shelf fungi, or bracket fungi (order Polyporales) have tubes beneath the cap as within the boletes, however they're not in agaric simply dissociable

layer. Polypores typically grow on living or dead trees, typically as harmful pests. Several of them renew growth annually and therefore turn out annual growth layers by that their age is calculatable. Examples include dryad's saddle (*Polyporus squamosus*), the *Fistulina hepatica* (*Fistulina hepatica*), the sulfur flora (*P. sulphureus*), the artist's flora (*Ganoderma applanatum*, or *Fomes applanatus*), and species of the genus *Trametes*. The clavarias, or club fungi (e.g., *Clavaria*, *Ramaria*), are shrublike, clublike, or coral-like in growth habit. One coral fungus, the cauliflower flora (*Sparassis crispa*), has planar clustered branches that lie close, giving the look of the vegetable cauliflower. The cantharelloid fungi (*Cantharellus* and its relatives) are club-, cone-, or trumpet-shaped mushroomlike forms with associated enlarged high bearing coarsely collapsed ridges on the side and dropping on the stalk. Examples embrace the extremely prized edible *Cantharellus cibarius* (*C. cibarius*) and therefore the horn-of-plenty mushroom (*Craterellus cornucopioides*). Puffballs (family *Lycoperdaceae*), stinkhorns, earthstars (a reasonably puffball), and bird's nest fungi are typically treated with the mushrooms. The morels (*Morchella*, *Verpa*) and false morels or orchels (*Gyromitra*, *Helvella*) of the phylum *Ascomycotina* are popularly enclosed with actuality mushrooms due to their form and fleshy structure; they tally a deeply collapsed or honeycombed conical sponge at the highest of a hollow stem. Some are among the foremost extremely prized edible fungi (e.g., *Morchella esculenta*). Another cluster of ascomycetes includes the cup fungi, with a concave or dishlike mature structure, typically extremely colored. different uncommon forms, not closely associated with actuality mushrooms however usually enclosed with them, are the jelly fungi (*Tremella* species), the jelly fungus or Jew's ear (*Auricularia auricular-judae*), and therefore the edible truffle. Mushrooms are freed from cholesterol and contain little amounts of essential amino acids and B vitamins. However, their chief price is as a specialty food of delicate, refined flavour and agreeable texture. By recent weight, the common commercially mature mushroom is quite ninety percent water, but three p.c supermolecule, but five p.c macromolecule, but one p.c fat, and regarding one p.c mineral salts and vitamins.

Poisoning by wild mushrooms is common and should be fatal or turn out

simply gentle duct disturbance or slight aversion. It's necessary that each mushroom supposed for uptake be accurately known (see mushroom poisoning).

CHAPTER 2

LITERATURE SURVEY

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. Currently, the square measure 3 comparisons of the simplest classification algorithms in data processing, namely: call Tree (C4.5), NaïveBayes and Support Vector Machine (SVM). The study methodology used is experiment with power-assisted tool of maorihen that has been testing within the comparison of the 3 algorithms. To conduct the testing, it use the mushroom knowledge of Agaricus and agaric family. The mushroom study were taken from the Audubon Society guidebook to North yank Mushrooms, in UCI machine learning repository. Results of the testing indicate that the C4.5 algorithmic rule has a similar accuracy level to the SVM by 100% but from the speed facet method of the C4.5 algorithmic rule is quicker than the SVM. Comparative classification algorithmic rule testing accuracy in previous data processing has not been done and supported the results of testing of the 3 best classification algorithms within the data processing. The C4.5 algorithmic rule has the very best accuracy compared to the other two fashionable classification algorithms, and in terms of process speed.

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. The existing preprocessing approaches and network models supported convolutional neural network are improved and fine tuned to understand the popularity of enoki mushroom caps. Experimental results demonstrate that CNN-driven classification application has higher recognition rate for enoki mushroom caps, that provides a crucial reference for the appliance of enoki mushrooms in agricultural automation production and helps to optimize yield and increase productivity. In this paper, we have a tendency to planned a deep learning based mostly multi species enoki mushroom image classification rule. First, a feature extraction framework supported deep convolutional neural network is made and is

optimized by zero low-level formatting coaching and fine-tuning. We have a tendency to believe that through the classification of enoki mushroom, we will improve the assembly potency of the merchandise and scale back the price, creating the agricultural production additional economical.

.

Author: Shuhaida 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. The Principal element Analysis (PCA) algorithmic program is employed for choosing the simplest options for the classification experiment victimization call Tree (DT) algorithmic program. The classification accuracy, constant metric, and time taken to make a classification model on a regular Mushroom dataset were measured. The activity feature of 'odour' was designated because the highest hierarchal feature that contribute to the high classification accuracy. This paper conferred the methodology and results for mushroom classification experiment supported their activity options like characteristics, population, and environment

Author: AlirezaMasoudian and Kenneth A.Mclsaac.(2013)

Mushrooms should be classified as healthy or unhealthy to make sure correct handling and maximize crop yield. To unravel the matter of identification, a quick and non-destructive method, Support Vector Machine (SVM), is applied to boost the popularity accuracy and potency of the golem. Initially, a median filter is applied to get rid of the inherent noise within the coloured image. Our preliminary results for mushroom classification square measure promising and therefore the experiments distributed on the info set highlight quicker computation time and a better rate of accuracy, reaching over nineteenth mistreatment this method, which might be used in real world state of affairs. Our experimental result on the each coaching and testing information set shows that our projected approach performs well mistreatment SVMs kernel functions and SIFT feature methodology. The SVMs classification methodology was the framework for this approach but different classification algorithms like the Adabooss classifier may be studied for accuracy enhancements. To reason the SIFT options and visual vocabularies the VLFeat open supply library was used inMatlab. but to implement this algorithmic rule for mushroom harvest golem the OpenCV library with

C++ are used as OpenCV runs quicker than Matlab.

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) distinction qualitative analysis combined with gradable cluster analysis was accustomed study the species of Amanita. The results show that their spectra were terribly similar. To reveal the differentiation of those spectra, FTIR distinction spectra were obtained by subtracting the spectrum of Amanita hemibaphavar. The results show that the spectra of various species of Amanita were similar. FTIR distinction spectra of various species exhibit evident variations within the region between 1800 and 1300 cm^{-1} , that was accustomed do gradable cluster analysis. All samples were divided into 3 clusters in keeping with the similarity or unsimilarity of the distinction spectra, each cluster is pictured one species. only one sample was known erroneously. This study has incontestible the flexibility of FTIR qualitative analysis together with gradable cluster analysis to discriminate Amanita mushrooms at species level.

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. Dehydration mushroom dried by explosion puffing drying exhibited higher color values, texture crisp, rehydration capability, lower apparent density than those of hot air drying. Dehydration mushroom dried by explosion puffing drying showed the additional engaging business production technology. In these 3 drying ways, dehydration mushroom dried by phase drying exhibits the minimum volume change, getting ready to the recent mushroom, the better color values and rehydration capability, however, it prices additional and consumes massive amounts of energy. Compared with freeze drying, the standard of product processed by explosion puffing drying is analogous. And this method exhibited higher color values and texture crisp.

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. In future, completed system are going to be developed and can be utilized by individual or organization committed oyster agaric production. We hope, this technique are going to be associate degree alternative for user to get data on oyster agaric diseases, symptoms and treatment.

CHAPTER 3

AIM OF THE PROJECT

This project represent a search and development of an application on the classification on Mushrooms as edible and toxic by supplying Regression rule in agricultural domain. Mushroom may be a fleshy plant organ of some fungi arising from a gaggle of plant part buried in stratum. Most of the mushrooms belong to the Sub- Division: Basidiomycota and many belong to subdivision Ascomycota of Kingdom-Fungi. It is reported that there are about 50,000 known species of fungi and about 10,000 are considered as edible ones. Of which, about one hundred and eighty mushrooms can be tried for artificial cultivation and seventy are widely accepted as food. With the wide accessibility of subtle applied math computer code for prime speed computers, the employment of supplying regression is increasing.

3.1 DESIGN AND METHODOLOGY

3.1.1 HARDWARE REQUIREMENTS

- RAM : 8GB
- Processor : 2.4GHz
- Main Memory : 8GB RAM
- Hard Disk Drive : 1tb
- Keyboard : 104 keys

3.1.2 SOFTWARE REQUIREMENTS

- Front end : Python
- Dataset : csv
- IDE : Anaconda
- Operating System : Windows 10

CHAPTER 4

PROPOSED SYSTEM

4.1 DESCRIPTION OF THE PROPOSED WORK

In this project we classified the mushroom as edible and poisonous one. We used Logistic Regression Algorithm to classify the eatable and poisonous mushrooms based on the gill colour.

It is the binary way of classifying the mushrooms using Logistic Regression Algorithm. The main advantage of Logistic Regression Algorithm is easier to implement, interpret and very efficient to train.

4.2 BLOCK DIAGRAM

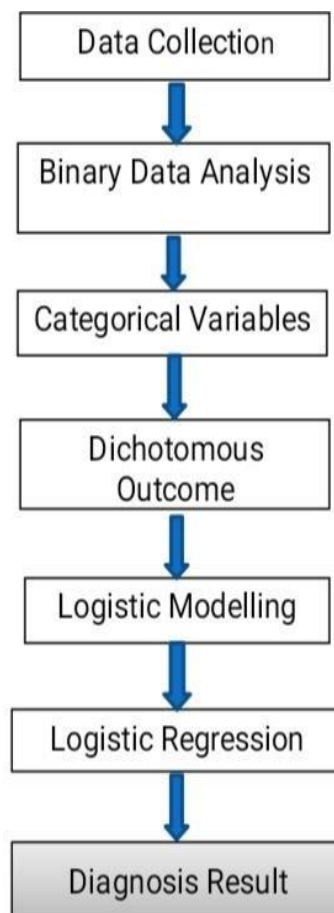


Fig 4.1 BLOCK DIAGRAM

4.3 IMPLEMENTATION ALGORITHM

- Reading the csv file of the dataset.
- Pandas `read_csv()` function imports a CSV file (in our case, 'mushrooms.csv') to DataFrame format.
- Examining the Data.
- After importing the data, to learn more about the dataset, we'll use `.head()` `.info()` and `.describe()` methods.
- Data Manipulation
- The data is categorical so we'll use LabelEncoder to convert it to ordinal. LabelEncoder converts each value in a column to a number.
- This approach requires the category column to be of 'category' datatype.
- By default, a non-numerical column is of 'object' datatype.
- From the `df.describe()` method, we saw that our columns are of 'object' datatype. So we will have to change the type to 'category' before using this approach.
- Let's look at the correlation between the variables.
- Usually, the least correlating variable is the most important one for classification. In this case, "gill-color" has -0.53 so let's look at it closely.
- Preparing the Data.
- Setting x and y axis and splitting the data into train and test respectively.
- Feature importance.
- By all methods examined before the feature that is most important is "gill-color".
- Predicting some of the `x_test` results and matching it with true i.e. `y_test` values using logistic regression.

4.4 LOGISTIC REGRESSION ALGORITHM

- The central mathematical concept that underlies Logistic regression is the logit—the natural logarithm of an odds ratio.
- Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability.
- The hypothesis of logistic regression tends to limit the cost function between 0 and 1.
- TYPES OF LOGISTIC REGRESSION:
 - 1.Binary (eg. Tumor Malignant or Benign)
 - 2.Multi-linear functions failsClass (eg. Cats, dogs)
- The simple logistic model has the form,
$$\text{thelogit}(x_1, x_2) = b_0 + b_1x_1 + b_2x_2.$$
- The probabilities $(x_1, x_2) = 1 / (1 + \exp(-f(x_1, x_2)))$.

4.5 DATASET

The dataset used in this project is mushrooms.csv contains 8124 instances of mushrooms with 23 features like cap-shape, cap-surface, cap-color, bruises, odor, etc. We have taken the dataset from KAGGLE NOTEBOOK.

The python libraries and packages we used in this project

- NumPy
- Pandas
- Seaborn
- Matplotlib
- Scikit-learn

We use the specifications like cap shape, cap color, gill color, etc. to classify the mushrooms into edible and poisonous.

CHAPTER 5

RESULT AND DISCUSSION

The following are the graphs that are obtained after training and testing the dataset.

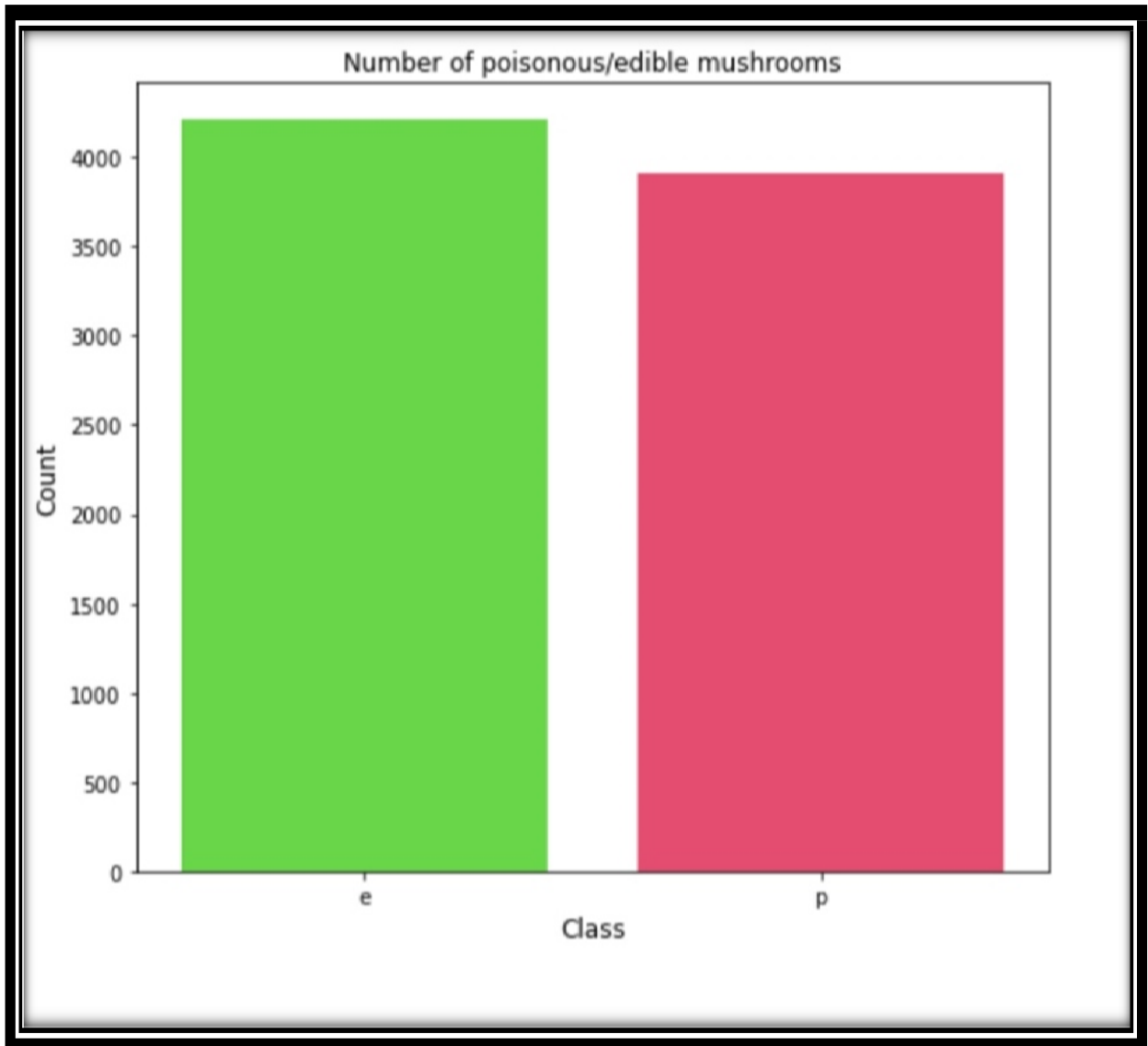


Fig 5.1 NUMBER OF POISONOUS AND EDIBLE MUSHROOM

The above graph clearly shows the numbers of poisonous mushrooms and edible mushrooms that are present in the 8124 samples.

Factor plot for various values of gill colour

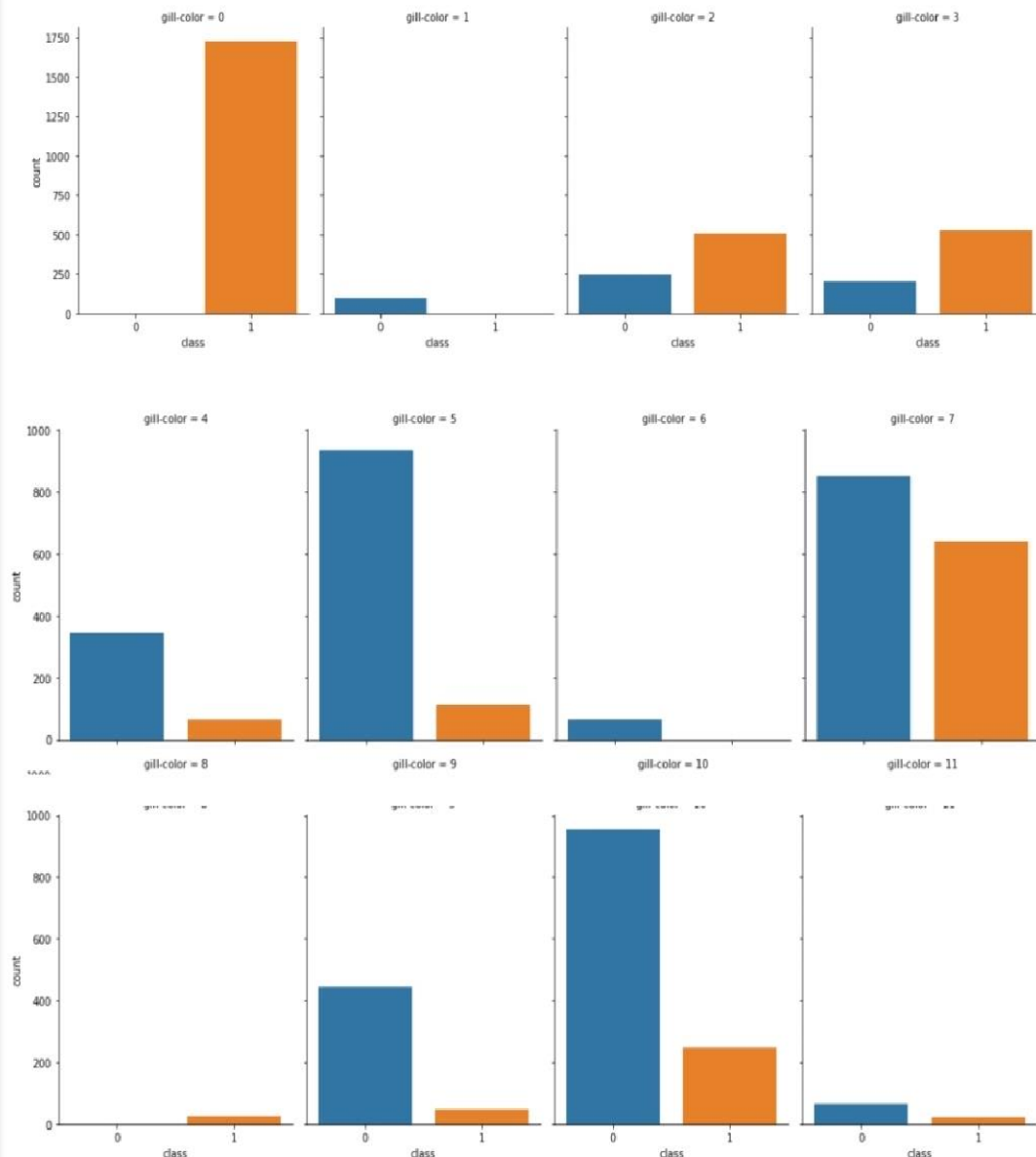


Fig 5.2 FACTOR PLOT FOR VARIOUS VALUES OF GILL COLOR

For each gill color the factor plot of graph are plotted for poisonous and edible mushrooms. The gill color include,

- Gill color = 0 = Orange

- Gill color = 1 = Green
- Gill color = 2 = Chocolate
- Gill color = 3 = Red
- Gill color = 4 = Black
- Gill color = 5 = Brown
- Gill color = 6 = Pink
- Gill color = 7 = White
- Gill color = 8 = Grey
- Gill color = 9 = Yellow
- Gill color = 10 = Buff
- Gill color = 11 = Purple

The plot shows that we could differentiate edible mushroom and poisonous mushroom well with Gill color.

Gill color 7 (White) shows that it could differentiate between edible and poisonous very well.

Correlation matrix

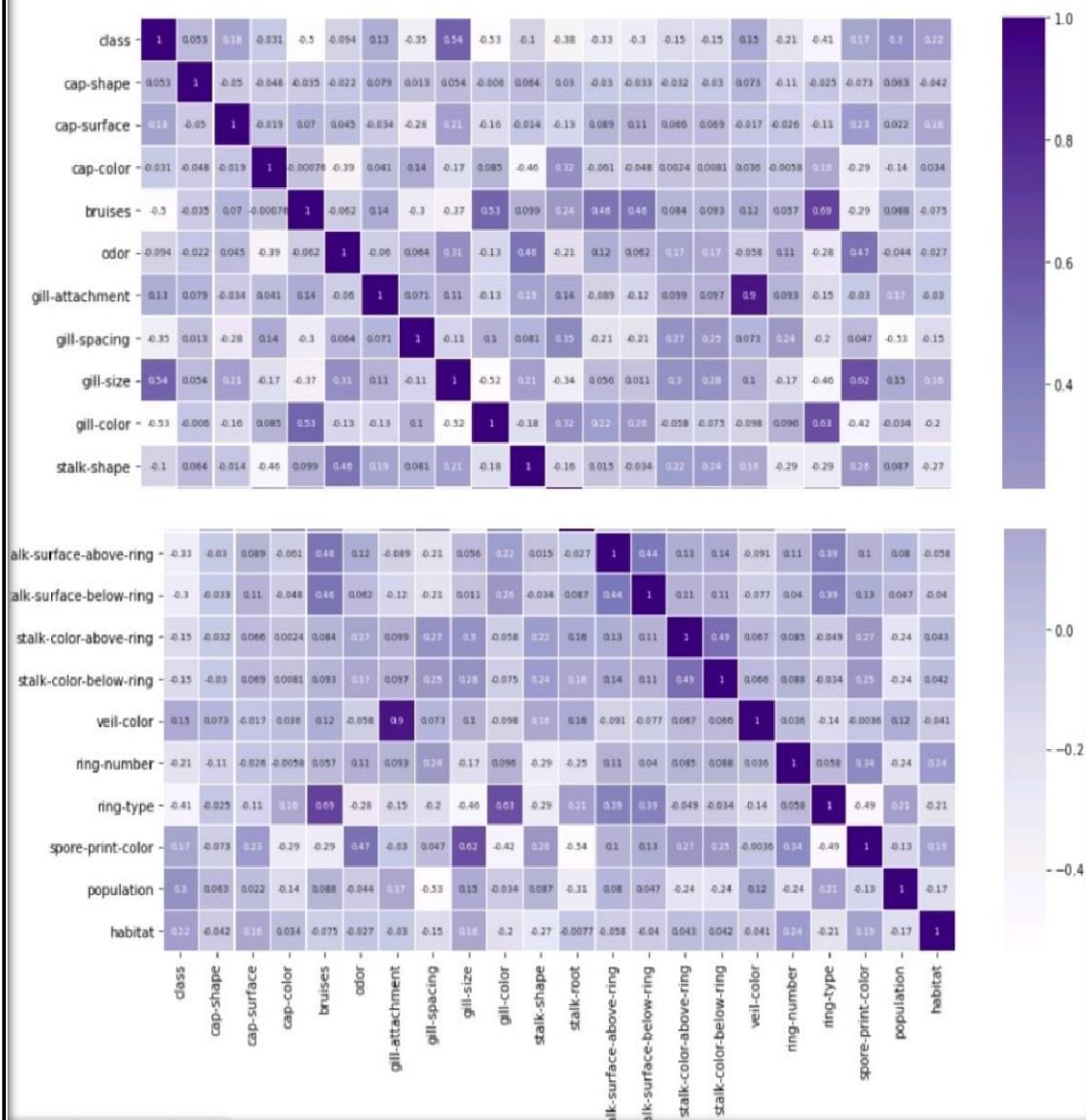


Fig 5.3 CORRELATION MATRIX

A correlation matrix is a table showing correlation coefficients between the variables. Each cell in the table will show the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis.

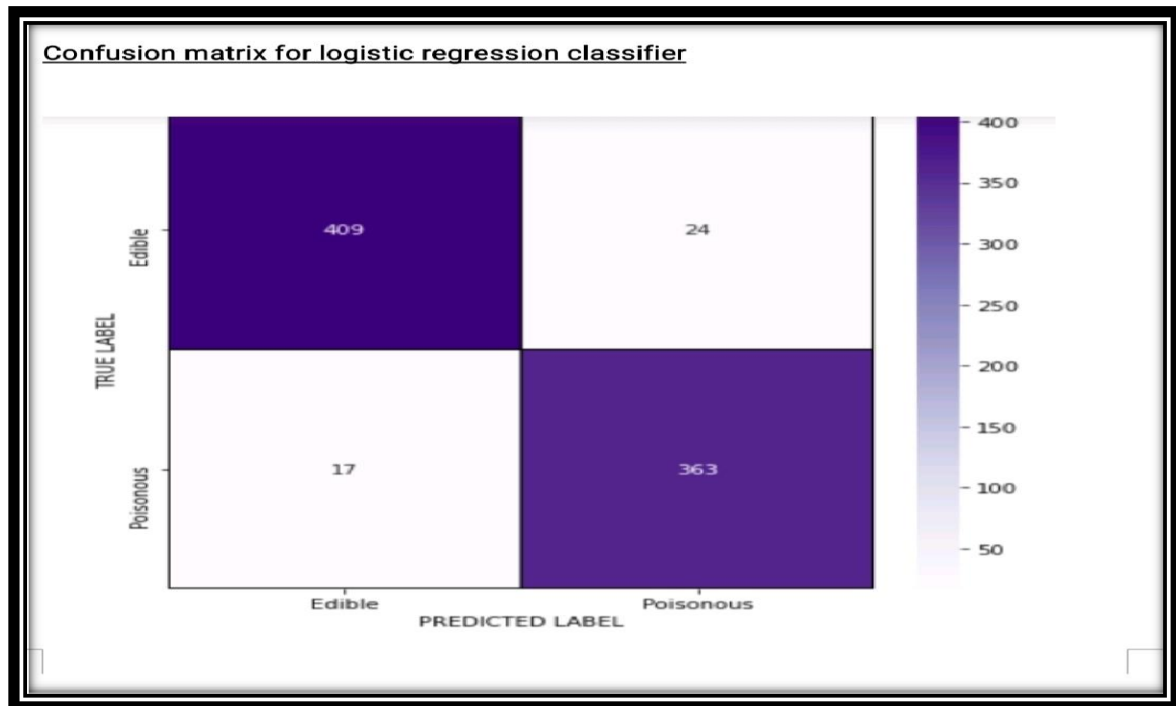


Fig 5.4 CONFUSION MATRIX FOR LOGISTIC REGRESSION CLASSIFIER

A confusion matrix is a table that is often used to find the performance of a classifier on a set of test data for which the true values are known.

The most basic terms include,

- True Positive (TP): These are cases in which we predict the edible mushrooms and they are not poisonous.
- True Negative (TN): We predict edible but they are poisonous mushrooms.
- False Positive (FP): We predict poisonous but they are edible mushrooms.
- False Negative (FN): These are cases in which we predict the poisonous mushrooms and they are not eatable.

Classification report for logistic regression classifier

Test Accuracy: 94.96%

```
)]: y_pred_lr = lr.predict(X_test)
print("Logistic Regression Classifier report: \n\n", classification_report(y_test, y_pred_lr))
```

Logistic Regression Classifier report:

	precision	recall	f1-score	support
0	0.96	0.94	0.95	433
1	0.94	0.96	0.95	380
accuracy			0.95	813
macro avg	0.95	0.95	0.95	813
weighted avg	0.95	0.95	0.95	813

Fig 5.5 ACCURACY PREDICTION

$$\text{Accuracy\%} = (\text{TP} + \text{TN}) / \text{Total} = 0.95 * 100 = 95\%$$

CHAPTER 6

CONCLUSION

In this project, we proposed a Logistic Regression Algorithm for Classification of Mushroom as Edible and Poisonous. From the confusion matrix, we can say that our train and test data is balanced. The proposed Logistic Regression Algorithm, using dataset can classify edible and poisonous mushroom and it achieved 95% accuracy with this dataset.






REFERENCES

1. Classification Algorithm For Edible Mushroom Identification – AgungWibowo, Yuri Rahayu, AndiRiyanto, TaufikHidayatulloh in 2018.
2. Quality Classification of Enoki Mushroom Caps Based On CNN - Jinhua Dong, LixinZheng in 2019.
3. Behavioural Features For Mushroom Classification - Shuhaida Ismail, Amy RosshaidaZainal, Aida Mustapa in 2018.
4. Application Of Support Vector Machine To Detect Microbial Spoilage Of Mushrooms - AlirezaMasoundian and Kenneth A.Mclsaac in 2013.
5. 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.

Document Information

Analyzed document	PR_15_4_21_to check_plg.docx (D101742857)
Submitted	4/15/2021 6:20:00 PM
Submitted by	S.Poornapushpakala
Submitter email	poornapushpakala.enc@sathyabama.ac.in
Similarity	4%
Analysis address	poornapushpakala.enc.sathya@analysis.urkund.com

Sources included in the report

W	URL: http://ohkubo-denki.co.jp/1-bhp-mrsts/atwgu---mushroom-classification-using-machin ... Fetched: 11/17/2020 11:52:07 AM		2
SA	REPORT-converted.docx Document REPORT-converted.docx (D58341331)		1
SA	Sathyabama Institute Of Science And Technology / Copy of Fully Completed Report (less).docx Document Copy of Fully Completed Report (less).docx (D101555440) Submitted by: narmadha.enc@sathyabama.ac.in Receiver: narmadha.enc.sathya@analysis.urkund.com		1
SA	ORANGE FRUIT thesis _rupali.docx Document ORANGE FRUIT thesis _rupali.docx (D75964534)		1
SA	Sudheer Report.docx Document Sudheer Report.docx (D91235331)		5

