Project Report

On

Binary Mushroom Classifier

Submitted in partial fulfilment of the requirements for the award of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

(Artificial Intelligence & Machine Learning)

by

Ms. V. Harshitha - 22WH1A6602
Ms. N. Jijnasa - 22WH1A6635
Ms. N. Prasanna - 22WH1A6643
Ms. M. Siri Chandana - 22WH1A6661

Under the esteemed guidance of

Ms. A Naga Kalyani

Assistant Professor, CSE(AI&ML)

BVRIT HYDERABAD College of Engineering for Women

(UGC Autonomous Institution | Approved by AICTE | Affiliated to JNTUH)

(NAAC Accredited - A Grade | NBA Accredited B.Tech. (EEE, ECE, CSE and IT)

Bachupally, Hyderabad – 500090

2024-25

Department of Computer Science & Engineering

(Artificial Intelligence & Machine Learning)

BVRIT HYDERABAD COLLEGE OF ENGINEERING FOR WOMEN

(Approved by AICTE, New Delhi and Affiliated to JNTUH, Hyderabad)

Accredited by NBA and NAAC with A Grade

Bachupally, Hyderabad – 500090

2024-25

CERTIFICATE

This is to certify that the major project entitled **"Binary Mushroom Classifier"** is a bonafide work carried out **by Ms. V.Harshitha(22wh1a6602), Ms. N. Jijnasa(22wh1a6635), Ms. N. Prasanna(22wh1a6643), Ms. M. Siri Chandana (22wh1a6661)** in partial fulfillment for the award of B. Tech degree in **Computer Science & Engineering (AI&ML), BVRIT HYDERABAD College of Engineering for Women, Bachupally, Hyderabad,** affiliated to Jawaharlal Nehru Technological University Hyderabad, Hyderabad under my guidance and supervision. The results embodied in the project work have not been submitted to any other University or Institute for the award of any degree or diploma.

Supervisor
Ms. A Naga Kalyani
Assistant Professor

Dept of CSE(AI&ML)

Head of the Department
Dr. B. Lakshmi Praveena
HOD & Professor

Dept of CSE(AI&ML)

DECLARATION

We hereby declare that the work presented in this project entitled **“”** submitted towards completion of Project work in IV Year of B.Tech of CSE(AI&ML) at **BVRIT HYDERABADCollege of Engineering for Women,** Hyderabad is an authentic record of our original work carried out under the guidance of **Ms. A Naga Kalyani, AssistantProfessor, Department of CSE(AI&ML).**

Sign with Date:

V.Harshitha

(22wh1a6602)

Sign with Date:

N.Jijnasa

(22wh1a6635)

Sign with Date:

N.Prasanna

(22wh1a6643)

Sign with Date:

M.Siri Chandana

(22wh1a6661)

ACKNOWLEDGEMENT

We would like to express our sincere thanks to **Dr. K. V. N. Sunitha, Principal, BVRIT HYDERABAD College of Engineering for Women**, for her support by providing the working facilities in the college.

Our sincere thanks and gratitude to **Dr. B. Lakshmi Praveena, Head of the Department, Department of CSE(AI&ML), BVRIT HYDERABAD College of Engineering for Women,** for all timely support and valuable suggestions during the period of our project.

We are extremely thankful to our Internal Guide**, Ms. A Naga Kalyani, Assistant Professor, CSE(AI&ML), BVRIT HYDERABAD College of Engineering for Women,** for her constant guidance and encouragement throughout the project.

Finally, we would like to thank our Major Project Coordinator, all Faculty and Staff of CSE(AI&ML) department who helped us directly or indirectly. Last but not least, we wish to acknowledge our **Parents and Friends** for giving moral strength and constant encouragement.

V.Harshitha (22wh1a6602)

N.Jijnasa (22wh1a6635)

N.Prasanna(22wh1a6643)

M.Siri Chandana (22wh1a6661)

ABSTRACT

The project aims to classify mushrooms as edible or poisonous using machine learning. A Random Forest Classifier is employed for its robustness and accuracy. The dataset is preprocessed with feature encoding, scaling, and handling missing values. Exploratory Data Analysis is conducted to uncover patterns and relationships. Various models are compared to identify the best-performing approach. Feature importance analysis highlights key factors influencing classification. Performance evaluation is carried out using standard classification metrics. The model is optimized to ensure reliability and scalability. This project contributes to improving food safety and preventing mushroom poisoning.

PROBLEM STATEMENT

Mushroom poisoning poses significant risks to human health and safety, making the accurate identification of edible and poisonous mushrooms critical. Traditional methods for mushroom classification rely on expert knowledge, which can be subjective and error-prone.

This project seeks to address the challenge of classifying mushrooms with high accuracy by employing a machine learning-based approach. The primary goals include:

1. Developing a classification model using key mushroom features to differentiate between edible and poisonous varieties.
2. Performing exploratory data analysis to uncover important patterns and relationships in the dataset.
3. Evaluating the effectiveness of the model through metrics like confusion matrices and classification reports, ensuring its reliability and scalability.

The ultimate objective is to create a robust and scalable solution that enhances food safety and minimizes the risks of mushroom poisoning.

DATA SET

Mushroom Dataset -Kaggle

https://www.kaggle.com/datasets/prishasawhney/mushroom-dataset

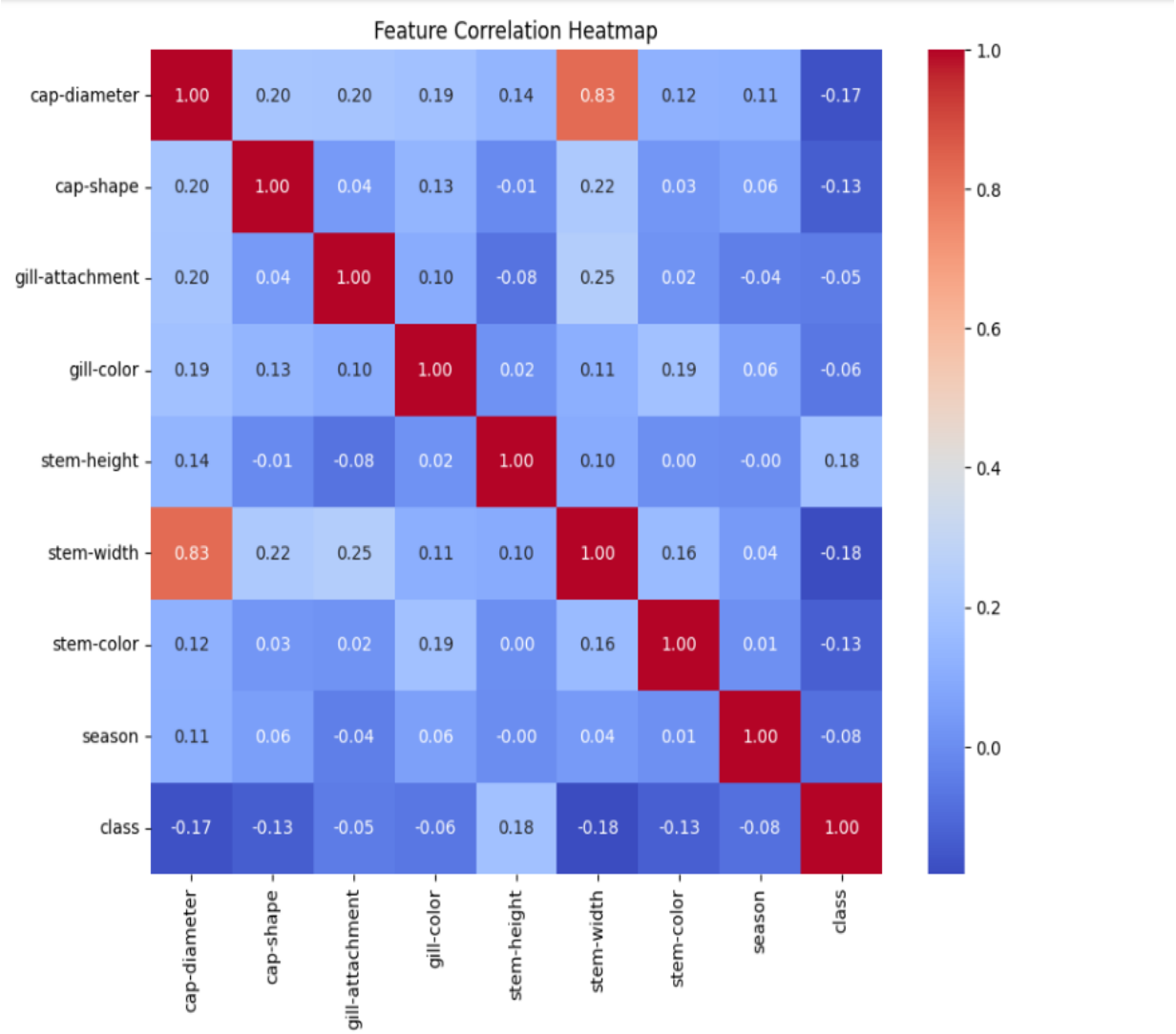
SOURCE CODE

```
# Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection
import train_test_split
from sklearn.ensemble import
RandomForestClassifier
from sklearn.metrics import
classification_report,
confusion_matrix,
accuracy_score

# Load the dataset
data =
pd.read_csv('mushroom_clean
ed.csv')  # Replace with your
dataset file path

# Data Preprocessing
# Encode categorical features
data_encoded =
data.apply(lambda col:
pd.Categorical(col).codes if
col.dtype == 'object' else col)
```

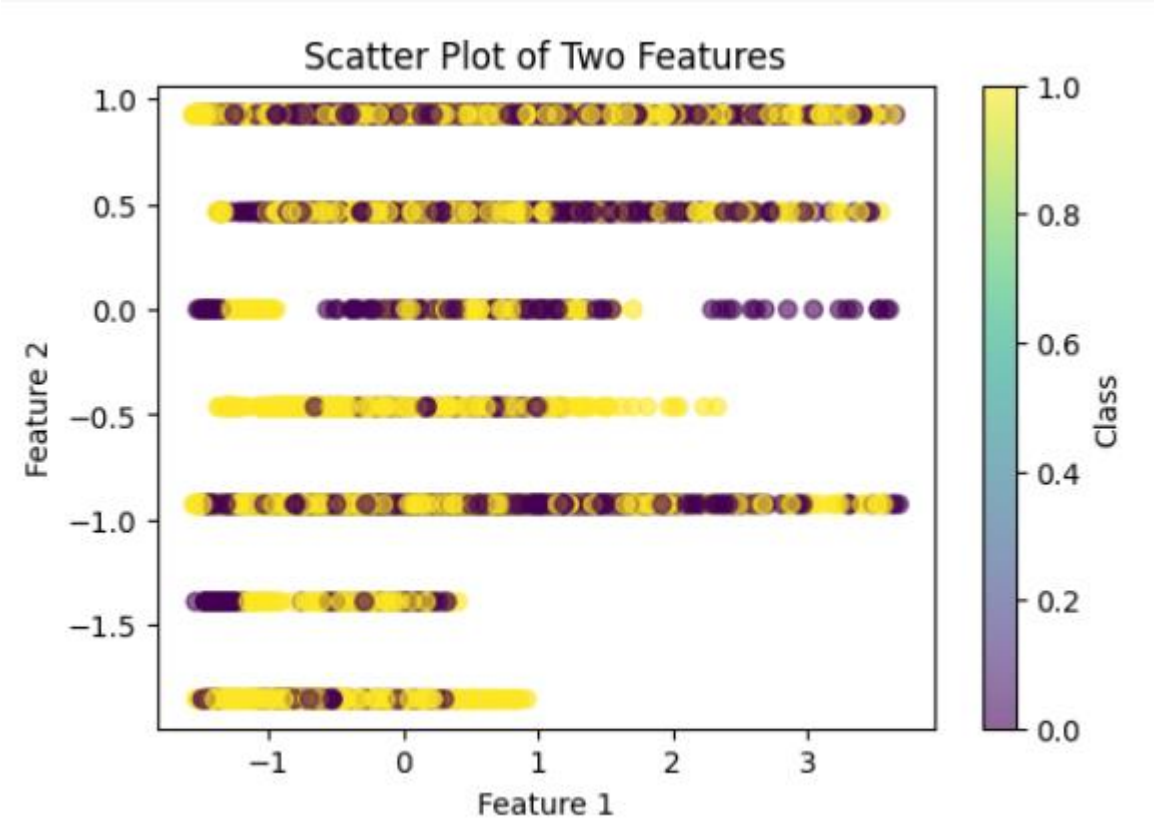
```
# Separate features and target
X =
data_encoded.drop(columns=['
class'])
y = data_encoded['class']

# Train-Test Split
X_train, X_test, y_train, y_test
= train_test_split(X, y,
test_size=0.2,
random_state=42, stratify=y)

# Train the Random Forest
Classifier
model =
RandomForestClassifier(rando
m_state=42)
model.fit(X_train, y_train)

# Make Predictions
y_pred =
model.predict(X_test)

# Evaluate the Model
print("Classification
Report:\n",
classification_report(y_test,
y_pred))
print("Confusion Matrix:\n",
confusion_matrix(y_test,
y_pred))
print("Accuracy:",
accuracy_score(y_test,
y_pred))

# Feature Importance
feature_importances =
```

```
pd.Series(model.feature_importances_, index=X.columns)
print("\nFeature Importances:\n", feature_importances.sort_values(ascending=False))

# Visualize Feature Importance
feature_importances.sort_values(ascending=False).plot(kind='bar', figsize=(10, 6), title='Feature Importance')
plt.show()
```

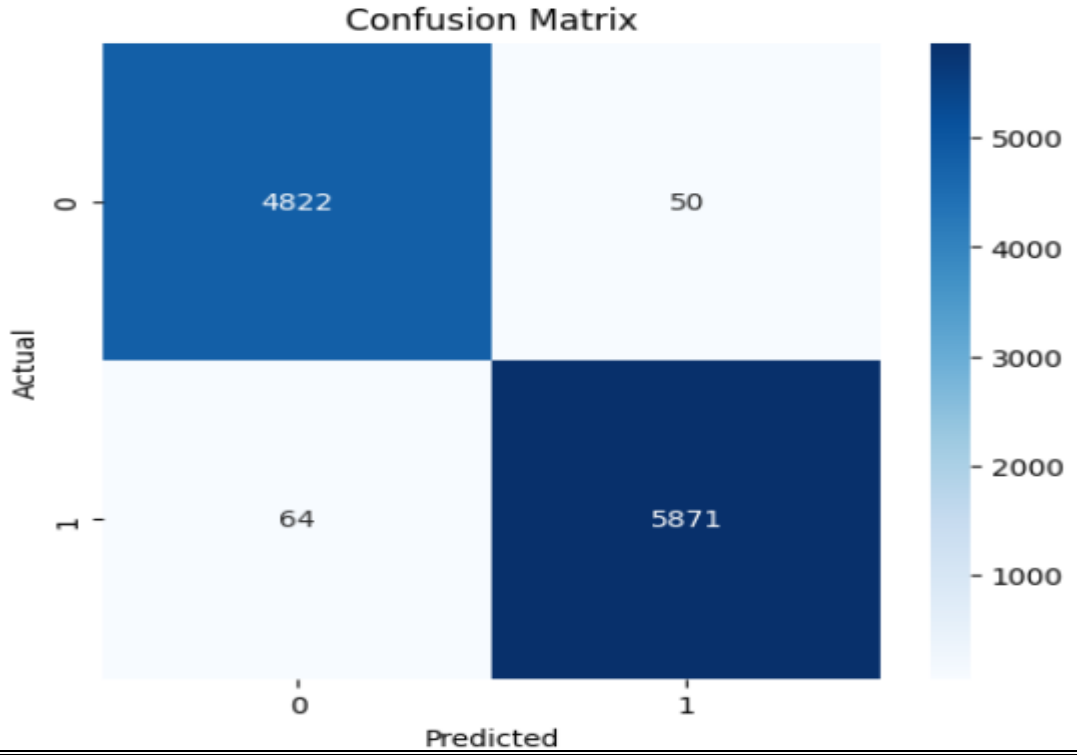
CORRELATION HEATMAP



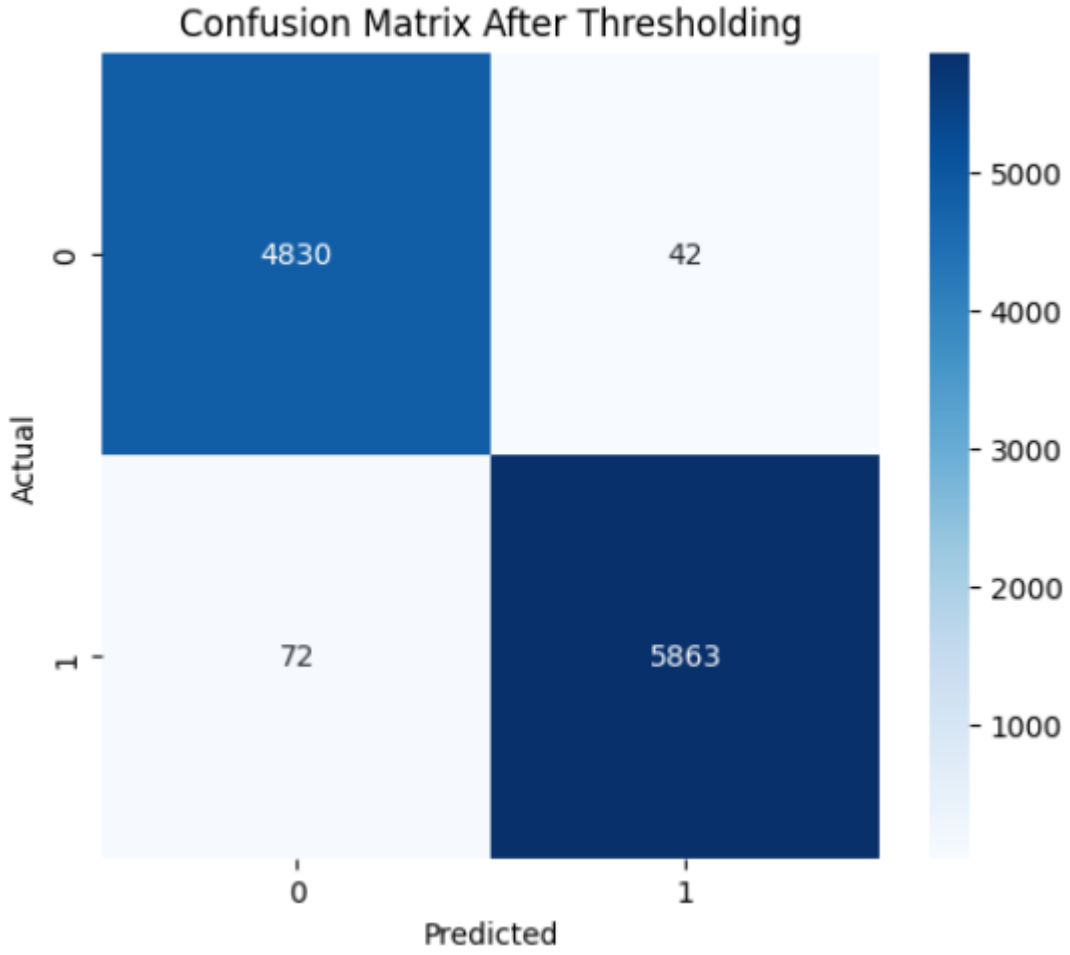
SCATTER PLOT OF TWO FEATURES



CONFUSION MATRIX



AFTER THRESHOLDING



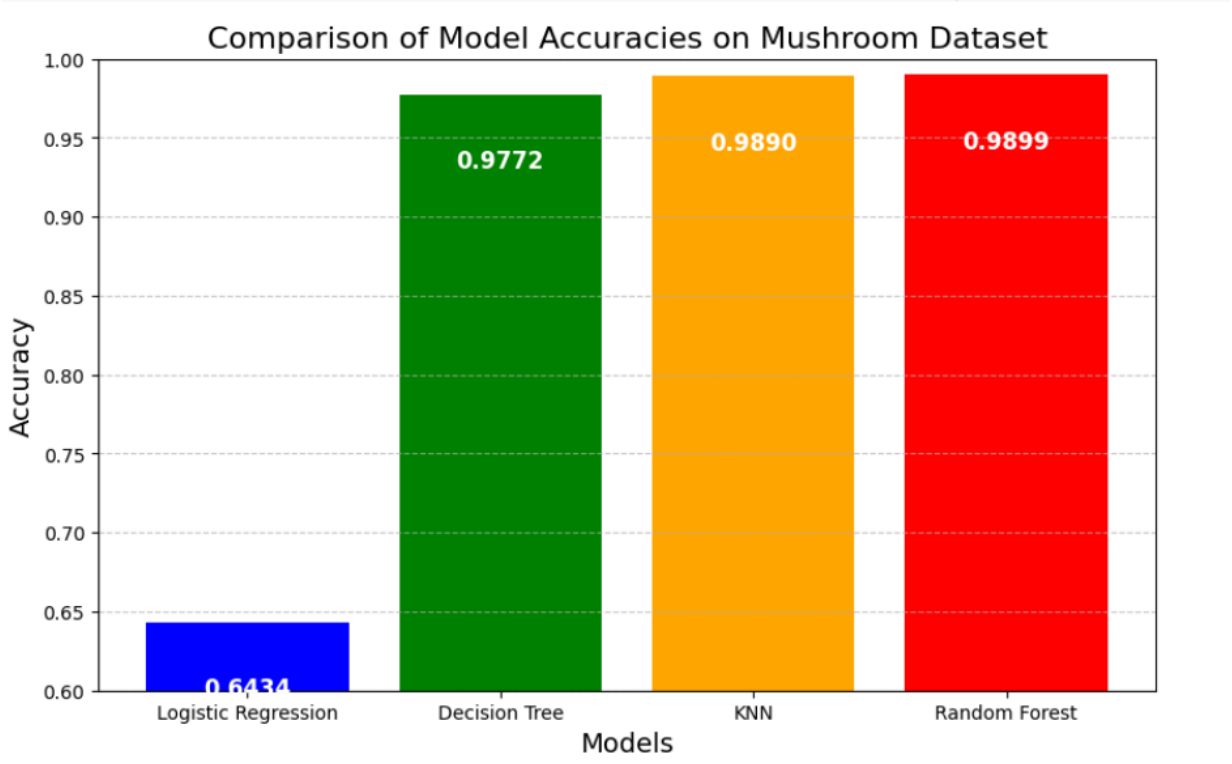
Classification Report:

	precision	recall	f1-score	support
0	0.99	0.99	0.99	4872
1	0.99	0.99	0.99	5935
accuracy			0.99	10807
macro avg	0.99	0.99	0.99	10807
weighted avg	0.99	0.99	0.99	10807

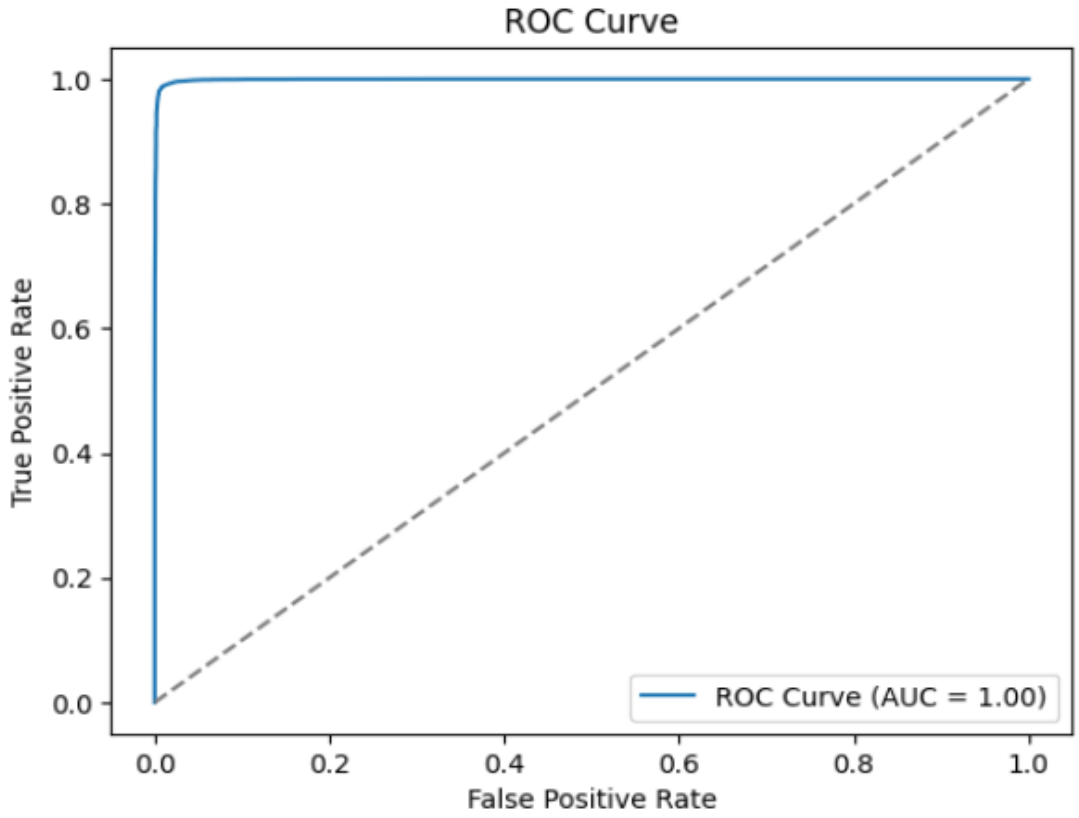
Confusion Matrix:
[[4822 50]
[64 5871]]

Accuracy: 0.9894512815767558

COMPARISION



ROC CURVE



GITHUB LINK:

https://github.com/siri-chandana-macha/Mushroom_Classification