1. Data Exploration

-------------------------

📦 Dataset Shape: (178, 14)

🏷 Target Labels: [0 1 2]

alcohol malic\_acid ash alcalinity\_of\_ash magnesium total\_phenols \

0 14.23 1.71 2.43 15.6 127.0 2.80

1 13.20 1.78 2.14 11.2 100.0 2.65

2 13.16 2.36 2.67 18.6 101.0 2.80

3 14.37 1.95 2.50 16.8 113.0 3.85

4 13.24 2.59 2.87 21.0 118.0 2.80

flavanoids nonflavanoid\_phenols proanthocyanins color\_intensity hue \

0 3.06 0.28 2.29 5.64 1.04

1 2.76 0.26 1.28 4.38 1.05

2 3.24 0.30 2.81 5.68 1.03

3 3.49 0.24 2.18 7.80 0.86

4 2.69 0.39 1.82 4.32 1.04

od280/od315\_of\_diluted\_wines proline target

0 3.92 1065.0 0

1 3.40 1050.0 0

2 3.17 1185.0 0

3 3.45 1480.0 0

4 2.93 735.0 0

**Inference:** The dataset consists of 14 features & 178 samples.

🔍 Data Types:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 178 entries, 0 to 177

Data columns (total 14 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 alcohol 178 non-null float64

1 malic\_acid 178 non-null float64

2 ash 178 non-null float64

3 alcalinity\_of\_ash 178 non-null float64

4 magnesium 178 non-null float64

5 total\_phenols 178 non-null float64

6 flavanoids 178 non-null float64

7 nonflavanoid\_phenols 178 non-null float64

8 proanthocyanins 178 non-null float64

9 color\_intensity 178 non-null float64

10 hue 178 non-null float64

11 od280/od315\_of\_diluted\_wines 178 non-null float64

12 proline 178 non-null float64

13 target 178 non-null int64

dtypes: float64(13), int64(1)

memory usage: 19.6 KB

🔢 Unique value count per feature:

target 3

nonflavanoid\_phenols 39

magnesium 53

alcalinity\_of\_ash 63

hue 78

ash 79

total\_phenols 97

proanthocyanins 101

proline 121

od280/od315\_of\_diluted\_wines 122

alcohol 126

flavanoids 132

color\_intensity 132

malic\_acid 133

dtype: int64

**Inference:** The dataset has 13 numerical & 0 categorical features.

📊 Summary Statistics:

min max mean std

alcohol 11.03 14.83 13.000618 0.811827

malic\_acid 0.74 5.80 2.336348 1.117146

ash 1.36 3.23 2.366517 0.274344

alcalinity\_of\_ash 10.60 30.00 19.494944 3.339564

magnesium 70.00 162.00 99.741573 14.282484

total\_phenols 0.98 3.88 2.295112 0.625851

flavanoids 0.34 5.08 2.029270 0.998859

nonflavanoid\_phenols 0.13 0.66 0.361854 0.124453

proanthocyanins 0.41 3.58 1.590899 0.572359

color\_intensity 1.28 13.00 5.058090 2.318286

hue 0.48 1.71 0.957449 0.228572

od280/od315\_of\_diluted\_wines 1.27 4.00 2.611685 0.709990

proline 278.00 1680.00 746.893258 314.907474

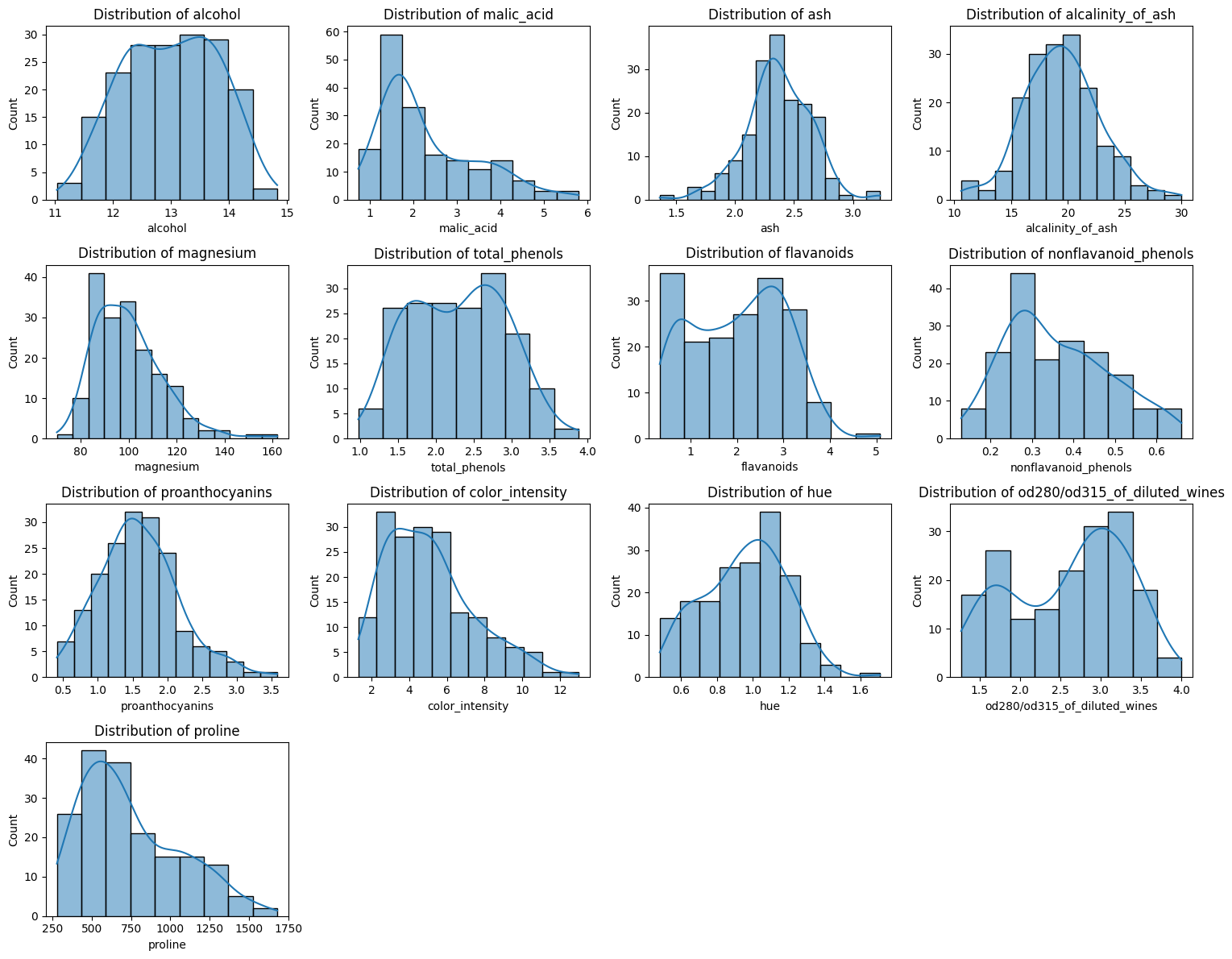
target 0.00 2.00 0.938202 0.775035

2. Exploratory Data Analysis

-----------------------------------

A pie chart with text on it

AI-generated content may be incorrect.



A screenshot of a graph

AI-generated content may be incorrect.



2.2 Checking Missing Values

-----------------------------------

Missing values per column:

alcohol 0

malic\_acid 0

ash 0

alcalinity\_of\_ash 0

magnesium 0

total\_phenols 0

flavanoids 0

nonflavanoid\_phenols 0

proanthocyanins 0

color\_intensity 0

hue 0

od280/od315\_of\_diluted\_wines 0

proline 0

target 0

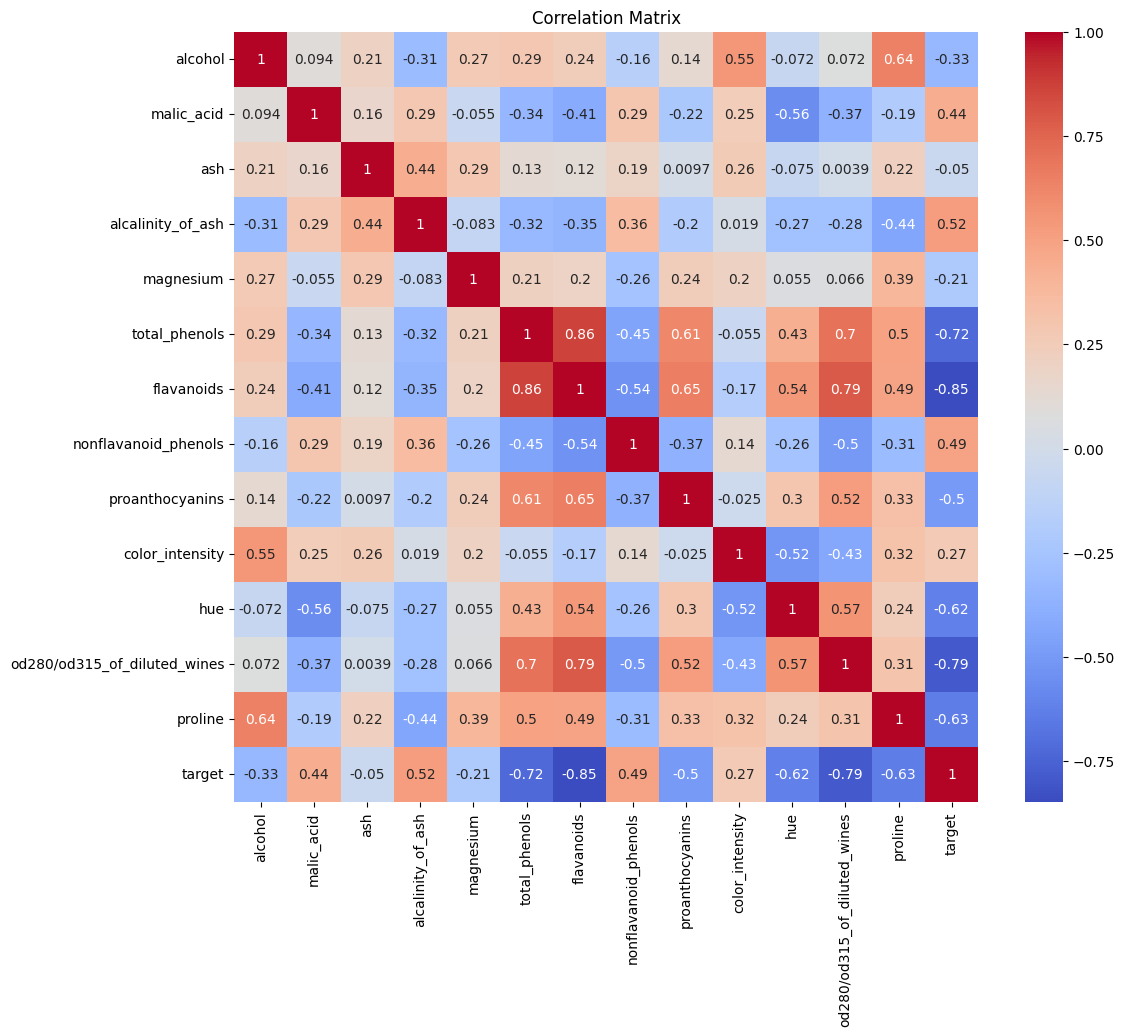
dtype: int64

A graph of a number of gray lines

AI-generated content may be incorrect.

2.3 Feature Correlation

------------------------------



3. Data Preprocessing

-------------------------

🔁 Checking for duplicate rows:

Number of duplicate rows: 0

✔ StandardScaler initialized

4. Data Manipulation

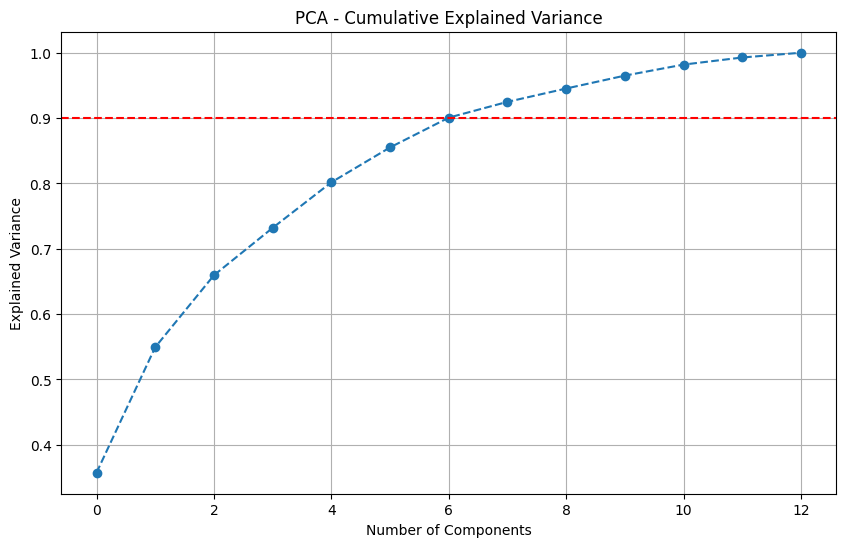
-------------------------

✅ Final Train/Test Split:

Train: (124, 13), Test: (54, 13)

5. Feature Selection / Extraction

----------------------------------------



6. Predictive Modeling (Random Forest + CV)

---------------------------------------------

✅ Cross-validated F1-Score (weighted): 0.7808

✅ Accuracy: 0.9259

✅ F1 Score (weighted): 0.9254

🎯 Actual Labels:

[np.int64(0), np.int64(1), np.int64(0), np.int64(0), np.int64(0), np.int64(0), np.int64(2), np.int64(1), np.int64(1), np.int64(2), np.int64(1), np.int64(1), np.int64(2), np.int64(1), np.int64(0), np.int64(2), np.int64(1), np.int64(0), np.int64(2), np.int64(2), np.int64(1), np.int64(2), np.int64(2), np.int64(2), np.int64(1), np.int64(2), np.int64(0), np.int64(1), np.int64(0), np.int64(1), np.int64(0), np.int64(1), np.int64(2), np.int64(1), np.int64(1), np.int64(2), np.int64(1), np.int64(1), np.int64(1), np.int64(0), np.int64(2), np.int64(0), np.int64(0), np.int64(0), np.int64(0), np.int64(1), np.int64(1), np.int64(0), np.int64(2), np.int64(0), np.int64(1), np.int64(1), np.int64(2), np.int64(0)]

🔮 Predicted Labels:

[np.int64(0), np.int64(1), np.int64(0), np.int64(0), np.int64(0), np.int64(0), np.int64(2), np.int64(1), np.int64(1), np.int64(2), np.int64(2), np.int64(1), np.int64(2), np.int64(1), np.int64(0), np.int64(2), np.int64(1), np.int64(1), np.int64(2), np.int64(2), np.int64(1), np.int64(2), np.int64(2), np.int64(2), np.int64(1), np.int64(2), np.int64(0), np.int64(1), np.int64(0), np.int64(1), np.int64(1), np.int64(0), np.int64(2), np.int64(1), np.int64(1), np.int64(2), np.int64(1), np.int64(1), np.int64(1), np.int64(0), np.int64(2), np.int64(0), np.int64(0), np.int64(0), np.int64(0), np.int64(1), np.int64(1), np.int64(0), np.int64(2), np.int64(0), np.int64(1), np.int64(1), np.int64(2), np.int64(0)]

📄 Classification Report:

precision recall f1-score support

0 0.94 0.89 0.91 18

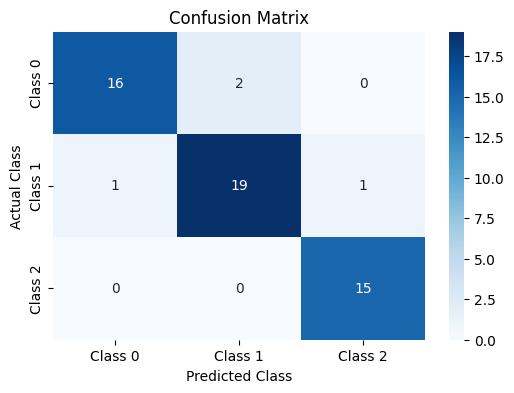
1 0.90 0.90 0.90 21

2 0.94 1.00 0.97 15

accuracy 0.93 54

macro avg 0.93 0.93 0.93 54

weighted avg 0.93 0.93 0.93 54



📈 AUC-ROC per class:

A graph with a line and a line

AI-generated content may be incorrect.

7. Feature Importances

-------------------------

A graph with blue bars

AI-generated content may be incorrect.

8. Simulate Noisy Test Data

------------------------------

🚧 Accuracy (with noise): 0.9074

🚧 F1 Score (with noise): 0.9072

9. Tuning Number of Estimators

------------------------------

🔍 Comparison of Different n\_estimators:

n\_estimators = 10: Accuracy = 0.9259, F1 Score = 0.9254

n\_estimators = 50: Accuracy = 0.9630, F1 Score = 0.9630

n\_estimators = 100: Accuracy = 0.9630, F1 Score = 0.9630

n\_estimators = 200: Accuracy = 0.9630, F1 Score = 0.9630

10. Final Evaluation Summary

------------------------------

n\_estimators Accuracy F1 Score

0 10 0.925926 0.925431

1 50 0.962963 0.963012

2 100 0.962963 0.963012

3 200 0.962963 0.963012

11. K-Nearest Neighbors (KNN)

-----------------------------------

Accuracy: 0.9444

F1 Score (weighted): 0.9447

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 18

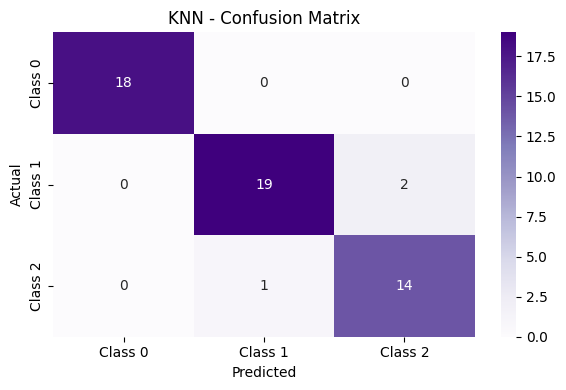
1 0.95 0.90 0.93 21

2 0.88 0.93 0.90 15

accuracy 0.94 54

macro avg 0.94 0.95 0.94 54

weighted avg 0.95 0.94 0.94 54



KNN AUC-ROC per class:

A graph of a function

AI-generated content may be incorrect.

12. Logistic Regression (OvR)

-----------------------------------

Accuracy: 0.9259

F1 Score (weighted): 0.9257

Classification Report:

precision recall f1-score support

0 0.90 1.00 0.95 18

1 1.00 0.86 0.92 21

2 0.88 0.93 0.90 15

accuracy 0.93 54

macro avg 0.92 0.93 0.92 54

weighted avg 0.93 0.93 0.93 54

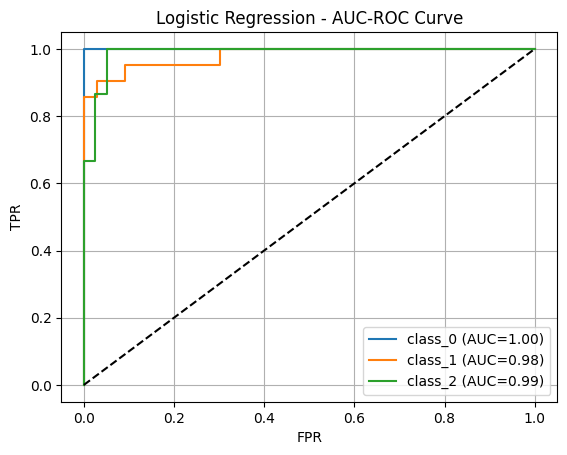
/usr/local/lib/python3.11/dist-packages/sklearn/linear\_model/\_logistic.py:1256: FutureWarning: 'multi\_class' was deprecated in version 1.5 and will be removed in 1.7. Use OneVsRestClassifier(LogisticRegression(..)) instead. Leave it to its default value to avoid this warning.

warnings.warn(

A diagram of a logistic regression

AI-generated content may be incorrect.

Logistic Regression AUC-ROC per class:



13. XGBoost Classifier

-----------------------------------

Accuracy: 0.9074

F1 Score (weighted): 0.9088

Classification Report:

precision recall f1-score support

0 0.78 1.00 0.88 18

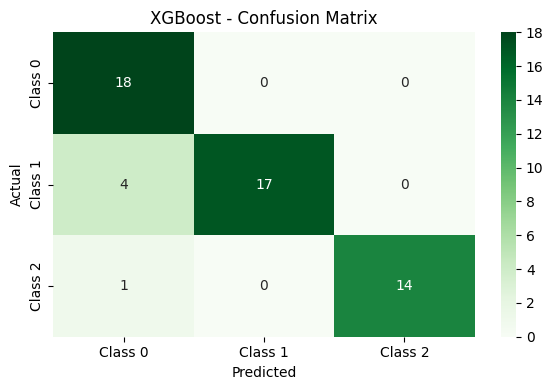
1 1.00 0.81 0.89 21

2 1.00 0.93 0.97 15

accuracy 0.91 54

macro avg 0.93 0.91 0.91 54

weighted avg 0.93 0.91 0.91 54



XGBoost AUC-ROC per class:

A graph of a curve

AI-generated content may be incorrect.

14. Support Vector Machine (SVM)

-----------------------------------

Accuracy: 0.9630

F1 Score (weighted): 0.9626

Classification Report:

precision recall f1-score support

0 0.95 1.00 0.97 18

1 1.00 0.90 0.95 21

2 0.94 1.00 0.97 15

accuracy 0.96 54

macro avg 0.96 0.97 0.96 54

weighted avg 0.97 0.96 0.96 54

A diagram of a graph

AI-generated content may be incorrect.

SVM AUC-ROC per class:

A graph with a line and a line

AI-generated content may be incorrect.

16. Multi-Layer Perceptron (MLP)

-----------------------------------

Accuracy: 0.8889

F1 Score (weighted): 0.8897

Classification Report:

precision recall f1-score support

0 0.78 1.00 0.88 18

1 0.94 0.81 0.87 21

2 1.00 0.87 0.93 15

accuracy 0.89 54

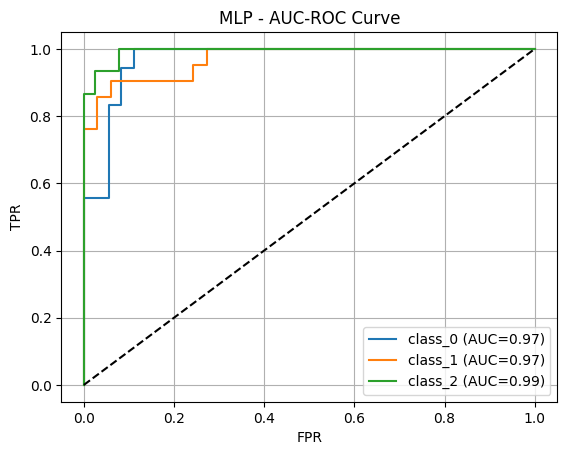
macro avg 0.91 0.89 0.89 54

weighted avg 0.91 0.89 0.89 54

A graph of confusion matrix

AI-generated content may be incorrect.

MLP AUC-ROC per class:



17. Final Model Comparison Summary

----------------------------------------

Model Accuracy F1 Score

0 SVM 0.962963 0.962586

1 KNN 0.944444 0.944663

2 Random Forest 0.925926 0.925431

3 Logistic Regression 0.925926 0.925660

4 XGBoost 0.907407 0.908835

5 MLP 0.888889 0.889651