TASK 5 -Exploratory Data Analysis (EDA)

By – Vani Goel

Email_ID- vanigoel.110@gmail.com

📊 Data Analysis Summary Report

1. 📋 Dataset Overview

- The dataset was examined using .info() and .describe(), providing insights into data types, null values, and statistical summaries.

- Categorical variables were analyzed with .value_counts() to understand class distribution.

---

2. 🔍 Exploratory Data Analysis (EDA)

✅ Pairplot

- Used to examine pairwise relationships and class separability.

- **Observation**: Some classes showed clear separation along specific feature dimensions, while others overlapped.

✅ Heatmap (Correlation Matrix)

- A heatmap of correlations was plotted using sns.heatmap().

- **Observation**:

    o Strong positive or negative correlations were observed between specific features.

    o Some features were weakly correlated, suggesting potential redundancy.

✅ Histograms

- Displayed distribution of numerical features.

- **Observation**:

    o Skewness in certain features indicated potential need for normalization.

    o Some features had outliers or long tails.

✅ Boxplots

- Showed spread and outliers for individual features.
- **Observation**:
    - Several features contained significant outliers.
    - Boxplot patterns differed across classes, suggesting class-dependent distributions.

✅ Scatterplots

- Visualized relationships between two variables at a time.
- **Observation**:
    - Useful for detecting clusters or linear/nonlinear relationships.
    - Some class-wise scatterplots showed visual separation.

---

3. 📈 Key Relationships and Trends

- Positively and negatively correlated feature pairs identified via heatmap.
- Class distribution was imbalanced in some cases, based on .value_counts().
- Certain features showed promise for class distinction based on pairplot and scatterplot analysis.

---

4. 📌 Key Observations

- Data is relatively clean with manageable missing values.
- Class overlap exists in some feature dimensions—may impact classification performance.
- Presence of outliers could influence model accuracy and may require treatment.
- Not all features are strongly correlated—suggesting each contributes unique information.

---

5. ✅ Actionable Recommendations

- Normalize or scale skewed features.
- Consider removing or capping extreme outliers.

- Use feature selection to retain only the most informative features.

- Address class imbalance if using supervised models.