ML_Lab Experiment 01

July 31, 2025

0.0.1 Lab Experiment 01

Objective: Perform various data preprocessing techniques like handling missing data and feature scaling.

Student Details

Field	Information
Name	*[Enter Your Full Name]*
USN	*[Enter Your USN]*
Section	*[Enter Your Section]*
Programme	B.Tech(H)
School	Computer Science and Engineering

[22]:
```
import platform
import psutil
from datetime import datetime
import sys
import getpass
import os

# Timestamp
timestamp = datetime.now().strftime('%Y-%m-%d %H:%M:%S')
# System Information
system_info = {
    "User Name": getpass.getuser(),
    "Timestamp": timestamp,
    "OS": platform.system() + " " + platform.release(),
    "Processor": platform.processor(),
    "RAM (GB)": round(psutil.virtual_memory().total / (1024 ** 3), 2),
    "Python Version": platform.python_version(),
    "Working Directory": os.getcwd()
}
# Print formatted results
print("  System Details\n" + "-"*40)
for key, value in system_info.items():
    print(f"{key:20}: {value}")
```

```
  System Details
----------------------------------------
User Name           : ashwinimathur
Timestamp           : 2025-07-31 10:45:03
OS                  : Darwin 24.5.0
Processor           : arm
RAM (GB)            : 8.0
Python Version      : 3.13.5
Working Directory   : /Users/ashwinimathur/Downloads/ML_Labs
```

0.0.2 Student Tasks

Step 01: - Import necessary Python libraries for data handling and preprocessing.

[23]:
```
# [ Edit this cell ] Write your code below this line
```

Step 02: - Load the dataset (CSV format is recommended) using pandas..

[24]:
```
# [ Edit this cell ] Write your code below this line
```

Step 03: - Explore the dataset: check shape, data types, sample rows, summary statistics, etc.

[25]:
```
# [ Edit this cell ] Write your code below this line
```

Step 04: - Detect missing values and any obvious anomalies (e.g., out-of-range values)..

[26]:
```
# [ Edit this cell ] Write your code below this line
```

Step 05: - Choose and perform strategies such as mean/median imputation or removal..

[27]:
```
# [ Edit this cell ] Write your code below this line
```

Step 06: - Choose and perform strategies such as mean/median imputation or removal..

[28]:
```
# [ Edit this cell ] Write your code below this line
```

Step 07: - Standardize or normalize your numeric columns..

[29]:
```
# [ Edit this cell ] Write your code below this line
```

Step 08: - Assign your input features to one variable (e.g., `X`) and output/target to another (e.g., y)..

[30]:
```
# [ Edit this cell ] Write your code below this line
```

Step 09: - Save the pre-processed/clean dataset as a CSV file to your system..

[ ]:
```
# [ Edit this cell ] Write your code below this line
```

Rubric: Data Preprocessing (10 Marks) For Course Instructor Use

Rubric for assessing an experiment involving missing data handling and feature scaling.

Criteria	Description	Max Marks	Marks Obtained
Missing Data Handling	Correct identification of missing values and application of an appropriate handling method, with basic justification.	3	
Feature Scaling	Appropriate scaling method selected and correctly applied, with a brief explanation.	3	
Code Implementation	Code is functional and syntactically correct, with clear output demonstrating preprocessing steps.	2	
Clarity, Insight & Viva	Clear justification of technique choices, reflection on preprocessing impact, and confident viva responses.	2	

Marks Obtained: _____ / 10

Course Instructor Signature: _____

Link for pandoc - Download the Software for export PDF for submission in GCR
https://github.com/jgm/pandoc/releases/tag/3.7.0.2