ML_Lab Experiment 02

July 31, 2025

0.0.1 Lab Experiment 02

Objective: Perform various data preprocessing techniques like handling outliers and encoding categorical variables

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

[ ]:
```
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}")
```

0.0.2 Student Tasks

Learning Objectives

1. Identify and treat outliers in structured datasets using statistical methods.
2. Encode categorical data into numerical methods that machine learning models can interpret.
3. Use Label Encoding and One-Hot Encoding using `scikit-learn`.
4. Improve data quality through encoding and outlier handling.

Step 01: - Import necessary Python libraries : Import pandas, numpy, scikit-learn, matplotlib, seaborn.

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

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

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

Step 03: - Detect and visualize outliers using boxplot visualization.

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

Step 04: - Perform the Statistical Outlier Detection (Z-score).

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

Step 05: - Encode Categorical Variables: Label encoding

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

Step 06: - Encode Categorical Variables: One-Hot Encoding

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

Step 07: - Save Cleaned & Encoded Dataset to your system..

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

0.0.3 Rubric: Data Preprocessing 02 (10 Marks)

For Course Instructor Use

A rubric for accessing experiment focused on outlier handling and categorical data encoding.

Criteria	Description	Max Marks	Marks Obtained
Outlier Detection & Removal	Visual identification via plots; accurate usage of Z-score method to filter outliers.	3	
Categorical Encoding	Proper application of Label Encoding for binary data and One-Hot Encoding for nominal data.	3	
Code Quality & Execution	Code runs without errors, is well-organized, and follows best practices; outputs are clearly displayed.	2	

Criteria	Description	Max Marks	Marks Obtained
Documentation & Viva Insight	Brief comments, explanations, and ability to answer questions during viva on technique choices.	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