SECB3203-01(PROGRAMMING FOR BIOINFORMATIC)

Semester 01 2025/2026
Section 01

Progress

Faculty of Computing

Dataset:
https://www.kaggle.com/datasets/uciml/mushroom-classification

Student Name	Matric Number
NURUL NATALIA BINTI ROSNIZAM	A23CS0165
ADLINA NARISYA BINTI ISMAIL	A23CS0033
NAZATUL NADHIRAH BINTI SABTU	A23CS0144

1.0 Software and Hardware Requirements

1. Software Requirements

   - Operating System : Windows 11

   - Programming Language : Python 3.13

   - Development Environment : Visual Studio Code

   - Libraries : pandas, numpy, matplotlib, scikit-learn

   - Dataset Source : Kaggle

2. Hardware

   - **Processor (CPU):** The central processing unit is responsible for executing all program instructions and handling computations. A modern processor allows data operations and analysis tasks to be performed quickly.

   - **Memory (RAM):** Sufficient RAM is needed to load the dataset and perform multiple data operations without slowing down the system.

   - **Storage:** Storage is needed to save the dataset, code and results safely.

   - **Graphics Processing Unit (GPU):** Having a GPU can accelerate plotting and rendering of visualizations in some environments.

Component	Specification
Processor (CPU)	Intel® Core(™) i5-8350U or equivalent
Memory (RAM)	16.0 GB / 32.0 GB
Storage	At least 1 GB free space
GPU	8 GB

2.0 Flowchart of Proposed Approach

```
                    ┌─────────┐
                    │  Start  │
                    └─────────┘
                         │
                         ▼
               ╱─────────────────╲
              ╱ Collect Mushroom   ╲
              ╲     Dataset        ╱
               ╲─────────────────╱
                         │
                         ▼
   ┌──────────────┐   ┌──────────────────┐
   │ Load Dataset │──▶│ Data Preprocessing│
   └──────────────┘   └──────────────────┘
                              │
                              ▼
                         ◇─────────◇        Yes   ┌────────────┐
                         │ Missing │─────────────▶│ Clean Data │
                         │  Data   │              └────────────┘
                         ◇─────────◇                    │
                              │ No                       │
                              ▼                          │
                   ┌──────────────────┐                  │
                   │ Exploratory Data │◀─────────────────┘
                   │  Analysis (EDA)  │
                   └──────────────────┘
                              │
                              ▼
   ┌──────────────────┐   ┌───────────────────┐   ┌──────────────────────┐
   │ Analyze Feature  │──▶│ Data Visualization│──▶│ Identify Patterns &  │
   │  Distributions   │   └───────────────────┘   │     Difference       │
   └──────────────────┘                           └──────────────────────┘
                                                              │
                                                              ▼
                                                   ┌──────────────────────┐
                                                   │  Model Development   │
                                                   └──────────────────────┘
                                                              │
                                                              ▼
                                                   ┌──────────────────────┐
                                                   │   Model Evaluation   │
                                                   └──────────────────────┘
                                                              │
                                                              ▼
                                                        ┌─────────┐
                                                        │   End   │
                                                        └─────────┘
```