UNIVERSITI TEKNOLOGI MALAYSIA
FACULTY OF COMPUTING
SEMESTER 1, SESSION 2025/2026

PROJECT PROGRESS 1
OBESITY LEVEL CLASSIFICATION

SECB3203 : PROGRAMMING FOR BIOINFORMATICS
SECTION 02

GROUP MEMBER:

1.	MUHAMMAD FARIHIN BIN SALEH	A25CS0102
2.	MUHAMMAD MIRZA HASIF BIN MOHD FAHMI	A25CS0108
3.	MUHAMMAD NAWFAL BIN MOHD SHAIFUDDIN	A25CS0109

LECTURER NAME   **:** DR. SEAH CHOON SEN

GROUP   **:** GROUP 07

TABLE OF CONTENTS

1.0 SOFTWARE AND HARDWARE REQUIREMENTS

**SOFTWARE REQUIREMENTS**

1. **Python 3.8+**

    Core programming language for data analysis, modelling, and visualization.

2. **Visual Studio Code or Google Colab**

    Primary code editor with Python extensions for development and debugging.

3. **Git & GitHub**

    Version control for collaborative coding, project tracking, and coding documentation.

4. **Required Python Libraries**

    Data Processing: `pandas`, `numpy`

    Machine Learning: `scikit-learn`, `xgboost`

    Visualization: `matplotlib`, `seaborn`

HARDWARE REQUIREMENTS

COMPONENT	REQUIREMENT
Processor	Intel Core i5 or AMD Ryzen 5
Memory (RAM)	8 GB
Storage	5 GB free space for datasets, code, and outputs
Operating System	Windows 10/11, macOS, or Linux

Table 1.1: Minimum Hardware Requirement

2.0   FLOWCHART OF THE PURPOSED APPROACH
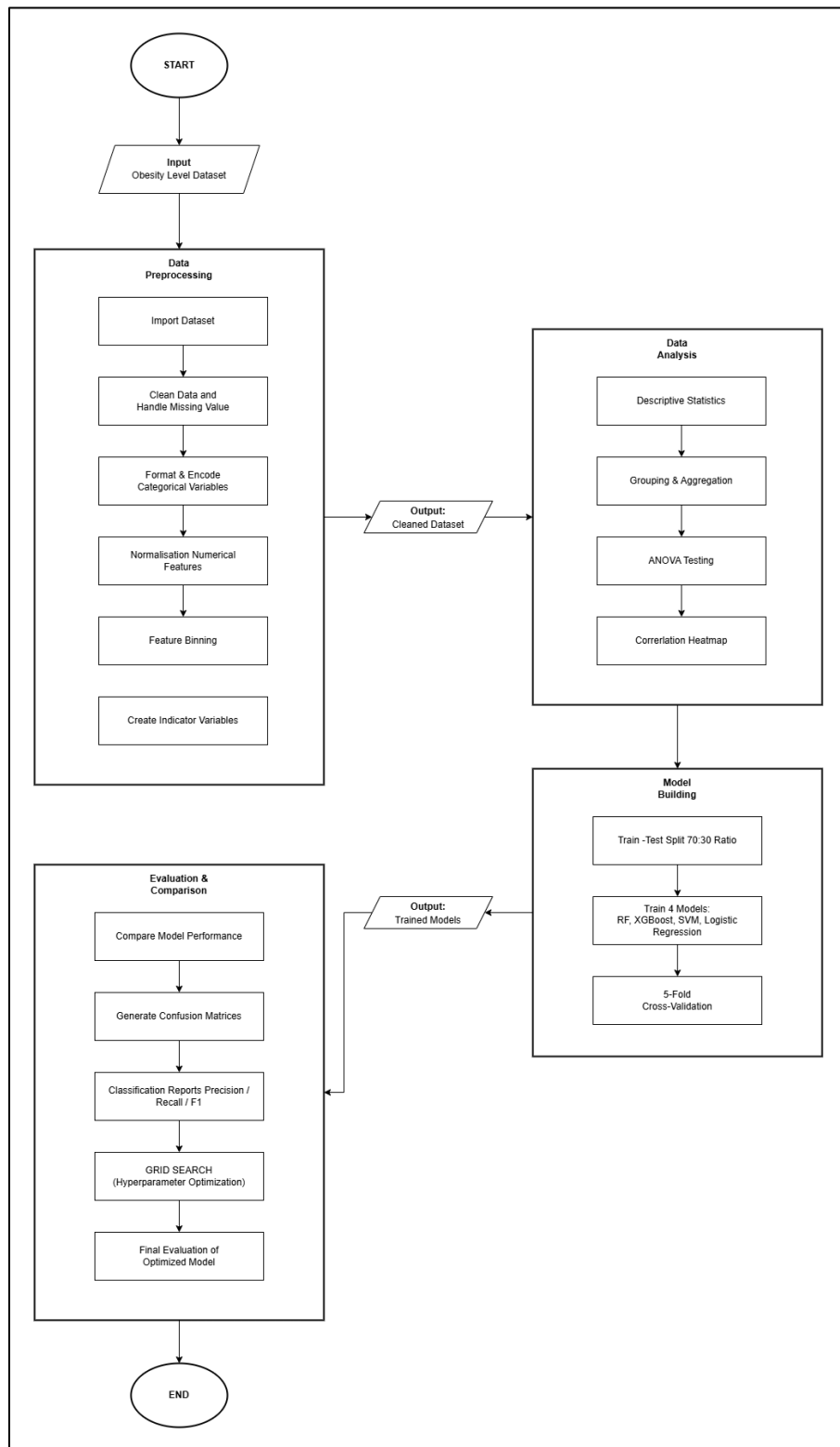
Figure 2.1: Flowchart for Obesity Level Classification

1. **Start and Input**
   - Project begins with dataset acquisition from Kaggle Obesity Levels dataset.

2. **Progress 2: Data Wrangling (Process)**
   - Complete preprocessing pipeline including cleaning, encoding, normalization, and feature engineering.
   - Output: Cleaned, structured dataset ready for analysis.

3. **Progress 3: Exploratory Data Analysis (Process)**
   - Statistical analysis to understand data patterns and relationships.
   - Includes grouping, ANOVA, and correlation visualization.

4. **Progress 4: Model Development (Process)**
   - Implementation of four classification algorithms with hyperparameter optimization.
   - Cross-validation ensures model robustness.

5. **Progress 5: Model Evaluation (Process)**
   - Comprehensive performance assessment using multiple metrics.
   - Visualization of results for clear comparison.

6. **Decision Point**
   - Evaluate if best model meets performance criteria.
   - If not satisfactory, return to data wrangling for refinement.

7. **End and Output**
   - Final model selection and complete project documentation.
   - Results compiled in final report for submission.