UTM UNIVERSITI TEKNOLOGI MALAYSIA
FACULTY OF COMPUTING
UTM Johor Bahru

SECB3203-01(PROGRAMMING FOR BIOINFORMATIC)

Semester 01 2025/2026
Section 01

Progress 4

Faculty of Computing

Dataset:
https://www.kaggle.com/datasets/miadul/tuberculosis-x-ray-dataset-synthetic

Project Title:
Tuberculosis Disease Classification Using Synthetic Chest X-Ray Images and Machine Learning Techniques

---

Student Name	Matric Number
NATIJAH BINTI HUDA	A23CS0142
NUR IMAN BINTI MOHAMAD ZAHARI	A23CS0158
LIANA DARWISYAH BINTI AZMAN	A23CS0102

LECTURER'S NAME : DR. SEAH CHOON SEN

SUBMISSION DATE :

4.0 Model Development

In this project, the focus is not on developing machine learning models from scratch, but on applying and comparing existing classification models to analyze the tuberculosis dataset obtained from Kaggle. These models are implemented using Python libraries such as scikit-learn to evaluate their effectiveness in classifying tuberculosis cases.

4.1 Application of Classification Models

Several well-established machine learning classification algorithms are applied to the dataset. These models are selected due to their reliability and suitability for binary classification tasks.

The dataset is divided into training and testing sets to ensure fair model evaluation.

- Features (X): Patient symptoms and extracted attributes
- Target (y): Tuberculosis classification label

4.2 Classification Models Used

The following classification models are applied:

- Logistic Regression
- Decision Tree
- Random Forest
- Naive Bayes

Each model is trained using the same dataset split to ensure consistency and fairness during comparison.

4.3 Model Training and Prediction

All models are trained using the training dataset and then used to generate predictions on the testing dataset. This process allows the models to learn patterns from the data and evaluate their ability to classify tuberculosis cases accurately.

4.4 Visualization for Model Comparison

Visualization techniques such as performance comparison charts are used to compare the results of different models. These visualizations help in identifying differences in model behavior and overall performance.

4.5 Prediction and Decision Making

The trained classification models are used to make predictions on unseen data. These predictions provide insights into the effectiveness of machine learning models in supporting tuberculosis classification and decision-making processes.