### High Level Document (HLD) for Thyroid Disease Prediction Project

**1. Introduction:**

* **Project Name:** Thyroid Disease Prediction Using Machine Learning
* **Objective:** To develop a machine learning model that accurately predicts thyroid disease based on patient data.

**2. System Architecture:**

* **Data Collection:**
  + Source: Hypothyroid dataset
  + Format: CSV
* **Data Preprocessing:**
  + Handling missing values
  + Encoding categorical variables
  + Normalizing numerical features
* **Model Building:**
  + Algorithms: Logistic Regression, SVM, KNN, Decision Tree, Random Forest
  + Training and validation using cross-validation
* **Model Evaluation:**
  + Metrics: Accuracy, Precision, Recall, F1-Score
  + Confusion Matrix analysis
* **Deployment:**
  + Saving the best model using pickle
  + Integration with a web application for user interaction

**3. Data Flow:**

* **Input:** Patient data (age, sex, medical history, etc.)
* **Process:** Data preprocessing → Model training → Model evaluation
* **Output:** Prediction of thyroid disease (negative, compensated hypothyroid, primary hypothyroid, secondary hypothyroid)

**4. Components:**

* **Data Preprocessing Module:** Handles data cleaning and transformation.
* **Model Training Module:** Trains various machine learning models.
* **Evaluation Module:** Evaluates model performance.
* **Deployment Module:** Saves and deploys the best-performing model.

**5. Tools and Technologies:**

* **Programming Language:** Python
* **Libraries:** Pandas, NumPy, Scikit-learn, Imbalanced-learn, Matplotlib, Seaborn
* **Environment:** Jupyter Notebook