# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 15 March 2024 |
| Team ID | SWTID1720447482 |
| Project Title | THYROID CLASSIFICATION |
| Maximum Marks | 2 Marks |

**Data Collection Plan Template**

| Section | Description |
|---|---|
| Project Overview | The goal of the project "Thyroid Classification using RandomForest, XGB Classifier, SVC Model" is to create and evaluate three machine learning models—RandomForest, XGB Classifier, and SVC—for the purpose of diagnosing thyroid problems. The project will identify the most accurate model to help healthcare providers make well-informed diagnoses and improve patient outcomes using Python in Google Colab. |
| Data Collection Plan | The data is collected from the Kaggle. |
| Raw Data Sources Identified | Essential parameters including age, sex, blood pressure, cholesterol, na_to_k, and medication are included in this dataset. The construction and assessment of machine learning models—RandomForest, XGB Classifier, and SVC—for the precise categorization of thyroid diseases will be made possible by the use of this extensive dataset. |

**Raw Data Sources Template**

| Source Name | Description | Location/URL | Format | Size | Access Permissions |
|---|---|---|---|---|---|
| Dataset 1 | Dataset contains 200 rows and 6 columns and each column has values of Age,Sex,BP,Cholesterol,Na_to_K,Drug. | https://drive.google.com/file/d/1tE3OdfeHV-Rl931leEgXlaI7p0fLylL5/view?usp=sharing | CSV | 5.89 KB | Public |