# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 08 August 2025 |
| Skill Wallet ID | **SWUID20250188325** |
| Project Title | Predictive Pulse: Harnessing Machine Learning for Blood Pressure Analysis |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Report:**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan:**

| Section | Description |
|---|---|
| Project Overview | The machine learning project aims to predict hypertension stage prediction based on patient information. Using a dataset with features such as age, gender, income, and medical history, the objective is to build a model that accurately classifies blood pressure stage (approved or denied), facilitating efficient and informed decision-making in the health assessment process. |
| Data Collection Plan | Search for datasets related to hypertension stage predictions, medical information, and patient details.<br>Prioritize datasets with diverse demographic information. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as Gender, |

| | |
|---|---|
| | Age, History, Patient, Severity and Blood pressure-related details for machine learning analysis. |

**Raw Data Sources Report:**

| Source Name | Description | Location/URL | Format | Size | Access Permissions |
|---|---|---|---|---|---|
| Kaggle Dataset | The dataset comprises patient details (age, gender), medical metrics (systolic, diastolic readings), and hypertension stage prediction outcomes. | BP Analysis using Machine Learning | CSV | 166.46 kB | Public |