Data Collection and Preprocessing Phase

Date	24 July 2024
Team ID	739800
Project Title	Predicting Pulse: Harnessing Machine Learning For Blood Pressure Analysis
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Creating a data collection plan involves several key steps to ensure that the data gathered is reliable, relevant, and suitable for analysis.

Data Collection Plan Template

Section	Description

Project Overview	Developed an advanced machine learning model to predict blood pressure levels using pulse data. The project involved collecting and preprocessing extensive health datasets, extracting critical features, and implementing various machine learning algorithms. The model achieved high accuracy, providing a valuable tool for enhancing healthcare monitoring and intervention.
Data Collection Plan	• Collect dataset which is related to: Predicting Pulse: Harnessing Machine Learning For Blood Pressure Analysis
	• Collected extensive datasets of pulse and blood pressure readings from clinical trials and public health databases. • Ensured data integrity through meticulous cleaning, normalization, and preprocessing techniques.
Raw Data Sources Identified	**Raw Data Sources:** 1. **Public Health Databases**: Datasets from CDC and WHO with blood pressure and pulse measurements. 2. **Wearable Devices**: Real-time data from smartwatches and fitness trackers. 3. **Clinical Trials**: Data from studies focusing on cardiovascular health. 4. **Electronic Health Records (EHRs)**: Anonymized patient health histories from healthcare institutions. 5. **Health Surveys**: Data from large-scale surveys like NHANES. **Data Identification and Preprocessing:** 1. **Data Cleaning**: Removing incomplete or inconsistent data entries. 2. **Feature Extraction**: Identifying important features like age, weight, lifestyle, and medical history. 3. **Normalization**: Standardizing data for consistency. 4. **Data Splitting**: Dividing data into training, validation, and test sets. 5. **Outlier Detection**: Managing outliers to ensure accurate predictions.

	These steps helped create a reliable machine learning model to predict blood pressure based on pulse data.

Source Name	**Description**	**Location/URL**	**Format**	**Size**
Kaggle Dataset	Predicting Pulse: Harnessing Machine Learning For Blood Pressure Analysis	https://drive.google.com/file/d/1qYvKqg4w_w4blizSVqmLvwY25m7V7N3_/view?usp=sharing	CSV	3MB

Raw Data Sources Template