

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	739785
Project Title	A Comprehensive Measure of Well- Being:The Human Development Index Using Machine Learning
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 Human Development Index (HDI) is a composite statistic that measures key dimensions of human development, including life expectancy, education, and income per capita. Traditional HDI calculation methods have limitations in capturing the complexities of human development. This project aims to leverage machine learning techniques to predict HDI more accurately and provide an interactive web application for users to make HDI predictions based on their inputs.

Data Collection Plan	<ul style="list-style-type: none"> ● Search for datasets related to Human Development Index ● This report outlines the data collection plan and identifies the raw data sources necessary for predicting the Human Development Index (HDI). The aim is to ensure comprehensive data acquisition to enhance the accuracy and reliability of the HDI prediction model .
Raw Data Sources Identified	<p>The raw data sources for this project include datasets obtained from Kaggle & UCI, 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,</p>

	marital status, income, and loan-related details for machine learning analysis.
--	---

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset contains the different feature related to the Human development from day to day.	Dataset - Google Drive	CSV	75 kB	Public