Data Collection and Preprocessing Phase

Date	21 June 2024
Team ID	739812
Project Title	Eudaimonia Engine: Machine Learning Delving into Happiness Classification
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

Elevate the 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 aim of the Eudaimonia Engine project, "Machine Learning Delving into Happiness Classification," is to develop a sophisticated machine learning model capable of accurately classifying and predicting happiness levels based on a variety of data inputs. The goal is to contribute to the understanding of well-being, offering valuable tools and knowledge to individuals and organizations to enhance overall happiness and quality of life.
Data Collection Plan	• Search for datasets related to info avail, house cost, school quality, police trust, street quality, events • 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 info avail,

	house cost, school quality, police trust, street quality, events details for machine learning analysis.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises details of (info avail, house cost, school quality, police trust, street quality, events)	https://www.kaggle.com/datasets/priyanshusethi/happiness-classification-dataset	CSV	**729B**	Public