

Data Collection and Preprocessing Phase

Date	21 June 2024
Team ID	739940
Project Title	Gem Valuation Revolution: Predicting Diamond Prices With Artificial Neural Networks
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

This data collection plan and raw data source identification report outline a systematic approach to gathering and preparing datasets for predicting diamond prices using machine learning techniques. By leveraging diverse sources and adhering to legal and ethical guidelines, stakeholders can access comprehensive, reliable data to enhance predictive modeling accuracy and decision-making in the diamond industry.

Data Collection Plan:

+Section	Description
Project Overview	This project overview outlines a structured approach to predicting diamond prices using machine learning techniques, emphasizing innovation, accuracy, and strategic impact within the gem valuation sector.
Data Collection Plan	<ul style="list-style-type: none"> Download datasets directly from online repositories or databases in standard formats like CSV, Excel, or JSON. Use APIs provided by diamond databases or marketplaces to retrieve real-time or batch data on diamond attributes and prices.
Raw Data Sources Identified	<ul style="list-style-type: none"> Kaggle offers various datasets related to diamonds, sourced from different contributors and sources. Look for datasets that include attributes like carat weight, cut, color, clarity, dimensions, and price. <p>Example: Diamonds Dataset on Kaggle</p>

Source Name	Description	Location/URL
Kaggle Dataset	The dataset comprises applicant details (carat,cut,colour,clarity,depth,table,price).	https://drive.google.com/file/d/1i_RjmwRtAvk82MNCQ-B2Vtson73jytpU/view



Raw Data Sources Report: