Data Collection and Preprocessing Phase

Date	15 October 2024
Team ID	739739
Project Title	Predicting Diamond Prices With ANN Using Deep Learning
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	This project aims to predict diamond prices using machine learning models based on attributes such as carat, cut, color, clarity, and geometric properties like depth, table, and dimensions. By analyzing historical data, the project seeks to uncover key factors influencing pricing and develop a tool that delivers accurate price estimates. This can benefit jewelers, buyers, and industry stakeholders by providing data-driven insights for valuation and decision-making.
Data Collection Plan	The data collection will focus on gathering datasets with key diamond attributes, including carat, cut, colour, clarity, depth, table, and price. Primary sources include publicly available datasets on

	platforms like Kaggle .The collected data will be Preprocessed to ensure quality and relevance for building predictive models.
Raw Data Sources Identified	**Raw Data Sources Identification:** 1. **Kaggle Datasets** - Pre-compiled datasets specific to diamond pricing.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions	
Kaggle Dataset	This dataset consist of information regarding predicting diamond prices.	https://www.kaggle.com/code/karnikakapoor/diamond-price-prediction/input	CSV	3 GB	Public	