Data Collection and Preprocessing Phase

Date	14th July 2024
Team ID	739959
Project Title	Sentiment Analysis of Commodity News (Gold)
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

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 Template

Section	Description
Project Overview	In order to analyse news stories on gold commodities and forecast market mood, this project creates a machine learning model. The model is trained using text analysis and natural language processing (NLP) on a tagged dataset of news items about gold. Accurately categorizing news stories as neutral, negative, or favourable regarding their attitude toward gold is the aim in order to provide insights into market opinions and trends. The research determines the best method for gold sentiment analysis by comparing the performance of several machine learning algorithms. The result is a sentiment analysis tool that can help traders, investors, and market analysts make well-informed decisions based on sentiment in the market. This tool may have an effect on risk management and investing strategies in the gold commodity market.

Data Collection Plan	• Search for datasets related to Sentiment Analysis of Commodity News (Gold). • Prioritize datasets with diverse demographic information.
Raw Data Sources Identified	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.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	This dataset consists of various parameters like news headlines etc.	https://www.kaggle.com/datasets/ankurzing/sentiment-analysis-in-commodity-market-gold	CSV	1.95 MB	Public