Data Collection and Preprocessing Phase

Date	04 November 2024
Team ID	SWTID1726834817
Project Title	Fake News Analysis in social media using NLP
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	This project aims to develop a machine learning model capable of accurately classifying news articles as "Fake" or "Real." By leveraging natural language processing techniques, the model will analyze textual features and identify patterns associated with fake news. The goal is to contribute to the fight against misinformation and promote informed decision-making.
Data Collection Plan	Search for datasets related to title, texts and labels of news.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle , the popular platform for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as title, texts and labels of the news for machine learning analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	The dataset comprises of titles, texts and labels ('REAL' / 'FAKE'). .	https://drive.google.com/file/d/1g46T2a8gu-1m1hOoKOLS_h6HK9Es6abQ/view	CSV	29.2 MB	Public