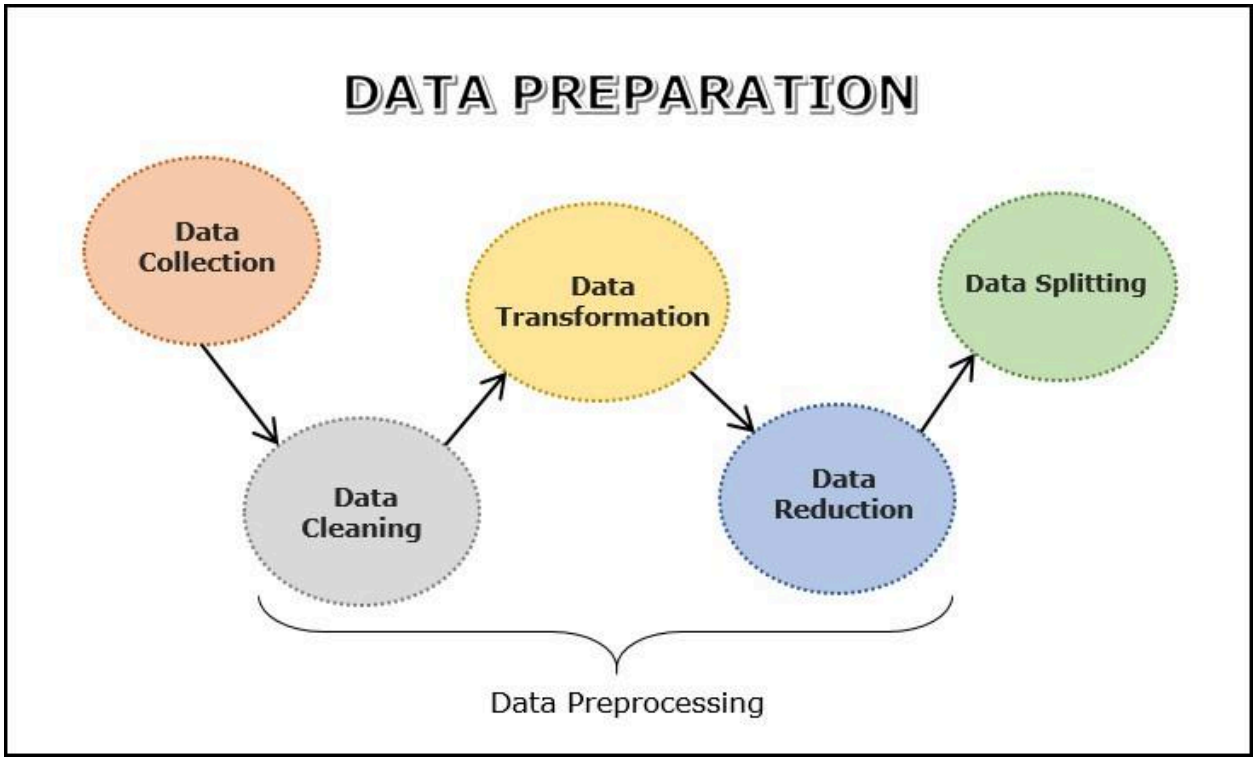
INDIVIDUAL TASK 2

Data preparation steps

Introduction



Steps in Data Processing:

1. Data Collection: Gathering raw data from various sources, such as surveys, sensors, databases, online sources, or manual input.
    - Example: Collecting customer feedback through surveys or obtaining sales data from a company's database.
2. Data Preparation: Cleaning and organizing the raw data to remove errors, fill in missing values, and ensure consistency. This step is crucial for ensuring that the data is accurate and ready for analysis.

- Example: Removing duplicate entries, correcting typos, and converting data into a consistent format (e.g., date formats).

3. Data Input: Converting the prepared data into a format that can be processed by computer systems or analytical tools.
    - Example: Entering data into a spreadsheet or uploading it to a database or data processing software.
4. Data Processing: Transforming the input data into a more useful format through calculations, aggregations, sorting, filtering, or other operations. This stage may involve complex algorithms, statistical analysis, or machine learning models.
    - Example: Calculating the average sales per month, categorizing customer feedback into positive, neutral, or negative, or applying a machine learning model to predict future trends.
5. Data Output/Interpretation: Presenting the processed data in a readable and actionable form, such as reports, charts, graphs, or dashboards. This step may also involve interpreting the results to draw conclusions or make recommendations.
    - Example: Generating a sales report that shows trends over time, creating a dashboard to monitor key performance indicators (KPIs), or summarizing the results of a survey.
6. Data Storage: Saving the processed data in a structured format for future access, analysis, or use. This could involve storing data in databases, data warehouses, or cloud storage systems.
    - Example: Storing the processed data in a relational database or data warehouse for future retrieval and analysis.

Conclusion :

In conclusion, the data preparation phase was critical in transforming raw, unstructured information into a high-quality dataset suitable for analysis. By

addressing missing values and normalizing features, we mitigated potential biases and ensured the model's robustness. This groundwork not only improves the accuracy of our current findings but also establishes a scalable framework for future data ingestion. Ultimately, the quality of our insights remains directly tied to the rigor of this preparation process."