**Data Cleaning and Preparation Methods**

1. **Handling Missing Data**

* Remove Missing Values
* Fill Missing Values (Imputation)
* Predictive Imputation (KNN Imputer)

1. **Handling Duplicate Data**

* Remove Duplicates

1. **Handling Outliers**

* Using Z-score
* Using IQR Method
* Capping Outliers

1. **Handling Inconsistent Data**

* Standardizing Text (Lowercase, Removing Spaces)
* Correcting Typos (Using Fuzzy Matching)

1. **Data Type Conversion**

* Encoding Categorical Variables

1. **Feature Scaling**

* Min-Max Scaling
* Standardization (Z-score Normalization)

1. **Text Data Cleaning**

* Lowercasing: Convert all text to lowercase.
* Removing Punctuation & Special Characters: Use regular expressions.
* Stopword Removal: Remove common words like "the," "is," "and."
* Stemming & Lemmatization: Convert words to their base forms (e.g., “running” → “run”).
* Tokenization: Split text into individual words or phrases.

1. **Feature Extraction using TF-IDF**
2. **Compute Cosine Similarity**
3. **Encode Target Variable**