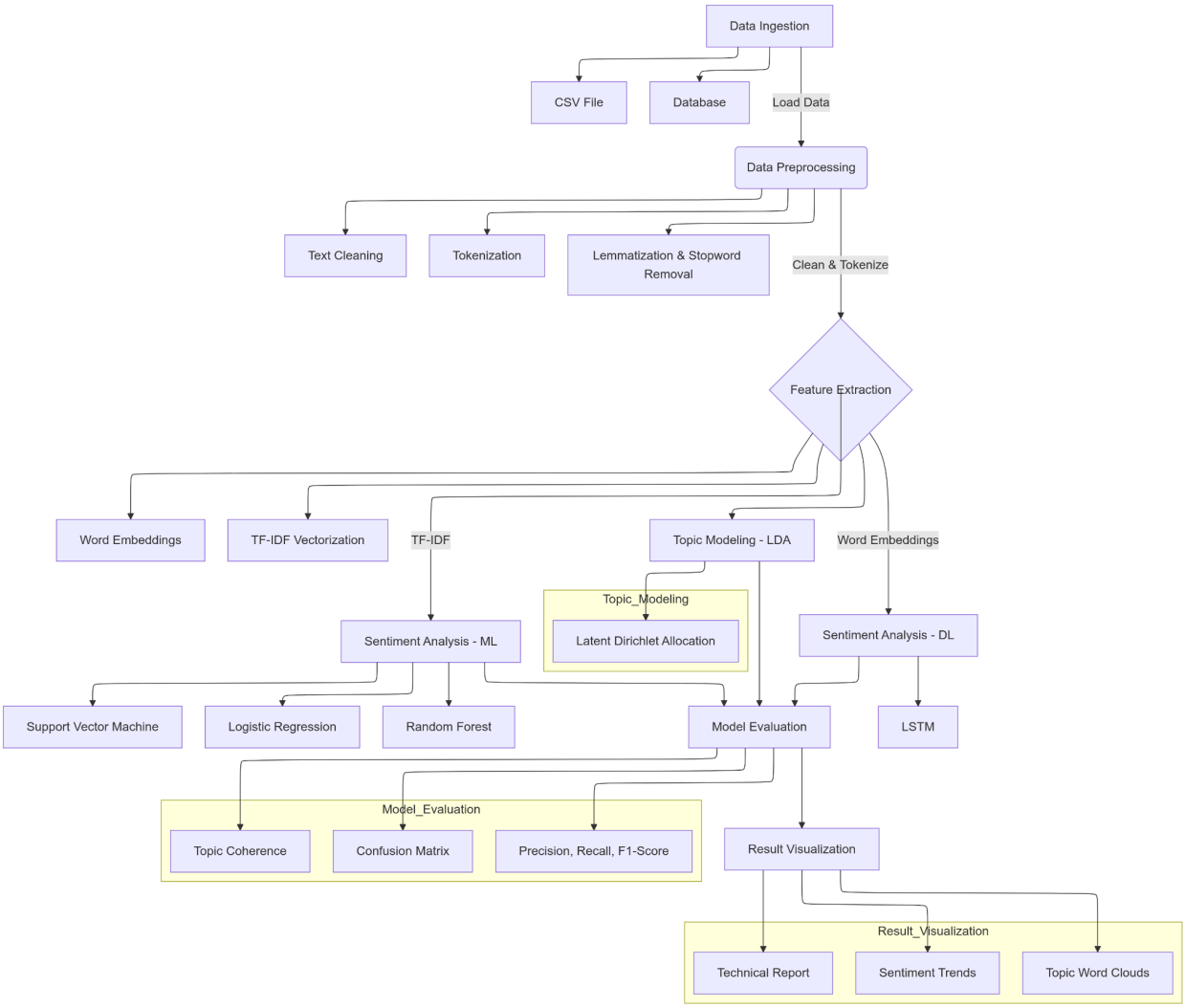
1. High-Level Design (HLD)

The HLD usually provides an overview of the project architecture, including the main components, data flow, and overall system design.

Tasks Identified from HLD:

- **Requirement Analysis**: Identifying project requirements, including data sources, sentiment analysis scope, and topic modeling goals.
- **Data Collection and Preprocessing**: Gathering raw data from various sources (e.g., social media, reviews) and preprocessing it by cleaning, tokenizing, and normalizing text data.
- **Sentiment Analysis Module**: Developing a module to classify sentiments (positive, negative, neutral) using ML models or rule-based approaches.
- **Topic Modeling Module**: Implementing a topic modeling component (e.g., LDA, NMF) to extract dominant themes/topics from the data.
- **Data Storage**: Storing processed and analyzed data in a structured format for further analysis and reporting.
- **Visualization and Reporting**: Creating dashboards and reports to visualize sentiment trends and topic distributions.



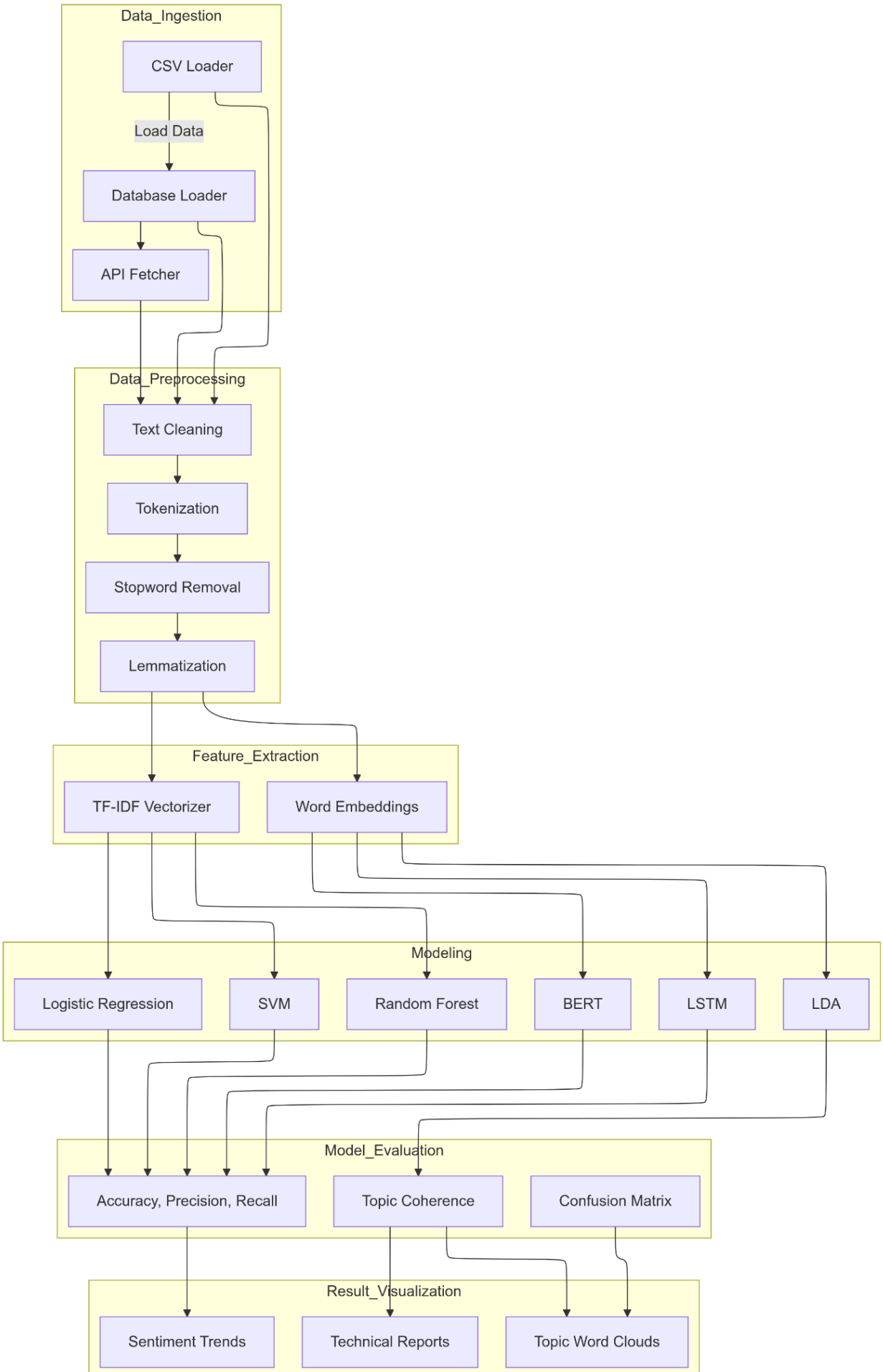
2. Low-Level Design (LLD)

The LLD provides detailed technical descriptions of the modules and processes, specifying how each component of the system is implemented.

Tasks Identified from LLD:

- **Data Ingestion Pipeline**: Implementing pipelines to fetch and load data from various sources into the system.
- **Text Preprocessing Techniques**: Implementing techniques such as stop word removal, stemming, lemmatization, and part-of-speech tagging.

- **Model Selection and Training**: Choosing appropriate models (e.g., SVM, Naive Bayes for sentiment analysis) and training them with labeled datasets.
- **Hyperparameter Tuning**: Optimizing model performance through parameter tuning and cross-validation.
- **Model Evaluation**: Evaluating models using accuracy, precision, recall, F1-score, and confusion matrix metrics.
- **Integration with Topic Modeling**: Merging sentiment analysis outputs with topic modeling to provide comprehensive insights.
- **Error Handling and Logging**: Implementing error handling, logging mechanisms, and monitoring to ensure robust system operations.

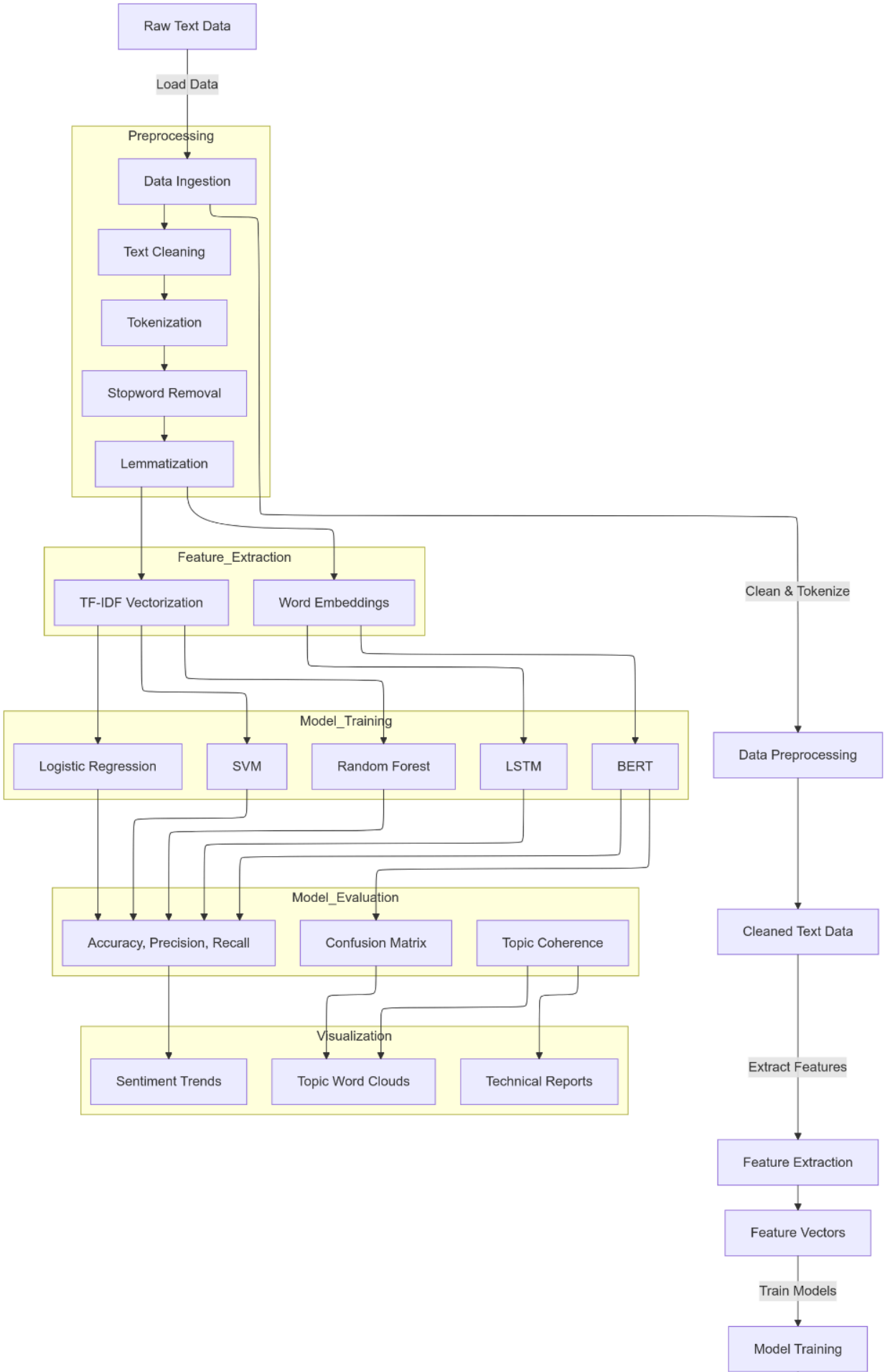

Data_Ingestion
CSV Loader
Load Data
Database Loader
API Fetcher
Data_Preprocessing
Text Cleaning
Tokenization
Stopword Removal
Lemmatization
Feature_Extraction
TF-IDF Vectorizer
Word Embeddings
Modeling
Logistic Regression
SVM
Random Forest
BERT
LSTM
LDA
Model_Evaluation
Accuracy, Precision, Recall
Topic Coherence
Confusion Matrix
Result_Visualization
Sentiment Trends
Technical Reports
Topic Word Clouds

3. Data Flow Diagram (DFD)

The DFD visualizes the flow of data through the system, highlighting key processes, data stores, and interactions.

Tasks Identified from DFD:

- **Data Input and Validation**: Validating input data for correctness and completeness before processing.
- **Data Transformation**: Converting raw data into a structured format suitable for analysis.
- **Sentiment Processing**: Analyzing sentiments from data using predefined workflows and models.
- **Topic Extraction**: Extracting topics using algorithms and linking them with sentiment data.
- **Result Compilation and Storage**: Aggregating results and storing them for reporting and visualization.


Raw Text Data
Load Data
Preprocessing
Data Ingestion
Text Cleaning
Tokenization
Stopword Removal
Lemmatization
Feature_Extraction
TF-IDF Vectorization
Word Embeddings
Model_Training
Logistic Regression
SVM
Random Forest
LSTM
BERT
Model_Evaluation
Accuracy, Precision, Recall
Confusion Matrix
Topic Coherence
Visualization
Sentiment Trends
Topic Word Clouds
Technical Reports
Clean & Tokenize
Data Preprocessing
Cleaned Text Data
Extract Features
Feature Extraction
Feature Vectors
Train Models
Model Training