📘 LOW LEVEL DESIGN (LLD)

1. Data Ingestion

- Input Format: CSV file
- Data Fields:
    - Date
    - Symbol
    - Open
    - High
    - Low
    - Close
    - Volume
    - Market Capitalization

---

2. Data Preprocessing

The following preprocessing steps are applied:

- Convert date column to datetime format
- Sort data by symbol and date
- Handle missing values by removing incomplete records
- Normalize numerical features using StandardScaler

---

3. Feature Engineering

New features are created to improve volatility prediction:

- **Daily Returns**: Percentage change in closing price
- **Rolling Volatility (Target Variable)**: 7-day rolling standard deviation
- **Moving Averages**: 7-day and 14-day averages
- **Liquidity Ratio**: Volume divided by market capitalization
- **Bollinger Bands**: Upper and lower volatility bands
- **Average True Range (ATR)**: Measures price movement range

---

4. Model Design

- Model Type: Supervised Regression
- Algorithm: Random Forest Regressor
- Input: Engineered numerical features
- Output: Predicted volatility value

---

5. Model Training

- Dataset split into training (80%) and testing (20%)
- Model trained using training data
- Hyperparameter tuning using GridSearchCV

---

6. Model Evaluation

Performance is evaluated using:

- Root Mean Squared Error (RMSE)

- Mean Absolute Error (MAE)
- R^2 Score

---

7. Deployment (Optional)

A Streamlit web application allows users to input feature values and receive predicted volatility output.