3.1. Problem Definition

- **Objective**: Predict future FOREX exchange rates or trends with optimised accuracy.
- **Scope**: Focus on short-term, medium-term, or long-term predictions based on specific trading strategies.
- **Key Metrics**: Use metrics like Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and directional accuracy (e.g., % of correct trend predictions) to evaluate performance.

3.2. Data Collection

- **Source Data**:

    Historical Market Data: Exchange rates, volume, and volatility (from financial APIs like Alpha Vantage, OANDA, or Yahoo Finance).

    Macroeconomic Indicators: GDP, inflation rates, interest rates, etc.

    Market Sentiment Data: News articles, social media sentiment, and financial reports.

- **Frequency and Granularity**:

    Minute, hourly, or daily data depending on the prediction horizon.

3.3 Data Preprocessing

- **Data Cleaning**:

    Remove missing, duplicate, or anomalous values.

    Address outliers using statistical methods or domain knowledge.

- **Feature Engineering**:

    Technical Indicators: Moving averages, RSI, Bollinger Bands, etc.

    Sentiment Scores: Extract sentiment polarity using Natural Language Processing (NLP) tools.

    Lagged Features: Include lagged exchange rates as predictors for time-series patterns.

- **Normalization**:

    Scale features to a standard range (e.g., Min-Max Scaling or Standardization) for model compatibility.

3.4. Hybrid Model Development

3.4.1. Model Selection

- Choose models based on their strengths for specific aspects of prediction:

    Long-Term Trends: Statistical models like ARIMA.

    Non-Linear Dependencies: Deep learning models like LSTM or GRU.

    Short-Term Volatility: Ensemble models like Gradient Boosting (XGBoost, LightGBM).

    Market Sentiment: NLP models (e.g., BERT or VADER) for textual data.

3.4.2. Model Integration

1. **Parallel Hybridisation**:

    Run models in parallel on the same input features.

    Combine their outputs using techniques like:

    > Weighted averaging.

    > Meta-model stacking.

2. **Sequential Hybridisation**:

    Use one model's output as input to another.

    Example: Use ARIMA to forecast trends, then pass residuals to an LSTM for fine-tuning.

3.4.5. Training and Validation

Data Splitting:

    Split into training, validation, and test sets (e.g., 70/15/15).

Use time-series cross-validation for temporal consistency.

- **Hyperparameter Tuning**:

  Use techniques like grid search or Bayesian optimisation for individual models and the hybrid ensemble.

- **Avoid Overfitting**:

  Implement regularisation (e.g., dropout for neural networks).

  Early stopping based on validation loss.

3.4.6. Performance Evaluation

- Evaluate models on test data using:

  Prediction Accuracy: MSE, RMSE, or R-squared.

  Directional Accuracy: Percentage of correct upward/downward trend predictions.

  Economic Metrics: Simulate trading strategies and calculate profits/losses.

- Compare hybrid models to baseline models (e.g., single LSTM or ARIMA).