

Assignment 2: Multi-Dimensional Return Forecasting and Portfolio Management

Project Objective:

Each team has to build a predictive system for a 6-stock Indian equity universe (RELIANCE, HDFCBANK, INFY, TAMO, BHARTIARTL, HUL). The goal is to synthesize structured financial data, unstructured text sentiment, and macroeconomic indicators into a robust machine learning pipeline to forecast returns and manage a forward-looking portfolio.

Data:

- **Market Data (OHLCV):** Fetch daily adjusted prices and volume for the period **Jan 1, 2020, to Dec 31, 2025**. Keep Oct 2025 -Dec 2025 data separate for final forward testing.
 - Source: Yahoo Finance.
- **Fundamental Data:** Extract quarterly metrics (P/E, Debt/Equity, ROE, EPS), etc. Use **a suitable method** to align quarterly data with daily market timestamps.
 - Source: MoneyControl
- **Macro Indicators:** Inflation, Integrate daily/monthly USD-INR rates, India 10Y Bond Yields, and Crude Oil prices, etc.
 - Sources: RBI data, Yahoo Finance
- **Alternative Data (Sentiment):** Scrape or API-fetch financial news headlines. Use a pre-trained transformer model to generate daily sentiment polarity scores.
 - Sources: Google News API + FinBERT

Feel free to use other data and data sources.

Core Requirements

1. Fetch appropriate data
2. Do necessary pre-processing and feature engineering
3. Train a model to predict the returns, report with relevant metrics and validate the predictions.
4. Make sure your model/models are not overfitting and working for all these symbols.
5. Combine the 6 stocks and create a portfolio using their predicted daily returns and report the performance of the portfolio for 3 months in future. You can choose any method for finding your portfolio weights.

Notes:

- **Eliminate Look-ahead Bias:** Ensure that data at time T only uses information available before T . All features (especially sentiment and macro) must be appropriately lagged.
- **Stationarity & Scaling:** Transform non-stationary price data into log returns. Normalize features using techniques like Robust Scaling to handle financial outliers.
- **Model Validation:** Utilize Time-Series Walk-Forward Validation (Nested Cross-Validation). Standard K-Fold shuffle is strictly prohibited as it destroys the temporal order of financial data.
- **Overfitting Safeguards:** Implement feature selection (e.g., Recursive Feature Elimination) and regularization to prevent the model from "memorizing" noise.
- **Cross-Symbol Robustness:** The model must demonstrate stable performance across all 6 tickers; avoid "over-tuning" to a single high-beta stock like TAMO.

Deliverables:

1. Zip file with well documented codes and data files.
2. 3-pager note consisting
 - a. Data and Feature Engineering
 - b. Model architecture and validation
 - c. Feature Importance
 - d. Model Results
 - e. Performance metrics like **Sharpe Ratio**, **Maximum Drawdown**, and **Hit Ratio** (directional accuracy) and Equity Curve