

FINAL YEAR PROJECT (1)

Roll # B21120206024

**Financial Computer Simulation (FM 617)
BS-Financial Mathematics (4TH Year)**

**Submitted to,
Sir Daniyal Hussain**

**Submitted by,
Jaweria Irfan**

CODING METHODOLOGY

1 .DATA LOADING AND PREPROCESSING

- STOCK PRICE DATA IS LOADED FROM A CSV FILE WITH COLUMNS: DATE, OPEN, HIGH, LOW, CLOSE, VOLUME.
- DATA IS SORTED BY DATE AND MISSING VALUES FORWARD-FILLED.
- FEATURE ENGINEERING PERFORMED INCLUDES:
 - LOGARITHMIC RETURNS ('LOGRETURN') TO MEASURE DAILY RETURNS.
 - MOVING AVERAGES ('MA10', 'MA20') AS TREND-FOLLOWING INDICATORS.
 - ROLLING VOLATILITY ('VOLATILITY') OVER 10-DAY WINDOWS AS A RISK METRIC.
 - ROWS WITH MISSING VALUES AFTER FEATURE CONSTRUCTION ARE DROPPED TO ENSURE CLEAN DATA.

2 .EXPLORATORY DATA ANALYSIS (EDA)

- VISUALIZED CLOSING PRICE OVER TIME.
- PLOTTED HISTOGRAM AND KDE OF LOG RETURNS TO ANALYZE STATISTICAL PROPERTIES.
- GENERATED CORRELATION HEATMAP TO UNDERSTAND RELATIONSHIPS BETWEEN FEATURES.



3.MODELING

- Linear Regression model predicts closing price using engineered features (MA10, MA20, Volatility).
- ARIMA time series model forecasts future closing prices based only on closing price history.
- Data split preserves time order (train on first 80%, test on last 20%).

4.EVALUATION

- Models evaluated with RMSE, MAE, and R^2 metrics
- Generated plots comparing actual vs predicted or forecasted prices.



FINANCIAL INTERPRETATION

- **Logarithmic Returns:** Provide continuous compounding returns for better statistical properties in analysis.
- **-Moving Averages:** Capture momentum and smooth out short-term noise; useful for trend signals.
- **Volatility:** Quantifies recent price variability and market risk, influencing expected returns.
- **Model Insights:**
 - Linear Regression relates trending features and volatility to predict future prices, capturing market influences.
 - ARIMA captures time-dependencies and repeated patterns in price data to forecast future values.
- **Practical Usage:**
 - Helps investors/analysts understand price dynamics.
 - Supports decision making for trading and risk management.



CODE WORKFLOW

1. Load CSV Data:

- Read and parse dates, sort for time series consistency.

2. Preprocess Data:

- Fill missing values.
- Calculate log returns, moving averages, volatility.
- Remove any incomplete rows.

3. EDA Visualization:

- Plot price series, return distribution, and feature correlations.

4. Model Training and Prediction:

- Linear Regression: Train on engineered features; predict closing prices.
- ARIMA: Fit on closing price time series; forecast future prices.

5. Evaluation and Visualization:

- Compute error metrics.
- Plot actual vs predicted/forecasted closing prices.

6. Output Results:

- Print metrics in console.
- Display graphical plots.

RESULTS

EDA Observations:

- The stock shows typical price fluctuations and trend changes.
- Log returns appear approximately normally distributed.
- Strong correlation seen among closing price and moving averages, moderate correlation with volatility.

Linear Regression Performance:

- High R^2 (~ 0.95) indicates good fit.
- Low RMSE and MAE show accurate prediction on test data.
- Predictions closely follow actual prices.

ARIMA Performance:

- Reasonable forecasting ability for short-term horizon (e.g., 30 days).
- Higher errors and sometimes negative R^2 indicate difficulties modeling volatile stock prices.
- Forecast captures basic trends but less responsive to sudden changes.

Visualizations:

- Actual vs Predicted plots validate model quality.
- Linear Regression smoother and closer to true values.
- ARIMA forecasts show general future trajectories.