STOCK PRICE PREDICTION FOR AAPL

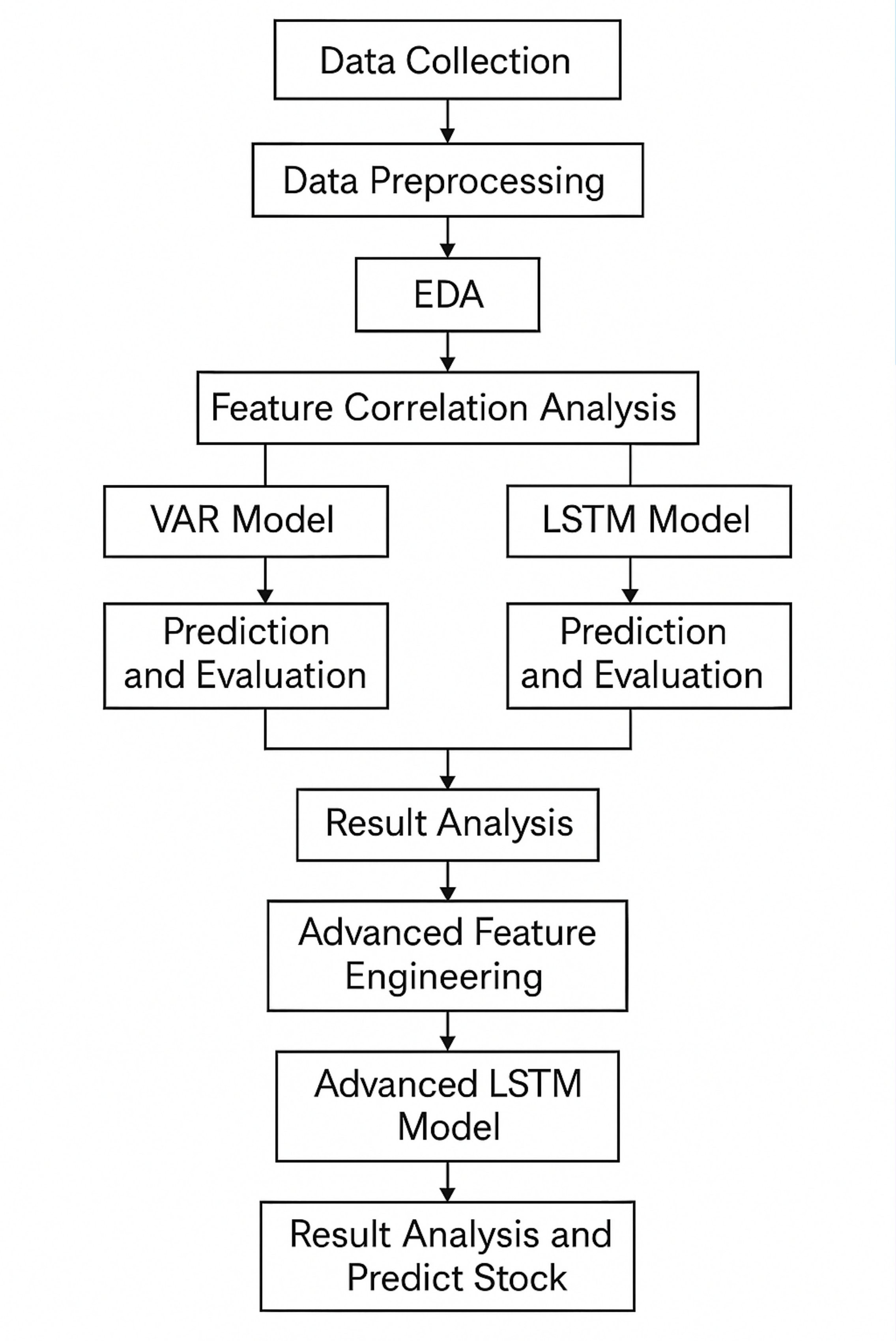
Final Presentation

CRN 23847 04/23/2025

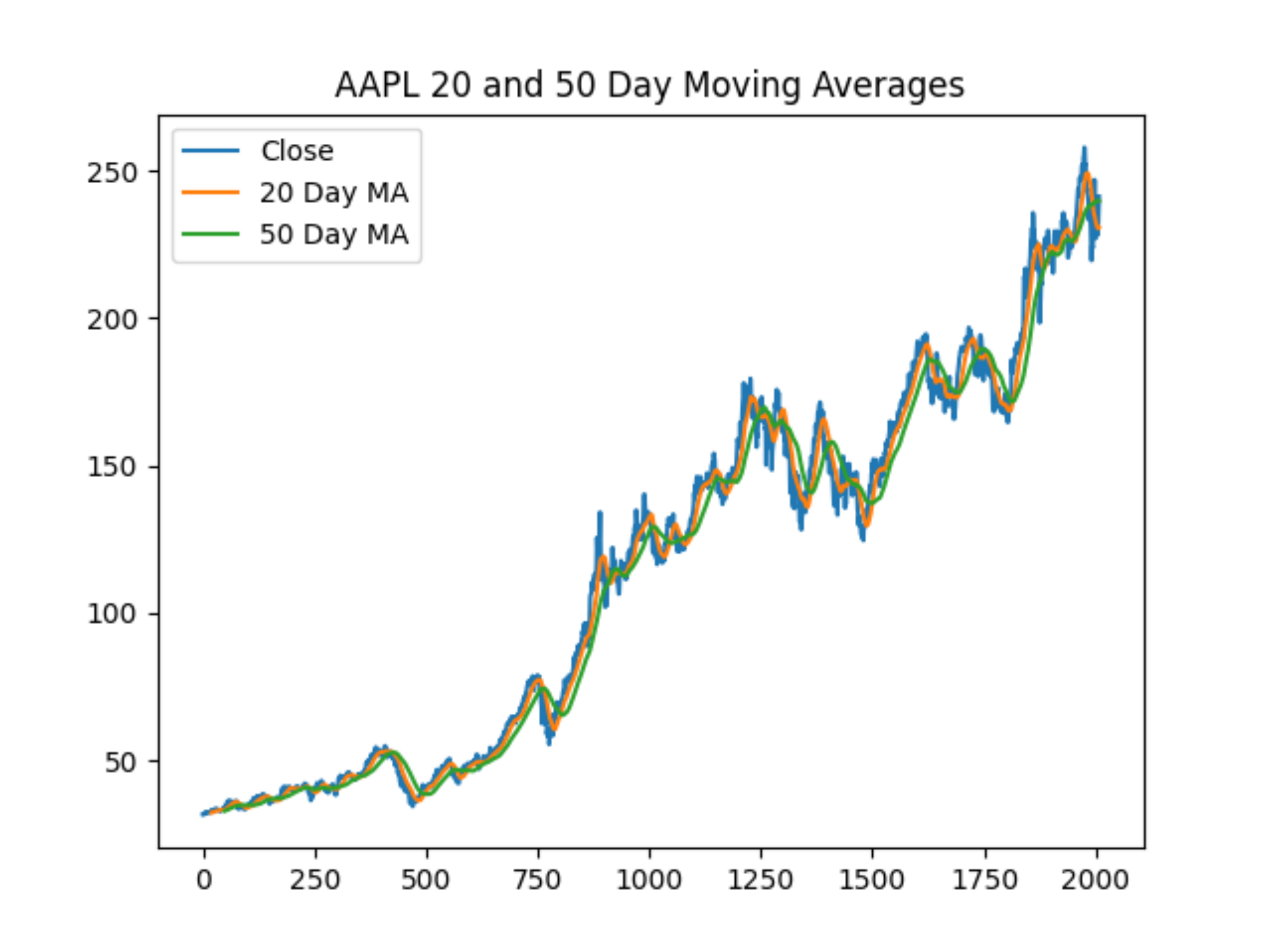
University of Central Missouri

Project Data Flow Overview

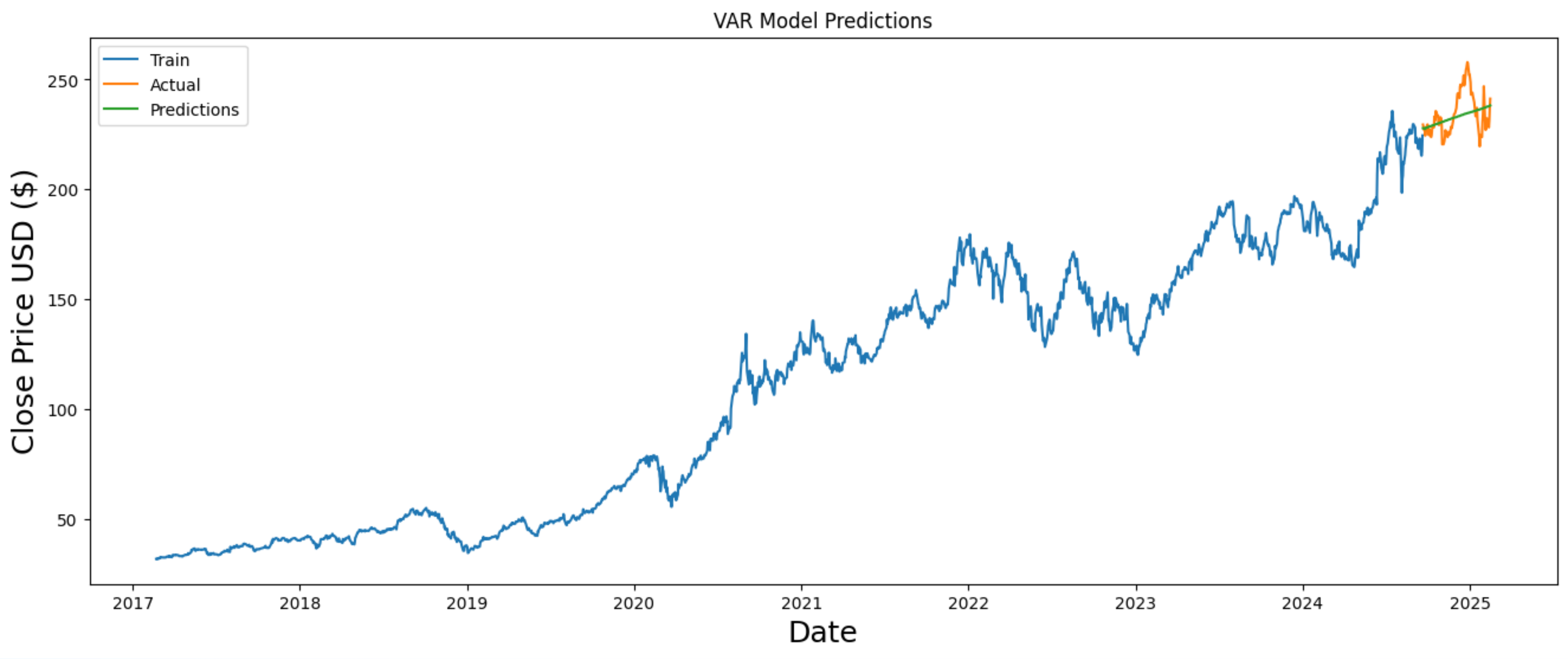
- Data Collection: Historical stock data for AAPL, MSFT, GOOG, AMZN via Yahoo Finance (yfinance API)
- Data Validation: Format checks, missing value handling, date consistency, stock integrity verification
- Data Visualisation & EDA: Time series plots, correlation matrix, moving averages, volume trends
- VAR Model: Baseline linear model capturing inter-stock correlations (multivariate time series)
- LSTM model: Univariate model on AAPL close prices to model long-term dependencies
- Advanced Feature Engineering: Rolling averages, momentum indicators, lag features to enrich input data
- Advanced LSTM Model: Deep LSTM layers with regularization, trained on engineered features
- Conclusion: Hybrid performance evaluation using MSE, RMSE, R^2



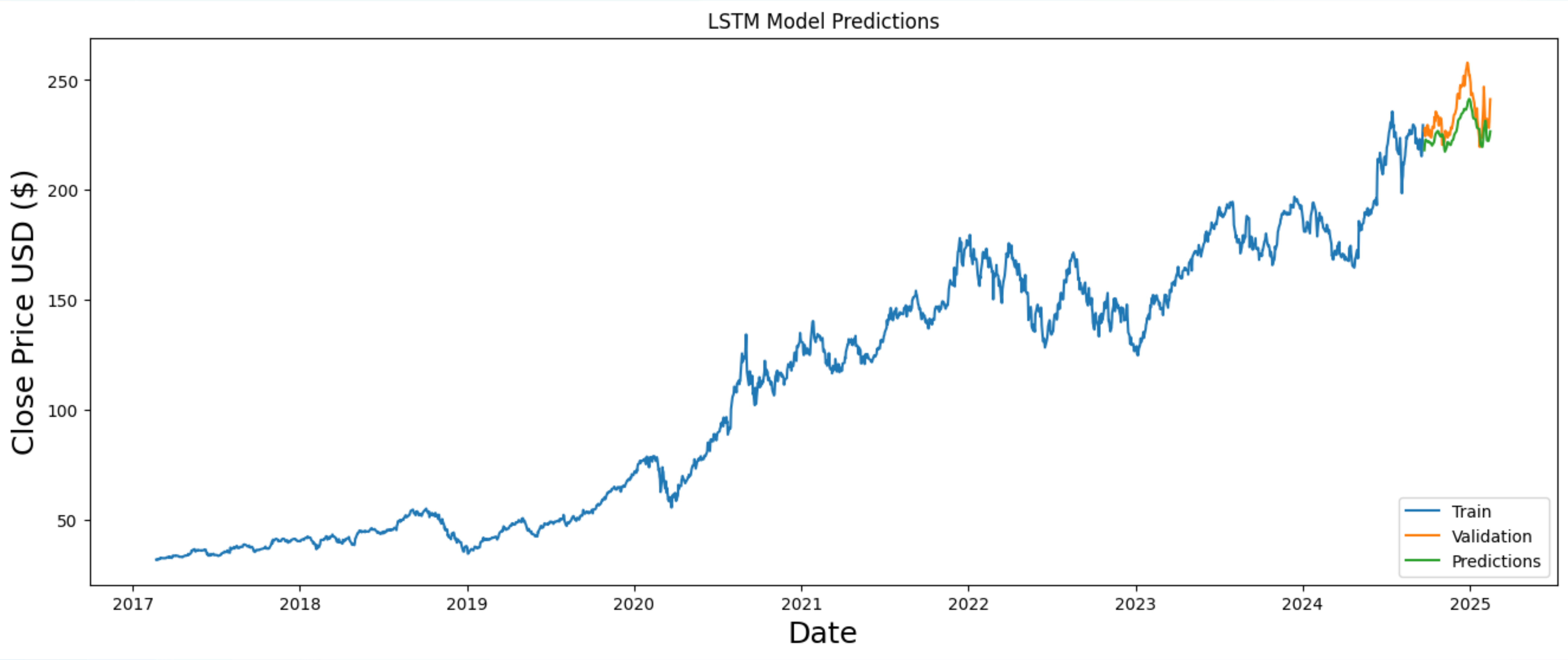
- Data Collection:
  - Source: Yahoo Finance via yfinance API
  - Stored as Pickle for efficient reuse
- Data Validation:
  - Checked data types, range, missing/null values
  - No NaNs, consistent datetime index
  - Ensured shape: (14,063 rows × 7 columns)
- EDA & Visualization:
  - Closing Price, Volume, Daily Returns, Correlation Matrix
  - Used rolling averages & cumulative returns to observe trends



- Why VAR?
  - Captures linear interdependencies across multiple stock features (Close, Open, High, Low, Volume)
- Methodology:
  - Data normalized with MinMaxScaler
  - Train-Test Split: 95% train, 5% test (chronological split)
  - Best lag selected using Akaike Information Criterion (AIC)
- Evaluation Metrics:
  - MSE, RMSE, MAE, R^2 used to assess prediction accuracy
- Tools: statsmodels, sklearn, pandas, Google Co-lab



- Why LSTM?
  - Captures long-term dependencies and nonlinear patterns in stock time series
- Architecture:
  - 2 LSTM layers with Dropout (10%)
  - Dense layers with L2 regularization
  - Trained for 50 epochs with EarlyStopping
- Training Strategy:
  - 60-day lookback window
  - 95% Train / 5% Test split
  - MinMaxScaler normalization
- Metrics:
  - MSE, RMSE, MAE, R^2 Score

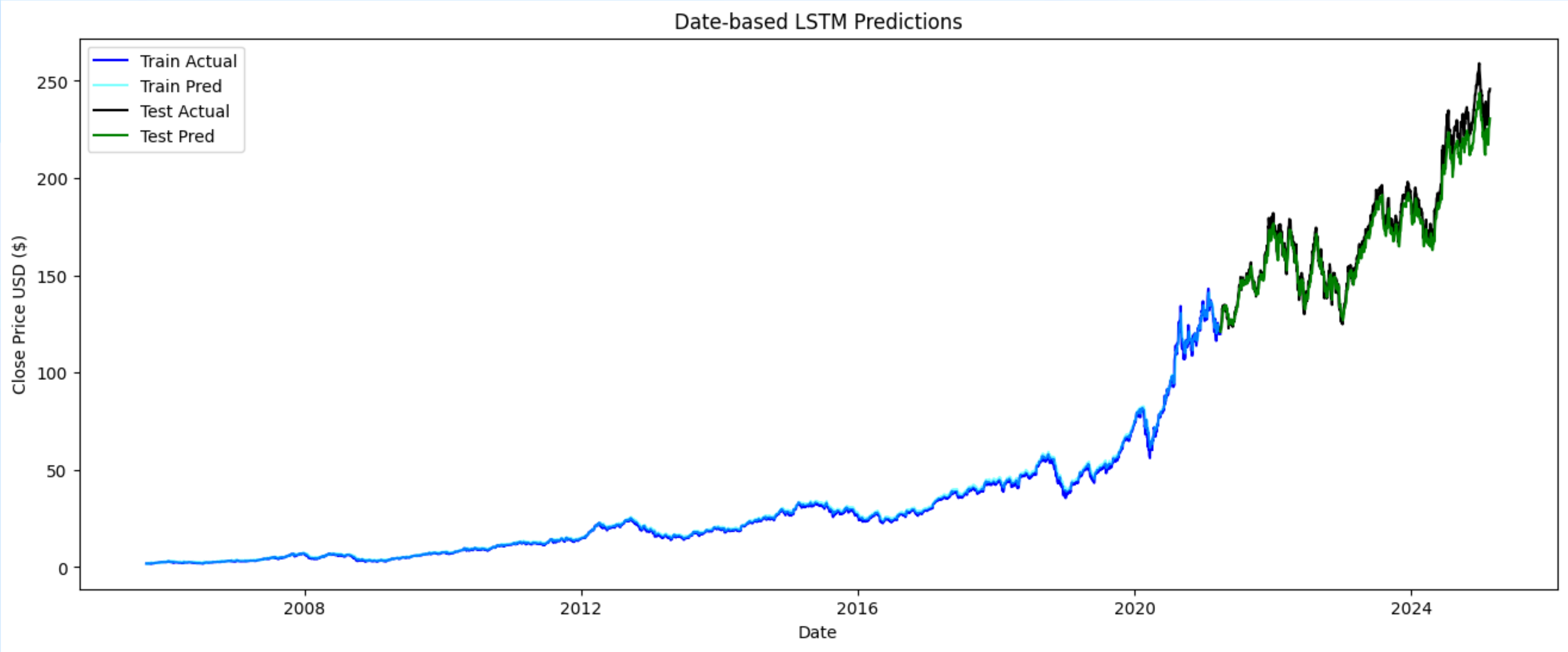


Advanced Feature Engineering

- Engineered Features:
  - RSI (Relative Strength Index)
  - EMA – Fast (20), Medium (100), Slow (150)
  - Next Day Close Price, Daily Target (Adj Close – Open)
  - Binary Target Classification (Price ↑ or ↓)
  - Date-based: Year, Month, Day of Week
- Feature Set:
  - 11-dimensional input including indicators and temporal context
  - 30-day rolling window (backcandles) to feed sequential data into LSTM
- Output:
  - X: Feature sequences (for model input)
  - y: Next-day adjusted close price

- Key Improvements in Advanced LSTM:
  - Integrated engineered features (RSI, EMAs, date parts)
  - Increased sequence dimensionality (11 features vs 1)
  - Model input: 30-day rolling window of multivariate data
- Model Architecture:
  - LSTM Layer (150 units) → Dense Layer → Linear Activation
  - Trained for 50 epochs with Adam optimizer and EarlyStopping
  - Trained for 50 epochs with EarlyStopping

- Performance Metrics:
  - RMSE ↓, MAE ↓, R^2 ↑ (compared to basic LSTM)
  - Closer predictions across volatility phases



Final takeaway

- VAR captured linear stock relationships but struggled with volatility
- Basic LSTM improved trend prediction using sequential memory
- Advanced LSTM outperformed both with engineered features and deeper learning
- Our hybrid approach boosts accuracy, making it valuable for financial forecasting
- Future extensions: news sentiment, technical indicators, live model deployment

Harshith Reddy Gundra. 700780724
Nithin Aaron Kommireddy. 700752900
Bhargav Sai Allam. 700752883
Varshith Thoutu. 700759296