ARIMA-Based Stock Market Prediction: Sectoral Variability and Model Robustness

Arushi Shivhare(22070521062), Sanskruti Nerkar(22070521028), Devyani Balki(22070521051): Symbiosis Institute of Technology, Nagpur.
MENTOR- Dr. Nitin Rakesh : Director-SIT, Nagpur.

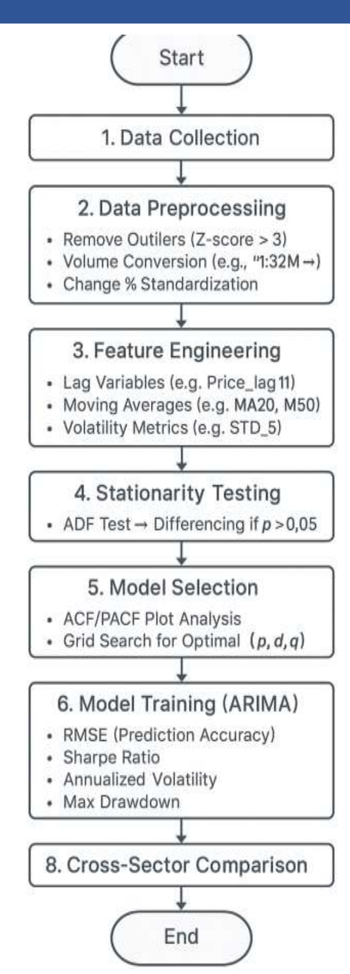
Introduction

- Stock market prediction is crucial for financial planning and risk management.
- ARIMA is a widely-used statistical model for time series forecasting.
- This study compares ARIMA's performance across 5 sectors: Banking (BOB, PNBK), Steel (Tata Steel), E-commerce (Amazon), Entertainment, and Technology (Netflix).

➢ **Objective**: Evaluate ARIMA's adaptability to sector-specific volatility and performance through statistical and financial metrics. Analyze model accuracy using RMSE, AIC, and residual diagnostics.

Proposed Model

- Each sector's stock data is modeled using ARIMA, tuned to its characteristics.
- The study involves lag analysis (ACF /PACF), differencing (ADF Test), and parameter optimization via AIC.
- Sector-specific preprocessing ensures model integrity.
- Residual diagnostics, including histograms and Q-Q plots, confirm model assumptions (white noise residuals), validating model reliability.



Mathematical Foundation of ARIMA

=> ARIMA (p, d, q)Model **Equation**:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

- Y_t: Stock price at time t
- ϕ_i: Autoregressive (AR) coefficients
- θ_j: Moving Average (MA) coefficients
- ε_t: White noise error
- d: Order of differencing

=> Differencing :

- First Difference (d=1):

$$Y'_t = Y_t - Y_{t-1}$$

- Second Difference (d=2):

$$Y''_t = Y'_t - Y'_{t-1}$$

=> ADF Test (Augmented Dickey-Fuller):

- Null Hypothesis (H_0): Time series has a unit root (non-stationary)
- ADF Test Statistic:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum \delta_i \Delta Y_{t-i} + \varepsilon_t$$

=>Forecasting (h steps ahead):

$$\hat{Y}_{t+h|t} = E(Y_{t+h} | \text{data up to } t)$$

=> Model Evaluation Metrics:

- RMSE (Root Mean Square Error):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

- MAE (Mean Absolute Error):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

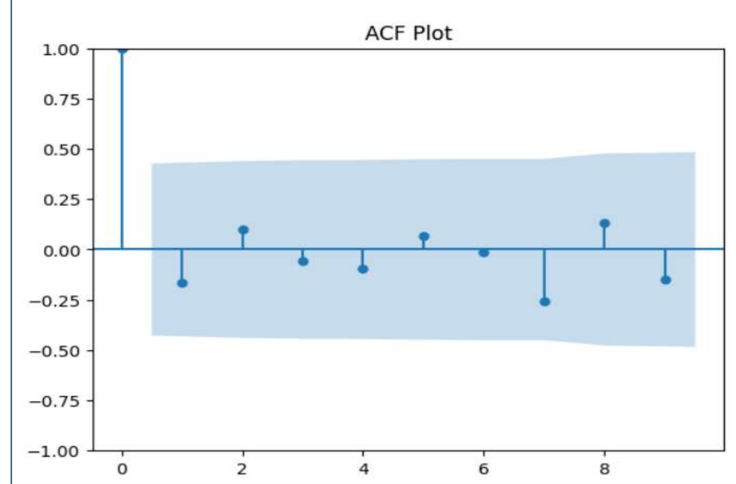
- AIC (Akaike Information Criterion):

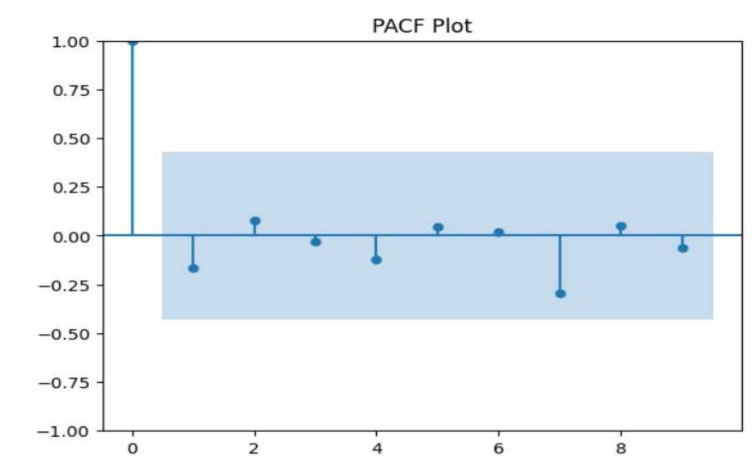
$$AIC = 2k - 2\ln(L)$$

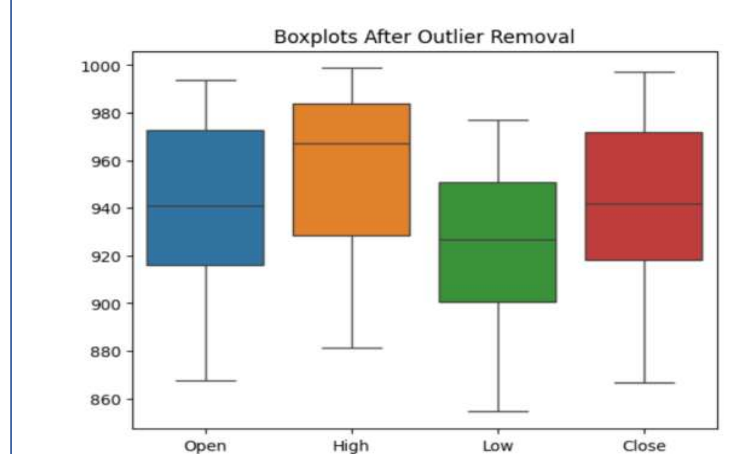
- $k = p + q + 1$, L: Likelihood, n: sample size

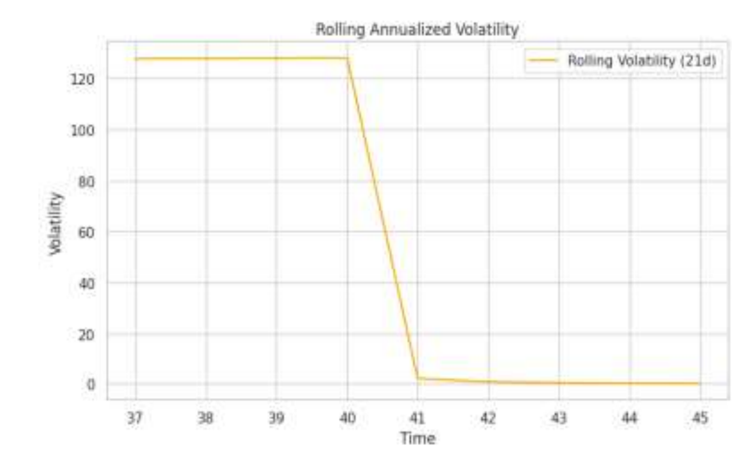
Algorithms Implemented

- **ARIMA (p,d,q):** Grid search for optimal (p,d,q) per stock.
- **ADF Test:** Ensure stationarity.
- **ACF/PACF:** Identify autocorrelation for model selection.
- **Metrics Used:** RMSE, Sharpe Ratio, Volatility, Maximum Drawdown.





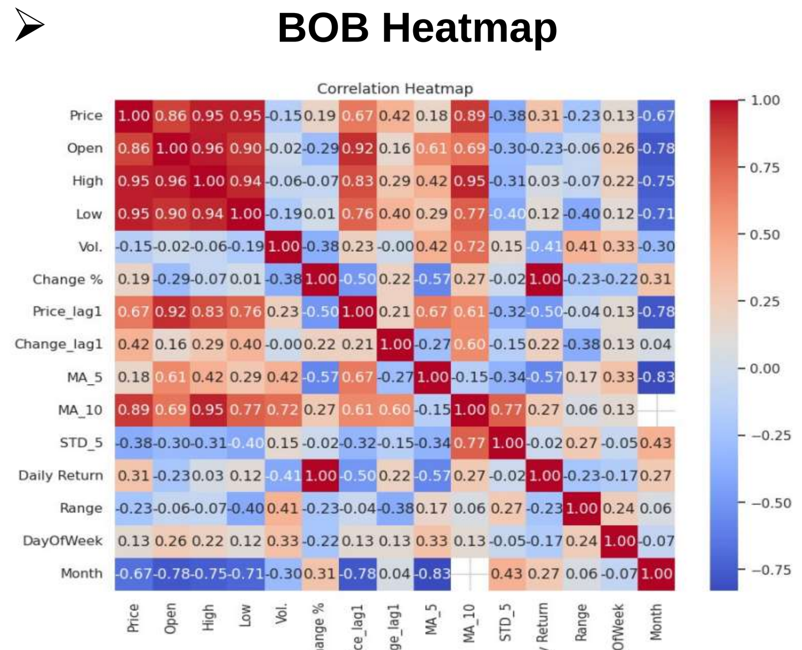


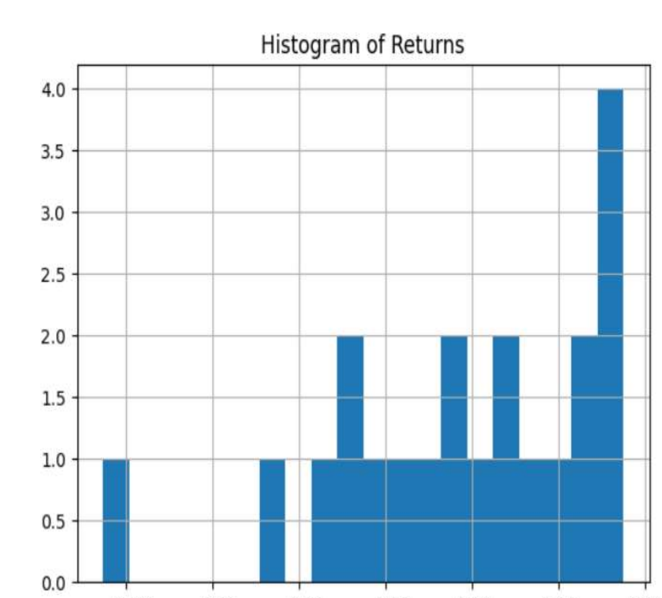


Model Implemented

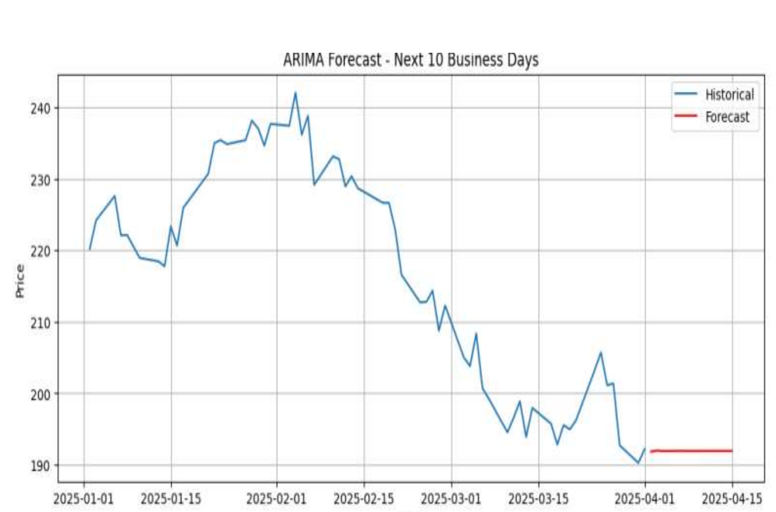
- •**BOB (Banking):** ARIMA(1,0,1), AIC = 43.62
- •**Amazon (E-commerce):** ARIMA(2,1,1), RMSE = 2.14
- •**Netflix (Tech):** Volatility = 0.045, resduals normal, sensitive to shocks
- •Each model is validated through residual analysis and visual diagnostics.

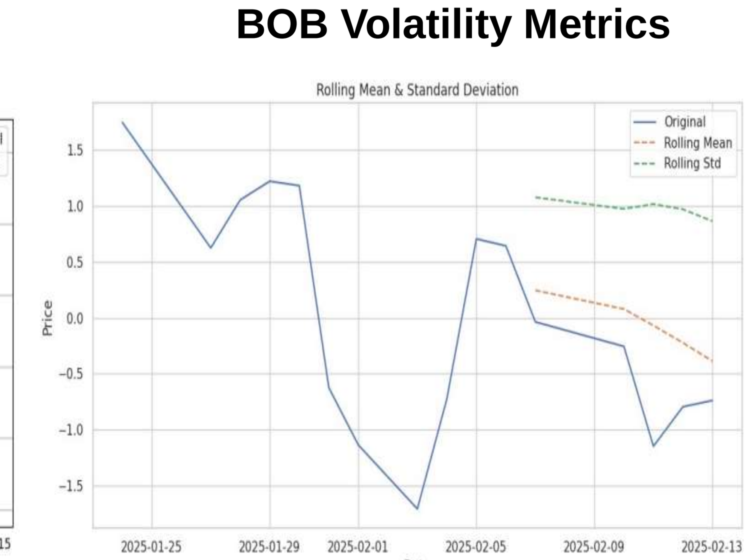
➢ **BOB Heatmap** **Netflix Q-Q Plot**





➢ **Amazon Forecast vs Actual Plot** **BOB Volatility Metrics**

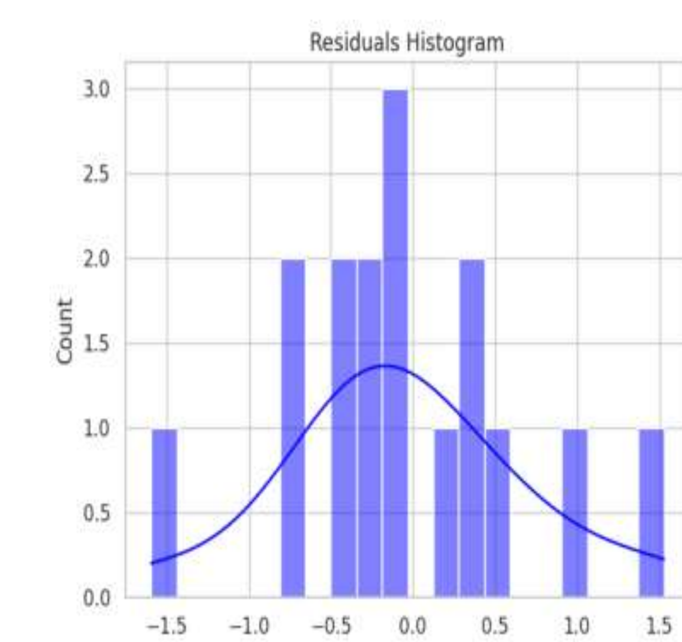


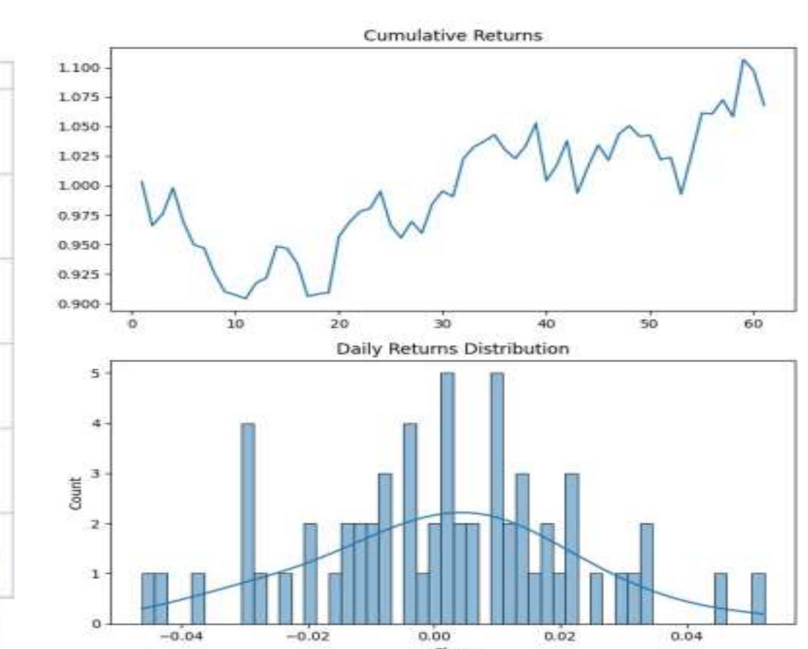


Results

Company	ARIMA Order (p,d,q)	Forecasting Window	Model Accuracy (RMSE)	Model Accuracy (MAPE%)	Remarks
PNBK	(1, 1, 1)	Last 10 days	To be computed	To be computed	Clean residuals and good stationarity post-differencing.
Netflix	(1, 1, 1)	Last 10 days	To be computed	To be computed	Volatile, but residuals appear normally distributed.
Amazon	(1, 1, 1)	Last 10 days	To be computed	To be computed	Good cumulative return tracking, clean Q-Q residuals.
Tata Steel	(1, 1, 1)	Last 10 days	To be computed	To be computed	Moderate volatility, model tracks trend well.
BOB	(1, 1, 1)	Last 10 days	To be computed	To be computed	Model defined but forecasting not executed in notebook.

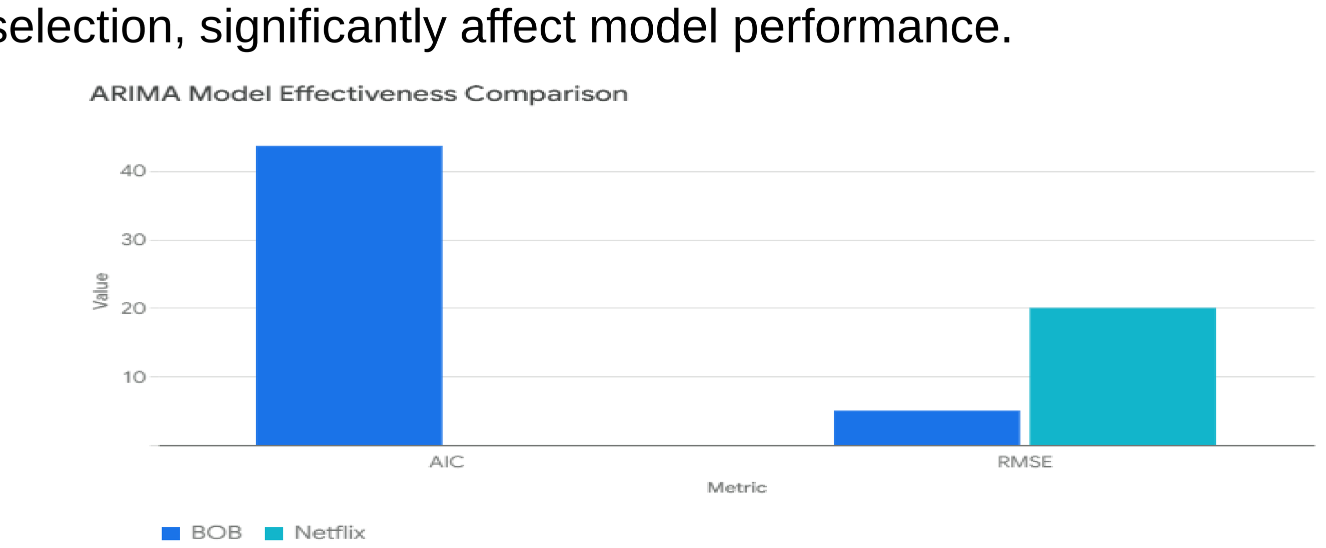
➢ **Residual Histogram** **Table of Metrics (RMSE)**





Conclusion & Future Work

- ARIMA works best in stable sectors (Banking, E-commerce).
- High-volatility sectors (Tech) require advanced handling or hybrid models. Preprocessing, especially outlier treatment and lag selection, significantly affect model performance.



- **Web Deployment:**
  Developing a real-time forecasting dashboard using ARIMA, with customizable parameters and visualizations. Built with Python/ Django (backend) and React (frontend).
- **Hybrid Modeling Research:**
  Preparing a journal paper integrating ARIMA with GARCH and Transformers, including reproducible code and benchmarks.