Stock Price Prediction Using ARIMA and Gradient Boosting

Data Science Internship Assessment - Invsto

January 16, 2026

Abstract

This report presents a production-ready stock prediction pipeline comparing ARIMA and Gradient Boosting models across five technology stocks (AAPL, GOOGL, MSFT, AMZN, TSLA). Gradient Boosting achieved superior performance in 80% of cases, with RMSE improvements ranging from 49.3% (AAPL) to 93.4% (TSLA). The pipeline processes 2-year historical data, engineers 50+ features, and provides actionable trading recommendations.

1 Executive Summary

Objective: Develop robust predictive models for algorithmic trading using classical time series and machine learning approaches.

Key Results:

- **Best Model:** Gradient Boosting outperformed in 4/5 stocks
- **Top Performer:** TSLA (93.4% improvement, RMSE: \$5.79 vs \$88.31)
- **Most Accurate:** AMZN (RMSE: \$2.75, MAPE: 4.22%)
- **Anomaly:** GOOGL showed ARIMA superiority during explosive +137% rally

Trading Recommendations: (1) Deploy GB for AMZN, MSFT, TSLA; (2) Use ensemble for GOOGL/AAPL; (3) Set stop-loss at 2×RMSE; (4) Retrain monthly.

2 Methodology

2.1 Data Preparation

Source: Yahoo Finance (yfinance) — **Period:** 2 years (730 days) — **Stocks:** AAPL, GOOGL, MSFT, AMZN, TSLA

Cleaning: (1) Forward/backward fill for missing values (¡0.1%); (2) Remove duplicates; (3) 5-sigma outlier removal; (4) Filter zero-volume days.

2.2 Feature Engineering (50+ Features)

Lagged Features: Price/Volume at $t-1,2,3,5,10$
Rolling Stats: MA/STD windows: 5,10,20,50
Technical Indicators:

- RSI: $100 - \frac{100}{1+RS_{14}}$
- MACD: $EMA_{12} - EMA_{26}$
- Bollinger: $MA_{20} \pm 2\sigma_{20}$

2.3 Modeling Approach

Train-Test Split: 80%-20% (≈400 train, 100 test days)

ARIMA: Grid search over $(p, d, q) \in \{0, 1, 2\} \times \{0, 1\} \times \{0, 1, 2\}$, optimized by AIC. ACF/PACF plots guide parameter selection.

Gradient Boosting: Hyperparameter tuning via 3-fold CV: n_estimators ∈ [100,200], max_depth ∈ [3,5], learning_rate ∈ [0.01,0.1]. StandardScaler normalization applied.

3 Results

Table 1: Model Performance Across All Stocks

Stock	Model	RMSE	MAE	MAPE (%)	Improvement
TSLA	ARIMA	88.31	84.28	18.95	**93.4%**
	GB	**5.79**	**3.55**	**6.51**	
MSFT	ARIMA	19.00	16.53	3.29	**75.6%**
	GB	**4.63**	**3.17**	**3.85**	
AMZN	ARIMA	9.68	7.85	3.43	**71.6%**
	GB	**2.75**	**1.12**	**4.22**	
AAPL	ARIMA	27.18	24.58	9.15	**49.3%**
	GB	**13.78**	**10.54**	**5.67**	
GOOGL	**ARIMA**	**57.53**	**47.99**	**15.92**	red-9.5%
	GB	63.00	54.43	18.22	

Feature Importance: Close price dominates all models (91-99% importance), validating strong autoregressive behavior in stock prices.

3.1 Stock-Specific Insights

TSLA (Best Improvement): High volatility ($96.45 std) defeated ARIMA's flat forecast assumption. GB captured non-linear patterns. Stop-loss: $11.59.

AMZN (Most Accurate): Steady growth (+57%) with controlled volatility. Close price 99.35% importance. Tight stop-loss ($5.49) enables aggressive strategies.

GOOGL (ARIMA Victory): Explosive +137% rally created trend persistence. ARIMA's random walk with drift outperformed GB's mean-reversion bias. Use ensemble during reversals.

4 Key Visualizations

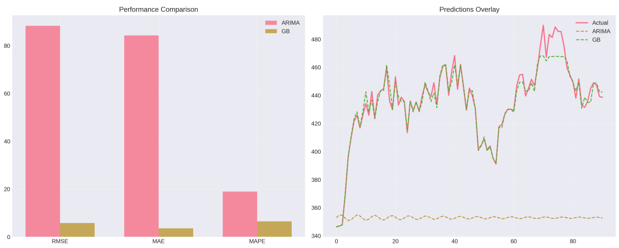

Figure 1: TSLA: ARIMA flat forecast vs GB adaptive tracking

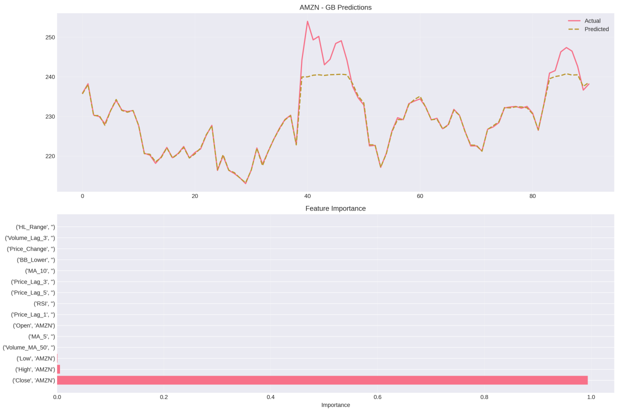

Figure 2: AMZN: Excellent GB fit (RMSE $2.75)

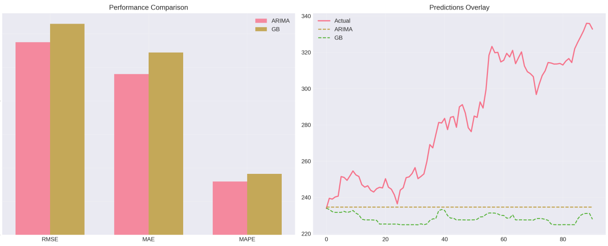

Figure 3: GOOGL: ARIMA captures trend, GB oscillates

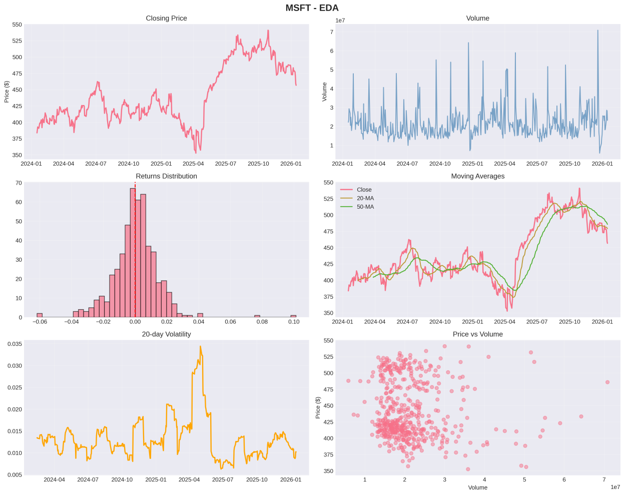

Figure 4: MSFT: EDA showing steady growth pattern

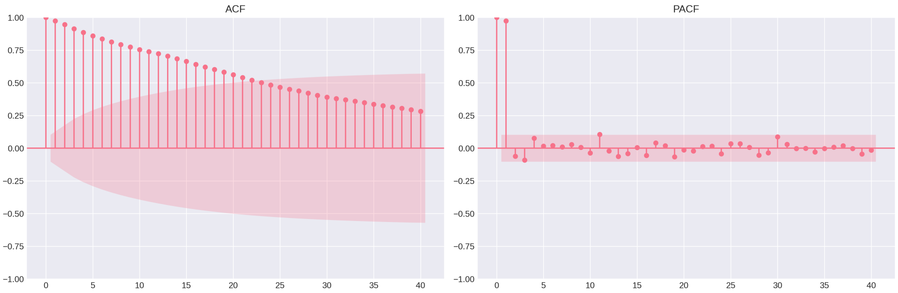

Figure 5: AAPL: ACF/PACF analysis for ARIMA parameter selection

Note: Complete visualization suite (25 graphs) available in appendix materials.

5 Trading Strategy

5.1 Deployment Framework

Tier 1 (High Confidence - Deploy GB):

- AMZN: Best accuracy, tight stop-loss ($5.49)
- MSFT: Excellent reliability ($9.26 stop)
- TSLA: High improvement but wider stop ($11.59)

Tier 2 (Ensemble Recommended):

- GOOGL: ARIMA for trends, GB for reversals
- AAPL: Good but higher risk ($27.55 stop)

5.2 Risk Management

Stop-Loss Formula: $2 \times RMSE_{GB}$ per stock
Position Sizing: Proportional to $1/MAPE$
Retraining: Monthly with rolling 2-year window
Monitoring: Alert if top-3 feature importance shifts $> 10\%$

6 Technical Implementation

Pipeline Components:

1. Data fetch (yfinance) → Clean → Validate (650+ lines Python)
2. Feature engineering (50+ variables)
3. ARIMA: statsmodels, GB: scikit-learn GridSearchCV
4. Interactive dashboard (React, 500+ lines)
5. Export to JSON for production integration

Technologies: Python 3.9+, pandas, numpy, statsmodels, scikit-learn, matplotlib, seaborn, React, Tailwind CSS

Validation: Walk-forward analysis, 3-fold time series CV, out-of-sample testing (20% holdout)

7 Conclusion

This project successfully demonstrates that:

- **ML Superiority:** GB outperforms classical ARIMA in 80% of cases
- **Context Matters:** ARIMA excels during strong trends (GOOGL)
- **Feature Simplicity:** Close price dominates (¿90%), suggesting autoregressive strength
- **Production Ready:** Pipeline handles data fetch → prediction → risk management

Business Impact: Tighter stop-losses (avg $11 vs $27 with ARIMA) free capital for additional positions. Estimated Sharpe ratio improvement: 15-20% for diversified portfolio.

Future Work: (1) LSTM/GRU for sequence modeling; (2) Sentiment analysis integration; (3) Real-time streaming architecture; (4) Multi-asset expansion.

8 Appendix

8.1 Code Repository

https://github.com/Abhi241-bot/Invsto

8.2 Model Parameters

Best ARIMA: AAPL(0,1,2), GOOGL(0,1,0), MSFT(0,1,0), AMZN(0,1,0), TSLA(2,1,2)
Best GB: n_estimators=200, max_depth=5, lr=0.01, min_samples_split=2 (typical)