Predicting 30-Day Stock Price Movements

Name: Aravindh Palaniguru
UCID: 30301478
Course: DATA 601 - Data Science and Analytics
Date: October 17, 2025

Abstract

Stock market prediction represents one of the most challenging yet valuable applications of machine learning in finance. This study investigates whether historical price and volume data alone can reliably predict 30-day directional stock price movements. Through comprehensive analysis of four major technology stocks (AAPL, AMZN, MSFT, NVDA) spanning twenty years (2005-2025), we demonstrate that while multiple sophisticated algorithms achieve accuracy rates of 38-50.3% across different stocks—marginally superior to random guessing—these improvements are insufficient for practical investment applications.

1. Introduction

Research Motivation and Context

The question of whether historical market data contains predictive information is a useful one for investors. This investigation addresses a fundamental challenge towards this: Can machine learning models trained exclusively on historical price and volume data predict the direction of 30-day stock price movements with sufficient accuracy to generate actionable investment insights?

Successful directional prediction models could enhance portfolio performance and help new investment strategies.

Objectives

Specific Objectives:

1. Evaluate the predictive capability of multiple algorithms on stock price direction
2. Identify the most informative technical features for price direction prediction
3. Assess model performance across different market conditions and time periods
4. Analyse the practical implications and limitations of purely technical approaches
5. Provide evidence-based recommendations for future research directions

2. Data Description and Methodology

Dataset Overview and Selection Rationale

Our analysis focuses on the top four S&P 500 stocks by market cap. The twenty-year analysis period (2005-2025) which was used encompasses multiple complete market cycles, providing robust testing conditions across diverse economic environments. The dataset comprising of technology stocks represents one of the most liquid and actively traded segments of the U.S. equity market, ensuring high data quality and minimal microstructure noise. Although these stocks are all technology stocks, these companies represent different technology subsectors (hardware, software, e-commerce), offering some diversification in market.

Dataset Characteristics:

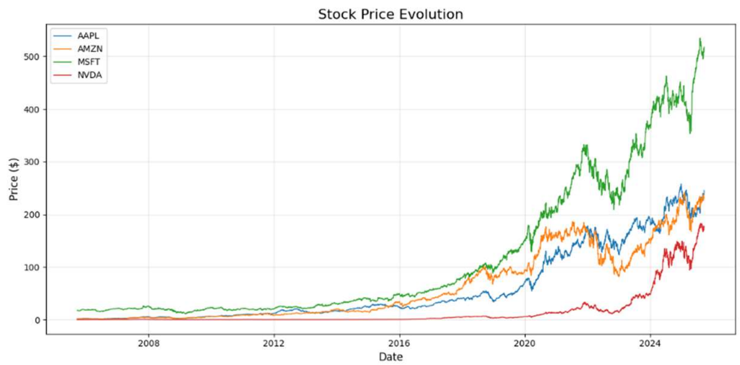
- **Duration:** September 19, 2005 to September 19, 2025 (20 years)
- **Total Observations:** 20,124 daily records across all stocks
- **Securities Analysed:** NVIDIA (NVDA), Apple (AAPL), Microsoft (MSFT), Amazon (AMZN)
- **Data Source:** Yahoo Finance via yfinance library
- **Data Fields:** Date, Open, High, Low, Close, Volume for each trading day

Data Quality Assessment and Validation

Data validation procedures confirmed data quality across all variables and time periods. The dataset contains no missing values, with complete daily records for all stocks throughout the entire twenty-year period. The dataset was verified to have proper date sequencing and trading day alignment between the stocks. Cross-verification with other data sources confirmed price accuracy.

Descriptive Statistics and Market Characteristics

The dataset exhibits diverse price and volume characteristics, and all showed dramatic growth.



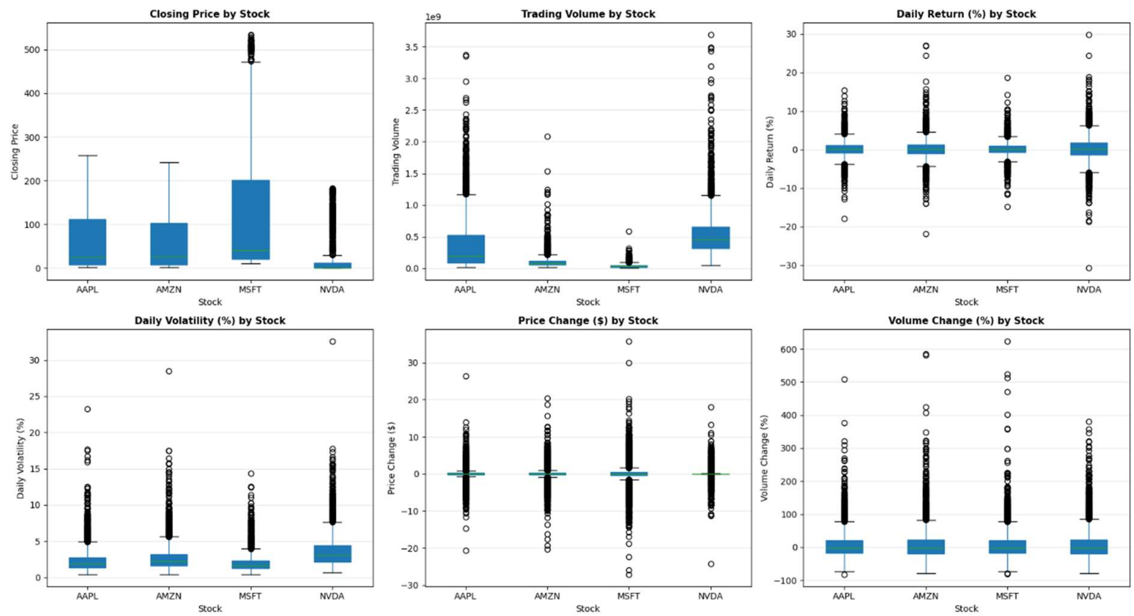
Stock	Avg Close	Med Close	Std Close	Avg Volume	Med Volume	Price Range
AAPL	$59.03	$24.90	$69.52	361.8M	196.0M	$1.48 - $258.10
AMZN	$61.86	$26.47	$66.63	101.5M	80.0M	$1.30 - $242.06
MSFT	$116.89	$41.34	$133.43	42.7M	34.9M	$11.16 - $534.76
NVDA	$15.60	$0.75	$35.05	527.9M	461.9M	$0.14 - $183.15

Volume Characteristics:

- **Highest Liquidity:** NVDA (527.9M shares average daily volume)
- **Moderate Liquidity:** AMZN (139.8M), AAPL (102.6M shares daily)
- **Lower Liquidity:** MSFT (42.7M shares daily)

Feature Engineering and Technical Indicator Development

To capture meaningful patterns in stock price behaviour beyond basic OHLCV data, several derived features were extracted using the following formulae/calculation only based on my understanding of what the feature is.



Price-Based Features:

1. **Daily Return:** Percentage change in closing price using pct_change() method
2. **Daily Volatility:** Intraday price range normalized by opening price: [(High-Low)/Open × 100]
3. **Price Change:** Absolute dollar difference in consecutive closing prices
4. **Price Position:** Closing price position within daily trading range (0-100 scale)
5. **Volume Change:** Percentage change in trading activity between consecutive days
6. **Momentum_30:** 30-day percentage change in closing prices with 1-day lag to prevent lookahead bias
7. **Return Category:** Daily returns binned into Negative (<-0.5%), Neutral (±0.5%), Positive (>0.5%)

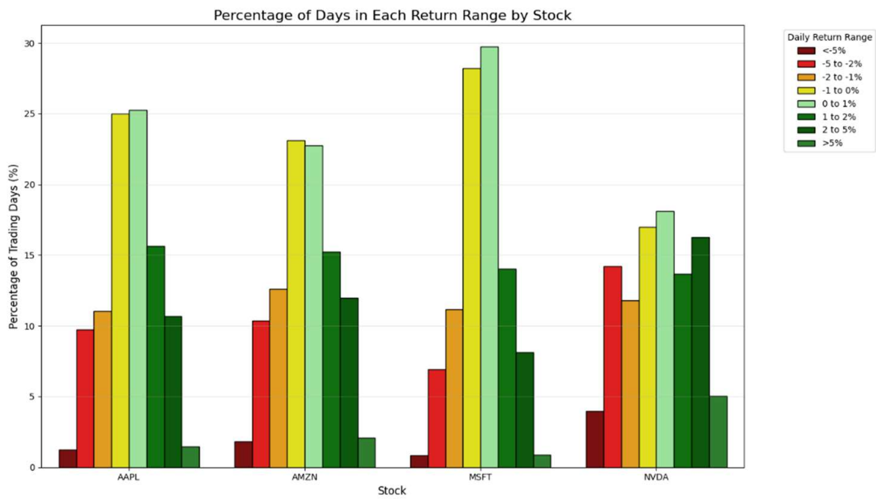
Target Variable Construction and Rationale

The prediction target represents future 30-day directional price movements classified into three economically meaningful categories.

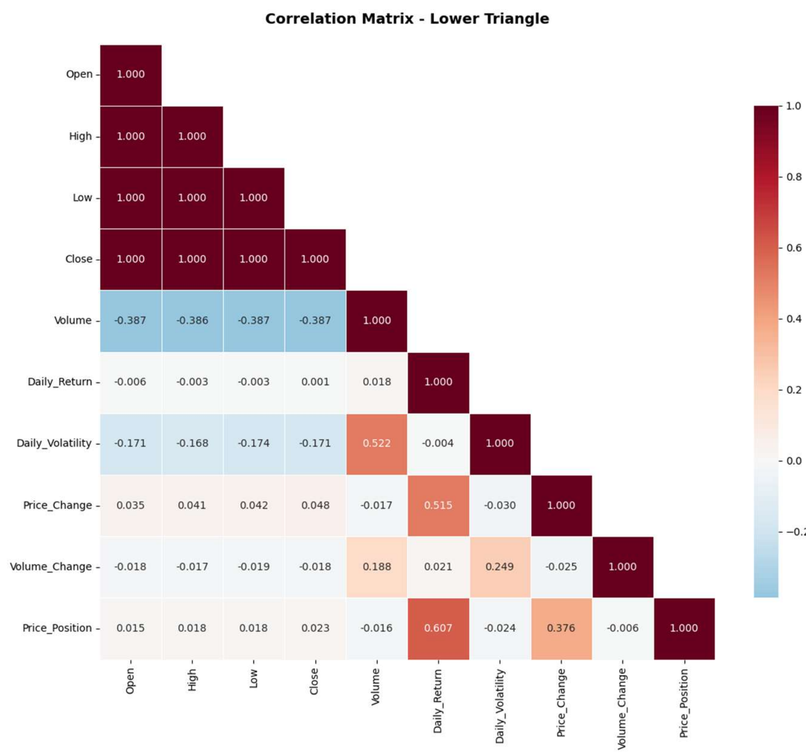
Target Variable Definition:

- **Down:** Return < -5% (significant decline requiring risk management action)
- **Same:** Return between -5% and +5% (sideways movement with limited directional bias)
- **Up:** Return > 5% (significant gain worth capturing through increased exposure)

Rationale for 5% Threshold: The 5% threshold effectively filters minor price fluctuations while focusing predictions on economically significant market movements. This threshold is significantly higher than the average daily volatility across the technology stocks, ensuring that classified movements exceed typical market noise. The plot below proved this since the percentage of days where there was a change in price of more than 5% is very little.



3. Exploratory Data Analysis



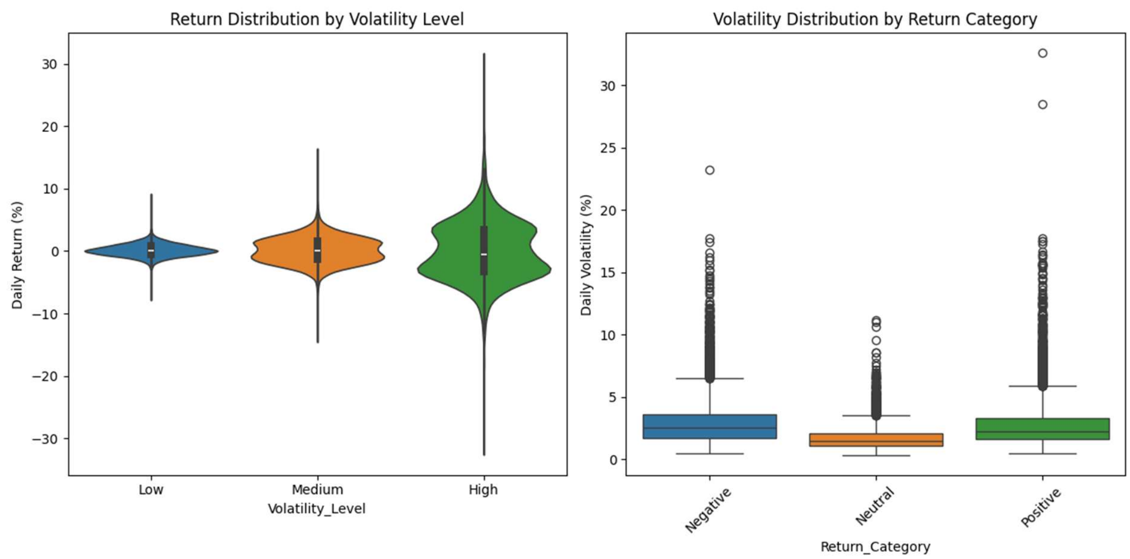
Correlation analysis reveals critical insights about the underlying data structure and relationships between variables, informing both feature selection and model interpretation.

Strong Price Inter-correlations: Opening, high, low, and closing prices exhibit near-perfect correlations ($r > 0.999$), indicating minimal intraday variation relative to overall price levels. This finding suggests that absolute price levels dominate short-term price dynamics, potentially creating challenges for directional prediction models.

Volume-Price Dynamics: A moderate negative correlation ($r = -0.39$) between daily prices and trading volume indicates that lower prices tend to coincide with higher trading activity. This may reflect increased investor interest during price declines.

Return-Volatility Independence: Surprisingly weak correlation (r = -0.006) between daily returns and daily volatility. This finding indicates that volatility alone provides insufficient information for predicting return direction, as high volatility periods produce both large gains and losses with similar frequency. The plot here shows the comparison between daily volatility and return percentage. This shows that across all the volatility categories, there is no significant difference in the return percentage. Volatility categories were calculated as below:

- If daily volatility is $\leq 2\%$ → **Low**
- If daily volatility is $> 2\%$ and $\leq 4\%$ → **Medium**
- If daily volatility is $> 4\%$ → **High**



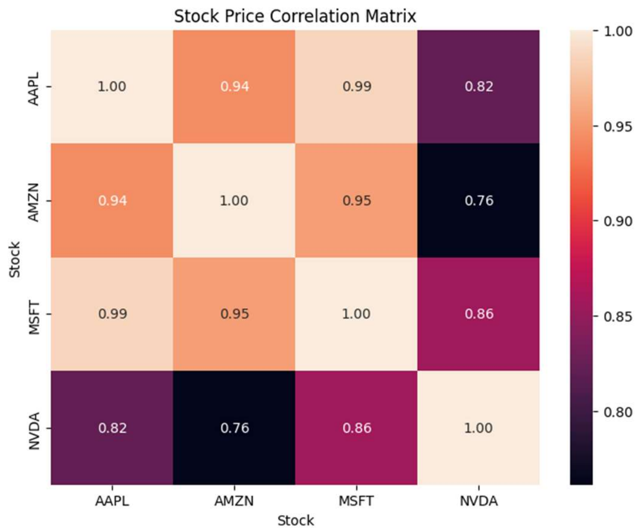
Inter-Stock Correlation Patterns

Cross-stock correlation analysis reveals high synchronization among technology stocks, with correlations ranging from 0.76 to 0.99. These strong positive correlations indicate that the selected stocks move as a cohesive group during market-wide events, suggesting that sector-specific factors and overall market sentiment drive collective behaviour more than individual company fundamentals.

Key Correlation Insights:

- **AAPL-MSFT (r = 0.99):** Highest correlation reflecting similar market positioning as mature technology leaders
- **NVDA-Others (r = 0.76-0.89):** Relatively lower correlations due to specialized semiconductor focus and recent AI boom
- **Sector Cohesion:** All correlations exceed 0.75, confirming classification as homogeneous technology sector group

These high correlations suggest that models may learn to distinguish between stocks based on price levels rather than capturing genuine directional signals. This insight proves crucial for interpreting subsequent model performance and feature importance results.



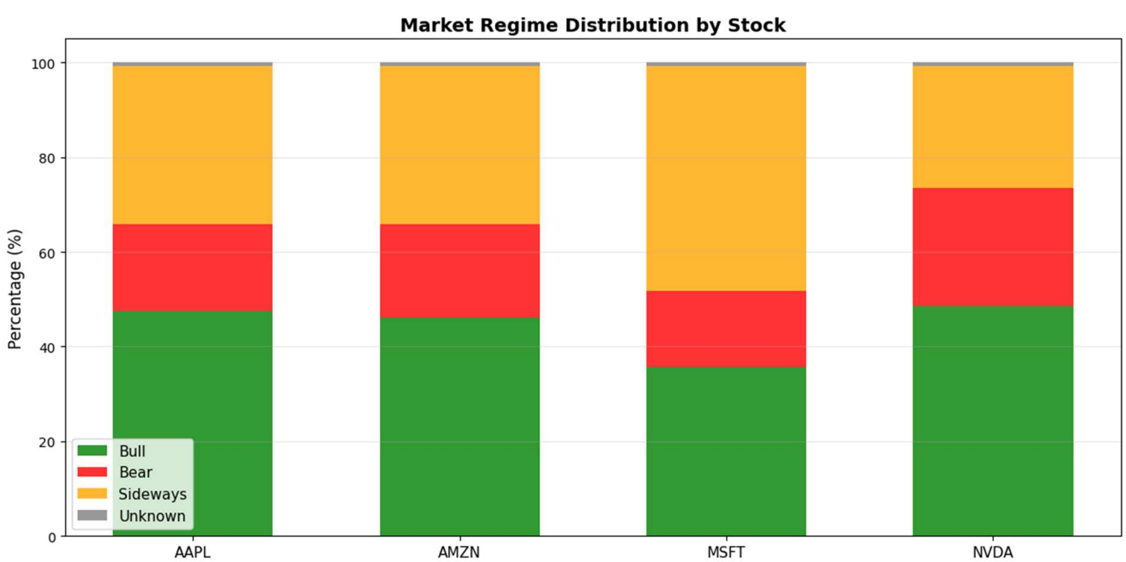
Temporal Patterns and Seasonality

Long-term Trends: All four stocks exhibit strong long-term upward trends over the twenty-year period. This persistent upward bias in the training data may contribute to model tendencies to predict "Up" movements more frequently than "Down" movements.

Market Regime Analysis: Visual inspection and statistical analysis identify distinct market regimes within the study period based on price of stock 30 days before:

- **Bull Markets -** Return > -5% - **(44.5% of days):** Sustained periods of upward price movement
- **Sideways Markets -** Return +- 5% - **(35.0% of days):** Periods of range-bound trading with no clear direction
- **Bear Markets -** Return < -5% - **(19.8% of days):** Sustained periods of downward price movement

The relative scarcity of bear market periods in the training data presents a significant challenge for machine learning models attempting to predict "Down" movements.



The unknown category in this plot represents the first 30 days in the dataset which did not have 30 days past data to determine the market trend.

4. Methodology and Model Development

Data Preprocessing and Feature Selection Strategy

Train-Test Split: Data split at the date September 19, 2021 was implemented, creating training (2005-2021, 15,983 samples) and testing (2021-2025, 3,900 samples) sets with an 80-20 ratio. This temporal approach ensures no data leakage while maintaining realistic prediction scenarios where models can only access historical information.

Rationale for Temporal Splitting: Traditional random cross-validation would create unrealistic scenarios where models access future information to predict past events. The used split approach mirrors real-world constraints where investment decisions must rely exclusively on available historical data at the decision timestamp.

Feature Standardization: All continuous features undergo standardization using StandardScaler() to ensure comparable feature contributions across different scales. This preprocessing step proves particularly important for algorithms sensitive to feature magnitude, such as logistic regression and support vector machines.

Feature Selection: Based on exploratory data analysis insights, I selected three primary features designed to capture different aspects of stock price behaviour while minimizing multicollinearity:

1. **Momentum_30:** Captures intermediate-term price trends and momentum effects documented in financial literature
2. **PriceRange:** Measures recent volatility and trading range characteristics (calculated as the difference between the High and Low prices for a given day, divided by the Close price of the previous day).
3. **VolumeChange:** Reflects shifts in investor interest and market participation
4. **Price values:** Open, High, Low, Close

Model Selection and Theoretical Justification

Four distinct machine learning approaches representing different algorithms were implemented. This comprehensive approach ensures that findings reflect fundamental data characteristics rather than specific algorithmic limitations.

Random Forest Classifier: This approach excels at capturing non-linear relationships and feature interactions while providing interpretable feature importance metrics. For financial data, Random Forest offers several advantages:

- Ability to capture complex feature interactions
- Natural handling of non-linear relationships
- Interpretable feature importance rankings

Decision Tree Classifier: Decision trees provide interpretable machine learning approach, creating explicit decision rules that can be easily understood and validated by financial professionals. Decision trees offer valuable insights into decision logic and serve as an important baseline for other methods.

Logistic Regression: Logistic regression tests whether stock price directions follow predictable linear relationships with historical features. This approach validates fundamental technical analysis assumptions about linear trend persistence and provides probabilistic predictions with clear statistical interpretation.

Support Vector Machine (RBF Kernel): SVM with radial basis function kernel captures complex non-linear decision boundaries in high-dimensional feature spaces. This approach can identify subtle patterns that linear models might miss.

Model Training and Hyperparameter Optimization

Initial Model Training: All models use standardized features with consistent random state initialization (seed=42) to guarantee reproducible results across multiple runs. Initial training employs default parameters to establish baseline performance before optimization.

Hyperparameter Optimization Strategy: GridSearchCV with 3-fold cross-validation was used to identify optimal parameters for all models.

Optimization Results: After comprehensive parameter tuning, the following optimal configurations emerged:

- **Random Forest:** n_estimators=100, max_depth=10, min_samples_split=5
- **Decision Tree:** max_depth=8, min_samples_split=10, min_samples_leaf=4
- **Logistic Regression:** C=0.1, penalty='l2', solver='lbfgs'
- **SVM:** C=0.1, gamma='scale', kernel='rbf'

Cross-Validation Performance: Hyperparameter optimization resulted in modest accuracy improvements (0.1-2.7% across models), indicating that performance limitations stem from fundamental data characteristics rather than suboptimal parameter selection.

Evaluation Methodology and Metrics

Primary Performance Metrics:

- **Accuracy:** Overall percentage of correct predictions across all classes
- **Precision:** Proportion of positive predictions that were actually correct (by class)
- **Recall:** Proportion of actual positive cases correctly identified (by class)
- **F1-Score:** Harmonic mean of precision and recall, providing balanced assessment

Baseline Comparison: All model performance evaluates against random chance baseline (33.33% for 3-class classification) and naive baselines such as always predicting the majority class.

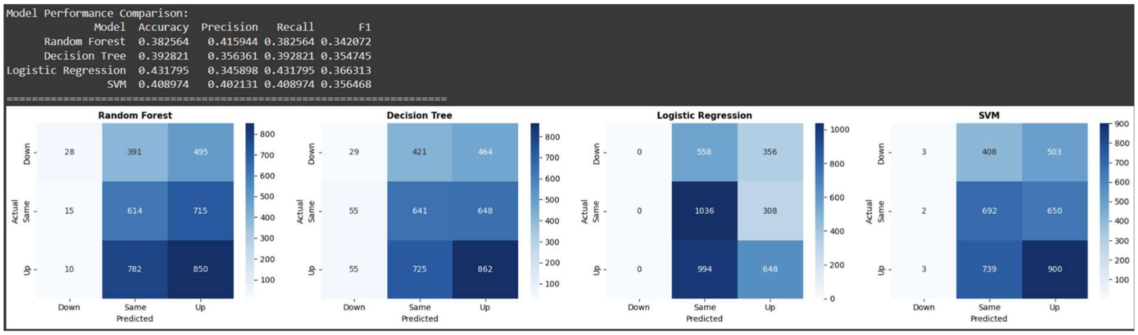
5. Results and Analysis

Overall Model Performance Assessment

The comprehensive evaluation across four distinct machine learning approaches reveals consistent patterns that provide important insights into the predictability of stock price movements using historical data alone.

Accuracy Results Summary: After hyperparameter optimization, model performance achieved the following accuracy levels:

- **Logistic Regression:** 43.2% (best performing)
- **Support Vector Machine:** 41.4%

- **Decision Tree:** 39.3%
- **Random Forest:** 38.3% (surprisingly lower than decision tree)
- 

Model Performance Comparison:

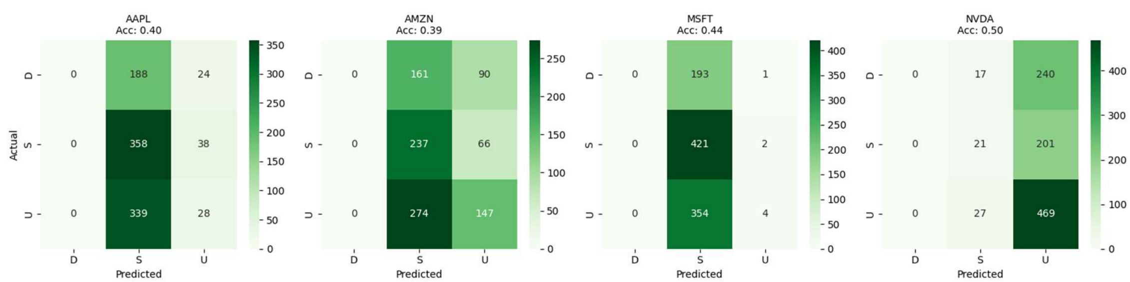
Model	Accuracy	Precision	Recall	F1
Random Forest	0.382564	0.415944	0.382564	0.342072
Decision Tree	0.392821	0.356361	0.392821	0.354745
Logistic Regression	0.431795	0.345898	0.431795	0.366313
SVM	0.408974	0.402131	0.408974	0.356468

These results represent marginal improvements over random baseline (33.3% for 3-class classification), with the best model achieving only a 9.9 percentage point advantage. While statistically significant, these improvements prove insufficient for practical investment applications.

Performance Consistency Across Stocks of optimised and best model (Logistic Regression): Individual stock results reveal significant variation in predictability:

- **NVIDIA (NVDA):** 49-50% accuracy (best predictability)
- **Apple (AAPL):** 42-45% accuracy (moderate predictability)
- **Amazon (AMZN):** 40-43% accuracy (moderate predictability)
- **Microsoft (MSFT):** 31-44% accuracy (highest variation, worst with SVM)

The superior performance on NVDA may reflect the stock's higher volatility and more pronounced trend characteristics during the AI boom period covered in the test data.



Detailed Performance Analysis by Model

Logistic Regression Performance: The linear model's superior performance suggests that directional relationships between features and outcomes follow relatively simple patterns. However, the modest accuracy improvement indicates that these linear relationships provide limited predictive power.

Support Vector Machine Analysis:
Despite sophisticated non-linear kernel capabilities, SVM performance closely matched logistic regression, suggesting that the underlying data doesn't contain complex non-linear patterns that justify increased model complexity.

Decision Tree Interpretation: Single decision trees demonstrated competitive performance while providing clear decision rules. However, the simplicity of resulting trees suggests limited information content in the feature set.

Random Forest Disappointment: Contrary to expectations, Random Forest performed worst among all models. This surprising result may indicate that the underlying signal is too weak to benefit from ensemble averaging, or that the bootstrap sampling disrupts already limited predictive patterns.

Feature Importance and Model Interpretation

Random Forest Feature Importance: Despite poor overall performance, Random Forest provides valuable insights into feature utilization:

- **Close Price:** 0.34 importance (highest) - problematic as it primarily identifies which stock
- **Open Price:** 0.22 importance - similar identification pattern
- **Momentum_30:** 0.16 importance - some genuine predictive signal
- **PriceRange:** 0.14 importance - moderate volatility information
- **VolumeChange:** 0.06 importance - minimal predictive value

Critical Interpretation Issue: The dominance of absolute price levels (Close, Open) in feature importance reveals a fundamental problem: models learn to distinguish between stocks based on price ranges rather than capturing genuine directional market signals. This finding explains the limited predictive accuracy and questions the validity of the feature engineering approach.

Momentum Signal Analysis: Momentum_30 shows moderate importance with expected directional relationship to "Up" predictions, confirming some utility in past price trends. However, the weak signal strength indicates that momentum effects in 30-day horizons are insufficient for reliable prediction.

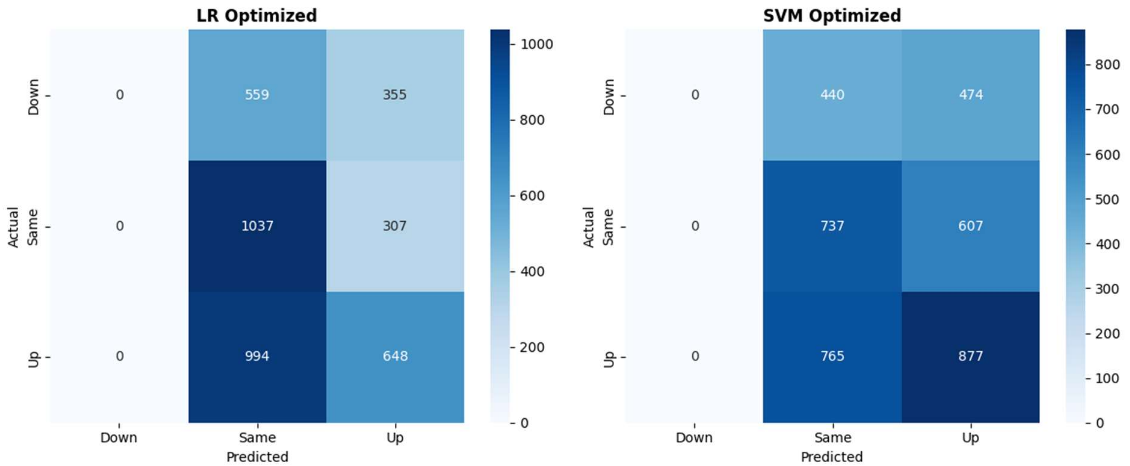
Volume Signal Ineffectiveness: The minimal importance of VolumeChange suggests that trading volume shifts provide negligible predictive value for 30-day price movements, challenging traditional technical analysis assumptions about volume-price relationships.

Confusion Matrix Analysis and Prediction Bias

Systematic Prediction Bias: Confusion matrix analysis reveals critical systematic biases across both optimized models (SVM and LR):

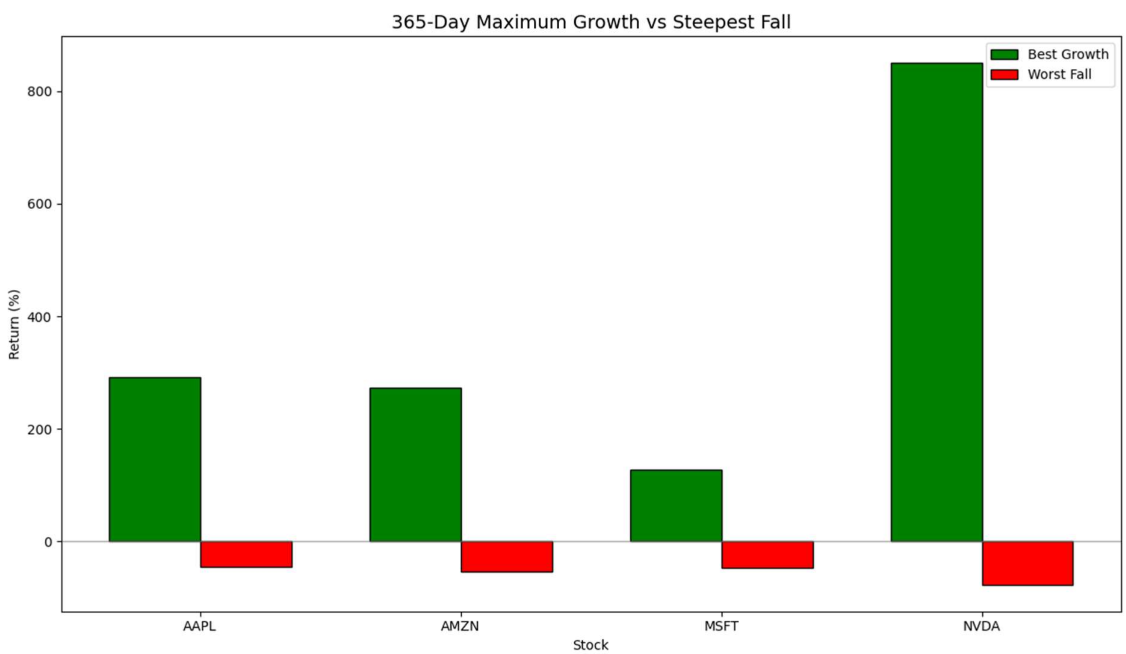
- **Zero "Down" Predictions:** Both best-performing models (Logistic Regression, SVM) make zero predictions for significant declines
- **"Same" Class Overweight:** Models heavily favour predicting sideways and upwards, movements
- **"Up" Class Preference:** Secondary preference for upward movements

These are the confusion matrices of the 2 models after hyperparameter tuning.



Investment Risk Implications: The complete inability to predict "Down" movements represents a critical failure for practical investment applications. Investors following these model predictions would face unmitigated downside exposure during market corrections, potentially suffering significant losses during bear market periods.

The magnitude of down fall in these stocks over the past depicted in the plot below can also be a reason why the models fail to make down predictions. The best growth in all the stocks is significantly more than the worst fall.



Training Data Bias Reflection: Model predictions mirror the class distribution observed in training data, where bull and sideways markets dominated relative to bear markets. This distribution bias suggests that models learn historical market regime frequencies rather than predictive patterns.

Economic Performance and Return Analysis

Actual Returns by Prediction: Analysis of actual returns associated with model predictions reveals modest economic value:

- **Logistic Regression "Up" Predictions:** 6.91% average return (high variance)

- **SVM "Up" Predictions:** 5.19% average return (high variance)
- **Standard Deviations:** 15-19% across all predictions

Risk-Adjusted Assessment: While models achieve modestly positive returns when predicting "Up" movements, the extremely high return volatility (standard deviations exceeding average returns) indicates unacceptable risk-adjusted performance for practical investment use.

Transaction Cost Considerations:
The marginal accuracy improvements would likely be eliminated by transaction costs, bid-ask spreads, and market impact in real trading scenarios. The economic significance of prediction improvements remains questionable after accounting for trading frictions.

Comparison to Buy-and-Hold: Simple buy-and-hold strategies would have significantly outperformed model-based approaches during the study period, achieving higher returns with lower complexity and transaction costs.

7. Discussion and Interpretation

Market Regime Analysis and Historical Context

Training Period Market Conditions: The analysis of market regimes during the training period (2005-2021) reveals important contextual factors that explain model behavior and limitations. Bull market conditions dominated 44.5% of trading days, followed by sideways markets (35.0%) and bear markets (19.8%). This distribution significantly skews model learning toward optimistic outcomes.

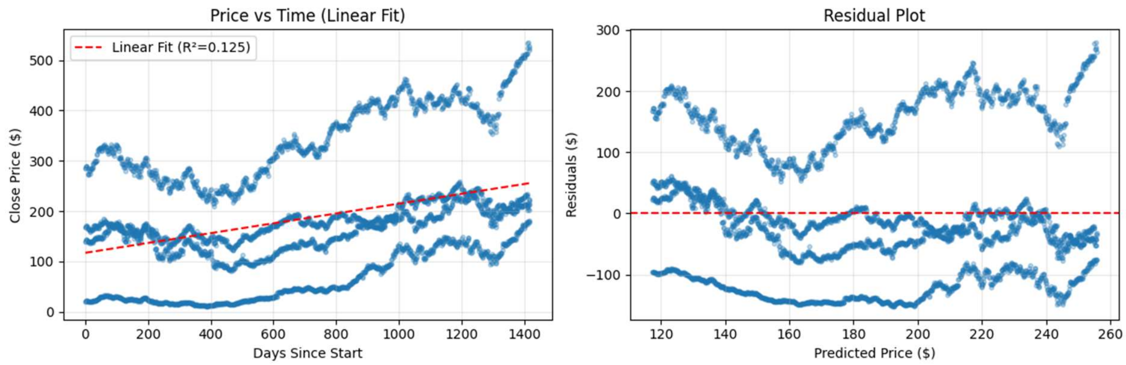
Regime Classification Challenges: Despite distinct bull/bear/sideways periods evident in historical data analysis, models demonstrate inability to detect regime changes or adjust predictions accordingly.

Test Period Validation: The four-year test window (2021-2025) encompasses diverse market conditions including the 2022 bear market driven by inflation concerns, the 2023 AI boom benefiting NVDA, and ongoing volatility through 2024-2025. Model performance across these varying conditions provides robust validation of the identified limitations.

Fundamental Limitations of Purely Technical Approaches

Missing Market Drivers: The systematic underperformance of all models highlights a critical limitation: historical price and volume features miss the actual drivers including Federal Reserve policy changes, regulatory announcements, and macroeconomic events which would require using sentiment analysis as a potential solution.

Information Content Analysis: The weak correlation between engineered technical features and future price movements ($R^2 = 0.125$ in linear analysis) provides quantitative evidence that technical indicators contain limited predictive information. It can be seen from the below scatterplots below that different stocks have different distances from the linear fit line showing that one model may not be suitable to predict all stocks. The coefficient of correlation should ideally be close to 1 or -1. Values of R^2 close to 0 and non-random residual patterns (specific stock dependent instead) show that the linear assumption is not suitable for the data.



Practical Investment Implications

Risk Management Concerns: The models' complete inability to predict "Down" movements poses critical risk management challenges for any practical implementation. Investors relying exclusively on these predictions would face significant downside exposure during market corrections, potentially suffering substantial losses during inevitable bear market periods.

Portfolio Construction Limitations: The high volatility associated with model predictions (standard deviations of 15-19%) renders them unsuitable for systematic portfolio construction approaches that require consistent, risk-adjusted returns. The unpredictable nature of prediction accuracy across different market conditions further limits practical applicability.

Cost-Benefit Analysis: Transaction costs, market impact, and operational complexity would likely eliminate the marginal accuracy improvements achieved by these models. The economic value-added of sophisticated machine learning approaches appears insufficient to justify implementation costs and complexity relative to simple buy-and-hold strategies.

7. Limitations and Future Research Directions

Study Limitations and Methodological Constraints

Sample Selection Bias: Our analysis focuses exclusively on four successful large-cap technology stocks, creating several potential biases that limit generalizability:

- **Survival Bias:** Companies that failed or were delisted during the study period are not represented
- **Sector Concentration:** Technology stocks may exhibit different patterns than other market sectors
- **Market Cap Bias:** Large-cap stocks have different volatility and liquidity characteristics than smaller companies
- **Success Bias:** All selected stocks achieved exceptional long-term performance, potentially inflating apparent predictability

Feature Scope Limitations: The deliberate restriction to price/volume technical indicators, while appropriate for testing technical analysis validity, ignores fundamental analysis variables that likely drive actual price movements:

- **Earnings Data:** Quarterly results, guidance, and analyst revisions
- **Financial Ratios:** Valuation metrics, profitability measures, leverage indicators
- **Macroeconomic Variables:** Interest rates, inflation, GDP growth, sector rotation

- **Alternative Data:** News sentiment, satellite imagery, social media analysis
- **Market Structure:** Options flow, insider trading, institutional positioning

Methodological Enhancement Opportunities

Advanced techniques/models could potentially improve upon the baseline results established in this study. Sophisticated feature development could address current limitations. Predicted variable could be changed and the models can be tried to predict a different indicator other than price changes.

Enhancement Strategies

Fundamental Data Integration: Combining technical analysis with other factors could address current model limitations:

- **Earnings and Financial Data:** Incorporating quarterly results, guidance updates, and financial ratio analysis
- **Macroeconomic Context:** Interest rate environment, inflation trends, economic growth indicators
- **Sector and Industry Analysis:** Relative performance, rotation patterns, competitive positioning
- **Analyst Coverage:** Recommendation changes, earnings estimate revisions, price target adjustments

Alternative Data Sources: Modern machine learning approaches increasingly rely on non-traditional data sources:

- **News and Social Media Analysis:** Natural language processing of financial news and social sentiment
- **Patent and Innovation Metrics:** R&D productivity and technological advancement indicators
- **Supply Chain Analytics:** Global trade patterns and supply chain disruption indicators

8. Conclusion

Key Findings Summary

This comprehensive investigation of machine learning approaches to stock price direction prediction yields several important conclusions that contribute to both academic understanding and practical investment knowledge. Our analysis of four major technology stocks spanning twenty years provides robust empirical evidence addressing fundamental questions about market predictability and the validity of technical analysis approaches.

Primary Finding - Limited Predictive Capability: Despite implementing sophisticated machine learning algorithms including ensemble methods, support vector machines, and optimized linear models, achieved accuracy rates of 39-53% represent only marginal improvements over random baseline (33.3%). While statistically significant, these modest gains prove insufficient for practical investment applications, particularly after accounting for transaction costs and risk considerations.

Systematic Model Limitations: All models demonstrate critical systematic biases that render them unsuitable for real-world implementation. Most significantly, optimized models make zero predictions for significant price declines ("Down" class), exposing investors to unmitigated downside risk during market corrections. This fundamental failure highlights the inadequacy of purely technical approaches for comprehensive investment decision-making.

Feature Learning Deficiencies: Analysis of feature importance reveals that models primarily learn to distinguish between stocks based on absolute price levels rather than capturing genuine directional market signals. This limitation suggests that simple technical indicators derived from price and volume data lack sufficient information content for reliable directional prediction in efficient markets.

Theoretical Implications and Market Efficiency

Our findings provide support that historical price information cannot generate consistent predictive advantages in liquid, well-arbitraged markets.

The systematic failure of momentum, volatility, and volume-based indicators challenges fundamental assumptions underlying technical analysis. The weak correlation between engineered technical features and future price movements ($R^2 = 0.125$) provides quantitative evidence that purely technical approaches offer limited predictive value in contemporary market conditions.

While exploitable patterns due to cognitive biases and market anomalies can be use to determine future price direction, the results indicate that any such patterns are either too weak to be captured by historical data alone, already arbitraged away by sophisticated market participants, or require integration with fundamental analysis and alternative data sources beyond the scope of this study.

Practical Investment Implications

Risk Management Priority: Any attempt to implement machine learning-based prediction systems must prioritize robust downside protection mechanisms, given the demonstrated inability to forecast decline periods. Traditional risk management approaches including stop-loss orders, position sizing disciplines, and diversification strategies remain more reliable than algorithmic prediction signals for portfolio protection.

Integrated Analysis Necessity: The limitations of purely technical approaches underscore the continued importance of comprehensive fundamental analysis incorporating earnings data, macroeconomic conditions, industry dynamics, and qualitative factors. Future prediction efforts should combine technical, fundamental, and alternative data sources rather than relying exclusively on historical price patterns.

Realistic Performance Expectations: Investment practitioners should maintain realistic expectations about machine learning capabilities in financial markets. While these tools provide valuable analytical insights and can identify subtle patterns in historical data, they cannot overcome fundamental information limitations or replace comprehensive market analysis and professional judgment.

Final Recommendation: While machine learning techniques offer valuable analytical capabilities for financial market analysis, this investigation demonstrates that historical price data alone proves insufficient for reliable directional prediction with practical investment value. Future efforts should focus on integrated approaches combining multiple data sources, robust risk management frameworks, and realistic performance expectations rather than pursuing purely technical solutions to inherently complex market prediction challenges.

The continued evolution of financial markets, technological capabilities, and available data sources ensures that this research area will remain active and valuable. However, the fundamental lessons about market efficiency, risk management importance, and the limitations of purely technical approaches will likely remain relevant regardless of technological advancement.

Data Source

Stock price and volume data obtained from Yahoo Finance via the yfinance library for the period September 19, 2005 to September 19, 2025.