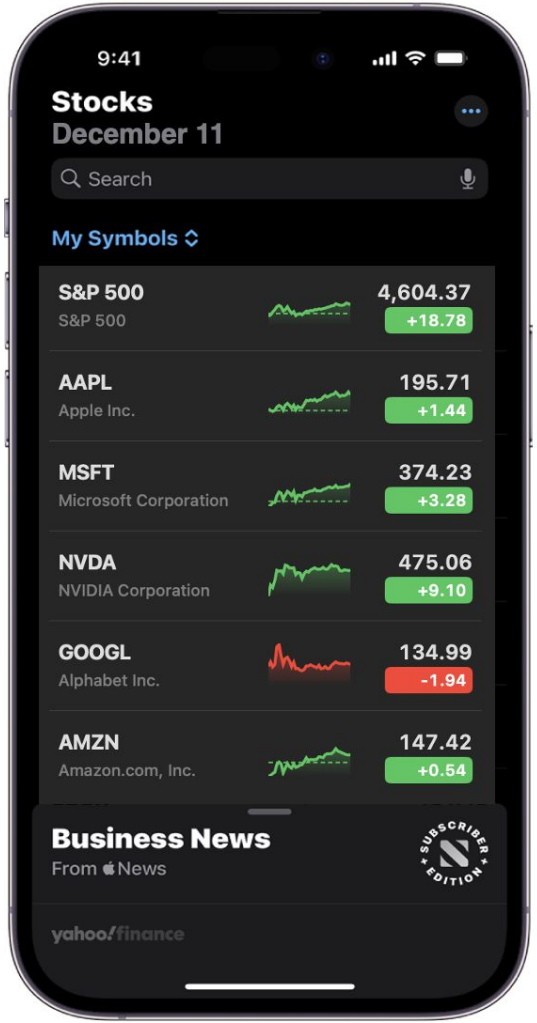
Predicting Future Stock Returns

Mateo, Jay, Adi, Vadym

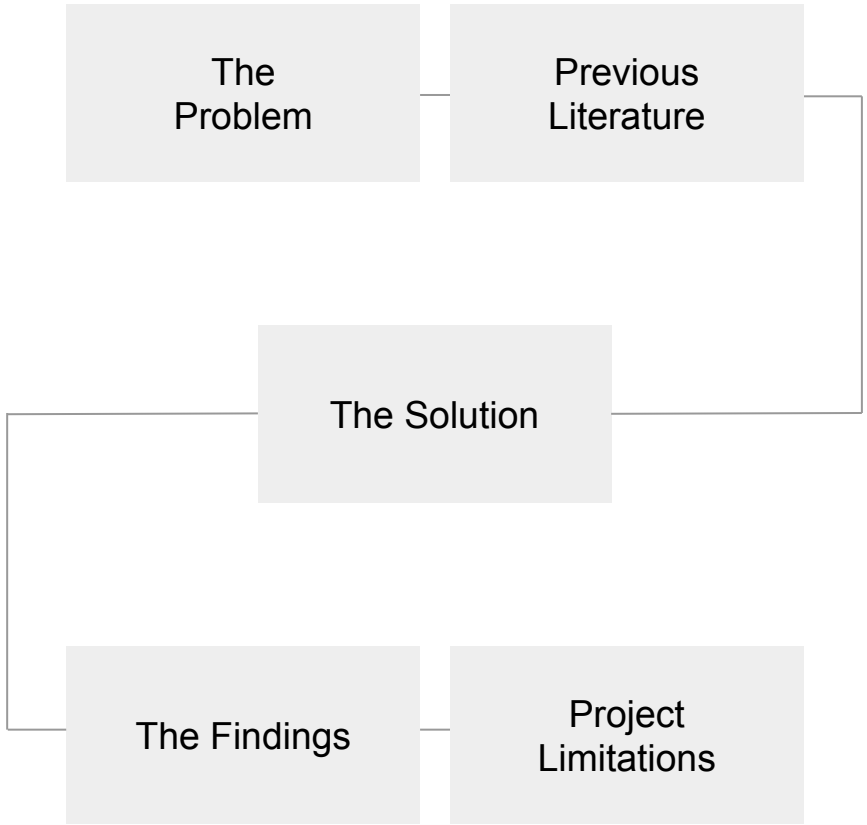
The Focus



Market benchmark

Analyzed Tech Stocks

The Agenda


The Problem
Previous Literature
The Solution
The Findings
Project Limitations

Business Objectives

- Accurate Investment Decisions
- Market Trend Analysis
- Investor Confidence Boost
- Risk Mitigation
- Portfolio Optimization

Problem Framing

Goal:

Predict Next Day Returns

Data:

2007-2023 Analysis

Focus:

Feature Engineering

Challenge:

Market Volatility & Model Inaccuracy

Strategy:

Utilize LM, XgBoost, Bagging, SVM, Lasso, Random Forest

Previous Work

Bollen, Mao, Zeng (2011)

Twitter **sentiment analysis** for predicting Dow Jones movements, 87.6% accuracy.

Heaton, Polson, Witte (2016)

Financial indicators & unconventional data, **Convolutional Neural Networks**.

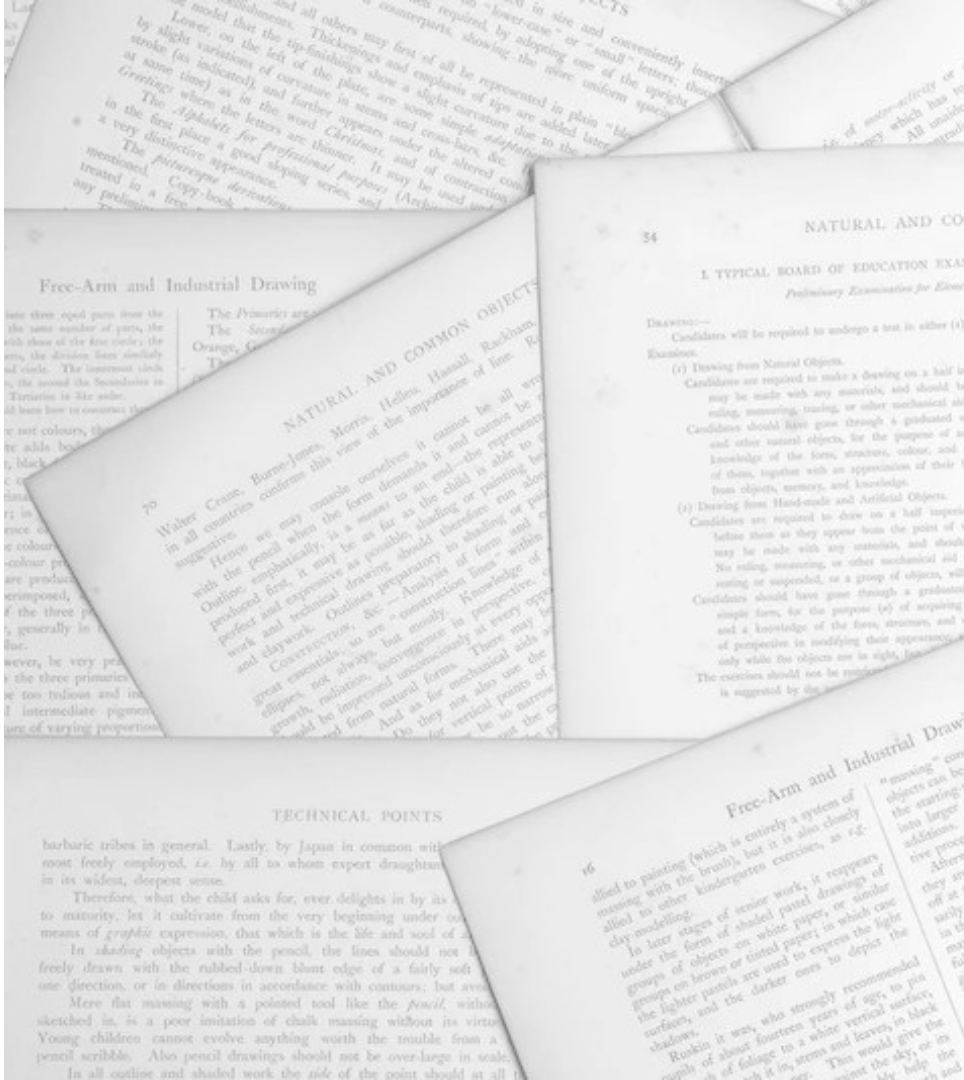
Krauss, Do, Huck (2017)

Used **Deep Neural Networks** and **Random Forests**, achieving Sharpe ratio > 4 (high)

Zhang, Zohren, Roberts (2020)

Analyzed high-frequency trading data, **LSTM** networks.

No publicly available predictors from these papers.



The Data & Scope

21,315 sample

25 variables

01/03/07 - 12/07/23 timeframe

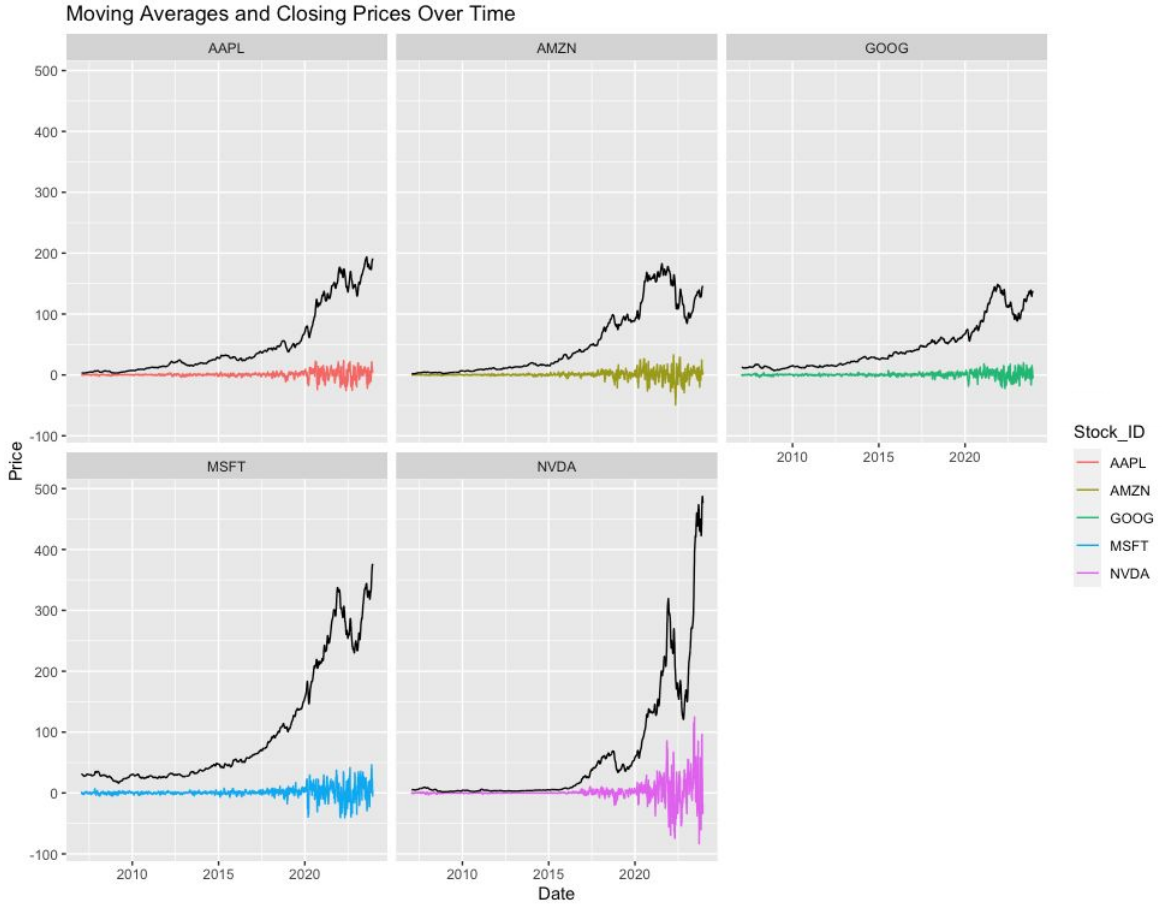
Quantmod, TTR, PerformanceAnalytics packages packages for data retrieval

Dates, daily & future returns, SMA, SD, RSI, Bollinger Bands, Beta data features

Incorporated S&P 500 variables: Daily Return, 14-day MA, 30-day MA, RSI additional data

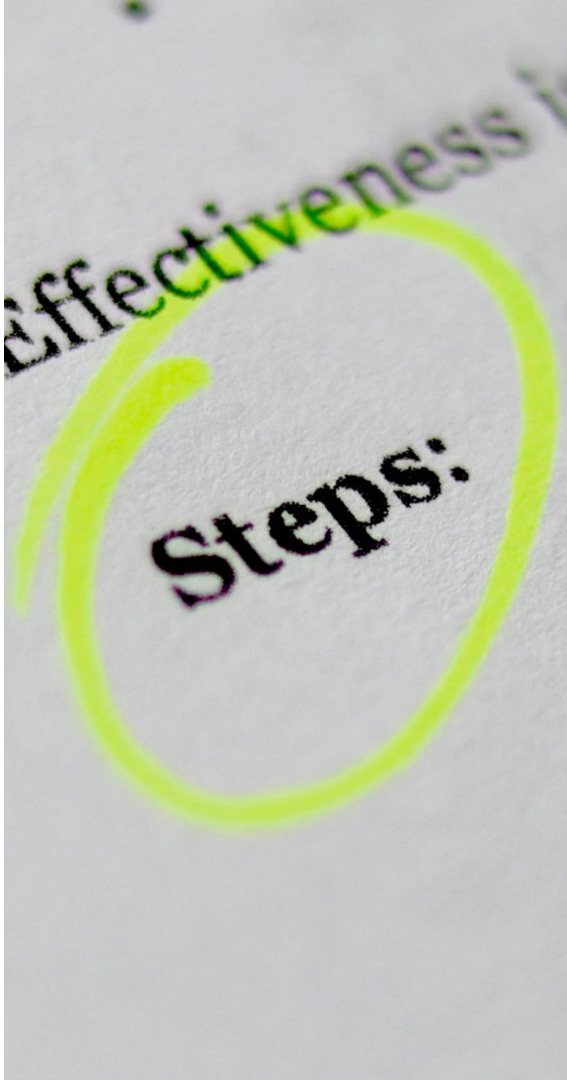
```{r}
```

Data Overview



Methodology

1. Test various models
2. Tune parameters where needed
3. Analyze each model on Apple data
4. Apply all models to all other stocks
5. Reevaluate models & Adjust parameters



Models Applied

XGBoost

Efficient for complex data, risk of overfitting.

Linear Regression

Simple, fast, but may oversimplify.

Random Forest

Handles non-linear data, robust to noise.

Bagging

Reduces overfitting, good for noisy data.

SVM

Effective in high-dimensional spaces, kernel-dependent.

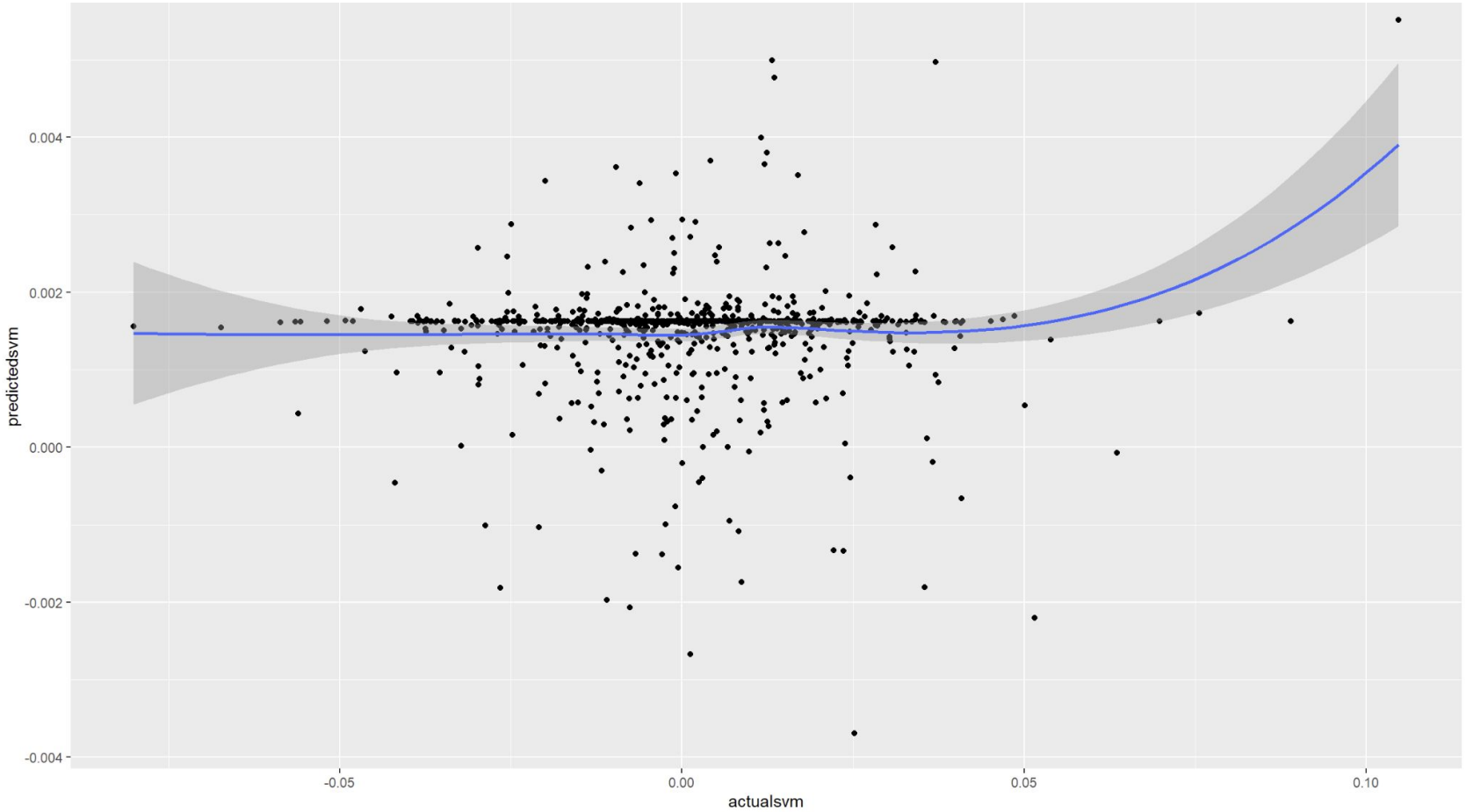
Lasso Regression

Performs feature selection, prone to bias.

Findings:

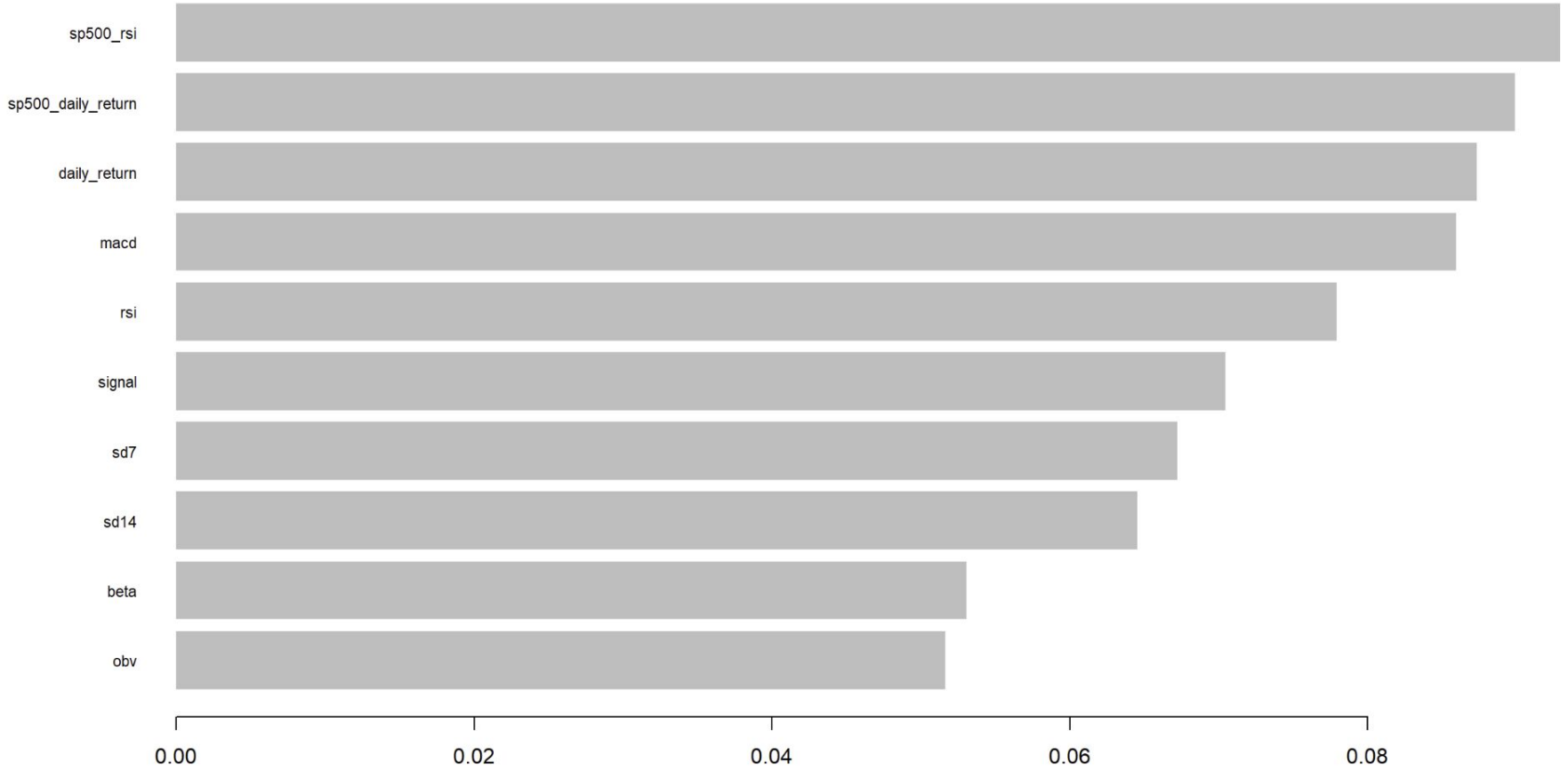
Apple Inc
NASDAQ: AAPL

SVM (51.77%)



Findings:

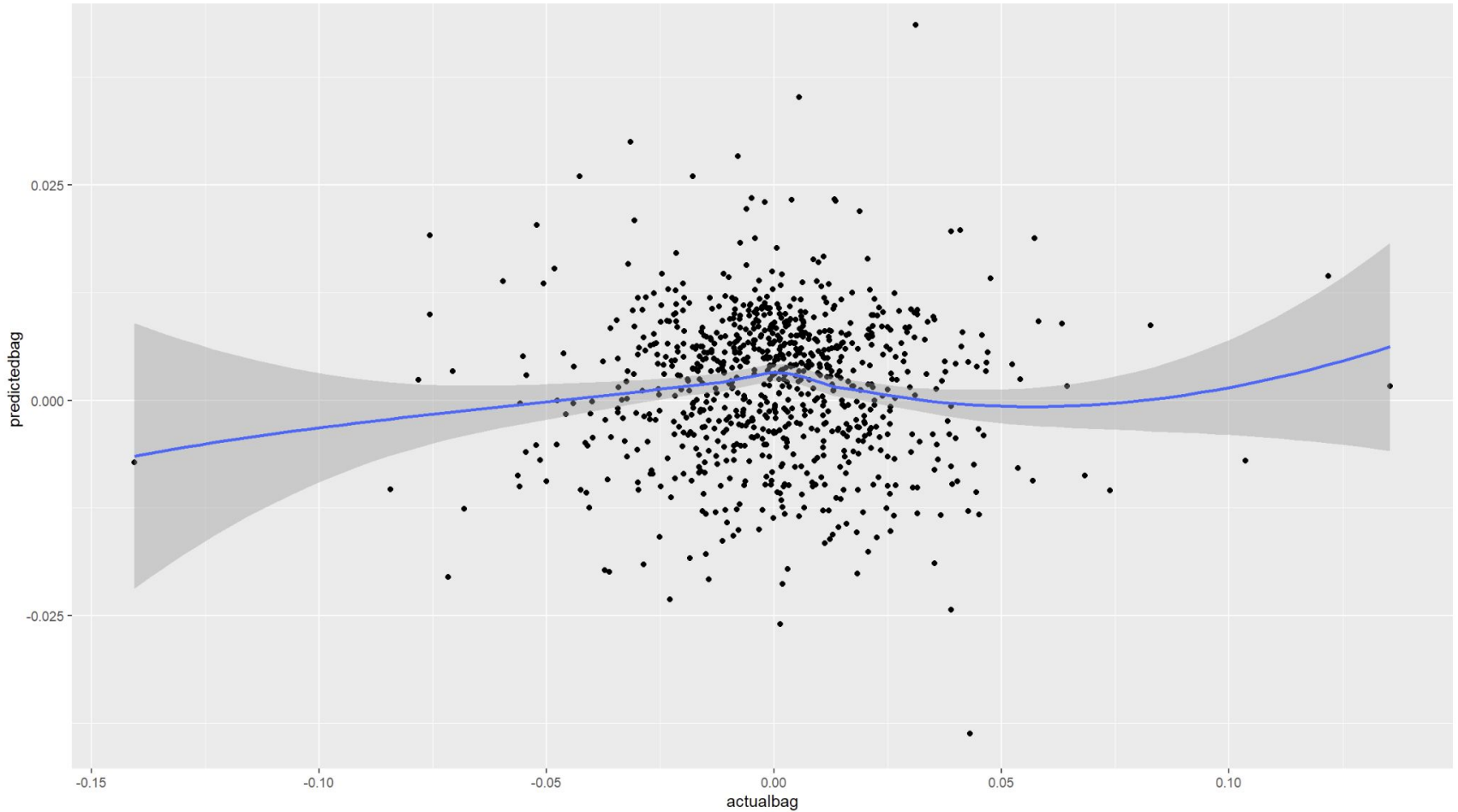
Apple Inc
NASDAQ: AAPL



Findings:

Amazon.com Inc
NASDAQ: AMZN

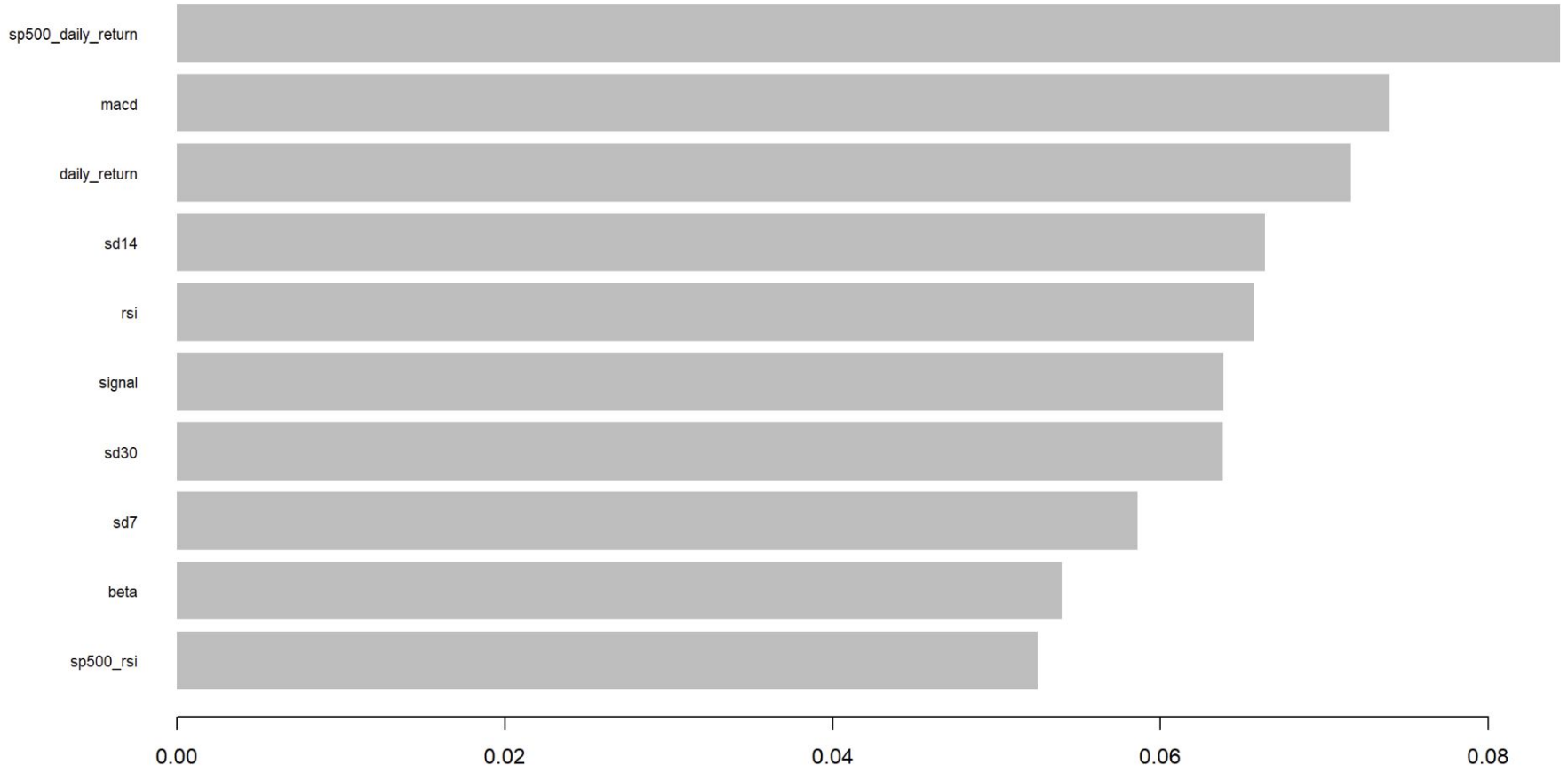
Bagging (50.35%)



Findings:

Amazon.com Inc

NASDAQ: AMZN

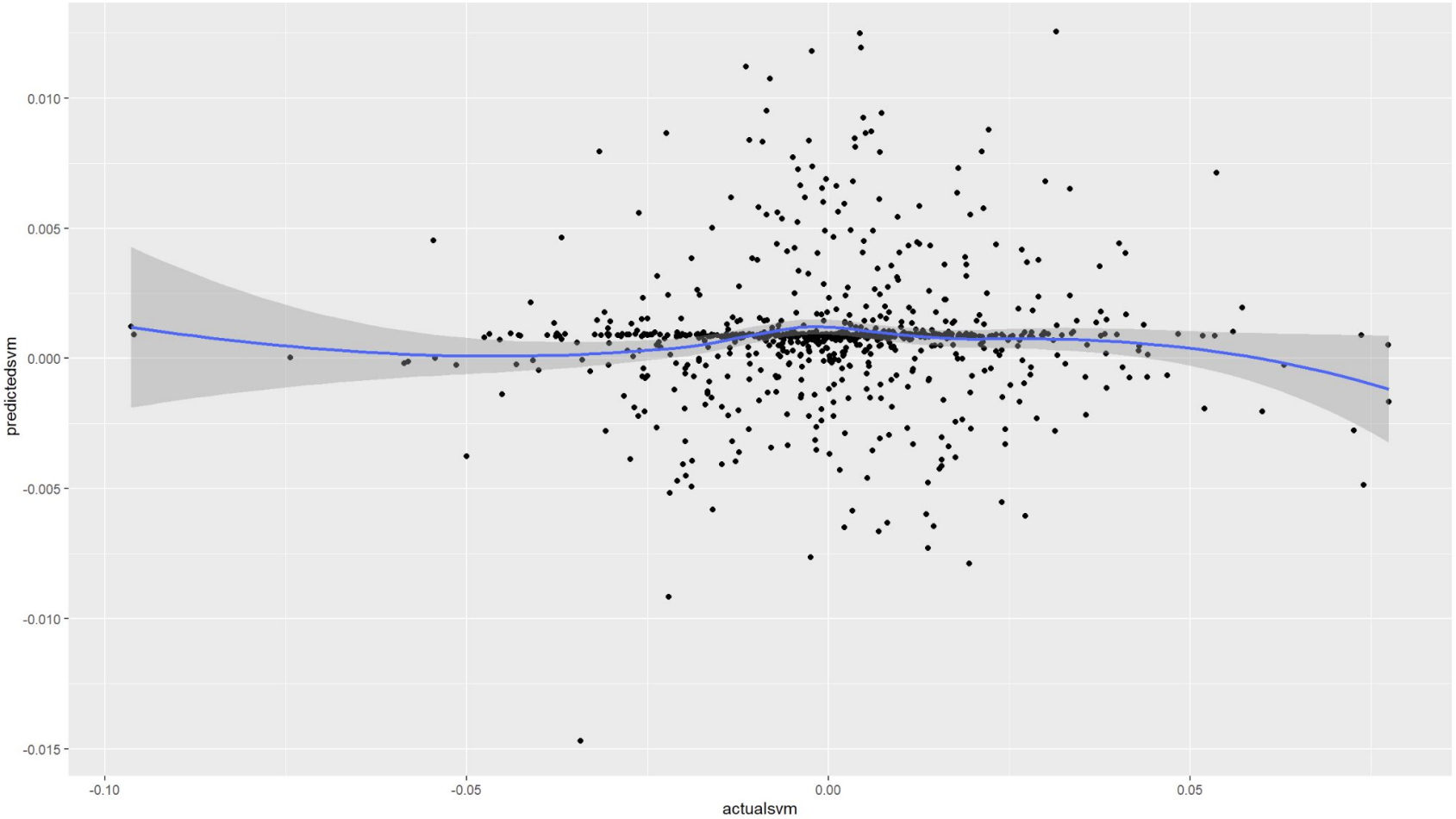


Findings:

Alphabet Inc Class A

NASDAQ: GOOGL

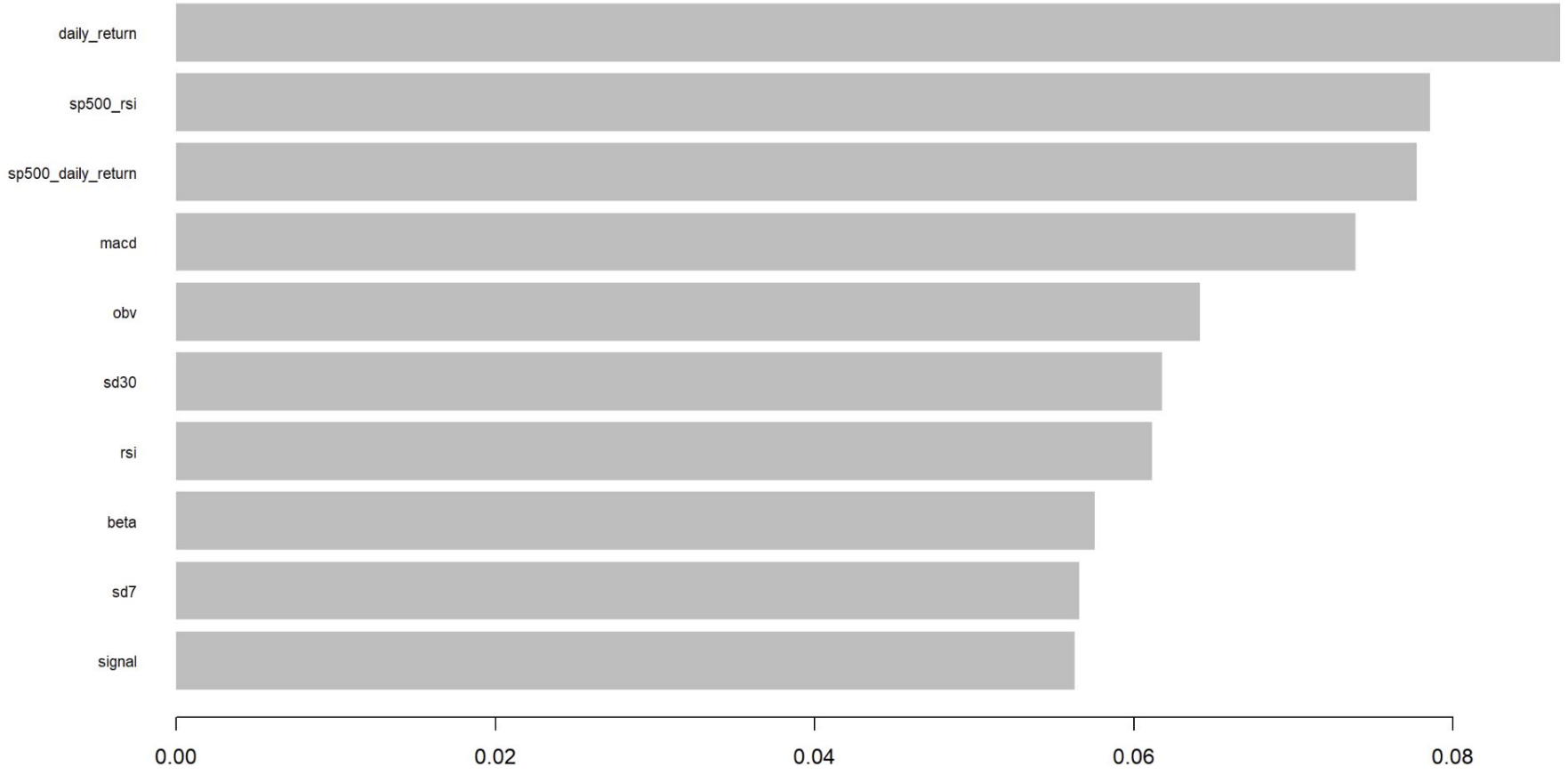
SVM (52.13%)



Findings:

Alphabet Inc Class A

NASDAQ: GOOGL

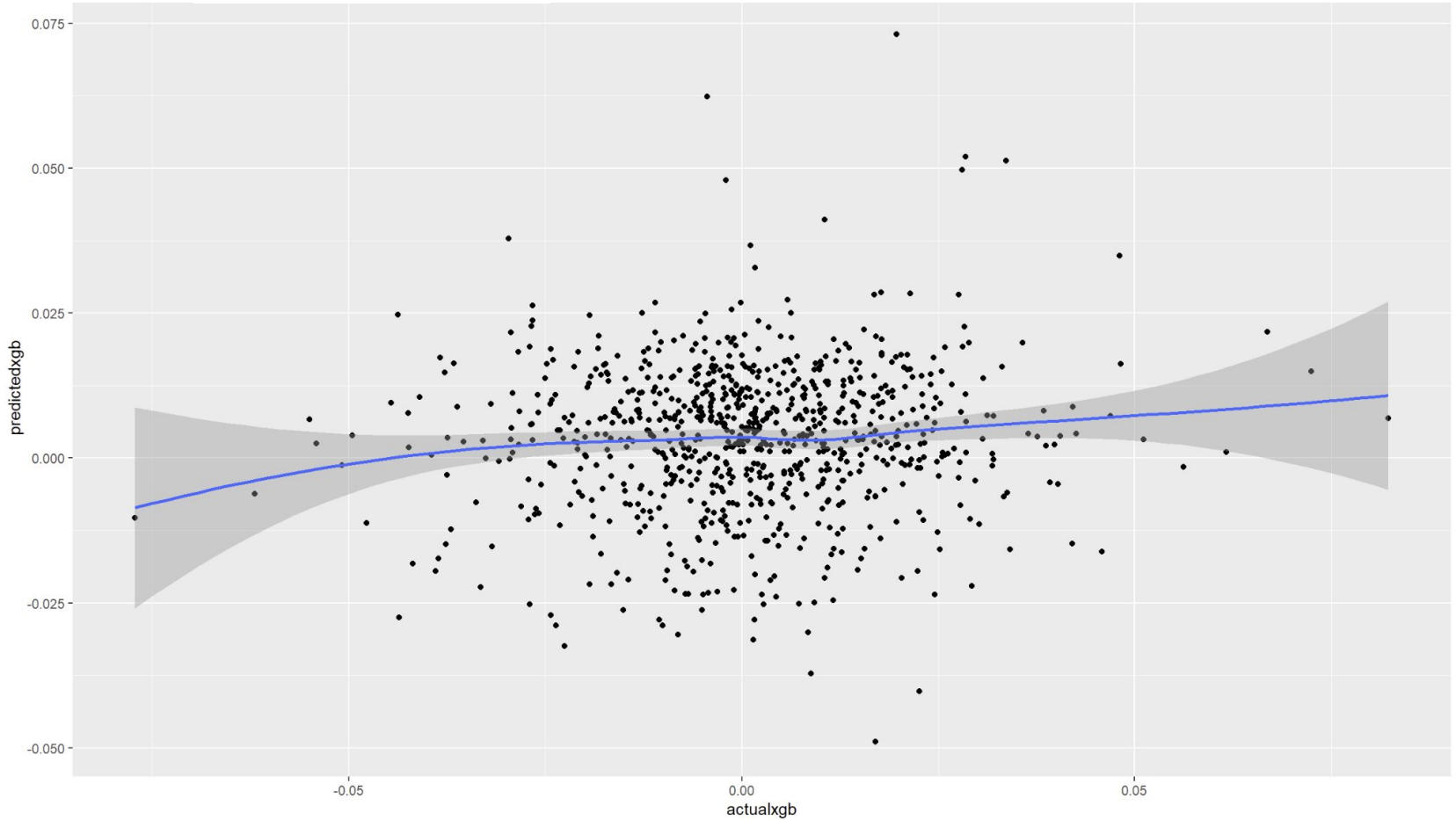


Findings:

Microsoft Corp

NASDAQ: MSFT

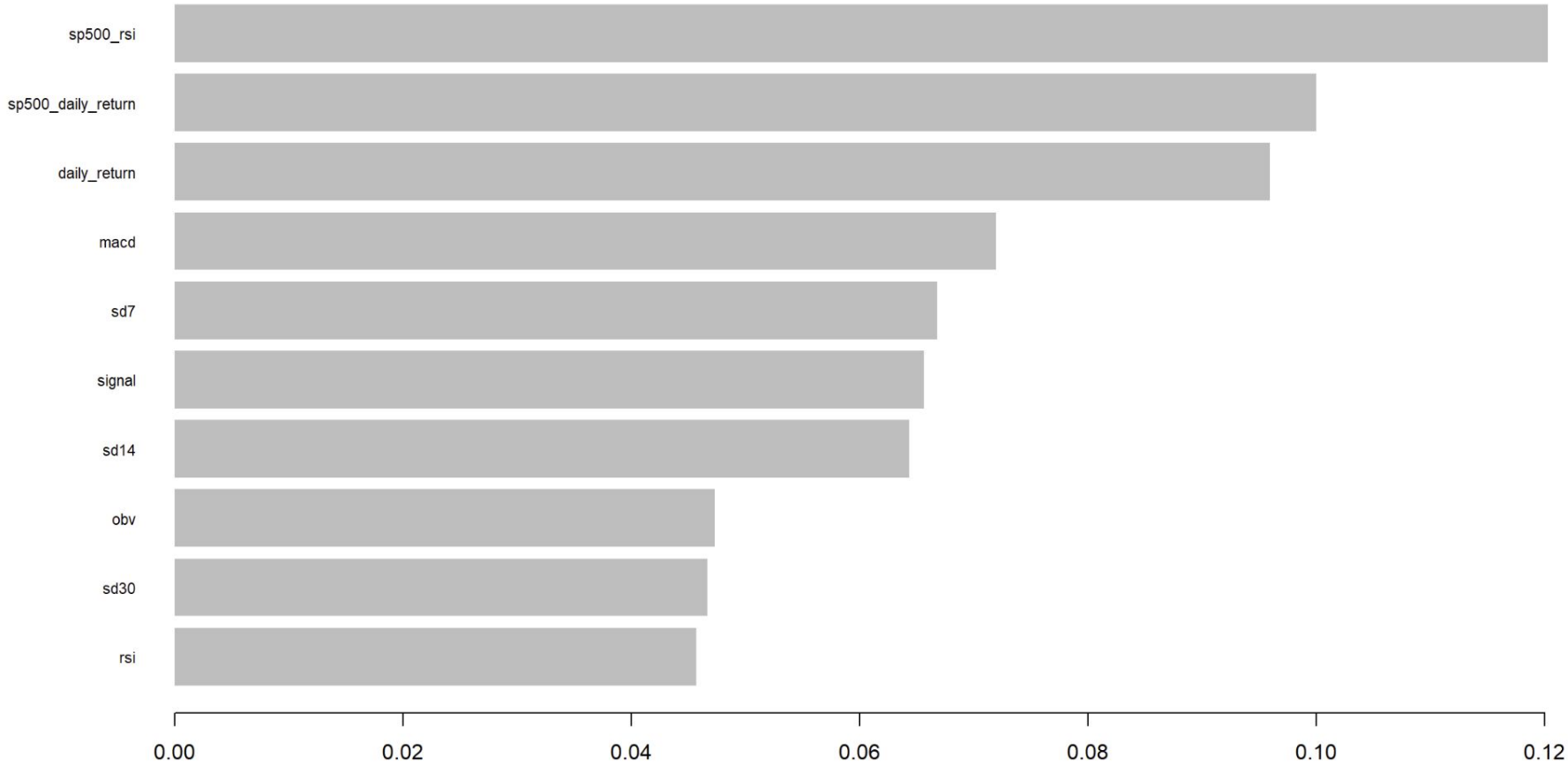
XGBoost (51.06%)



Findings:

Microsoft Corp

NASDAQ: MSFT

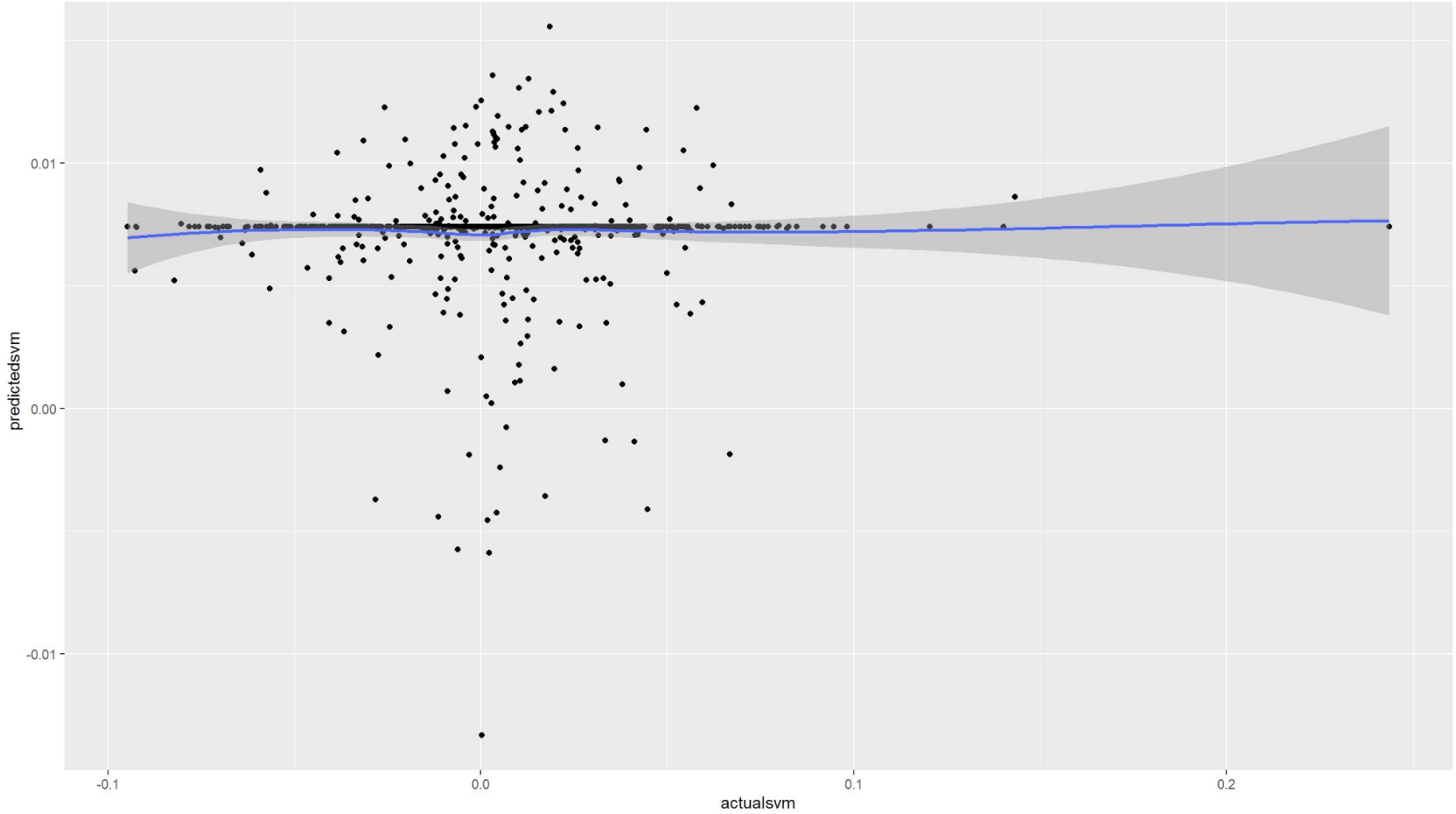


Findings:

NVIDIA Corp

NASDAQ: NVDA

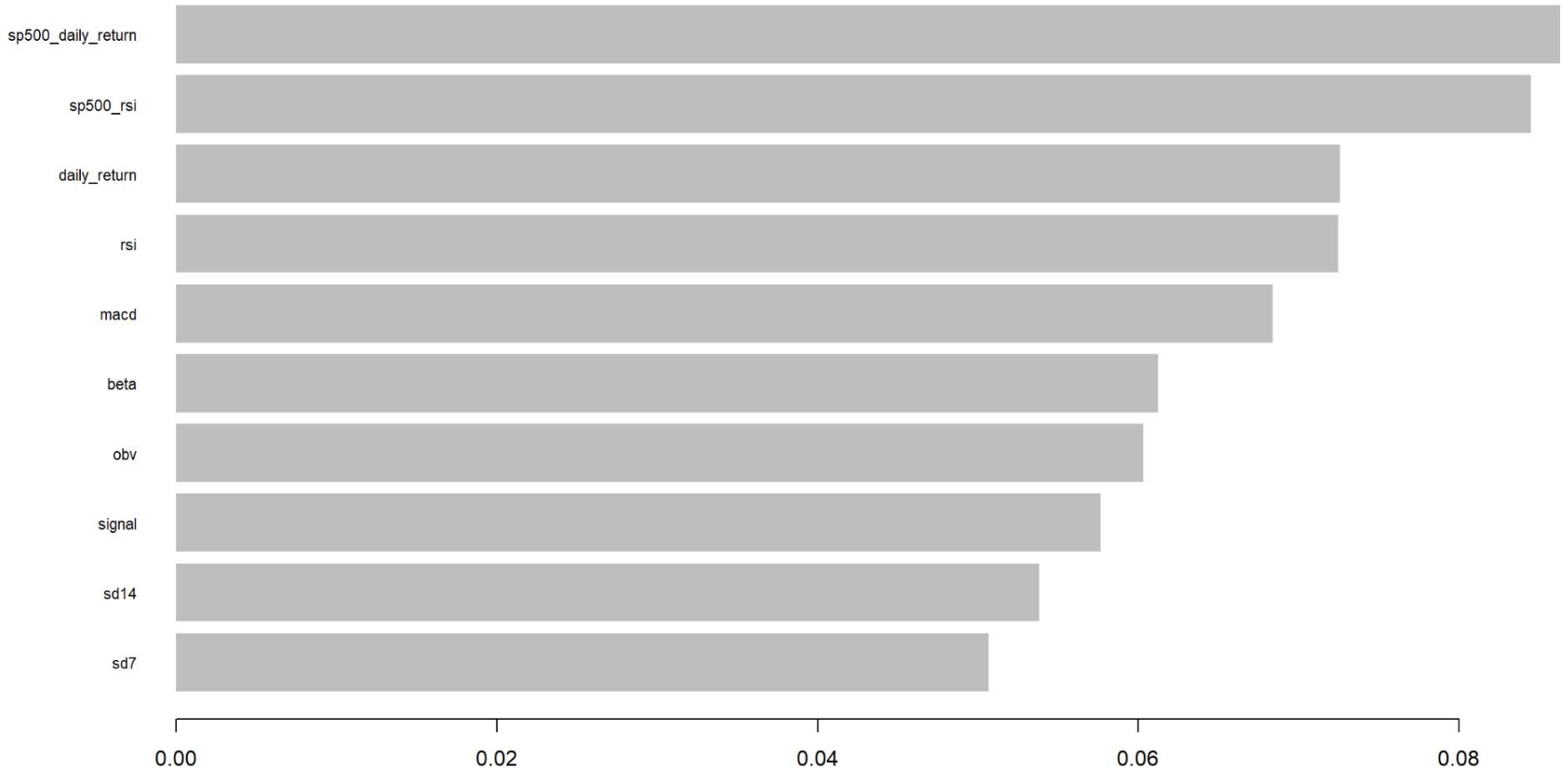
SVM (52.72%)



Findings:

NVIDIA Corp

NASDAQ: NVDA



Overall Model Summaries

Accuracy

	Linear	xgboost	Random Forest	Bagging	Lasso	SVM
AAPL	-	0.4704	0.4882	0.4894	-	**0.5177**
AMZN	-	0.4976	0.4976	**0.5035**	-	0.4669
GOOG	-	0.4894	0.4752	0.4752	-	**0.5213**
MSFT	-	**0.5106**	0.4941	0.4929	-	0.5083
NVDA	-	0.4976	0.5000	0.5059	-	**0.5272**

Specificity

	Linear	xgboost	Random Forest	Bagging	Lasso	SVM
AAPL	-	**0.8809**	0.8065	0.8561	-	0.0323
AMZN	-	0.2434	**0.4193**	0.3952	-	0.2096
GOOG	-	0.8747	**1.0000**	1.0000	-	0.2130
MSFT	-	0.3447	**0.9709**	0.9636	-	0.2500
NVDA	-	**0.1959**	0.1832	0.1908	-	0.0102

Sensitivity

	Linear	xgboost	Random Forest	Bagging	Lasso	SVM
AAPL	-	0.0971	0.1986	0.1558	-	**0.9594**
AMZN	-	**0.7425**	0.5731	0.6079	-	0.7146
GOOG	-	0.1454	0.0067	0.0067	-	**0.7964**
MSFT	-	0.6682	0.0415	0.0461	-	**0.7535**
NVDA	-	0.7594	0.7748	0.7792	-	**0.9757**

RMSE

	Linear	xgboost	Random Forest	Bagging	Lasso	SVM
AAPL	0.0226	0.0314	0.0285	0.0282	0.0444	**0.0190**
AMZN	0.0242	0.0271	**0.0248**	0.0251	0.0466	0.0250
GOOG	0.0204	0.0290	0.0315	0.0332	0.0421	**0.0199**
MSFT	0.0203	0.0213	0.0295	0.0291	0.1044	**0.0180**
NVDA	0.0368	0.0419	0.0457	0.0520	0.0737	**0.0333**

Similar Conclusions

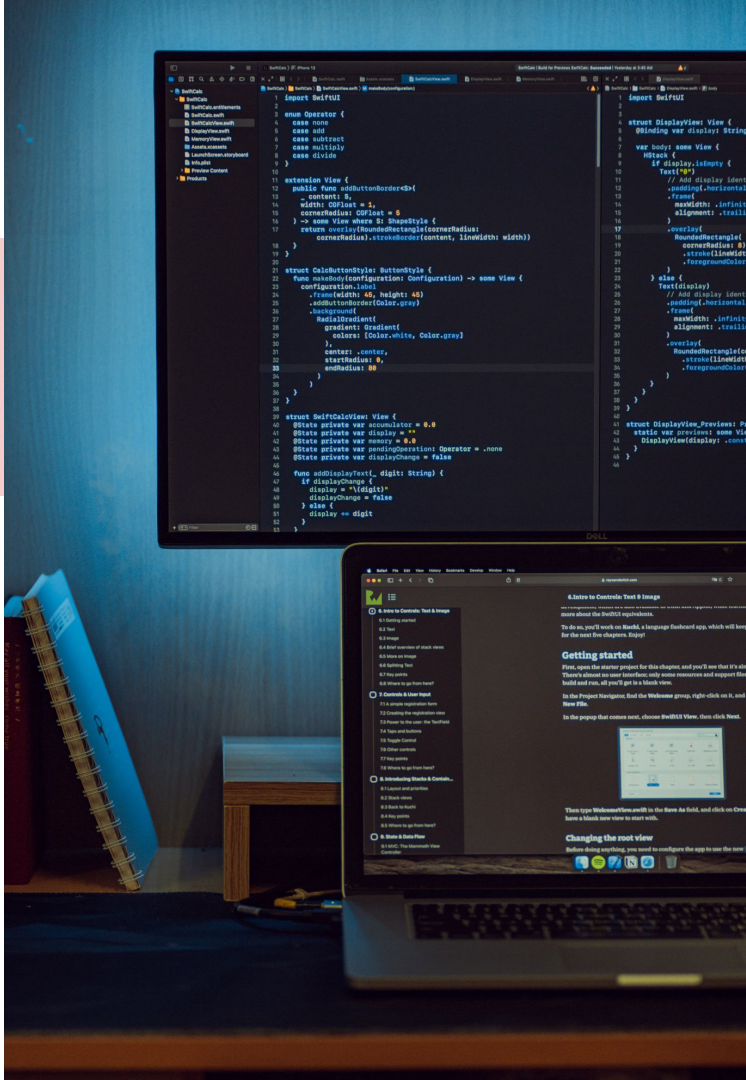
S V M

Model

72% - 99%

Best Performance - High Accuracy

- Lead to the best performance statistics (accuracy, precision, and recall)
- Outperformed other models, especially for S&P index price movement



Johnson, Jaya. 2023. Machine Learning for Financial Market Forecasting. Master's thesis, Harvard University Division of Continuing Education.

Project Constraints and Potential

Limitations

Market Unpredictability

No sentiment analysis or more complex models included.

Model Limitations

Variable Scope

Limited number of variables, tuning complexity

Model Scope

Future Opportunities

Sentiment Analysis:

Twitter, Bloomberg, Financial Times, The Economist

Advanced ML Models:

Adopt LSTM and Bi-LSTM

Expand Number of Features

Custom Ensemble Model

Thank you!