Algorithmic Stock Price Forecasting

Samuel Hughes
Khoury College of Computer Science
Northeastern University
hughes.sa@northeastern.edu

Muhammad Salman
College of Engineering
Northeastern University
salman.mu@northeastern.edu

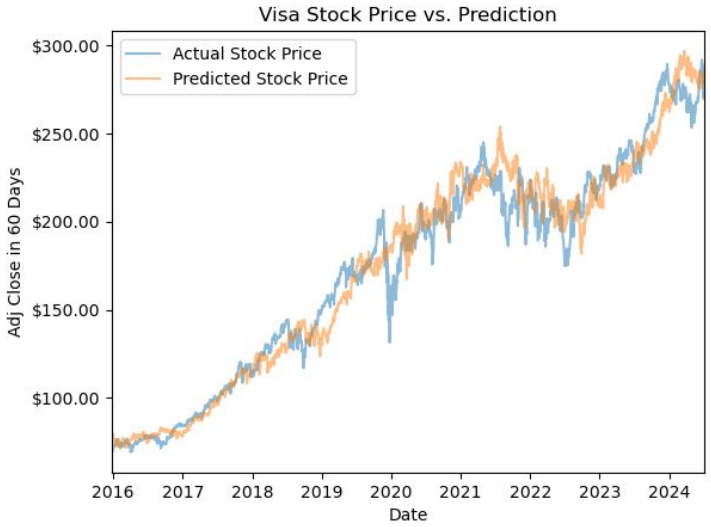
Problem Statement

Goals

- Use state-of-the-art machine learning to forecast future stock prices
- Identify key factors that influence stock price
- Visualize and validate stock price predictions

Motivation

- Algorithmically validated process
- Clear real-world application
- Potential for financial success



Data Overview

Target

- Predict price increase/decrease of given stock 60 business days into future

Scope

- Dow 30: 30 high profile United States companies (e.g. Amazon, Apple, Tesla)
- Retroactively train and predict from October 2015 until September 2024
- Bolster predictive ability via use of external data sources

Financial Data

Traditional Features

- Adjusted closing price
- Trading volume
- Recent stock price gains and losses
- Aggregations of stock price in past 60 business days

Rolling Window Features

- Mean adjusted close and volume in 10 day windows
- Arithmetic differences of adjacent windows

External Factors

Macroeconomics

- Unemployment rate
- Gross domestic product (GDP)
- 10-year treasury yield rate

Congressional Trading

- Reported congressional purchases in past 90 calendar days
- Reported congressional sales in past 90 calendar days

Methods

Models

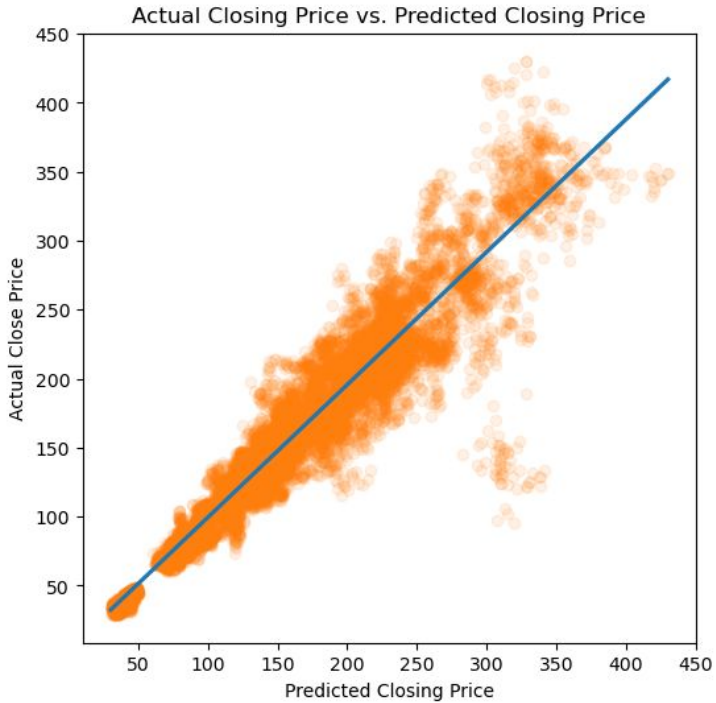
- Gradient boosting (XGBoost, LightGBM)
- Neural networks (Dense neural network, LSTM)

Evaluation

- Hyperparameter tuning to bolster predictive performance
- Train and validate on 22 stocks
- Test on 8 stocks
- Minimize mean absolute error (MAE)
- Compare against naive forecast as baseline

Performance Comparison

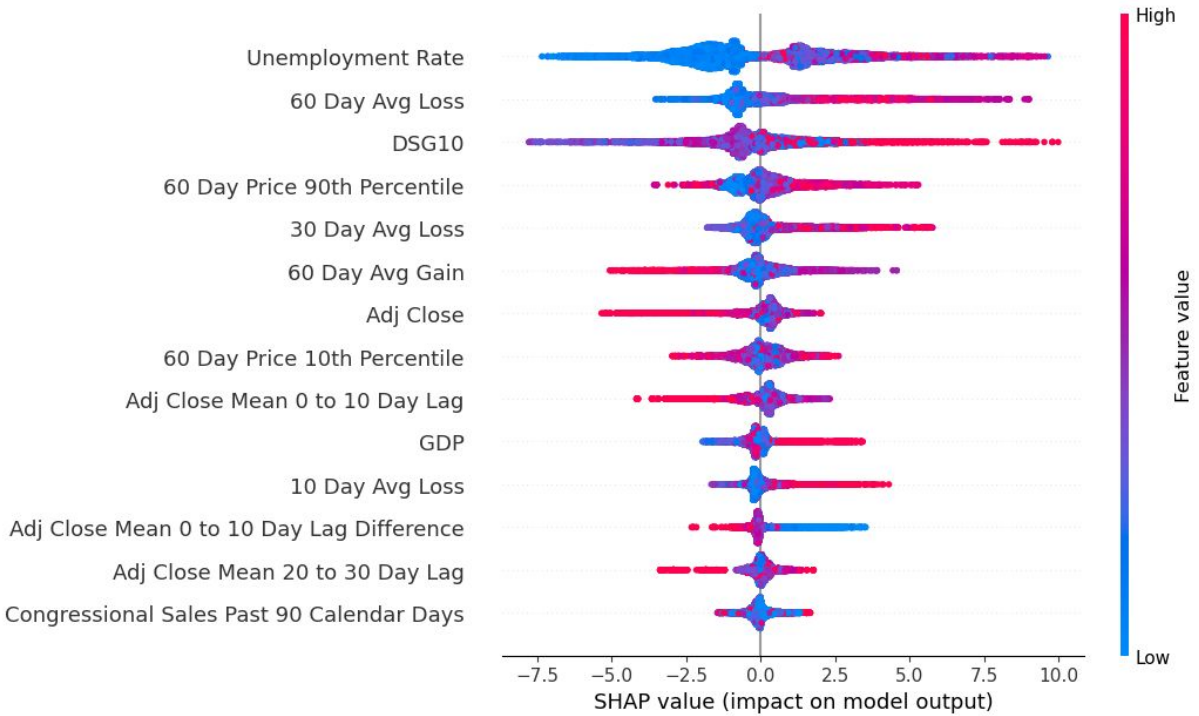
Model	Mean Absolute Error	Improvement Over Baseline
LightGBM	12.78	6.82%
XGBoost	13.14	4.21%
Dense Neural Network	13.21	3.70%
LSTM	13.40	2.32%



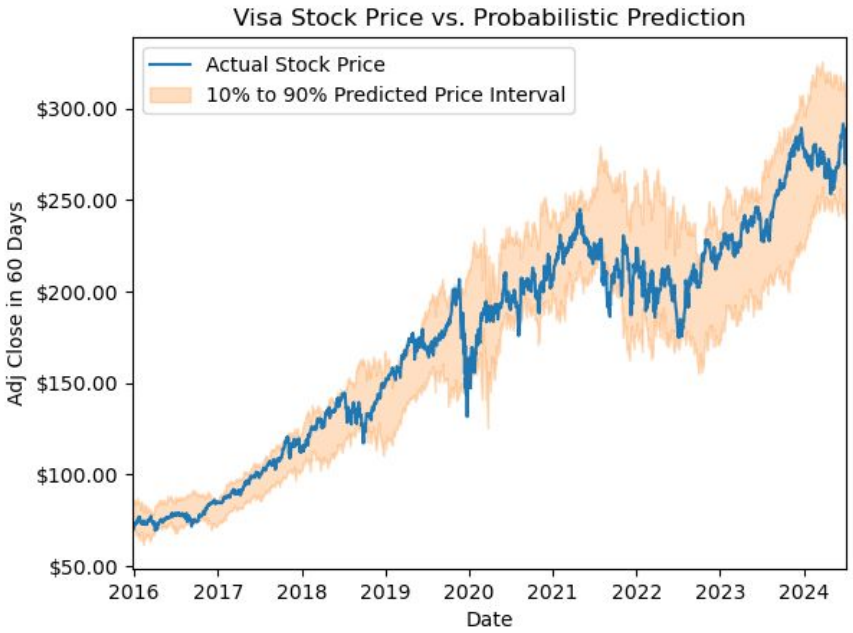
Feature Importance

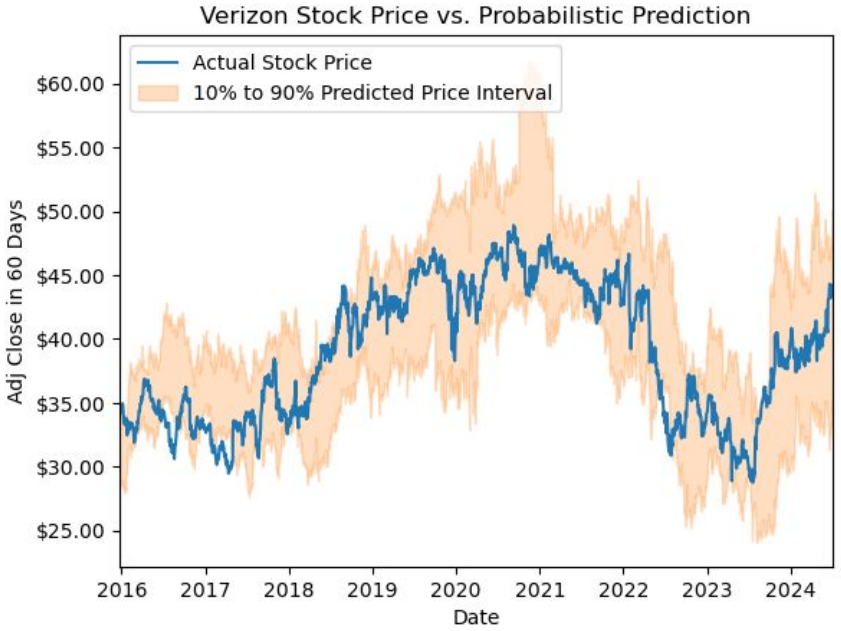
Insights

- Unemployment rate is great indicator
- Recent price decrease may mean rebound is imminent
- Congressional influence is mixed



Probabilistic Forecasts





Discussion

Achievements

- Outperformed benchmark by 6.8%
- Identified driving factors in prediction
- Aided analysis with probabilistic forecast

Novelty

- The use of congressional data and focus on state-of-the-art supervised learning models distinguishes our project from others

Future Work

- Include features based around company and financial news sentiment
- Expand analysis beyond Dow 30 stocks

Thank You!