#128

MARKET MAVEN

A tool for stock price prediction and visualization

Jacob Wright
jwright383@gatech.edu

Robert Carlton
rcarlton7@gatech.edu

Alex Shropshire
ashropshire6@gatech.edu

Manav Kotharia
mkotharia1@gatech.edu

Motivation

The impact of stock price movements is of critical importance to developing sound economic policies, understanding global and individual investment trends, and securing the financial future of governments and individuals.

The United States has the largest stock market globally, with the capitalization of listed companies **exceeding $40T dollars** (2020). It is estimated that nearly **60% of total worldwide equity market value is in US listed stocks** (2023). In 2023, **61% of US adults have money invested in the stock markets**, thus identifying better approaches to stock price prediction and helping investors understand what impacts stock prices is crucial.

Data Innovation

*Current research approaches to stock price prediction rely on business fundamentals or technical stock movements, occasionally combined with social data. Our approach **combines a variety of fundamental, technical and social data.***

Our quest for improved stock price prediction started with the search for better data. We began by considering how both the stock market and companies are evaluated.

Technical analysis, the study of stock price movements and their patterns through the use of candle stick charts, is one well known and widely used approach.

Another approach, **fundamental analysis,** evaluates economic data to draw conclusions about the direction of stock prices. These approaches are occasionally combined with **social sentiment** to increase the robustness of stock price predictions.

Our approach leverages all three types: technical, fundamental and social sentiment data, along with a framework called PESTEL for additional data. The PESTEL framework outlines 6 external factors that influence and affect firm performance - political, economic, social, technological, environmental and legal factors.

To assemble such a broad array of data, we relied on multiple sources. Technical data, such as stock open/close prices were accessed through freely available APIs including **Yahoo! Finance.** Fundamental (macro/micro economic) data was accessed from other free-to-use APIs including the **St. Louis Federal Reserve Economic Data API.** We also used a **paid API**, to access other data that fit within the PESTEL framework such as analyst ratings, stock tweets from Twitter/X and community investment portals, balance sheet and income statement financial ratios, technology investment, and even political stock transactions (e.g. buy/sell orders) - reflecting the political, economic, social, technological, environmental and legal factors that can impact a company and its stock.

» 47 unique data elements / stock
» 730 days of historical data / stock
» 6 categories of PESTEL data / stock

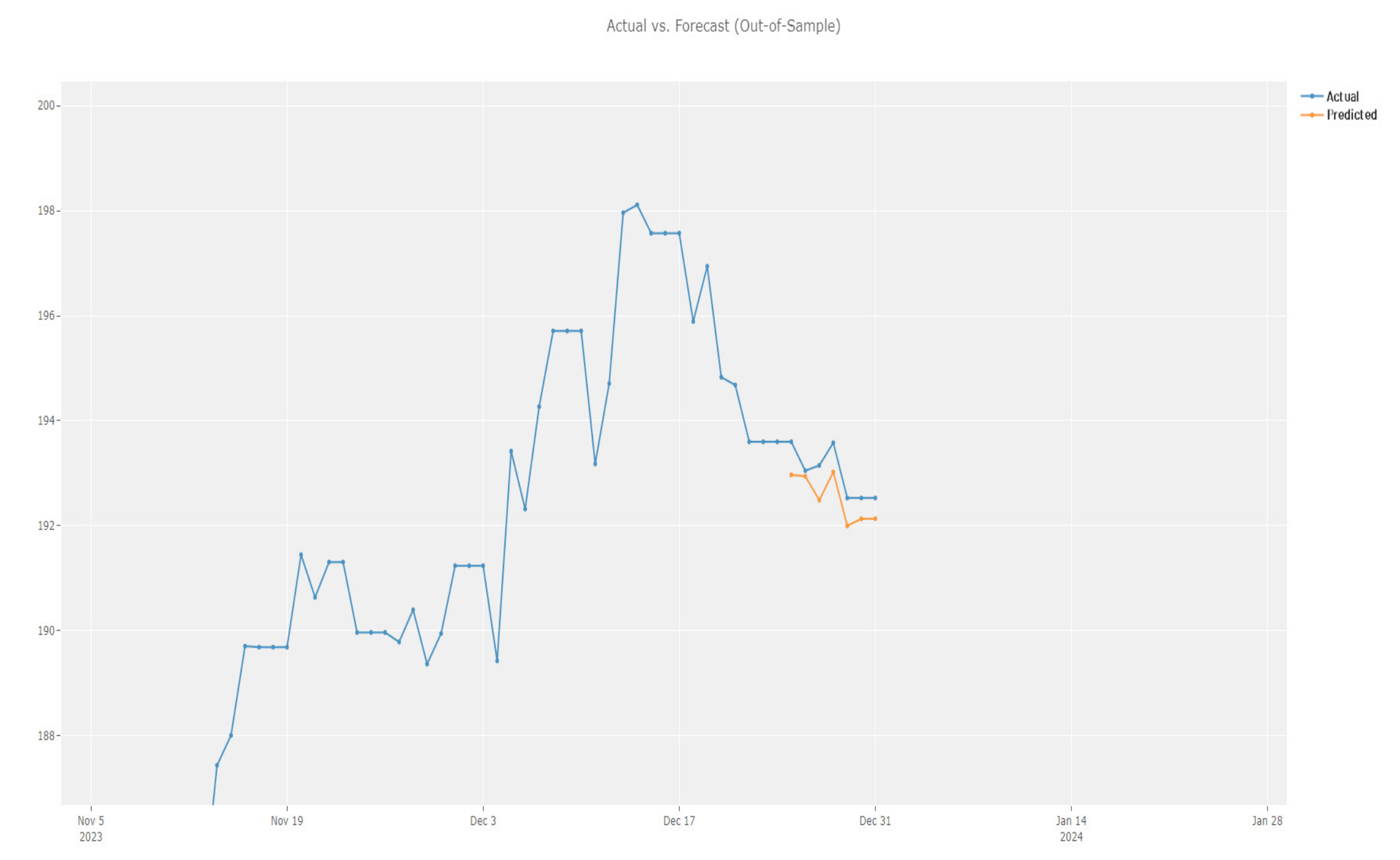

Figure 1: Time series chart showing actual stock price and forecast price using only hyperparameter tuning and model selection.

Ensemble models

*Many different modeling approaches have been used in an attempt to predict stock prices; including ARIMA time series, regression, and random forests. But, rarely have ensemble models been utilized, yet they show promise. Our approach **leverages AutoML techniques to build ensemble models** to predict stock price.*

Stock price prediction is a well studied problem. In the academic research, we found that a wide range of models have been used, including logistic regression, ARIMA, GARCH, ridge and linear regression, recurrent neural networks, LSTM networks, convolution neural networks, and reinforcement learning.

One key finding is that **rarely have ensemble models been studied and applied to stock price prediction.** Stock price prediction is a unique challenge and difficult to solve due to the high levels of noise in the data.

Recent advances in AutoML automated tools can substantially improve the speed and efficiency of data scientists. We leveraged the latest autoML tools available in Python and the PyCaret package to accelerate the selection, hyper-parameter tuning, and ensemble building process.

Our model development began with training across a wide variety of models including **ARIMA, Bayesian Ridge Gradient Boosting, Crostron, ElasticNet, Random Forest, AdaBoost and Light Gradient Boosting, most with Conditional Deseasonalizing and Detrending** using 5-fold cross validation. Model performance was evaluated using RMSE.

From this group, the top 3 models were selected and **5 iterations of hyper-parameter tuning** were completed to optimize RMSE. The optimized top 3 models were then blended into an ensemble model, weighting each individual model 33.3% , then further tuning to optimize model weights. The final model was used to predict stock price 1-to-7 days ahead.

» 8 time-series models plus blending
» 5 iterations of hyper-parameter tuned
» Blending optimization for ensemble weights
» RMSE performance comparison
» Predictions out to 7 days

Key Findings

Our analysis revealed some surprising insights on the data and the algorithms used to predict stock prices, notably around how different models perform on individual stocks and whether ensemble models are the best approach.

Several interesting findings came out of our modeling:

1. No single model had the best performance, based on RMSE, across all the stocks we analyzed. Our peer research revealed that prior academic modeling of stock prices focused on identifying and tuning a single model to predict many individual stocks price, however, our use of AutoML techniques allowed us to test many different models and revealed that **each stock has a model best suited to predicting its future stock price.**

2. During our literature review, we noticed that many different models had been utilized to predict future stock prices but rarely had ensemble models been used. We posited that ensemble models might have advantages over simpler models and tested ensemble models on each stock using PyCaret. To our surprise, **ensemble models, after hyper-parameter tuning and ensemble balancing, were not always the best performing models** based on RMSE. For example, Gradient Boosting returned the optimal RMSE for ticker AAPL (*RMSE= 4.1365*), while Bayesian Ridge Regression provided the optimal RMSE for GOOGL (*RMSE= 0.6662*), and ARIMA provide the optimal RMSE for tick TSLA (*RMSE= 1.0225*).

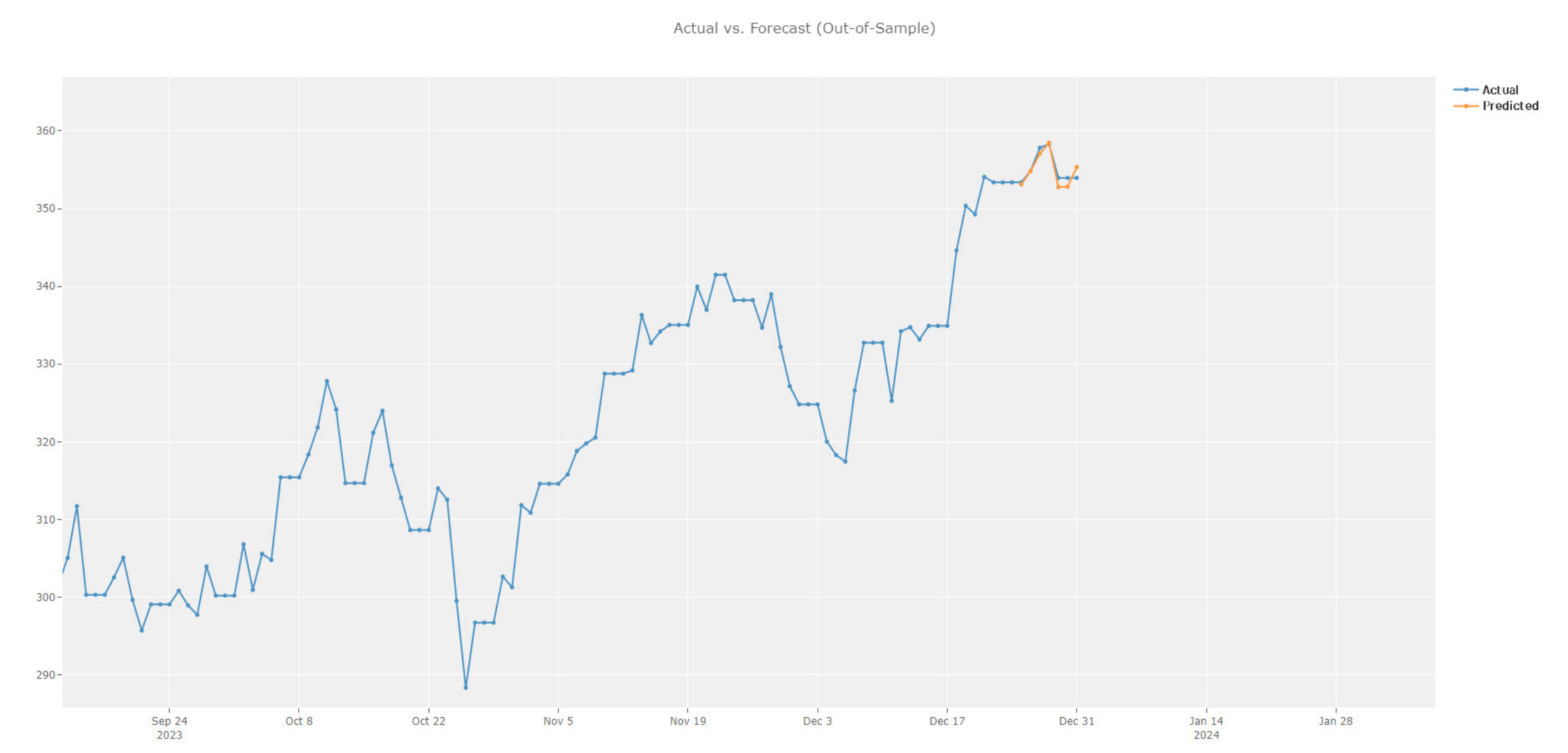

Figure 2: Time series chart showing actual stock price and forecast price using hyperparameter tuning and ensemble model.

Visualization Innovation

Stock price visualizations often consist of historical time series graphs or candlestick charts. However, providing insight into the models and foundational data gives investors more decision information about stock price predictions.

We developed a unique visualization (or dashboard) with 3 design goals:

1. Provide investors with the time series visualization they expect showing historical data and future stock price predictions. This type of visualization is common and easily interpretable, however, we've **added the ability to compare the stock price predictions from the models we utilized**.

2. Provide investors with a better understanding of the fundamental, technical, social and other data used to predict the stock price in an easy-to-understand format and layout. Investors can **see how each of the PESTEL components used to predict the future stock price behaves**.

3. **Provide investors with comparative data on similar stocks across the PESTEL components** (e.g. political, economic, social, technology, environmental & legal) to help them better understand the stock price prediction in the context of other similar stocks.

In our visualization, all of these elements are presented to the investor when they select a specific stock to view its future predicted price.

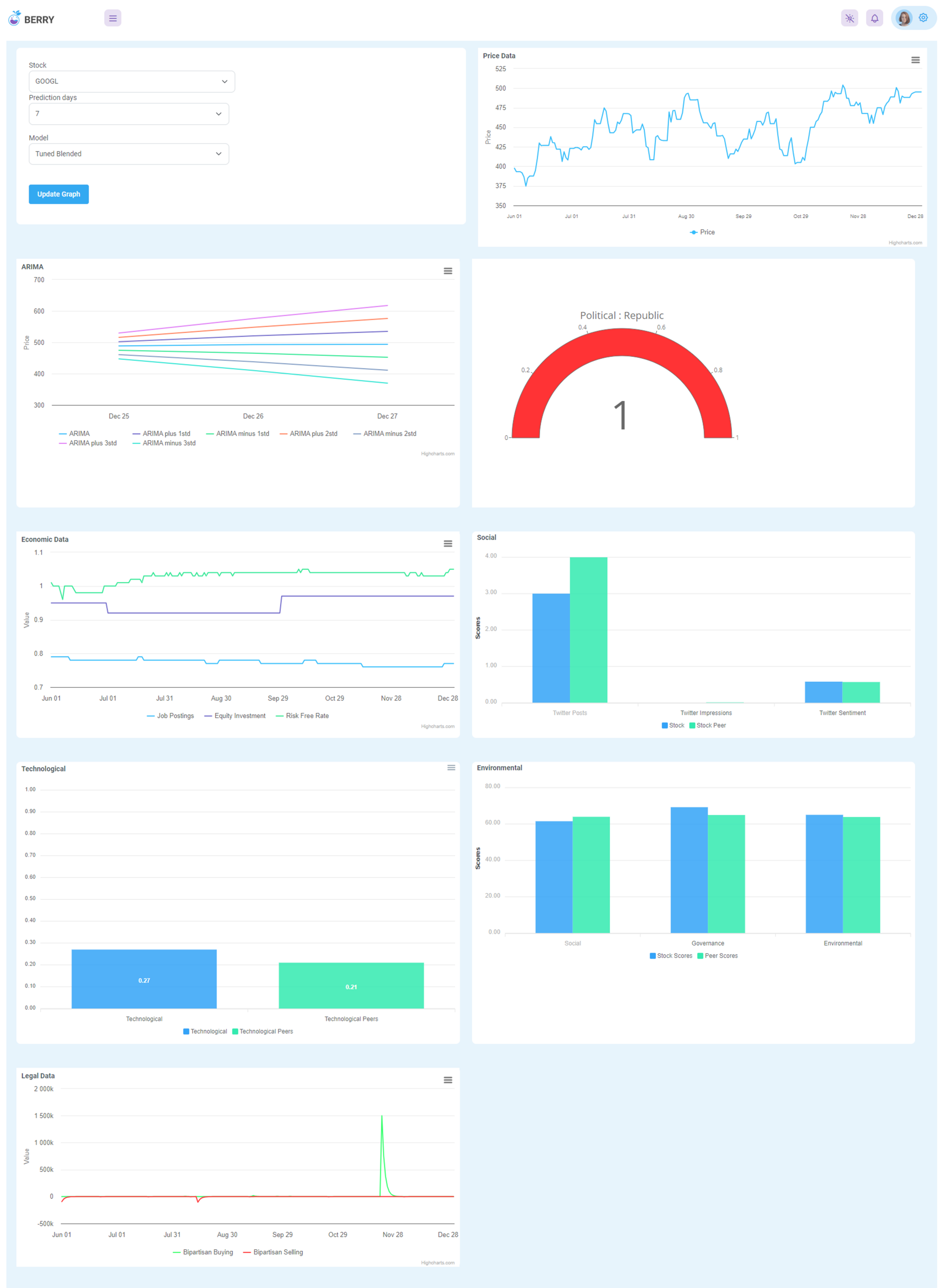
Stock: GOOGL
Prediction days: 7
Model: Tuned Blended

Figure 3: Visual interface showing price prediction, PESTEL metrics, and stock comparative data. Stock and prediction model are user selectable..

Future Directions

Our insights on how our research on stock price prediction could be extended and enhanced.

Our research into stock price prediction has yielded interesting insights in data selection, model optimization and the use of model selection using AutoML and PyCaret.

Opportunities to extend this research include determining if differing sets of input data for each stock might improve price predictions, understanding which type of data (technical, fundamental, social or PESTEL) might have a greater impact on the predictions, and the impact of longer prediction time frames (moving from a 7 day window to 30 or 90 days).