# Interim Report: Stock Market Predictor

Ely Fernandez, Lewi Gao, Dennis Grigoryev, Kirtan Patel, and Andrew Perrone

**Introduction:**

We are creating a service to help investors gain an edge in investing. Along with improving our skills in Python, learning to leverage neural networks, and utilizing web-scraping we want to design a stock market application for beginners and experienced investors to easily get an overview of the current and future market landscape. Coupling our skills with investing knowledge gives us a unique perspective on what potential users may want to have in a market-predicting application.

**Key Technologies:**

- yfinance API: Data Gathering
    - The yfinance API developed by Ran Aroussi is a free and open-source library. It allows access to financial data provided by Yahoo Finance and will be our main way of amassing market data.
- Beautiful Soup: Web Scraping
    - Beautiful Soup is a Python library that takes webpages (HTML or XML) and creates parse trees from them. This will be our data set for generating sentiment analysis.
- Streamlit: Frontend Design
    - Streamlit is an open-source Python library used to create graphs and other UI components. It will be the main way we create the front-end design.
- LSTM: Trend Recognition and Number Crunching
    - Long Short-Term Memory is a type of recurrent neural network that stores and recalls information for long periods and is designed to retain data over multiple time steps. It will be used to analyze stock data and generate projected prices.

- ChatGPT:  Sentiment Analysis
    - Created by OpenAI, ChatGPT is an LLM (large language model) that utilizes machine learning to write responses to queries like humans. It has the functionality to analyze text and generate statements on the overall sentiment of the text.

**Similar Apps:**

*Altindex:*

Altindex is a popular investment analysis site. It takes alternative data points to provide users with a more comprehensive view of an investment portfolio. It uses job postings, website traffic, customer satisfaction, app downloads, and social media trends to analyze stocks. The site itself, https://altindex.com/, trades stocks and provides users with a score using a diverse dataset. However, Altindex falls short as a stock predictor application because the investment site considers social media trends in its analysis, possibly creating bias in its predictions.

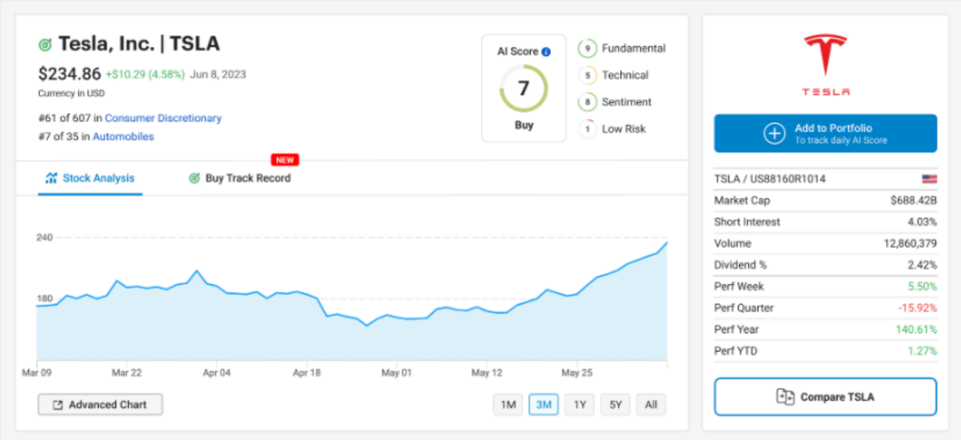
*Morningstar:*

Morningstar is one of the most influential financial services companies in the United States. Morningstar has a variety of services geared towards investors, with Morningstar Direct being one of their analysis tools. Two of its main modules are Direct Lens and Analytics Lab. Direct Lens allows users to build a profile and test how their profile reacts to stress under different scenarios. Morningstar models these outcomes based on the historic performance of the portfolio and by analyzing how predisposed a portfolio is to risk. Analytics Lab lets users freely write code to analyze all of Morningstar's data on a particular stock or portfolio. It is useful for both beginners and experienced coders, with options to pre-populate code using drop-down menus. The outcomes from these analyses can be shared across users. Morningstar Direct's biggest advantage is the sheer amount of data and manpower dedicated to creating the datasets. It also has a lot of flexibility in how users can manipulate data and how it is visualized, allowing users to customize the application to fit their needs.

However, it has a few major issues that make it difficult for a more casual user to navigate. First, Morningstar Direct is gated behind a steep licensing fee of $17,500 annually per person. For this reason, most of Morningstar Direct's clients are corporations or large firms and not individuals. Second, Morningstar Direct's specialty is in larger portfolios and fund analysis, which means that a novice user looking for specific stocks to invest in may be overwhelmed while using the application. [https://www.morningstar.com](https://www.morningstar.com)

*Danelfin:*

Danelfin is an analytics website that uses its AI-powered platform to help users predict stock prices and make the best choices for their investment strategies. Their AI platform combines machine learning and other quantitative finance to give investors a better platform for trading and choosing stocks. Specifically, they calculate probabilities to determine whether a stock will beat the market or not (whether or not it outperforms the predicted price for that company's earnings/share prices). This information is condensed into a score out of 10,  as seen in the example stock shown below. [https://danelfin.com/](https://danelfin.com/)


Tesla, Inc. | TSLA

$234.86 +$10.29 (4.58%) Jun 8, 2023

Currency in USD

#61 of 607 in Consumer Discretionary

#7 of 35 in Automobiles

AI Score 7 Buy

9 Fundamental

5 Technical

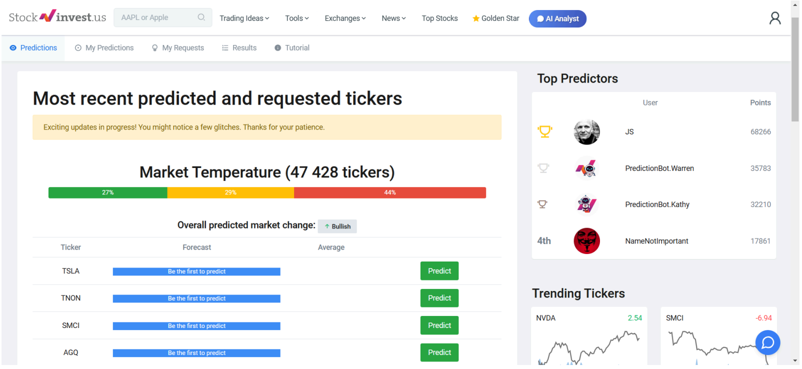
8 Sentiment

1 Low Risk

| | |
|---|---|
| TSLA / US88160R1014 | |
| Market Cap | $688.42B |
| Short Interest | 4.03% |
| Volume | 12,860,379 |
| Dividend % | 2.42% |
| Perf Week | 5.50% |
| Perf Quarter | -15.92% |
| Perf Year | 140.61% |
| Perf YTD | 1.27% |

*Stockinvest:*

Stockinvest uses a model and analysts' opinions to predict the direction of certain stocks and the market as a whole for the next day, which can take a few hours. However, it also has a feature where users can practice predicting the direction stocks will go on a moment-to-moment basis. It aggregates this data from all of its users showing their current opinion of the market. Stockinvest also provides investment ideas daily. Overall, Stockinvest provides a small number of features used for predicting the market and practicing predictions. https://stockinvest.us/



Most recent predicted and requested tickers

Exciting updates in progress! You might notice a few glitches. Thanks for your patience.

Market Temperature (47 428 tickers)

Overall predicted market change: ↑ Bullish

| Ticker | Forecast | Average | |
|---|---|---|---|
| TSLA | Be the first to predict | | Predict |
| TNON | Be the first to predict | | Predict |
| SMCI | Be the first to predict | | Predict |
| AGQ | Be the first to predict | | Predict |

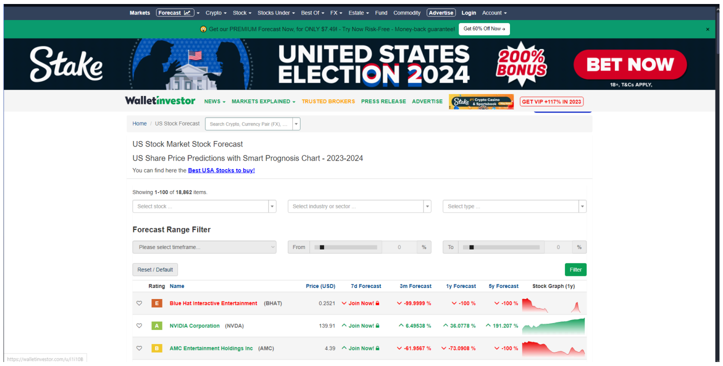
Top Predictors

| | | User | Points |
|---|---|---|---|
| | | JS | 68266 |
| | | PredictionBot.Warren | 35783 |
| | | PredictionBot.Kathy | 32210 |
| 4th | | NameNotImportant | 17861 |

Trending Tickers

*WalletInvestor:*

WalletInvestor's goal is to beat automated software and professionals and give people the resources necessary to navigate through a volatile market. WalletInvestor uses the history of traditional markets and their volatility data to forecast how a specific stock may change on a day-to-day basis. However, since it is an AI-based price predictor, the data it has does not accurately reflect real-world events and how they may influence the everyday market. The filter options do help investors to create a foundation in their investment but the homepage UI can overwhelm and create stress for those who are not ready to start fully creating their investments. Another roadblock with WalletInvestor is that to have a portfolio and keep track of the stocks you would want as an investor, the annual fee is close to $120, including taxes and fees for about a 70% accuracy rate. https://walletinvestor.com/stock-forecast#premium



US Stock Market Stock Forecast

US Share Price Predictions with Smart Prognosis Chart - 2023-2024

You can find here the Best USA Stocks to buy!

Showing **1-100** of **18,862** items.

**Forecast Range Filter**

| | Rating | Name | Price (USD) | 7d Forecast | 3m Forecast | 1y Forecast | 5y Forecast | Stock Graph (1y) |
|---|---|---|---|---|---|---|---|---|
| ♡ | E | Blue Hat Interactive Entertainment (BHAT) | 0.2521 | Join Now! | -99.9999 % | -100 % | -100 % | |
| ♡ | A | NVIDIA Corporation (NVDA) | 139.91 | Join Now! | 6.49538 % | 36.0778 % | 191.207 % | |
| ♡ | B | AMC Entertainment Holdings Inc (AMC) | 4.39 | Join Now! | -61.9567 % | -73.0908 % | -100 % | |

## Overall Design:

*Streamlit - Frontend Design:*

Streamlit will be used to host the website and will include various modules with graphs, tables, and current stock prices. Streamlit will streamline the process of creating a prototype UI. The web application will include a drop-down bar where we can select the stocks that we want to see. The tables will show the prices of the specific stock such as the opening price, closing price, and other prominent information. We will provide historical data spanning from 10 years out so users can have context for their decisions. Stock prediction will be shown as an overlay on top of the real-time price of the selected stock. These graphs and tables will use data from yfinance. The sentiment analysis feature will display text written by ChatGPT and a percentage that represents the amount of positive/negative words used to describe a stock. We will primarily be using Streamlit for frontend integration and the backend will primarily use Python.

*Neural Network (LSTM):*

We will use a neural network to predict stock market performance. A neural network is a series of algorithms that train a computer to process data and recognize patterns in a structured manner mimicking the human brain. It consists of nodes that are connected, similar to neurons. Neural networks are layered with nodes, input, output, and hidden layers in between. The nodes are connected to others and when activated send data to the next node when their outputs are above a specified value. These networks take time to train to become more accurate and precise. IBM lists the formula for individual nodes in a neural network as:

$$\sum w_i x_i + \text{bias} = w_1 x_1 + w_2 x_2 + w_3 x_3 \ldots + bias$$
$$where\ w = weight,\ x = input,\ bias = \ detection\ term$$

It is important to note that because stock markets are susceptible to change, it is a challenge to have accurate predictions. A Long Short-Term Memory (LSTM) neural network provides algorithms that can help overcome these challenges. It can process single data points within sequences of data. Recurrent Neural Networks (RNNs) often forget older data, and cannot remember in longer sequences, whereas an LSTM neural network can retain more information for longer periods, therefore analyzing patterns more efficiently. This issue is called the “Vanishing Gradient Problem.”

For our project, we have decided to implement an LSTM neural network. Here is the current LSTM we are testing, following steps from the article: Implementing Time Series Stock Price Prediction with LSTM and yfinance in Python. (cited below):

```
model = Sequential()
model.add(LSTM(units=50, return_sequences=True, input_shape=(x.shape[1], 1)))
model.add(LSTM(units=50))
model.add(Dense(units=1))
model.compile(optimizer='adam', loss='mean_squared_error')
```

Currently, we are trying to alter our program to make it run properly for an LSTM model. For our model, we will be reshaping the data to ensure it matches the expected input format for the LSTM layer. We will also be defining several time steps to control the amount of past information the model remembers/forgets. Here is a snippet of some code explained:

```
time_steps = 10
```

- our number of past data/observations to compare to

  *x, y = create_lstm_data(close_prices_scaled, time_steps)*

- creates input data (x) and target values (y) using a function defined earlier

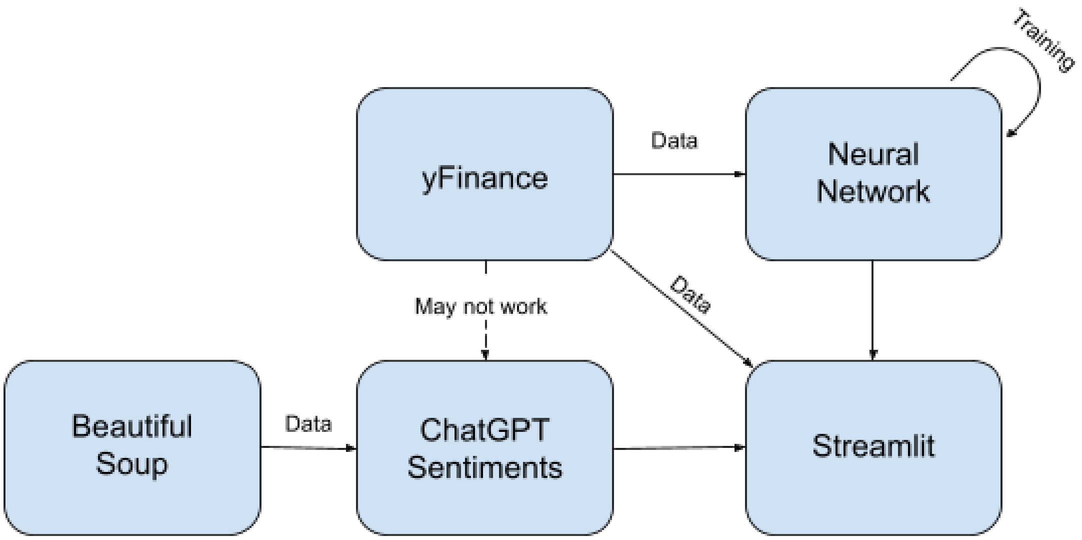
  *x = np.reshape(x, (x.shape[0], x.shape[1], 1))*

- reshapes the input data to fit the LSTM model requirements
- LSTMs expect input of the shape of (samples, time steps, and features)

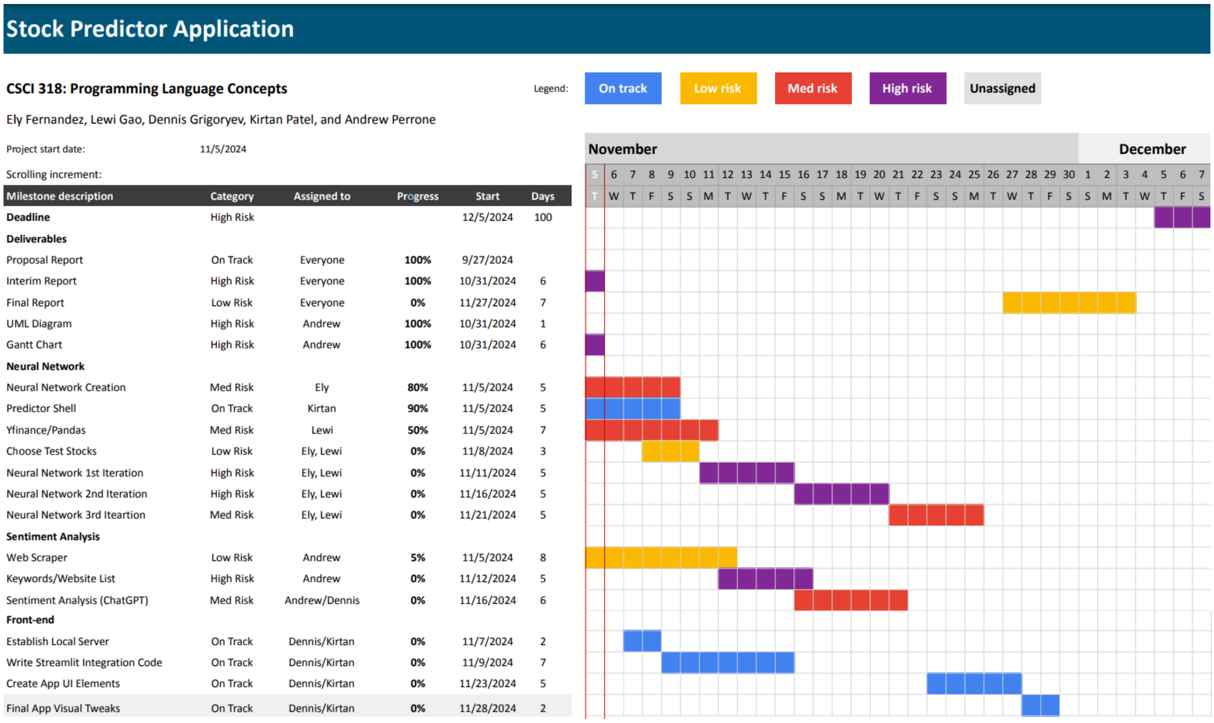
*Sentiment Analysis (BeautifulSoup and ChatGPT):*

To add a second dimension to stock analysis, our application will utilize sentiment analysis to gauge public opinion on different stocks, which is important because markets are vulnerable to public opinion. We will use BeautifulSoup to create a dataset of people discussing the stock market and use the data to create a sentiment analysis. BeautifulSoup converts the text on webpages into parse trees so it can be manipulated. To pull only the most relevant websites, we will only run BeautifulSoup on a curated list of websites and blogs known for discussing the stock market. To run the sentiment analysis, we will create lists of words with positive and negative associations. Once those lists are created, we can compute the ratio of positive to negative words in the dataset. Then, we will query ChatGPT with the list of words and request its opinion of the overall sentiment, using the result to judge whether a stock may trend positively or negatively through public opinion.

## UML Diagram:



## Implementation Calendar:



### Stock Predictor Application

**CSCI 318: Programming Language Concepts**

Legend: On track Low risk Med risk High risk Unassigned

Ely Fernandez, Lewi Gao, Dennis Grigoryev, Kirtan Patel, and Andrew Perrone

Project start date: 11/5/2024

Scrolling increment:

| Milestone description | Category | Assigned to | Progress | Start | Days |
|---|---|---|---|---|---|
| **Deadline** | High Risk | | | 12/5/2024 | 100 |
| **Deliverables** | | | | | |
| Proposal Report | On Track | Everyone | **100%** | 9/27/2024 | |
| Interim Report | High Risk | Everyone | **100%** | 10/31/2024 | 6 |
| Final Report | Low Risk | Everyone | **0%** | 11/27/2024 | 7 |
| UML Diagram | High Risk | Andrew | **100%** | 10/31/2024 | 1 |
| Gantt Chart | High Risk | Andrew | **100%** | 10/31/2024 | 6 |
| **Neural Network** | | | | | |
| Neural Network Creation | Med Risk | Ely | **80%** | 11/5/2024 | 5 |
| Predictor Shell | On Track | Kirtan | **90%** | 11/5/2024 | 5 |
| Yfinance/Pandas | Med Risk | Lewi | **50%** | 11/5/2024 | 7 |
| Choose Test Stocks | Low Risk | Ely, Lewi | **0%** | 11/8/2024 | 3 |
| Neural Network 1st Iteration | High Risk | Ely, Lewi | **0%** | 11/11/2024 | 5 |
| Neural Network 2nd Iteration | High Risk | Ely, Lewi | **0%** | 11/16/2024 | 5 |
| Neural Network 3rd Iteartion | Med Risk | Ely, Lewi | **0%** | 11/21/2024 | 5 |
| **Sentiment Analysis** | | | | | |
| Web Scraper | Low Risk | Andrew | **5%** | 11/5/2024 | 8 |
| Keywords/Website List | High Risk | Andrew | **0%** | 11/12/2024 | 5 |
| Sentiment Analysis (ChatGPT) | Med Risk | Andrew/Dennis | **0%** | 11/16/2024 | 6 |
| **Front-end** | | | | | |
| Establish Local Server | On Track | Dennis/Kirtan | **0%** | 11/7/2024 | 2 |
| Write Streamlit Integration Code | On Track | Dennis/Kirtan | **0%** | 11/9/2024 | 7 |
| Create App UI Elements | On Track | Dennis/Kirtan | **0%** | 11/23/2024 | 5 |
| Final App Visual Tweaks | On Track | Dennis/Kirtan | **0%** | 11/28/2024 | 2 |

**Conclusion:**

There are multiple stock market prediction applications available, but each application has its shortcomings in terms of creating a holistic analysis. Our project attempts to look at a combination of analyses that we have not been able to find in other existing applications. Most importantly, we want to run a sentiment analysis simultaneously with an LSTM neural network model to show public opinion of a given stock alongside a predictive model. Seeing how volatile public opinion is can influence a user's decision to invest, making our application easy to understand and navigate allows even the most novice of investors to make strategic decisions. By utilizing a combination of analyses and lowering the barrier to entry for users, our application attempts to make stock market prediction accessible to anyone looking to invest in the market.

# References

(n.d.). AltIndex: Alternative Data for Investing. Retrieved November 5, 2024, from https://altindex.com/

About Streamlit in Snowflake. (n.d.). Snowflake Documentation. Retrieved November 5, 2024, from https://docs.snowflake.com/en/developer-guide/streamlit/about-streamlit

*Chat GPT: What is it?* (n.d.). University of Central Arkansas. Retrieved November 5, 2024, from https://uca.edu/cetal/chat-gpt/

Guide, S. (2023, December 22). *Implementing Time Series Stock Price Prediction with LSTM and yfinance in Python*. Medium. Retrieved November 5, 2024, from https://medium.com/@deepml1818/implementing-time-series-stock-price-prediction-with-lstm-and-yfinance-in-python-a769a3fe9a7b

Guide, S. (n.d.). *yfinance Library - A Complete Guide - AlgoTrading101 Blog*. Algo Trading 101. Retrieved November 5, 2024, from http://algotrading101.com/learn/yfinance-guide

*How it Works | Stock Ratings Powered by AI. Pick Best Stocks to Buy.* (n.d.). How It Works | Stock Ratings Powered by AI. Pick Best Stocks to Buy. https://danelfin.com/how-it-works IBM. (n.d.). *What is a Neural Network?* IBM. Retrieved October 31, 2024, from https://www.ibm.com/topics/neural-networks

How to Do Sentiment Analysis (with Python). (2023, June 18). https://www.lupagedigital.com/blog/sentiment-analysis-python/

K. R. Reddy, B. T. Kumar, V. R. Ganesh, P. Swetha and P. K. Sarangi, "Stock Market Prediction Using Recurrent Neural Network," 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), Bhopal, India, 2022, pp. 1-4, doi: 10.1109/CCET56606.2022.10080154. (*US Stock Market Stock Forecast, US Stock Price Predictions With Smart Prognosis Chart 2023-2024*, n.d.)

US Stock Market Stock Forecast, US Stock Price Predictions with Smart Prognosis Chart 2023-2024. (2023). Walletinvestor.com. https://walletinvestor.com/stock-forecast#premium