

# Stock Predictor

Clay Roberts, Pete Acevedo, Pedro Alvarado, Awaleh Houssein, Mateo Rivas

**Problem Statement:** The stock market, including individual stocks, is volatile and has a tendency to deceive traders into investing just before an inevitable downturn. In this project, we intend to create an individual stock predictor that can provide valuable insight for making worthwhile stock predictions using a combination of machine learning techniques. Our goal is to reduce the risk of poor return on investment and boost confidence in an individual's stock holdings through the use of our predictive model.

**Methods:** The methods we plan to employ are Long Short Term Memory (LSTM), Moving Average Time Series analysis, and the possible use of an attention model to minimize reliance on Recurrent Neural Network (RNN), which are typically used in LSTM. RNNs tend to lose important relationships over long distances between data and take considerable computational resources to train.

**New Approach:** In the past, sentiment analysis has been implemented to assist with stock predictions. However, we instead propose using sentiment analysis to predict if a press release for a company was announced with the intention of announcing positive news, or if that press release was instead announced to mitigate the impact of a future negative press release. This approach could introduce valuable insight into an investment prior to committing financial resources. This could be a very interesting prospect, but it's not 100% clear if we will be including this novel approach in our final project. If we have sufficient time and resources after completing the initial problem, we intend to include this novel approach into our final project.

**Experimental Setup:** Historical stock data will be gathered and cross-validated to maintain a number of training and testing data sets. Data will be cleaned and dimensionally reduced to optimize predictions while keeping the model trainable on a standard personal computer. If we have the opportunity to pursue the novel approach, we will need to scrape press release information from a website such as Bloomberg. Outputs can return simple binary responses, either "invest" or "don't invest", or we can provide confidence percentages on investment. To check the validity of methods used, we will handle using a number of metrics, including:

1. Mean Absolute Percentage Error: This metric provides the average difference between predicted and actual stock values.
2. Root Mean Squared Error: This metric places additional focus on outlier events.
3. Sharpe Ratio: This metric allows us to compare the return on investments with the associated risk of the investment.

