# Stock Price Trend Prediction using Machine Learning

Mentor: Professor Wendy Tang
Names of Team Members, Michael Barrett & Jack Ye

Stony Brook University
Electrical and Computer Engineering

SATURN

## Background

Currently machine learning is being widely adopted in many areas and fields. One specific field in particular is the finance and investment sector. In regards to finance, machine learning is being used for tasks such as operation efficiency, product development, data mining, as well as analysis/prediction. When dealing with personal investment, we tend to think of the stock market, commodities, and more recently cryptocurrency. Each type of investment has their own benefits and drawbacks. Due to the large amount of available public data and ease of access, we decided to focus our project towards stock market trend prediction. Inherently the stock market is heavily influenced by public perception and as a result has shown high volatility. The intuitive logic is to use the mathematical analysis and learning capability of machine learning to identify key patterns to predict upcoming trends within the stock market.

## Objective

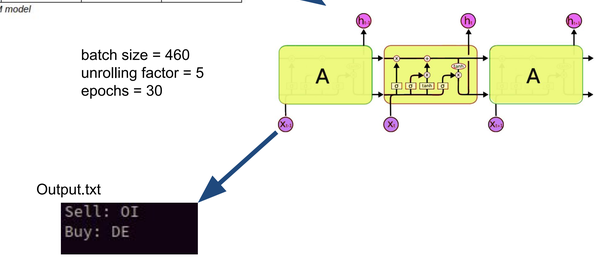
The goal of this project is to implement a stock market prediction system using machine learning in a Tensorflow environment. The main focus of the model is to reliably predict the upcoming trend for specified companies within the industrial, chemical, or agricultural sector in the stock market. The acceptable reliability of the model is at least an over fifty percent accuracy for binary trend prediction.

## Design Approach

| Date | Low | High | Close | Open |
|---|---|---|---|---|
| 2020-04-24 | 7.90 | 8.37 | 8.28 | 8.08 |
| 2020-04-23 | 7.80 | 8.62 | 7.92 | 8.10 |
| 2020-04-22 | 6.95 | 8.09 | 7.95 | 7.00 |
| ... | ... | ... | ... | ... |

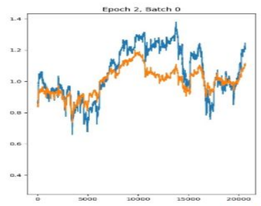
*example of data input format for LSTM model*

batch size = 460
unrolling factor = 5
epochs = 30

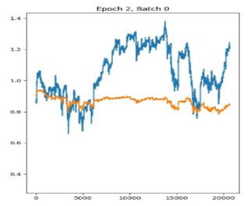



## CNN models

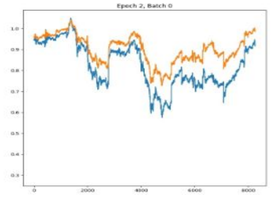
ReLU Activation With a 0.5 ratio of training/testing data



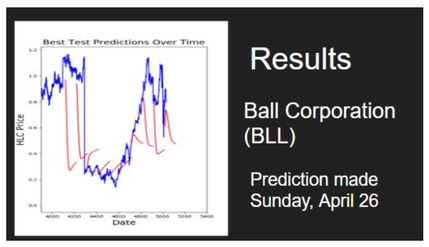

Tanh Activation With a 0.5 ratio of training/testing data



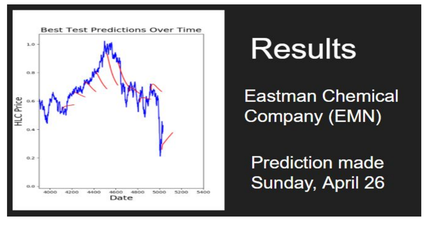

SELU Activation With a 0.8 ratio of training/testing data



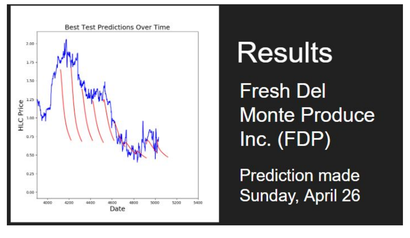

## LSTM model













## Result/Analysis

In analysis, closeness of the prediction line to the stock line was not considered. Rather, the prediction line was considered as a general binary trendline which was either an uptrend or a downtrend. Prediction lines that were not uptrends or downtrends were ignored.

Example of accuracy calculation:

For 100 day time windows on stocks in the agriculture / chemical / industrial sector, 34 correct predictions and 31 incorrect predictions resulted in an accuracy of 52.3%.

### Glossary

- Convolution Neural Network (CNN) - A feedforward neural network that uses weight and bias parameters for backward propagation to retrain itself.
- Machine learning - A type of artificial intelligence in which computers use huge amounts of data to learn how to do tasks rather than being programmed to do them.
- Long Short Term Memory networks (LSTM) - A type of neural network that uses time based data to analyze patterns that can be used for inference of future predictions. Indifferently treats the time difference between the data, by using states. Additionally it combines the weighted data before passing it to the following layer and updates itself after output data and comparison to ideal output.
- TensorFlow - An open-source software library for dataflow and differentiable programming across a range of tasks, often used for machine learning.

### Acknowledgements

Advisor: Professor Wendy Tang
Instructor: Professor Harbans Dhadwal

## Team SATURN