Stock Price Prediction Using Machine Learning Algorithms

Kate Catalena, Nader Ghasemi, Mohd Faisal Khan, Zebo Xiong

Texas A&M University

Introduction

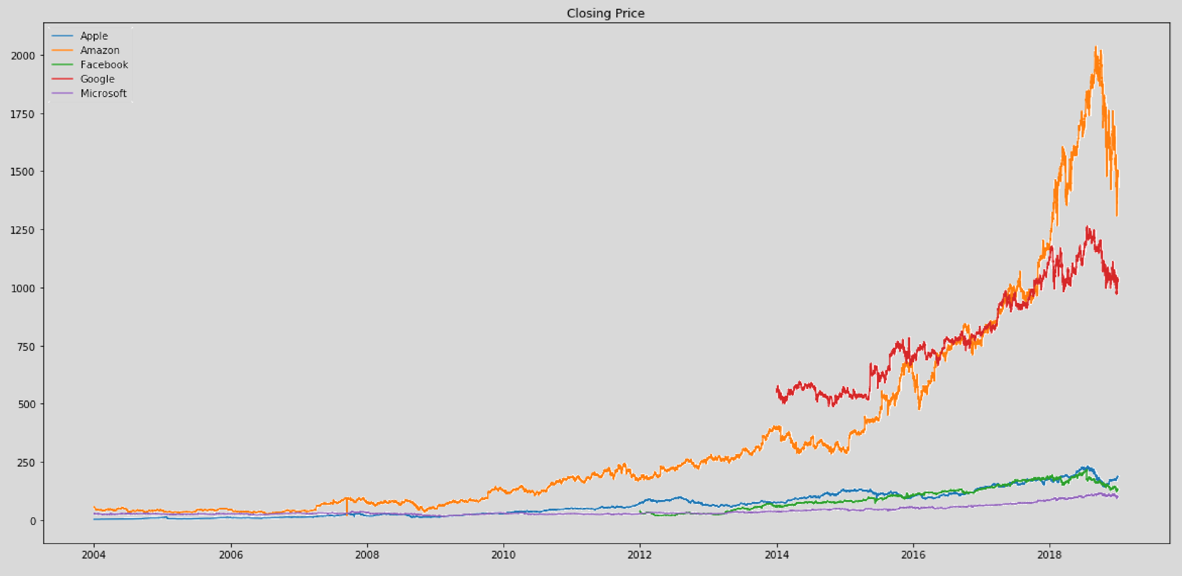
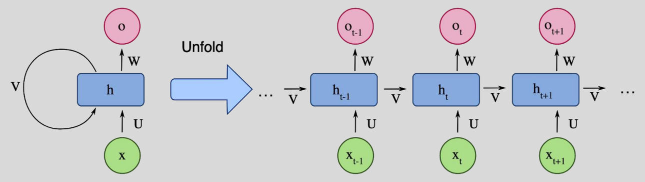


Fig 1: Variation of Stock prices of Big 5 Tech companies,

Predicting stock market trends has been and still is a very popular area of research. Stock market prices represent changes in the economy. Many factors go into determining what a stock market price will do in the future, which makes predicting future trends nearly impossible. In this research, we plan to take a survey of current machine learning techniques and apply them to financial data to analyze which techniques perform better in terms of predicting future stocks. This survey would be useful in understanding each method, including their strengths, weaknesses, and overall performance in predicting future stocks. The objective here is to, based on the historical daily/hourly data, we wish to make certain predictions on the next days' trends

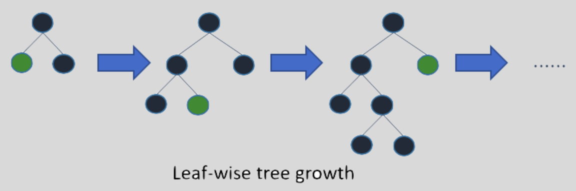
Methods

LSTM



Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images). A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

XGBOOST



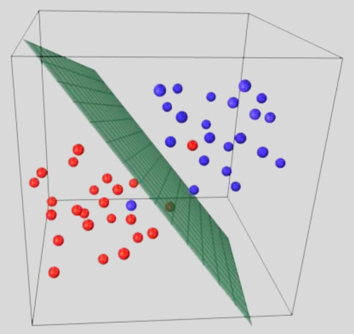
XGBoost is a decision tree based algorithm that uses decision trees. It uses parallelized tree building and tree pruning using depth first search, Using the data, the parameters like tree depth, learning rate, gamma and more factors are optimized on the training model and tested.

REGRESSION/SVR

Regression is one of the most basic, easy, and quick to implement ML algorithms. There are a variety of regression methods. For this research, we have implemented several of sklearn's regression methods: SVR, SGD, Polynomial, and several others.

Our main focus was on SVR regression, where we could vary several model parameters in order to pinpoint the model that would provide the best performance.

SVM



The specific kernel function we use in this study is the radial basis kernel (RBF).

•In this study we tested four features to predict stock price direction, price volatility, index momentum, sector volatility, and sector momentum and finally selected the stock momentum.

• Averaging some quantity over the past 5 minutes.

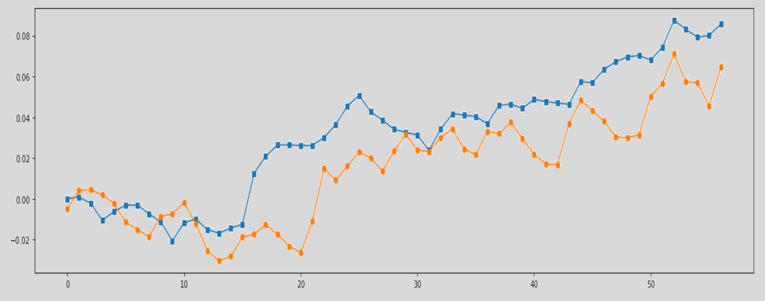
• Stock Momentum

Related-work

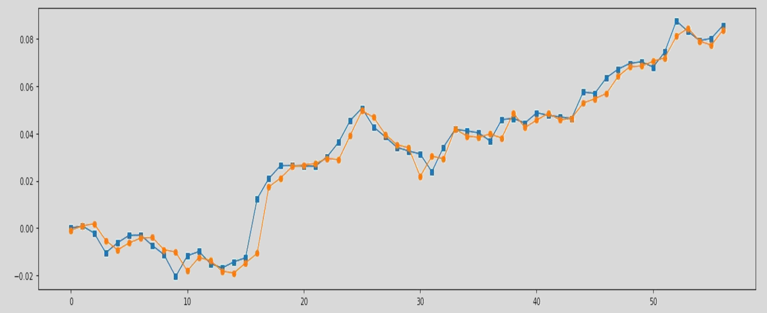
Leong et al. looked at incorporating a support vector machine to predict movements in future stock prices [1]. To solve the SVM optimization problem, they used minimum graph cuts, which allowed their SVM to better predict more complex tree structures. With a 3-fold cross-validation method, they were able to achieve 78% accuracy with their SVM, which indicates their model was successful in learning and predicting future stock market prices without overfitting. Huang et al. used a neural-network method to better predict Chinese stock prices. More specifically, they used a Bayesian-LTSM neural network method [2]. They use six indicators for input into their LTSM. For their data, they looked at 26 years' worth of data and then broke down each year into a week. With this model, they reported that, on average, their Bayesian-LTSM predicted stock market 25% percent better than a regular LTSM approach.

Experiments/results

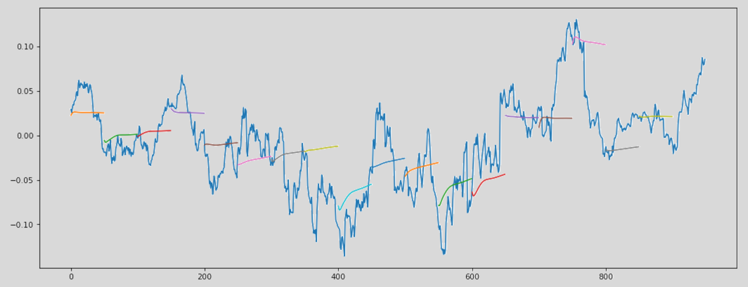
LSTM



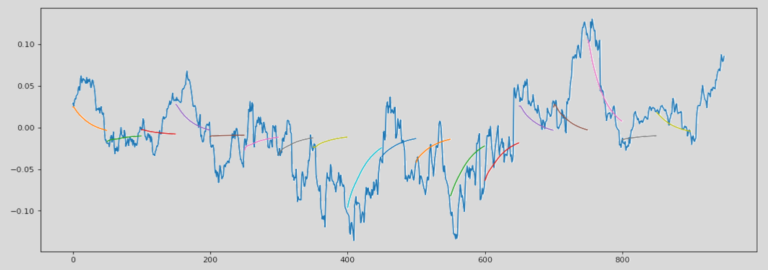
2 Epoch, 50 Batch size, point to point prediction



10 Epoch, 50 Batch size, point to point prediction



2 Epoch, 50 Batch size, sequence prediction



20 Epoch, 50 Batch size, sequence prediction

XGBOOST

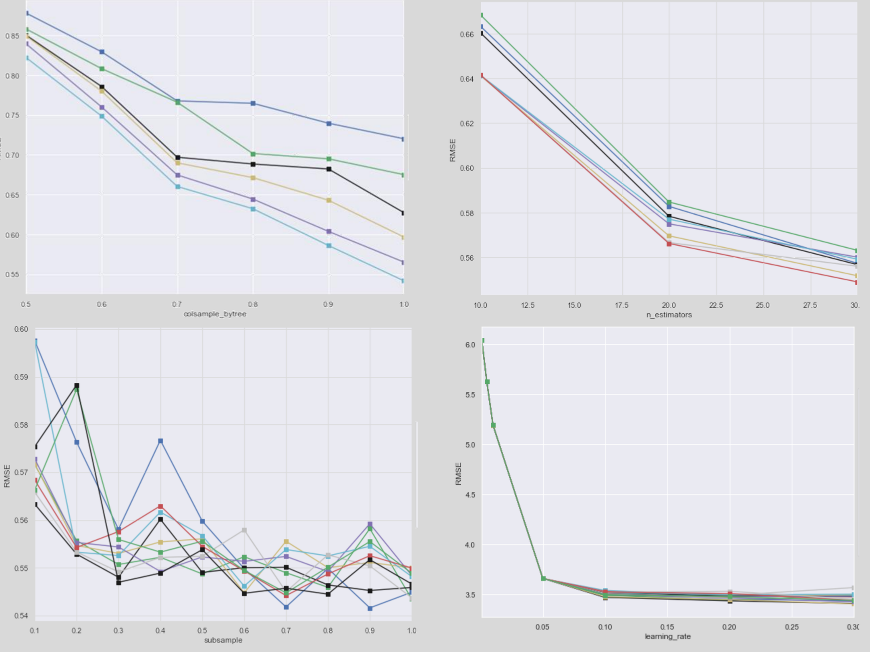


Fig 2: Variation of RMSE on different parameters; a: Colsample; b: n_estimators; c: Subsample; d: Learning Rate

Fig 3: Splitting of data. Blue=train, yellow=cross validate ;red=test

Fig 4: Predicted data plotted against test data

	RMSE	MSE	MAE
APPLE	0.837	0.7	0.47
FACEBOOK	1.397	1.95	0.72
AMAZON	8.398	70.52	4.42
MICROSOFT	0.383	0.146	0.22
GOOGLE	6.662	44.38	4.56

Table 1: Result of model on various data

REGRESSION/SVR

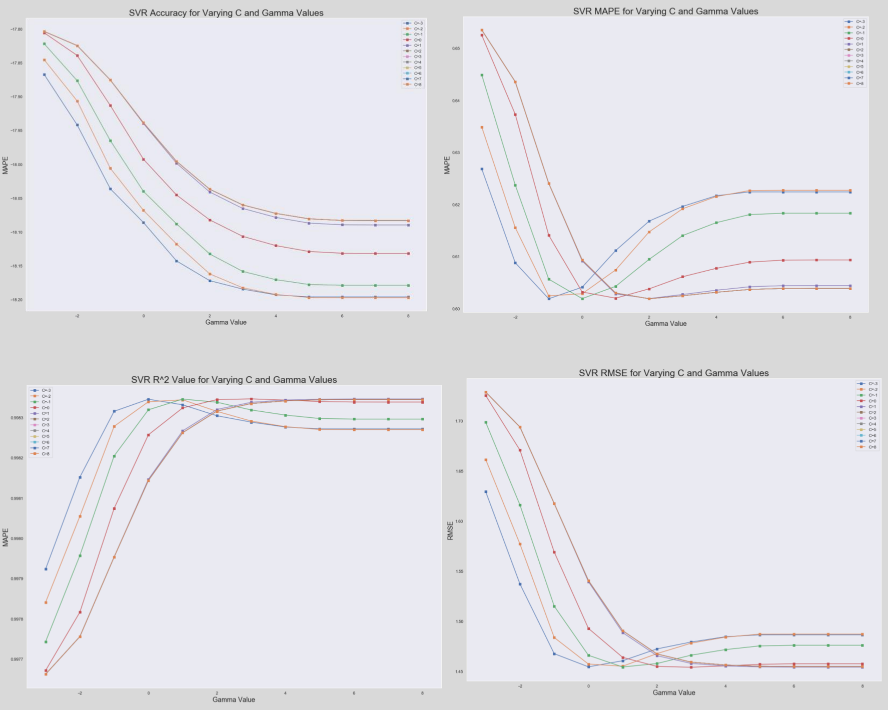


Fig 4: Variation of RMSE on different parameters

Degree	RMSE	MAPE
1	0.77	0.28
2	4.14	1.76
3	3823.97	1250.27
4	19.06	7.04

Table 2: Result of polynomial regression

SVM

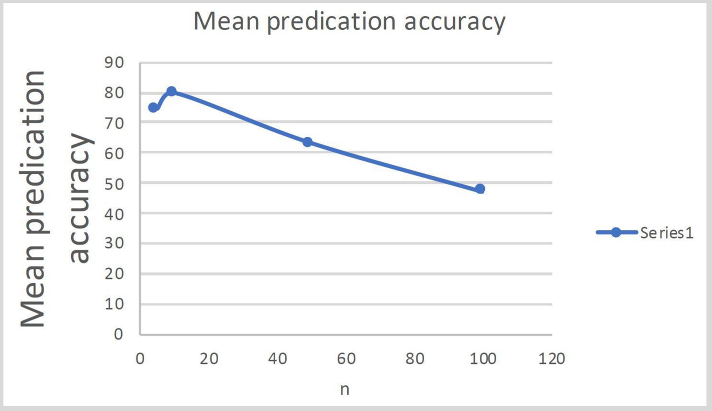


Table 1: Features used in SVM

Feature name	Description	Formula
σ_s	Stock price volatility. This is an average over the past n days of percent change in the given stock's price per day.	$\frac{\sum_{i=t-n+1}^{t} \frac{C_i - C_{i-1}}{C_{i-1}}}{n}$
Stock Momentum	This is an average of the given stock's momentum over the past n days. Each day is labeled 1 if closing price that day is higher than the day before, and -1 if the price is lower than the day before.	$\frac{\sum_{i=t-n+1}^{t} y}{n}$
σ_i	Index volatility. This is an average over the past n days of percent change in the index's price per day.	$\frac{\sum_{i=t-n+1}^{t} \frac{I_i - I_{i-1}}{I_{i-1}}}{n}$
Index Momentum	This is an average of the index's momentum over the past n days. Each day is labeled 1 if closing price that day is higher than the day before, and -1 if the price is lower than the day before.	$\frac{\sum_{i=t-n+1}^{t} d}{n}$

We let C_t be the stock's closing price at time t, where t is the current day, and define I_t as the index's closing price that day. The stock's directional change on a given day is labeled as $y \in \{-1,1\}$, and the index's directional change is defined as $d \in \{-1,1\}$. We use these features to predict the direction of price change between t and $t+m$, where

Discussion

In this project, we have demonstrated a machine learning approach to predict the stock market trend using different neural networks. The result shows how we can use historical data to predict stock movement with reasonable accuracy. Overall all the methods performed nicely for the given set of hourly data. We see some variation in the predicted data when the trend in the stock prices changes drastically and when we don't have enough history data to train our models with. For example, we tested 5 stocks using Xgboost and the RMSE and MSE is high for Amazon and Google. As seen in fig 1, Amazon data varies greatly and for Google we only have data after 2014.

If we use minute data for the stocks, it would be interesting to see the how those models compare to the hourly models.

Acknowledgements / References

[1] Carson Kai-Sang Leung, Richard Kyle MacKinnon, and Yang Wang. 2014. A machine learning approach for stock price prediction. In Proceedings of the 18th International Database Engineering & Applications Symposium (IDEAS '14), Ana Maria Almeida, Jorge Bernardino, and Elsa Ferreira Gomes (Eds.). ACM, New York, NY, USA, 274-277. DOI: https://doi.org/10.1145/2628194.2628211

[2] Biao Huang, Qiao Ding, Guozi Sun, and Huakang Li. 2018. Stock Prediction based on Bayesian-LSTM. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (ICMLC 2018). ACM, New York, NY, USA, 128-133. DOI: https://dl.acm.org/citation.cfm?id=3195170

Image Source: Google