## **Stock Market Analysis Using Machine Learning and Artificial Intelligence** 2023 Undergraduate Research Conference Advisor : Dr. Yonghong Tong Chance Caccamise

#### Introduction

Although not being entirely successful, investing and trading has always been a passion of mine that I will continue throughout my entire life. Therefore, I thought it was about time that I used the skills learned throughout my education to benefit myself in the long run. My analysis consists of two algorithms that use data from the S&P 500. Using 5 years worth of data starting from 2018 to 2022, my analysis uses multiple different indicators to dig deeper into the world of algorithmic trading.

#### Goal

The goal of the analysis is not to predict the exact price of the S&P 500. Although that would be great, it is very unrealistic. My goal for this project is to learn more about what indicators play a significant role in predicting the closing price and to be able to forecast future potential opening prices using Long Short Term Memory Neural Networks. Eventually, the goal is to use live data to day and option trade. Learning more about what goes into algorithmic trading is only the beginning and eventually, I will be able to passively earn a profit.

#### Future Work

This is only the beginning of a massive project. With the ultimate goal of creating an algorithmic trading bot, I have a lot of work ahead. Some of this work includes:

- Being proficient in finance and performing an excessive amount of fundamental analysis.
- Building multiple trading strategies based off of statistical analysis.
- Adding risk management techniques
- Backtesting and improving trading strategies using historical data.
- Deploying the trading bot with a small amount of capital and analyzing the results.
- Improving the model to increase overall profits.

## **XGBoost Regressor**

XGBoost is a powerful machine learning algorithm that trains a series of decision tree models,where each tree is designed to correct the errors of the previous tree. XGBoost Regressor uses this algorithm to predict continuous values. In the first chart, the daily high low, open and volume weighted average price are used to predict the close value for the next day. Out of these 4 variables, the daily open price is most significant when using 70% training data.

To differentiate the variables, I included 7 popular technical analysis indicators to the model. They are: RSI, CCI, MACD, EMA,SMA,VWAP and Bollinger Bands. After using 75% training data to test these indicators, I learned that of the 7 used indicators, RSI,CCI,MACD and Bollinger Bands are more significant.

Lastly, I combined all of the variables into one model using 80 training data to test overall significance. After doing so, the algorithm showed that the 4 technical indicators listed above are still most significant.



Open -High -Low -VWAP -0 200

	C.C
DCI	
RSI -	
CCI -	
_12_26_9 -	
BP_5_2.0 -	
_12_26_9 -	
BB_5_2.0 -	
EMA -	
VWAP -	
_12_26_9 -	
BU_5_2.0 -	
SMA -	85.0
(	0 100

RSI	-	
CCI		
BBP_5_2.0	-	_
)_12_26_9	-	
1_12_26_9	-	-
3BB_5_2.0		-
5_12_26_9		
EMA		-
VWAP	-	-1
3BU_5_2.0		106.
Open	-	97.0
Low	_	95.0
SMA	70	.0
High		)
	0 1	.00





- 1

[









#### **LSTM Neural Network**

Long Short Term Memory Neural Networks are a type of recurrent neural network where the output data is put back into the input. They contain an input gate, forget gate and an output gate to selectively remember or forget information from previous steps. Included in the chart to the right is the forecasted daily open price using 80 training data, 2 layers and 100 epochs. This network uses the daily high, low, open, close and volume to train the model. Each network uses the previous 14 days to predict the next day.

To see if indicators alone could be used to predict price, I ran the same neural network structure as stated above using the top indicators from the XGBoost Regressor calculation. By using the daily RSI, CCI, MACD, Bollinger Bands and Open prices, I was successfully able to forecast the next 120 trading days without the need of input data.

Once again, I combined all of the variables in attempt to improve accuracy of the predictions. After being trained with the past 5 years worth of data ranging from the daily high price and technical indicators, the network was able to produce the chart to the right. The result of this chart was very similar to the technical indicator forecast and should be improved in order to increase accuracy.

#### **Computer & Information Sciences Department**

# NIAGARA JNIVERSITY

