Neural Net Stock Trend Predictor

SJSU SAN JOSÉ STATE UNIVERSITY

Advisor:
Dr. Chris Pollett
Committee Members:
Dr. Robert Chun
Mr. Paul Thienprasit

By
Sonal Kabra

Agenda

- Purpose
- Introduction
- Review of Existing Work
- Prior Experiments
- Our Approach
- Neural Networks
- Models Developed
- Results
- Conclusion

Purpose

- This project consisted of experiments and implementations of several neural nets to predict Stock Market movement and indicates whether the stock under study should be -- Bought, Neutral, or Sold to generate profit.

- All of our neural nets were designed to predict stock prices for the following week.

Introduction

- Since the beginning of the stock market in 1817 in the United States, accurate stock prediction has been a goal of investors.
- One difficulty in accurately predicting stocks is the high number of variables on which they depend.

Introduction

- Our neural nets used financial data from Quandl.
- Below is some example data showing attributes this data has.

Date	Open	High	Low	Close	Volume (BTC)
1/19/2016	387.04	387.04	387.01	387.01	5.24961215
1/18/2016	382.11	390	375.01	387.1	19467.04404
1/17/2016	388.43	393.31	378.76	382.02	20715.0043
1/16/2016	359.16	392.5	352.5	388.4	74056.46019
1/15/2016	429.27	429.43	357.02	359.16	123870.5349
1/14/2016	431.09	435	426	429.25	13304.53779
1/13/2016	431.99	436.36	416	431.11	33149.40771
1/12/2016	449.26	449.98	427.01	432.04	20739.28438
1/11/2016	449.34	452.49	443.42	449.26	14809.00262
1/10/2016	449.23	450.26	441.01	449.35	13778.25597
1/9/2016	454	456.01	447.07	449.23	8950.479347
1/8/2016	458.41	465	446.55	454	29967.85309

Introduction

- Most investors follow two analytical methods:

- **Fundamental Analysis**
  - Studies company fundamental factors
  - Helps the investors to find the stocks worth investing

- **Technical Analysis**
  - Identifies the future uptrend or downtrend patterns.

Review of existing work

- **Stock Prediction is not a new concept.**
- **Kara et al. [1]**
  - Two models: a neural network and an SVM, each used to predict the direction of stock price index movement.
  - Both use ISE National 100 Index for the dataset
  - Both use a total of 10 technical analysis indicators
  - The neural network had an accuracy of 75.74% and the SVM had an accuracy of 71.52%.

Our Approach

- Instead of predicting Up/Down signals, it will predict stock trade signals namely “Buy, Sell or Neutral” for next week.

- Instead of combining all the technical indicator, neural net will train separately for each indicator.

Prior Experiments

- Whenever human invests in stocks, they try to study the past data to find the similar pattern.
- The earlier experiments used K-nearest neighbor and decision tree machine learning regression techniques.
- By using those techniques, earlier experiments predict the closing price of the same day.

K-nearest neighbor: Prior Experiments

- The algorithm states that the prediction values are similar for the objects that are in close proximity of each other.

- Thus, we can assume that the prediction values will be almost equal for such objects.

K-nearest neighbor: Prior Experiments

- KNN has a high error range for both stocks
- The predictions are completely off from the right prices.

```
Training for FB......
AVG score is  0.0584505431993
This is the result for FB......
Predicted adjusted close value for today is 126.51

Training for CSCO......
AVG score is  -0.284828431011
This is the result for CSCO......
Predicted adjusted close value for today is 30.32
```

Actual Prices: FB: 149.78 CSCO: 34.3

Decision Tree: Prior Experiments

- Helps to make predictions by mapping given observations to conclusions.
- Divide the information into small gatherings based on maximizing information gain.

```
Training for FB......
AVG score is  0.989572154988
This is the result for FB......
Predicted adjusted close value for today is 150.45

Training for CSCO......
AVG score is  0.98784596313
This is the result for CSCO......
Predicted adjusted close value for today is 33.88
```

Actual Prices: FB: 149.78 CSCO: 34.3

Our Approach

Technical Analysis for stock prediction

- Effective for short-term trading.
- Observes money flow, momentum, and volatility.
- Supplements in confirmation of trend or pattern
- 2 Types:
  - Leading Indicators
  - Lagging Indicators

Moving Average Crossover

- Is a lagging indicator.
- Formula:

$$SMA = \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n}$$

- 5-Day and 10-Day **Moving Average Crossover** Strategy

Relative strength Index

- It is a leading indicator
- It tells whether the given stock is overbought or oversold.
- Formula:

$$RSI = 100 - \left(\frac{100}{(1+RS)}\right)$$

- The project is using the 14-Day period for the RSI.
- The RSI value above 70 indicates an oversold region, while below 30 indicates the overbought region.

On Balance Volume

- It is used to find buying and selling trend of the stock.
- It calculates the positive and negative flow of the volume on its price.
- If the current closing price is more than the previous close price:

  Current OBV = Previous OBV + Current Volume

- If the current closing price falls below the previous close price:

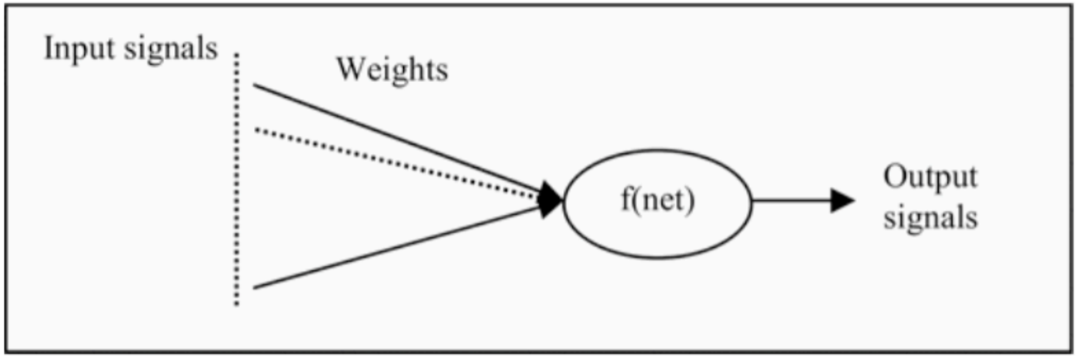
  Current OBV = Previous OBV - Current Volume

- Else it will just assign the previous OBV to current OBV.

Artificial Neural Network

- Are computational models that replicate the behavior and adapt the features of biological neural systems.

- Has thousands of artificial neurons just like the human brain has neuron nodes.



Artificial Neural Network

- Input layer comprises of nodes or all input features in the Training set.
- Hidden layer comprises of the node responsible for the processing and learning of data from the Input Layer.
- Output layer comprises of a class node.

Training Neural Network

- Backpropagation algorithm.
- The problem is set up as minimization of a loss function.
- RPROP Algorithm

Libraries Used

- Keras (Neural Network Library)
- Sklearn (Machine learning and data analysis library)
- Numpy (for mathematical calculations)
- Matplotlib (Plotting the results)
- Quandl API (Stock data)
- Pandas (storing stock data structure)

Dataset

- S&P 500 market
- Blankets a diverse set of multinational corporations
- Collect the dataset from Quandl.
- Python Quandl API

Data Preprocessing

- All the features in the data set are not in similar range.
- The values in datasets are normalized in the range of [-1,1].
- **The formula is:**

X_std = (X - X.min) / (X.max) - X.min)

Data Preprocessing

- The data is partitioned into the training (70% of the dataset), the validation (20%) data set, and test (10%)
- The 100 contiguous data points are randomly held from the generated dataset.
- The neural net is trained on around 800 stock data points and later tested on 100.

Moving Average Crossover Model

- 4-layer neural network.
- 30 input nodes: Three nodes for each day till ten days.
- Input features: 5-Day SMA, 10-Day SMA and Closing price of that day.
- 2 hidden layers: 60 and 60.
- The activation used is tanh.

RSI Model

- 4-layer neural network.
- 20 input nodes: Two nodes for each day till ten days.
- Input features: 14-Day RSI, and Closing price of that day.
- 2 hidden layers: 40 and 40.
- The activation used is tanh.

OBV Model

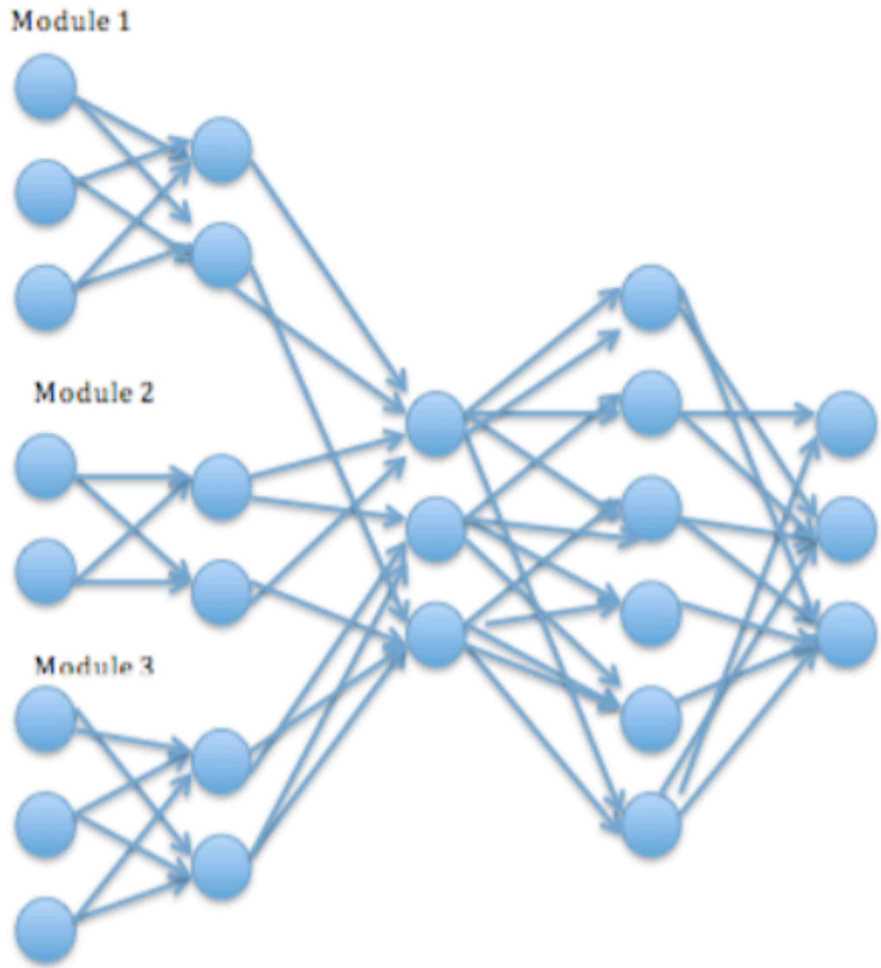
- 4-layer neural network.
- 30 input nodes: Three nodes for each day till ten days.
- Input features: On balance volume of the day, Volume of the day, and Closing price of that day.
- 2 hidden layers: 60 and 60.
- The activation used is tanh.

SJSU SAN JOSÉ STATE UNIVERSITY

Merged NN Randomized Model

- All the models are merged into final layer of the neural network as shown in the following figure.

- The whole architecture is trained together, instead of training each model differently.



Merged NN in Sequence Model

- The architecture is same as previous model.
- The test data set generated in this experiment is not random.
- The training is strictly forced to use the early days of stock data, and testing is done in recent days of stock data.

Results

- As this is a multi-classification problem ("Buy," "Sell," or "Neutral"), the accuracy metric used is a Confusion Matrix.

- Accuracy defined is the number of correctly classified points in comparison to the total number classifications made.

$$\text{Accuracy} = \frac{\text{True\_Buy} + \text{True\_Sell} + \text{True\_Neutral}}{\text{Total number of Observations}}$$

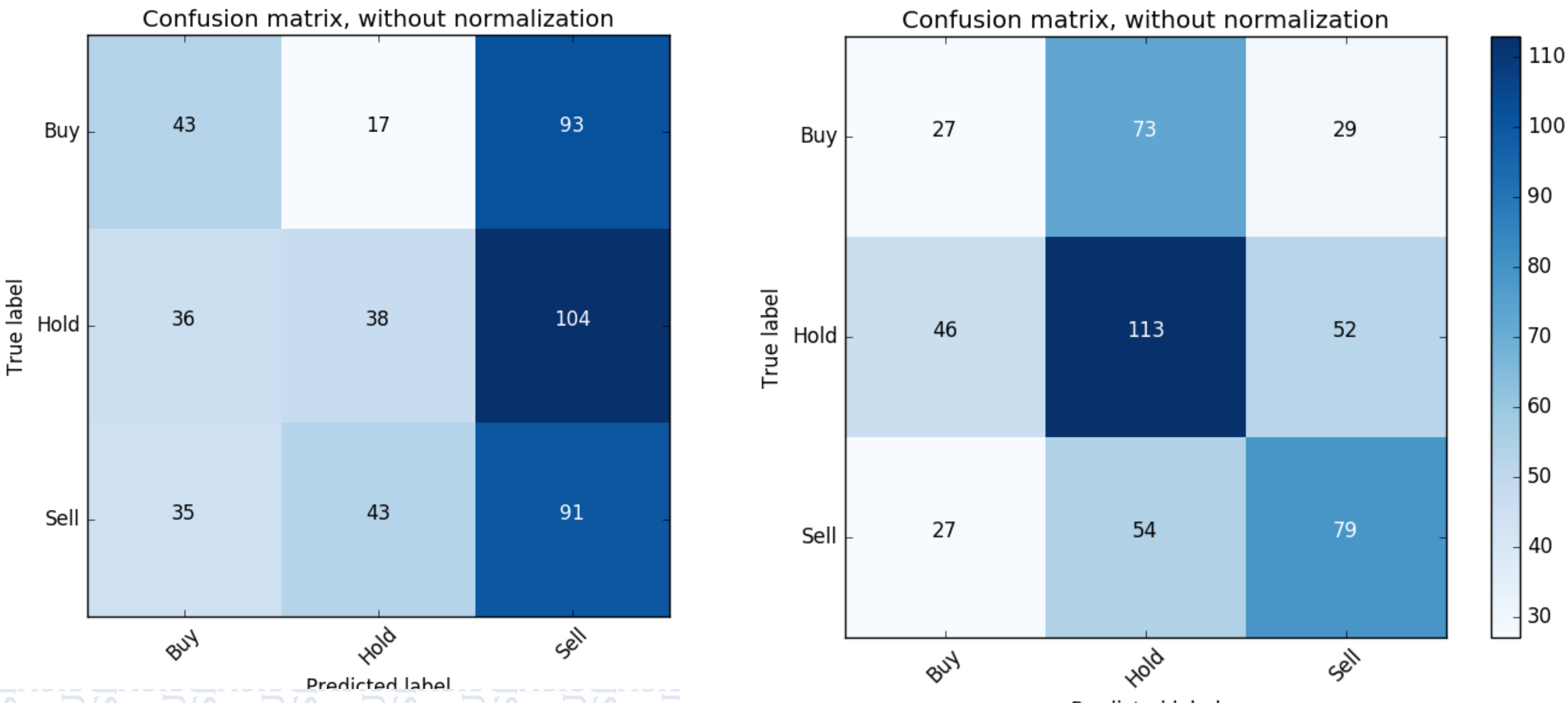
Results

- To calculate the profitability for each model following formula is used to calculate the normalized weekly return of a stock.:

$$\text{Return Threshold } (=1\%) * \frac{(\text{Total\_Positive} + \text{Total\_Negative}) - (\text{False\_Positive} + \text{False\_Negative})}{\text{total observations}}$$

- The average risk-free rate of weekly return is 0.035%

Moving Average Crossover



HD[Acc.37.60%(+/-3.47%)]
1.04

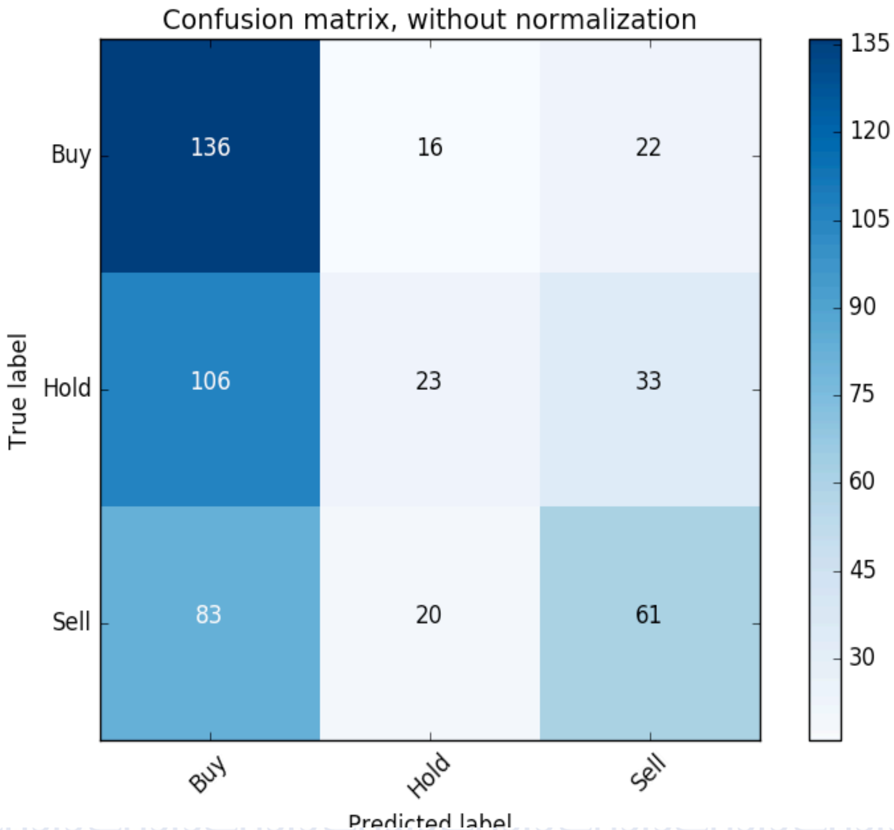
C [Acc. 47.00%(+/-4.16%)]
2.72

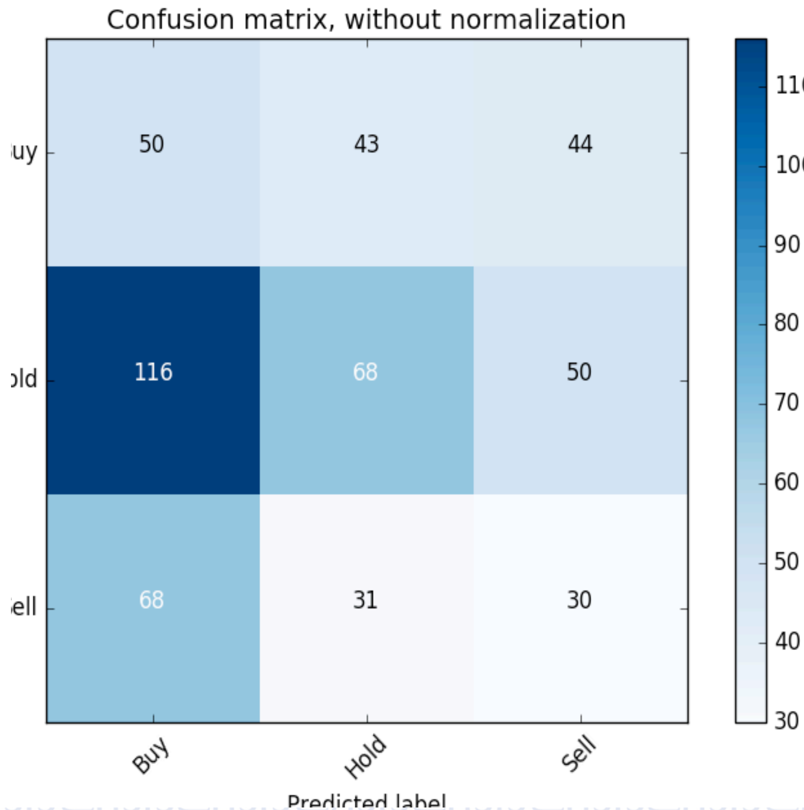
Moving Average Crossover

Stocks	Percentage weekly return
AAPL	0.066%
MCD	0.088%
XOM	0.076%
CAT	0.104%
C	0.1%
HD	0.012%

SJSU SAN JOSÉ STATE UNIVERSITY

RSI Model





CAT [Acc.37.40%(+/-6.41%)]
1.87

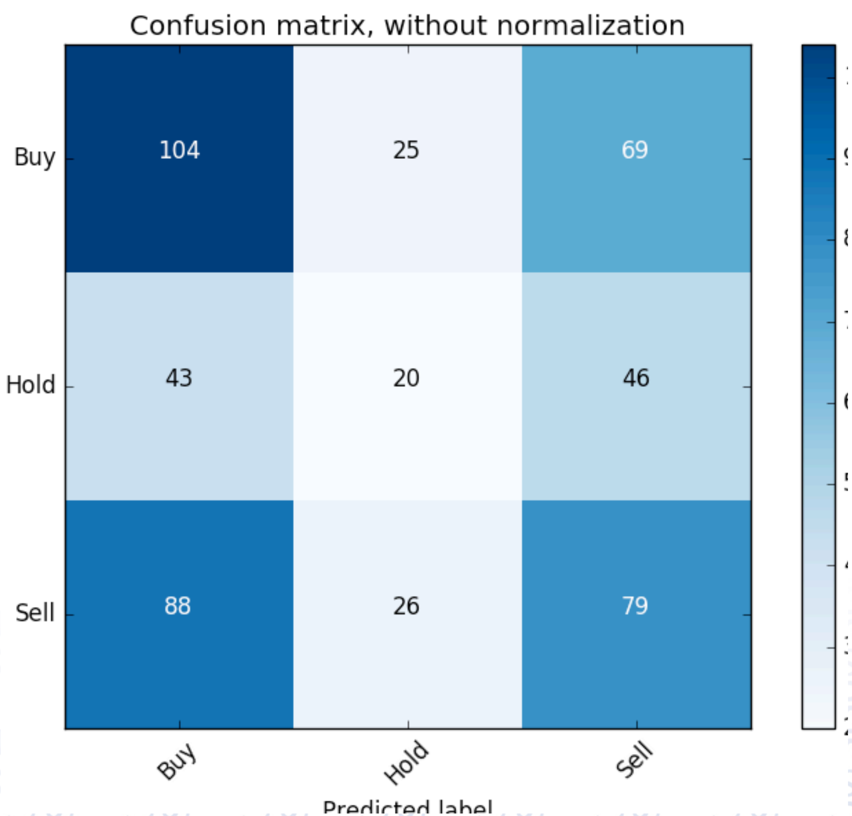
XOM [Acc.29.60%(+/-8.16%)]
0.71

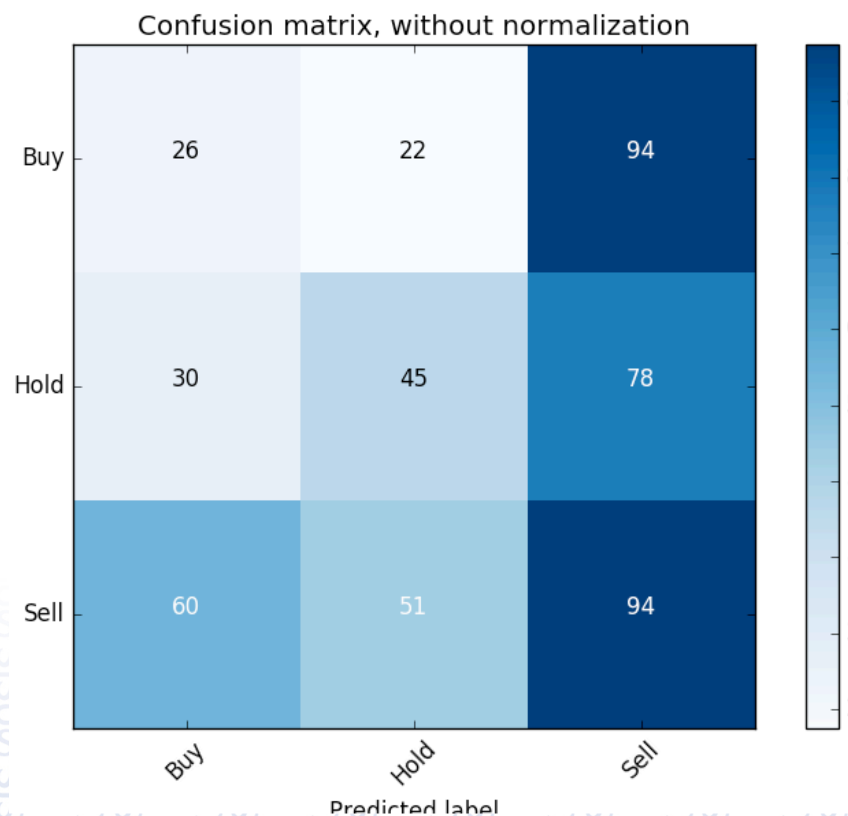
RSI Model

Stocks	Percentage weekly return
AAPL	0.126%
MCD	0.068%
XOM	-0.064%
CAT	0.184%
C	0.078%
HD	0.008%

SJSU SAN JOSÉ STATE UNIVERSITY

OBV Model





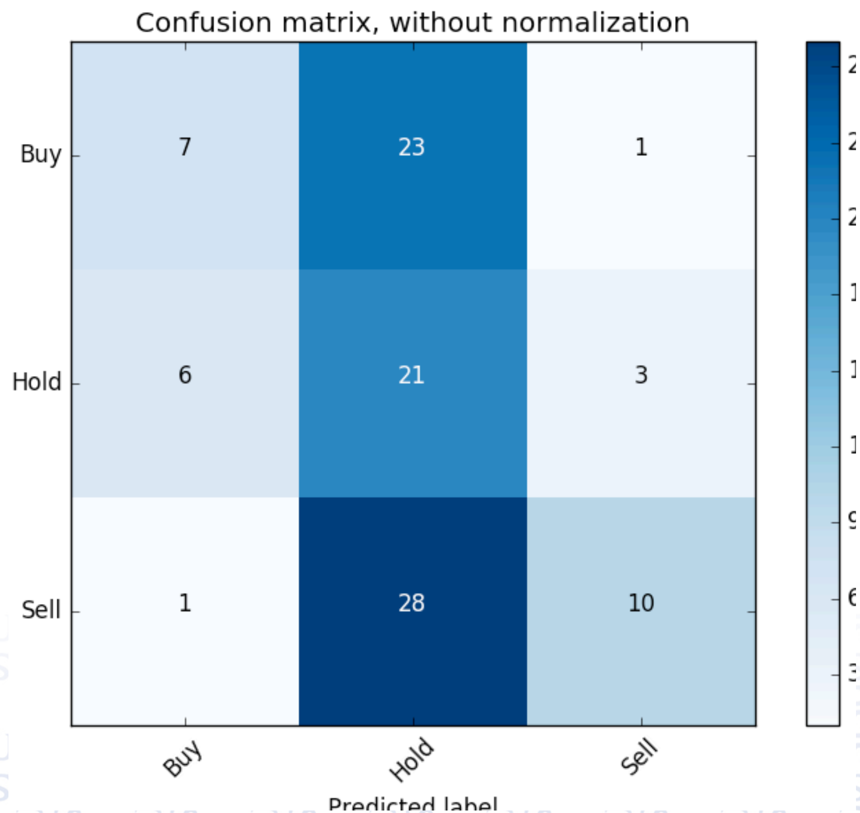
C [Acc. 46.80%(+/-4.58%)]
1.16

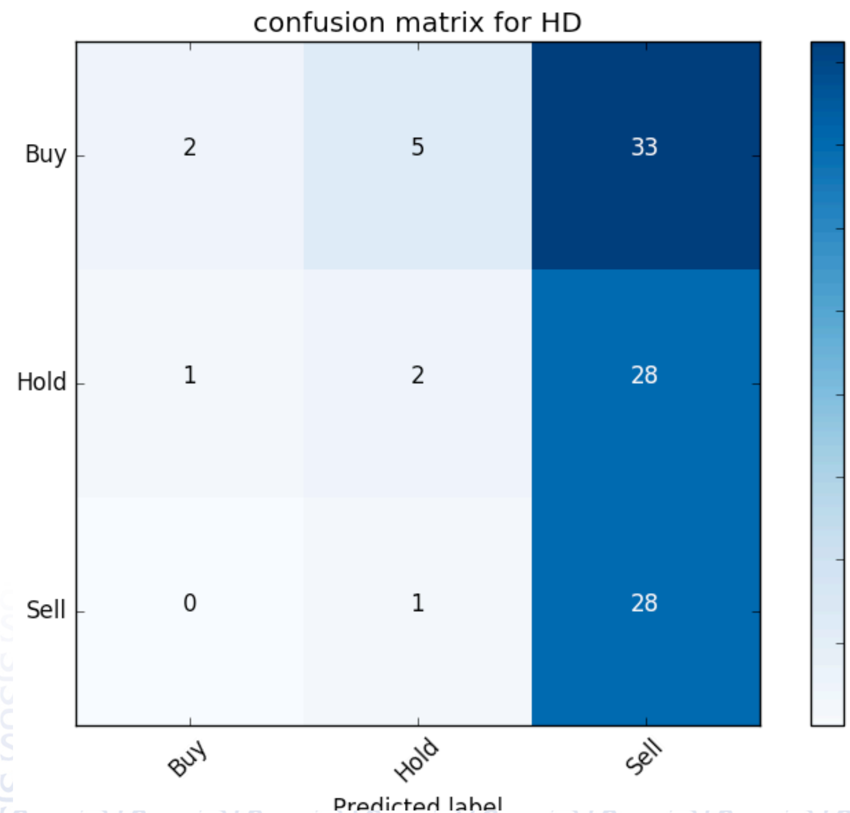
HD [Acc. 34.20%(+/-1.16%)]
0.77

OBV Model

Stocks	Percentage weekly return
AAPL	0.052%
MCD	-0.002%
XOM	-0.006%
CAT	0.028%
C	0.052%
HD	-0.068%

Merged NN Randomized Model





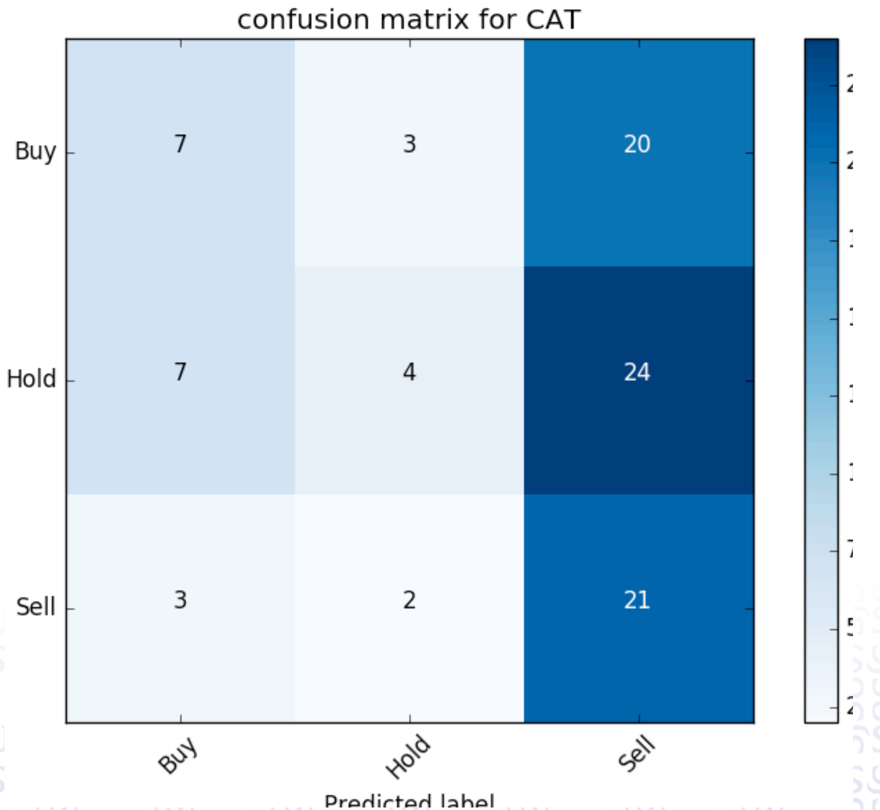
CAT [Acc.44.36%(+/-4.76%)]
8.5

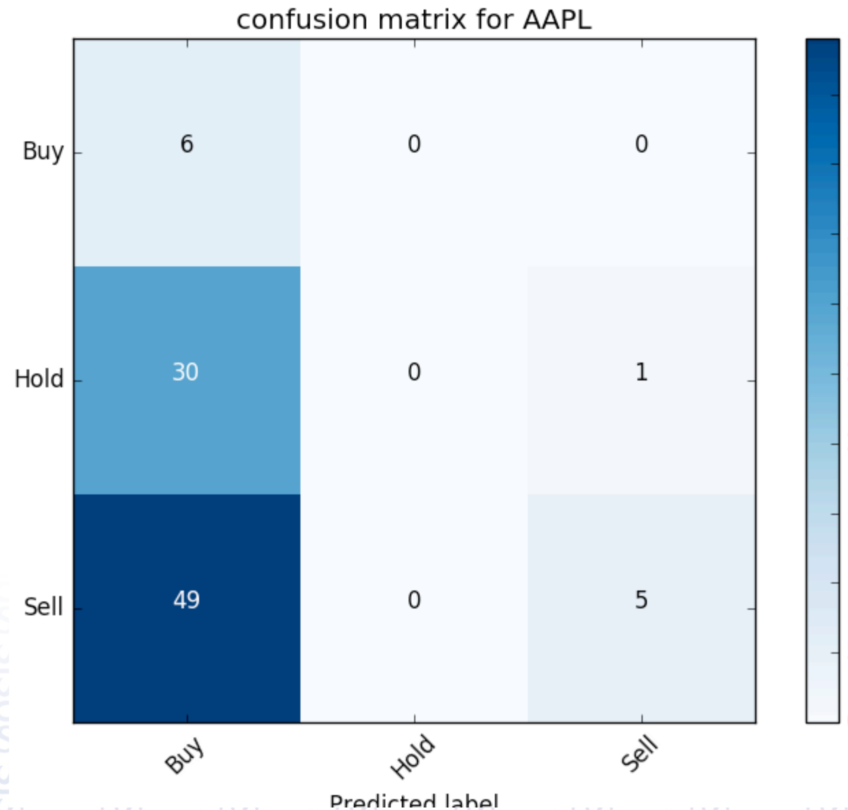
HD [Acc.34.26%(+/-1.76%)]
0.90

Merged NN Randomized Model

Stocks	Percentage weekly return
AAPL	0.19%
MCD	0.06%
XOM	-0.001%
CAT	0.15%
C	0.06%
HD	-0.03%

Merged NN In Sequence Model





CAT [Acc.37.53%(+/-3.72%)]
1.21

AAPL[Acc.26.22%(+/-5.88%)]
0.22

Merged NN In Sequence Model

Stocks	Percentage weekly return
AAPL	-0.38%
MCD	-0.36%
XOM	0.001%
CAT	0.184%
C	-0.01%
HD	-0.18%

Results

Why some models may performed poorly:

- Model Complexity
- Training Data
- Market Noise

Conclusion

- From the Confusion matrices for above simulations, Merged Model Randomized still gives better results than the Merged Model in Sequence.
- If we consider only moving average crossover model, then that model gives more returns than rest of them.
- Therefore, for future development one can surely use Moving average crossover model.

References

- Yakup Kara, Melek Acar Boyacioglu, and Ömer Kaan Baykan. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. *Expert systems with Applications*, 38(5): 5311–5319, 2011.
- Shunrong Shen, Haomiao Jiang, and Tongda Zhang. Stock market fore- casting using machine learning algorithms, 2012.
- E. F. Fama, K. R. French, "Common risk factors in the returns on stocks and bonds", *Journal of financial economics*, vol. 33, no. 1, pp. 356, 1993.

References

- D. G. McMillan, "Stock return dividend growth and consumption growth predictability across markets and time: Implications for stock price movement", *International Review of Financial Analysis*, vol. 35, pp. 90101, 2014.
- M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," *2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, Rajshahi, 2016, pp. 1-4.
- M. D. Godfrey, C. W. Granger, and O. Morgenstern, "The random-walk hypothesis of stock market behaviora," Kyklos, vol. 17, no. 1, pp. 1-30, 1964.
- J. Murphy, "Technical analysis of the financial markets, prentice hall, london," 1998

THANK YOU..!!!