An Intelligent Framework using Hybrid Social Media and Market Data, for Stock Prediction Analysis

Eslam Nader Desokey
College of Computing and Information Technology Arab Academy for Science and Technology ,Egypt

Email: EslamNader [AT] outlook.com

Amr Badr
Professor at Dept. of Computer Science - Faculty of Computers and Information - Cairo University ,Egypt

Abdel Fatah Hegazy
Dean College of Computing and Information Technology Arab academy for science and technology, Egypt

Abstract—**Stock market prediction is one of the most attractive places for investment. It is one of the most prevalent topics for research studies and commercial applications because of its dynamic and unpredictable nature.**

One of the most important sources of data can be found in social media such as Twitter or Facebook. This is because the data reveals to a great extent the social mood of the public and their future behavior, if the data is well analyzed using intelligent financial web-mining and stock prediction systems.

This framework builds on previous researchers' efforts to analyze and determine social media impact on stock market behavior, and hence predict prices movement.

We found that results reached by researchers in the last decade have notbeen satisfying to investors. In response, we are attempting to establish a greater correlation between market data and social media feeds in order to increase the accuracy of stock market prediction.

Keywords- Stock market prediction; social media; twitter mood; behavioral finance; ontology

I. INTRODUCTION

Financial market plays an important role in economics, especially the stock market. How to predict the stock price is a widely researched problem. Last decade, most methods were based on a stock or index's history data, using unclear hybrid models or neural networks to predict its future. Now, with social media booming, many researcher have started using the social media behavior of investors and the public to predict the stock market. Those researchers used data from news sources, Twitter, Facebook, LinkedIn, and even Google searches. The majority reached good results despite working with growing markets (e.g. Egypt and all Middle East markets) instead of well-established markets such as the United States. [1].

Stock market prediction has attracted much interest from the academic sector in addition to business. However, the question still persists: Can the stock market actually be predicted? Previous research on stock market prediction was based on random walk theory and the Efficient Market Hypothesis (EMH). According to the EMH, stock market data are largely driven by new information, i.e. news, rather than current and historical market data. However, news is unpredictable. Stock market prices will follow a random walk model and cannot be predicted with more than 50 percent accuracy [2].

In stock market analysis we use two methods. The first is fundamental analysis and the second is technical analysis.

a. Fundamental Analysis

Fundamental analysis s an assessment method for stocks that depends on the statistical data of a company. It includes audit reports, financial status of the stock, the quarterly balance sheets, the dividends, and policies of the stock where stocks are to be observed. Fundamental analysts try to study everything that can affect the stock's value, including macroeconomic factors (like the overall economy and industry condition) and stock specific factors (such as financial and management conditions). Fundamental analysis also includes the analysis of sales data, the strength and investment of the company, plant capacity, market competition, import and export volume, production indexes, price statistics, and the daily news or rumors about the company [3] [4].

b. Technical Analysis

Technical analysis is an evaluation method for securities, it is a statistical analysis that is derived from market activities, such as past prices and volumes. Technical analysts do not attempt to measure a security's intrinsic value, but instead use charts and other tools to identify patterns that can suggest future activity [4].

The rest of this paper is organized as follows. Section 2 will discuss the background of the stock exchange and social media, and its impact on our world. Section 3 will explain most important data sources and parameters used in most researches, then set ranking for those data sources and parameters. Section 4 will discuss in depth the related research. Finally, Section 5 proposes our framework in detail. The paper is concluded in Section 6.

II. BACKGROUND

Before beginning to discuss our framework, some concepts need to be defined in order to unify the background of ourfield “stock market prediction and social media”.

a. Stock Exchange

A stock exchange is an organized marketplace for securities featured by the centralization of supply and demand for the transaction of orders by member brokers for institutional and individual investors. In other words, it is an exchange which provides services for stock brokers and traders to buy or sell stocks, bonds, and other securities [4] [5].

b. Market Data

Market data are prices and trade-related data for a financial instrument reported by a trading venue such as a stock exchange. Market data allows traders and investors to know the latest prices and see historical trends for instruments such as equities, fixed-income products, derivatives, and currencies.[6]

The market data for a particular instrument includes the following:

- The identifier of the instrument (ticker symbol)
- The latest prices for bids and asks
- The time of the last trade
- Total volume traded
- Total volume of bids and asks
- Static data about the financial instrument that may have come from a variety of sources

The previously mentioned form of market data is used for intraday trading. For historical trading, we have other form of data as follows: [6]

- Open - This is the price of the first trade for the day.
- High - This is the highest price that the security traded during the day.
- Low - This is the lowest price that the security traded during the day.
- Close - This is the last price that the security traded during the day.
- Volume - This is the number of shares that were traded during the day
- Date - this is the trading date of the transaction.

c. Social Network

A social network is a social structure comprising of persons or organizations, which usually are represented as nodes, together with social relations, which correspond to the links among nodes. The social relation could be either explicit, such as kinship and classmates, or implicit, like, friendship and common interests [7].

d. Big Data

Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured. Moreover, big data may be as important to business – and society – as the Internet has become. Why? More data may lead to more accurate analysis.[15]

Presently, all business executives are speaking about platform three, which is built on a foundation of cloud and mobile computing, social media, and big data technologies. These new sources of data have a competitive advantage and all providers are seeking to use those technologies. As a result, the number of users and software is booming as IDC said. [14]

e. Social Media

Social media is the popular shape for big data implementations.

Social media includes computer-mediated tools that allow people to create, share, or exchange information, ideas, pictures, and videos in virtual communities and networks. Social media is defined as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, which allows the creation and exchange of user-generated content. Furthermore, social media depends on mobile and web-based technologies to create highly interactive platforms through which individuals and communities share, co-create, discuss, and modify user-generated content. They introduce substantial and pervasive changes to communication between businesses, organizations, communities, and individuals. These changes are the focus of the emerging field of techno self-studies.”[7]

f. Impact of Big Data in Social Media

These days, everything from making a purchase order, viewing certain sales items, to making online trading can be tracked online. Social media is a major factor in this subject. Popular sites like Facebook, Twitter, Instagram, Foursquare and Pinterest record millions of comments, likes, and chat streams on a daily basis.

IBM estimates that 2.5 quintillion bytes of new data are created every day. To put this into perspective, social media alone generates more information in a short period of time than existed in the entire world just several generations ago. Popular social media sites create massive quantities of data that, if translated properly by large-scale applications, would be any brand’s golden ticket into the minds of its consumers [8].

III. DATASOURCES AND COLLECTION

In this section, we will collect the most popular data sources for stock market prediction using social media and try to set our ranking based on the research results.

First, we should list the primary data source for stock market predication, which is the “market data” that we discussed in section 2. There are also other data sources which will be listed as follows:

a. News Media Data

Most of the related research uses news media as a data source for updated news about stock market and stores news headlines in a special database.

We have two types of news sources: The first is international financial news providers like Bloomberg, Reuters, CNN-Money, Yahoo finance and CNBC. The second type is local financial news providers. In Egypt, the most popular local providers are "Arab Finance and Mubasher Info. In China, the most popular local provider is guba.com.cn.

After selecting news media sources and gathering this data in a private database, implementation of sentiment analysis utilizes widespread data sets used for this purpose like Harvard-IV-4 Psychological Dictionary, financial positive and negative terms list (Bill McDonald) and more [9] [10].

Related research shows that the existence of negative mood seems to be more predictive of financial market values than positive mood.

b. Search Engine Data

Search engines like Google and Bing are used for gathering stock market data worldwide. This data is collected based on their search engine ranking and filtration. Some of researchers collect this data on either daily or weekly basis [11].

c. Social Media Data

The massive amount of social media data that has becomeavailable in recent years has provided significant research opportunities for social scientists, computer scientists, and data scientists. The majority of scientists use data coming from social media in order to make stock market predication more accurate. Most researchers use twitter for this prediction. [9]

However, reality is more complicated. There are many theoretical and methodological issues in predicting future outcomes using social media data that are far from being settled until now. Deeper studies and experiments are required to discover the true potential of social media as a reliable source of data. While prediction represents a problem in a wide variety of scholarly fields, social media-based forecasts today receive significant attention. We thus consider it appropriate to discuss and reflect on the promises of social media-based forecasts, as well as the perils and pitfalls it is plagued with, and strategies to address these problems [7].

Data Source	Using	Data Source	Using
Market Data	*****	Social Media Data	*****
News Media Data	**	Search Engine Data	*

IV. REVIEW OF RELATED WORKS

This core section will examine the most important research in this field, then focus on our framework.

a. Twitter mood predicts the stock market by Johan Bollen, Huina Mao, and Xiao-Jun Zeng

In this paper, the authors look into large-scale feed from Twitter (tweets) to identify which is related to the Dow Jones Industrial average DJIA closing value. They collected tweets from February 2008 to November 2008,then created subjectivity analysis using "OpinionFinder" and "Google-Profile of Mood States (GPOMS)". They used seven measures. First, they used OpinionFinder to find the positive or negative value. Then, they used GPOMS to identify the other six measures, the Calm, Alert, Sure, Vital, Kind, and Happy values.. Their goal was to study the impact of the public mood on the stock market. They require reliable, scalable and early assessments of the public mood at a time-scale and resolution appropriate for practical stock market prediction [2].

Finally they built the self-organizing Fuzzy Neural Network model that predicts DJIA values on the basis of two sets of inputs: (1) the past 3 days of DJIA values, and (2) the same combination of various permutations of their mood time series [2].

b. Predicting the French stock market using social media analysis by Vincent Martin

In this paper, the researchers apply the Bollen et al. work to the French stock market, using French tweets. They implement two analysis models in Twitter — sentiment analysis and subjectivity analysis — then use the results of these analysis to train their neural network.

Their neural network input are sentiment, subjectivity analysis and CAC40 closing values for two days. The single output value is predicting the closing value at day+1 for CAC40.

Finally, this work faced a significant obstacle, in that all available data set for semantic analysis is written in English, but they wanted to apply it to French tweets. Therefore, results have been far form expected values. So, they continue to work on improving their analysis as they stated in their conclusion [12].

c. Trading on Twitter: The Financial Information Content of Emotion in Social Media by Hong Keel Sul,Alan R. Dennis, Lingyao (Ivy) Yuan

In this work, the authors study the impact of emotions contained in tweets on stock price changes in the short term and long term (10 days). They built their work using two hypotheses.

To evaluate their hypotheses they collect data related to S&P 500 firms from Center for Research in Security Prices (CRSP) between March 2011 and February 2012. Also, collecting public tweets that contained financial information using the "$" symbol meant that the tweet contains investment information about a firm. After collecting data from tweets, they clean this database on one firm symbol per tweet. Now after cleaning they have 2,503,385 tweets. The researchers analyze these tweets using the Harvard-IV dictionary to get positive and negative emotions and add two Cumulated

Abnormal Return (CAR) variables as control variables for more reliability. They determine that each tweet related to t or t+1, based on tweet time if it's after 4 pm, will be t+1 [13].

We notice from comparing this paper with previous research that stock market predictions with social media are producing more accurate result. Also, we can add a new parameter for sentiment analysis, "Number of Followers".

d. Stock Market Prediction Based on Public Attentions: a Social Web Mining Approach by Ailun Yi

In their research, the authors analyze stock market prediction through a new method with social media to discover the correlations between social media behavior and stock price changes.

Their method depends on retrieving such information in a timely manner and analyzing the level of attention in on a large scale, which reveals interesting relationships to the stock prices.

They analyze this correlations using simple frequency counting, loose n-gram models and noun phrase expansion. They implement those models using a selected data set that contains three type of sources, Newswire, Blog and Twitter.

In light of the above considerations, the researchers set experiments mainly based on Twitter dataset, which consists of 61,756,056 posts crawled from February 2009 to June 2009. In testing, they use data from April to June for their prediction task because of the large number of tweets per day in this period.

After data collection, they filter collected data using noun phrasing, loose n-gram model and document frequency thresholding. Next, they apply the Hybrid Wrapper method on selected data for prediction, then start using ε-insensitive support vector regression algorithm for the learning phase of their model. Finally, they compare their results with moving average from technical analysis.

We notice from studying this work that the researchers use data from 2009, when social media was less prevalentand not yet developed. Now, social media data has more information than before. Therefore, in their future work I expect that they will get a greater correlation in their results [14].

V. PROPOSED PREDICTION MODEL

The proposed model comprises several stages as shown in Figure 1. The first stage is concerned with the knowledge of representation on our ontology. Next, for the social media data collection, we collect more than 1.5 million rows. Then go on to cleaning stage that using several algorithms, in order to make our data set more accurate and minimize rumors. The fourth stage merges our market data feed and social media data, after clearing, to reach one consolidated database. We start our final stage prediction engine in our rules-based system.

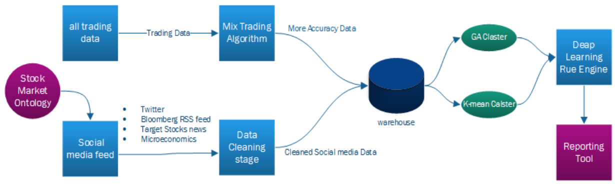

Figure 1: Proposed model

In the rest of this section we introduce each stage in more detail.

a. Knowledge Representation "Ontology"

In this stage we start by build our knowledge representation layer using ontology.

After searching public stock market ontology, we did not find any one build ontology that we were able to build our layer on. So, we build the first Egyptian stock market ontology, using our methodology with the following steps.

- Experts and our technical analysis identification
- Application knowledge capture techniques
- Identify Concepts
- Extract concepts constraints and the relationships between concepts
- Implementing our ontology using "Protégé"[1]

After implementing the ontology, we are able to extract the necessary knowledge to start our prediction engine.

b. Gathering Social Media Feed

In this stage we build our aggregators using the ontology system output 'XML files'. This contains all knowledge about a specific symbol in market. We build two aggregators, one for Twitter and the second built in Google search.

i. Twitter Feeder:

This program uses Twitter API to retrieve all tweets containing any keyword on the MXL File.

We use TweetSharp for .net developer to build our engine with the following structure.

---

[1] http://protege.stanford.edu/

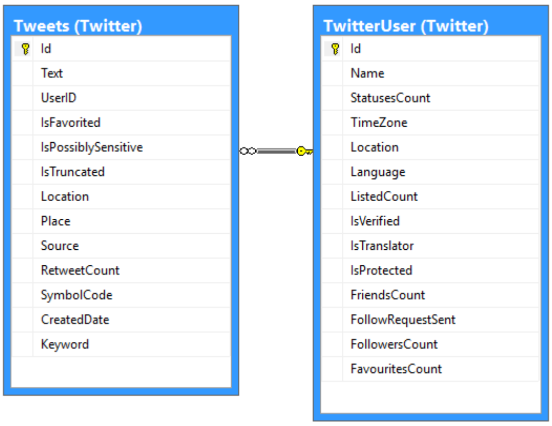

Figure 2 Twitter ERD

We collect 129,012 tweets from August 25, 20155 to October 23, 2015.

These tweets come from 67,238 unique users around the world.

ii. Google Feeder:

We faced a challenges in getting all updates and news containing any of our keyword, generated from pre-existing ontology on the following set sites.

- www.bloomberg.com
- www.reuters.com
- www.arabfinance.com
- www.mubasher.info

Other researchers have previously used RSS generated form those websites. But we find that RSS contains limited knowledge. Since we need all updates, we determine that Google is the best search engine crawler to use.

We use the Custom Search Engine feature from Google to build our Google Feeder that uses selected website as the base for the search engine. We search for the generated keyword from our ontology every six hours.

We use Google API Custom search v1 to build our engine with the following structure.

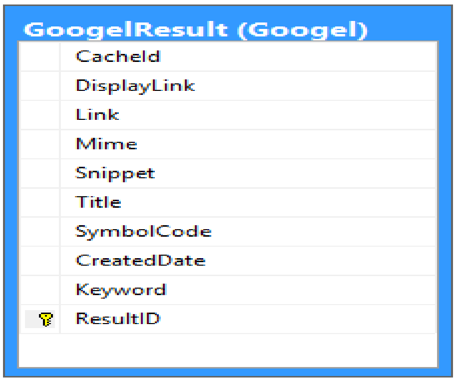

Figure 3 Google Structure

We collect 890,342 results from September 20, 2015 to October 23, 2015.

iii. Stock Market Trading Data

We get historical trading data for all symbols and indices form May 15, 2001 to October 8, 2015-10-08 and 523,401 register at EGX as the following structure

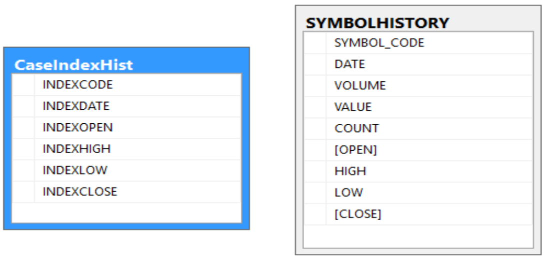

Figure 4 Market Data Structure

Also we use Market news, news that generated from the stock exchange itself. We collect 66,157 news items for the same period[2].All the following stages will deal with collected data form this stage.

c. Data Cleaning

At this stage we dig deeper in our collected data and do the following check on social media data.

a. Detect Arabic words on our data set and delete rows that contains Arabic. We previously agreed that our data set will be in English only. We use basic regular expression mask for .net developers.

b. Detect duplicated data on our dataset. We use general comparing in database ‘Like’ statement.

After those stages we delete 9,951 items from our dataset.

Now we have 1,025,945 rows for all texture dataset ready to use in our next step.

2We received all data related to the Egyptian Stock Exchange from Egypt for Information Dissemination (egID).

We use the Levenshtein Distance algorithm to establish a relation between our information base ontology and the collected dataset to keep track for our data impact.

We calculate the distance between our ontology keyword as shown in Figure 5 and collected data “tweets, Google search”.


```
<Predictor>
  <SymbolCode> EGS3C001C016 </SymbolCode>
  <SymbolNameE >Arabian Cement Company</SymbolNameE>
  <ReutersCode>ARCC.CA</ReutersCode>
<ReutersCode2>ARCC</ReutersCode2>
  <SectorNameE>Construction and Materials</SectorNameE>
  <INDEX>EGX30</INDEX>
  <CEO>Jose Vadillo</CEO>
  <StockMarket>EGX</StockMarket>
</Predictor>
```


Figure 5 Generated file form ontology

Therefore, we convert the huge textual data to numerical data in order to combine with our market data in one prediction engine.

Next, we calculate the mean for all calculated distances for each row, so as to own one value for distance between our ontology and collected dataset.

All Levenshtein data store with the structure as show in Figure 6

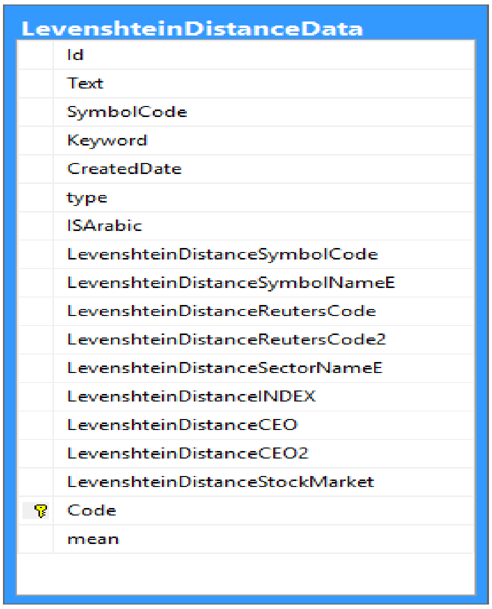

Figure 6 Levenshtein structure

We notice when we use the prediction engine that we have an error in result related to data accuracy. After searching in our data set, we found a significant duplicated data not discovered in the cleaning stage. Therefore, we returned to the cleaning stage and add a secondary cleaning stage after applying Levenshtein algorithm.

c. Using Levenshtein distance to re-clean our dataset form duplication by applying:

X= first row from dataset

Y= second row from dataset

N= number of Levenshtein distance in our case =8

a = min distance between two rows

If (a=0) this means that the two rows are identical and we must keep one and delete the other

We repeat the use of the above equation on our dataset to deep clean duplicated data after this operation and find that much of the duplicated data came from the same news or title but from difference sources so that we keep one of these rows.

Finally now we have 164,854 for all our textual dataset divided as.

Twitter	106,583
Google	53,870
News	4,401

Now we re-enter our textual data into our prediction engine.

At the next stage, we combine results from traditional research about market data.

a) **Stock Market trading**

In this stage, we use previous research in market data algorithms with significant investment to enhance this algorithms. We combine the most used algorithms employed in our market, in order to discover our objective from the researchdetermine if the symbol direction is “Buy, Sell, or Hold”.

Algorithms combined in our model:

- Moving Average ConvergenceDivergence – MACD
- Relative Strength Index – RSI
- Bollinger Band

To select the three algorithms, we create a survey from algorithms used by technical analysts in the market. We combine these algorithms in one graph to predict the symbol behavior.

We use the RSI as a base for our chart. We then update the other algorithm’s value to be relative to RSI and chart them as follows.

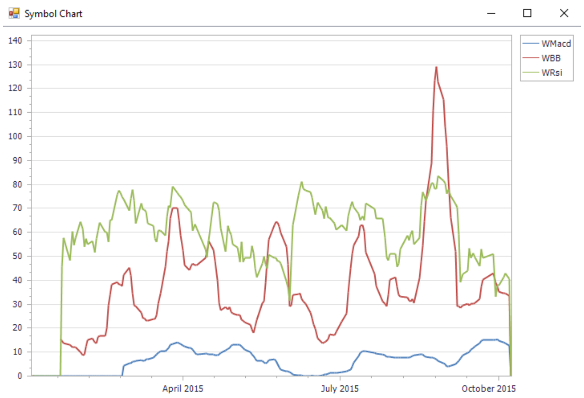

Figure 7 EGS02211C018 Combined Chart

From this chart, our expert technical analysis decides the symbol behavior "Buy, Sell, Hold" that will be used in predication model.

The following table describes a 186 trading day clustering based on our expert technical analysis result from combined the/a chart for one symbol.

Total	Hold	Buy	Sell
186	59	60	67

d. Prediction Engine

In this stage, we start our intelligent modeling to build our engine. We start with two types of cluster for our textual dataset, K-Mean and genetic algorithm (GA), to cluster our dataset for three clusters "Buy, Sell, Hold".

i. Apply K-Mean algorithm for Cluster

We use Dato machine leaning tools to cluster our textual dataset, using the following settings.

- Select all Levenshtein data as features for cluster "Symbol Code, Symbol Name, Reuters Code, Sector Name, INDEX, CEO, Stock Market"
- Number of clusters : 3
- Number of examples : 164,854
- Number of feature columns: 9
- Training method: elkan
- Number of training iterations: 7
- Batch size : 164,854
- Total training time (seconds) : 1.8036

Finally we get result from the K-Mean cluster as following

Cluster Result	Count
Hold	127,479
Buy	34,772
Sell	2,603

ii. Apply Genetic Algorithm for Cluster

The second way to cluster our textual dataset is to use GA. We use AForge.NET [3], an open source C# framework designed for developers and researchers in the fields of computer vision and artificial intelligence for image processing, neural networks, genetic algorithms, fuzzy logic, machine learning, and robotics.

We build our GA model as follows.

- Population site : 20
- Iteration :100
- Selection Method: Elite Selection
- Chromosome Length : 164,854
- Chromosome encoding : Permutation chromosome

Permutation encoding can be used in ordering problems, such as the traveling salesman problem or task ordering problem [15].

In permutation encoding, every chromosome is a string of numbers, which represents a number in a sequence [15].

We get best fitness value = 2 from this GA model with the following cluster.

Hold	Buy	Sell
146,591	12,001	6,262

e.Apply Deep Learning

Finally, we use deep learning techniques to create our neural networks model using Timothy Masters' deep earning project [4] with the following architecture:

i. This is a real-domain model

There are two unsupervised layers, not including input

- First hidden layer has five neurons
- Last hidden layer has five neurons

There are two supervised layers, including

- Hidden layer has six neurons

ii. Training model

Auto encode greedy training parameters.

Initial annealing iterations for starting weights = 100

Initial random range for starting weights = 1.00000

Gradient reduction max iterations = 500

Gradient reduction convergence tolerance = 0.0000500

Greedily trained layers will be fine-tuned after each addition

Unsupervised section weights will remain fixed, not fine-tuned by supervised training

iii. Supervised layer(s) training parameters

Initial annealing iterations for starting weights = 100

Initial random range for starting weights = 1.00000

Supervised optimization max iterations = 1000

Supervised optimization convergence tolerance = 0.0000500

[3] http://www.aforgenet.com/framework/

[4] http://www.timothymasters.info/Deep_learning.html

Table 1:Weights for unsupervised hidden layer 1

	1	2	3	4	5
Code	-0.0637	0.1103	-0.0841	0.0323	0.0976
Symbol Name	0.7395	0.4484	-0.0156	-0.5660	-0.0482
Reuters	-0.0097	-0.1021	-0.0717	0.0088	0.0590
Reuters 2	-0.1263	-0.1072	-0.1922	-0.0364	-0.0329
sector	-0.3824	-0.2100	-0.0163	0.2934	0.1953
index	-0.0910	-0.0166	-0.0702	0.0542	0.0595
CEO	-0.0632	0.3330	-0.1341	0.0376	0.0463
CEO2	-0.3956	0.4083	0.9893	-0.5194	-0.5371
Market	-0.1314	-0.0508	-0.1715	0.0236	0.0128
BIAS	**-0.9337**	**-4.2777**	**-1.4842**	**-0.7576**	**-2.0151**

Table 2:Weights for unsupervised hidden layer 2

	1	2	3	4	5
Neuron 1	-1.8388	2.0983	0.8491	2.3214	2.8816
Neuron 2	0.3853	-.3100	-2.6613	0.3628	-2.3519
Neuron 3	0.6032	1.2986	2.6995	1.9906	0.5955
Neuron 4	1.8885	0.5832	2.9159	0.8589	3.2262
Neuron 5	-1.3180	-.4273	-1.5365	-1.7263	0.1277
BIAS	**0.4453**	**1.0486**	**-0.4851**	**-0.7930**	**-1.1331**

Mean squared error and R-squared of target

MSE of target = 0.20260 RMS = 0.45011 RSQ = 0.00002

Table 3 :Total Win vs. Total Loss above and below various fractions

Threshold	Frac Gtr/Eq	Net	Frac Less	Net
0.141	0.990	24293	0.010	-232
0.147	0.950	23397	0.050	-1128
0.147	0.900	22165	0.100	-2360
0.148	0.800	19675	0.200	-4850
0.148	0.700	17230	0.300	-7295
0.148	0.600	14736	0.400	-9789
0.149	0.500	12207	0.500	-12318
0.149	0.400	9820	0.600	-14705
0.150	0.300	7414	0.700	-17111
0.150	0.200	4912	0.800	-19613
0.151	0.100	2468	0.900	-22057
0.151	0.050	1215	0.950	-23310
0.151	0.010	256	0.990	-24269

(For all tested cases, long ratio = Inf (net=24525.000) and short ratio = 0.0000 (net=-24525.000)

Note **Ratio =** Inf

f. Create Rule-Base Model

In the final stage of our model, we build our rule base model as follows.

Rule 1:

If point X in C1 then

Close Price =<Predict Deep>

And Decision =Hold, Buy, Sell

With CF =

Hold= 0.7 Buy=0.2 Sell = 0.1

Rule 2:

If point X in C2 then

Close Price =<Predict Deep>

And Decision =Hold, Buy, Sell

With CF =

Hold= 0.6 Buy=0.2 Sell = 0.2

Rule 3:

If point X in C2 then

Close Price =<Predict Deep>

And Decision =Hold, Buy, Sell

With CF =

Hold= 0.5 Buy=0.3 Sell = 0.2

VI. CONCLUSION AND FUTURE WORK

In this paper, we introduce the hypothesis that social media updates can increase the accuracy of stock market prediction. We create social media aggregators that work from September 20, 2015 to October 23, 2015 and acquired 1,025,945 transactions on the Internet, related to the Egyptian Stock Market .As our next step, we run several algorithms to clean our dataset then build our engine using GA and deep learning for neural networks. Our preliminary result are show in the rules base model. This engine needs more investigation using fuzzy model to help investors attain more accuracy results. In time, more development will occurin the social media field, and hence, more correlations can be deduced and more accurate predictions can be achieved. No doubt, the more we dig into the social media data mines, the more information we can deduce that will definitely enrich the stock analysis field and yield much better results.

ACKNOWLEDGEMENT

This research is supported by Egypt for Information Dissemination (EGID) and the Egypt stock Exchange (EGX) for constant consulting and support.

REFERENCES

[1] X. D. Zheng Chen, "Study of Stock Prediction Based on Social Network," *31,* 2013.

[2] H. M.-J. Z. Johan Bollen, "Twitter mood predicts the stock market," *7,* 2010.

[3] P. S. Chavan., "Parameters for Stock Market Prediction," *30,* 2013.

[4] 'echnical Analysis Definition," investopedia, 20 6 !015. [Online]. Available: ıttp://www.investopedia.com/terms/t/technicalanalys s.asp.

[5] "Glossary main," Egypt Stock Exchange, 28 6 2015. [Online]. Available: http://www.egx.com.eg/English/glossary_main.aspx .

[6] R. S. E.-f. a. N. M. D. Ahmed A. Gamil, "Stock Technical Analysis using Multi Agent and Fuzzy Logic," *29,* 2007.

[7] S. Y. a. S. Kak, "A Survey of Prediction Using Social Media," *14,* 2012.

[8] "Impact-of-big-data-in-social-media," opallios, 2 3 2015. [Online]. Available: http://www.opallios.com/impact-of-big-data-in-social-media/#sthash.u21QfQMB.dpuf.

[9] H. Mao, "Predicting Financial Markets: Comparing Survey News Twitter and Search Engine Data," *15,* 2011.

[10] W. S. Chan, "Stock Price Reaction to News and No-News," *41,* 2003.

[11] J. Masanès, Web Archiving, 2007.

[12] V. Martin, "Predicting the French stock market using social media analysis," *5,* 2013.

[13] A. R. D. (. Y. Hong Keel Sul, "Trading on Twitter The Financial Information Content of Emotion in Social Media," *9,* 2014.

[14] A. Yi, "Stock Market Prediction Based on Public Attentions: a Social Web Mining Approach," *35,* 2009.

[15] "3rd-platform," 2 5 2015. [Online]. Available: http://www.idc.com/prodserv/3rd-platform.