Stock Price Prediction Using Neural Network Models Based on Tweets Sentiment Scores

Anderson Rioba Ondieki[*], George Onyango Okeyo, Ann Kibe

School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
*Corresponding author: anderson.rioba@gmail.com

Abstract Stock Exchange Prediction using neural networks has been an interesting research problem whereby many researchers have developed a keen interest in prediction of future values and trends. Little research has been done to apply and improve prediction models based on newer and impactful variables to show that mining opinions and sentiments from the information shared in Twitter platform can be converted into statistical values and applied as inputs in a neural network together with other inputs to facilitate an improvement in the accuracy of predictions of stock prices and movements. In this research, two stocks were selected on the basis of their social media communication in twitter and this information was used as additional feature by deploying a supervised learning approach to compute daily company twitter sentiment score for improving prediction purposes in neural networks. The daily twitter sentiment scores were computed in a supervised learning algorithm by use of WordNet and Sentiwordnet lexicons for classification and scoring. Through experimentation with different sets of hidden layers and 70% training set. 15% validation set and 15 % test set, the research applied two Non Linear Autoregressive Neural Network with Exogenous Inputs (NARX) models which were trained using Levenberg-Marquadt back propagation. The results showed that adding lexicon based twitter sentiment scores as additional inputs to other company stock variables for stock price prediction improved the prediction accuracy and resulted to a more accurate NARX model.

Keywords: *artificial neural networks, accuracy, sentiment scores, Sentiwordnet, Wordnet*

Cite This Article: Anderson Rioba Ondieki, George Onyango Okeyo, and Ann Kibe, "Stock Price Prediction Using Neural Network Models Based on Tweets Sentiment Scores." *Journal of Computer Sciences and Applications*, vol. 5, no. 2 (2017): 64-75. doi: 10.12691/jcsa-5-2-3.

1. Introduction

Predictive modeling is the process by which a model is created or chosen to try to best predict the probability of an outcome. Being considered as one of the most common data mining tasks, it involves the process of taking historical data, identifying patterns in the data that are seen though some methodology and then using the model to make predictions about what will happen in the future [5].

Sentiment analysis is a new kind of text analysis which aims at determining the opinion and subjectivity of reviewers. [23] Normally when consumers want to make a choice between competing products, the reputation of the product forms a major part in their decision making process. Therefore applying sentiment analysis helps to unearth their thinking on particular products. Sentiment analysis also helps companies to evaluate what consumers like or dislike and thus take a step in improving their service thus improve their reputation.

Existing classification solutions of sentiments are normally segmented into two categories: machine learning methods by applying learning algorithms on known textual datasets and lexicon based approach which involves calculation of sentiment polarity on semantic orientation of words or sentences in a corpus of text. A hybrid approach can also be considered where lexicon based approaches are combined with machine learning algorithms to provide better classification accuracy and thus improve the classification of text [24].

The stock market provides a regulated place where brokers and companies may meet to make investments on neutral ground [2]. The stocks are listed and traded on stock exchanges which are entities of a corporation or mutual organization specialized in the business of bringing buyers and sellers of the organizations to a listing of stocks and securities together.

Since most of the listed companies on our stock exchange market are also constantly on communication in the social media sphere, consumers are increasingly voice their opinions about and experiences with these companies and their associated brands online. Therefore companies need to be aware of what is said about them in the public sphere as it may impact them either positively or negatively and have a direct influence on their values in the stock market thus impacting on either gaining returns or losing their investments value [11].

Use of sentiment analysis models on social media news as a predictive factor for prices in stock exchange companies would also be essential in knowing the mood and sentiment factor of companies and thus inform investors/traders on the company sentiment portrayal by

the public as a considerable factor in forming decisions on which company to invest in.

The process of sentiment analysis usually involves several steps inorder to achieve the objective of classification or scoring [24]. First, collection of data from several user generated sources such as social media network such as Facebook and Twitter, blogs and comments or news articles. They can be collected either manually or softwares that can be used to scrap the relevant information from the sources. Later on, text preparation is carried out to ensure the necessary textual information is extracted for classification purposes. Thereafter, sentiment is detected and grouped based on the level of objectivity. Sentiment classification is later done to group sentiment text based on multiple points such as positive, negative, neutral, bad etc. In other classification and scoring objectives, we can use lexical databases to obtain sentiment scores in a statistical form. Examples of the lexical databases include Sentiwordnet and Sentistrength. The final step is to get the results of the classification process. Key statistics include the accuracy of the classification model, precision, recall amongst others.

Sentiment analysis processes usually consist of three commonly use approaches: Machine Learning approaches, Lexical-based approaches and Hybrid approach [24]. Machine learning approaches have the advantage of being capable to create models that can be applied in specific contexts for varied purposes. Techniques applied in machine approach include but not limited to term presence and frequency, opinions and phrases and parts of speech data. In lexical-based approaches, they have the advantage of having a wider coverage of terms. They apply corpus-based approaches, and dictionaries and also they can be constructed manually. The hybrid approach combines both the lexicon-based approach and the machine learning approach. This is a huge advantage as it combines the advantages of both machine learning and lexical-based approach. It achieves higher accuracy levels in the classification process and provides a good symbiotic learning.

For purposes of calculating the sentiment value of the company tweets under study, Wordnet was used for applying research processes into the system.It is a lexical database that has relations between similar words. The relations include synonym, hyponym, hypernym, and also groups sentiment based on adjectives, adverbs , verbs and nouns. WordNet provides path similarity measure between sense which is a numerical value that tells how close two words are by their meaning. This way we could produce a numerical values for features telling how close words are to the feature word meanings. Unfortunately, WordNet does not include many words from the company tweets, but given information from WordNet it is possible to handle some interesting cases.

Predicting the stock price trend by interpreting the seemly chaotic market data has always been an attractive topic to both investors and researchers. Among those popular methods that have been employed, Artificial Neural Networks techniques are very popular due to their capacity of identifying stock trend from massive amounts of data that capture the underlying stock price dynamics. However, there is limited research done to analyze the impact of social media sentiments using sentiment

analysis models in enhancing prediction accuracy of the stock values based on the existing stock values and stock sentiment score. By stock sentiment score, this means with regards to the sentiments that a company has received; it may either be positive or negative sentiment which may make investors be bullish or bearish with regards to a particular stock.

The main objective of this research paper is to evaluate the impact of company-specific lexicon-based twitter sentiment data in improving the accuracy of predictions of stock exchange prices using NARX neural network.The research will therefore explain particular methods to obtain sentiment scores from social media by use of specific collection, classification and scoring algorithms and artificial neural network algorithm and also apply sentiment analysis model to social media sentiment on companies to develop a company-specific twitter sentiment score statistic and explore if adding sentiments as inputs features to a neural network will improve or decrease the prediction performance accuracy of the NARX neural network model.

The main contribution of this paper is a method to compute daily company twitter sentiment scoring model that would quantify the sentiment expressed in twitter into a statistic values form. A comparative analysis of prediction using technical indicators data of selected stocks and then later adding a twitter sentiment scoring model to produce lexicon-based sentiments scores to a recurrent NARX neural network is then be applied to check whether there will be an improvement in prediction accuracy. The research therefore hopes to prove that adding sentiment scores as additional inputs to a neural network will improve the prediction accuracy of the network.

The rest of the paper is structured as follows. Section II presents a review of related work. Section III describes the research methodology used, while section IV discussed the experiment results and analysis. The paper is concluded in section V.

2. Related Work

2.1. Stock Market Prediction using Artificial Neural Networks

The financial market movement's prediction has always been regarded as one of the most challenging tasks of time series prediction since the financial market is complicated, dynamic, and evolutionary, volatile and nonlinear. In addition, the affected factors in financial market include political events, general economic conditions, investors' expectations and psychology, and other financial market movements [7].

There has been a lot of research done on the predictive power of neural networks in carrying out prediction tasks. A multilayer perceptron (MLP) neural network model with the Back-Propagation algorithm was applied to predict the Saudi Arabia stock market prices [13]. Their proposed model results from their simulation demonstrated the viability of the proposed model in predicting Saudi Arabia stock markets. A trading model that applied artificial neural networks and wavelet de-

noising in the input time series was developed to predict S&P 500 index that would achieve relatively high rates of return over a long period of time [9].

Procedural neural networks for stock price prediction were applied where their model processed temporal information synchronously without slide time window, which is typically used in the well-known recurrent neural networks [6].

Feedforward neural networks with various experimentations of the network architecture, input data, and training methods were used to predict the movement direction of the next trading day of the Stock Exchange of Thailand (SET) index [18].

An autoregressive neural network where various error metrics for performance evaluation of the predictor by applying real data from national stock exchange of India to evaluate the accuracy of the method was applied to predict future national stock exchange of India index returns [12]. The results were not accurate but the use of better neural predictive systems and training methods for minimizing the prediction errors for the future work was suggested.

2.2. Sentiment Analysis in Stock Markets Predictions

In recent work on stock market prediction[3], Twitter messages from StockTwits was used to identify expert investors for predicting stock price rises by applying support vector machine (SVM) to classify each stock related message to two polarities - "bullish" and "bearish" and then identified experts according to their success. However, the prediction performance realized was still low.

A set of expression patterns to extract opinions was deployed and then mapped those features into different sentiment orientations using a psychometric instrument known as the GPOMS (Google profit of mood state) algorithm [4]. SOFNN (Self-Organizing Fuzzy Neural Network) was trained and showed that one of the six mood dimensions called "Calm" was a statistically significant mood predictor for the DJIA (Dow Jones Industrial Average) daily price up and down change. However, their research only predicted movements in the DJIA index but it wasn't applied on individual companies which would advise investors on which stocks to invest in.

It was discovered that economic analysis portrayed a relationship that existed between consumer sentiment and stock price movements [19]. In that study, features from Twitter messages were applied to capture public mood that was related to four Technology companies for predicting the daily up and down price movements of these companies' NASDAQ (National Association of Securities Dealers Automated Quotations) stocks. A model that combined features namely positive and negative sentiment, consumer confidence in the product with respect to 'bullish' or 'bearish' lexicon and three previous stock market movement days was applied. The features were then deployed in a Decision Tree classifier using cross-fold validation which in turn yielded accuracies of 82.93%, 80.49%, 75.61% and 75.00% in predicting the daily up and down changes of Google

(GOOG), Microsoft (MSFT), Apple (AAPL) and Amazon (AMZN) stocks respectively in a 41 market day sample.

Understanding and predicting the sentiment change of the public opinions would allow business and government agencies to react against negative sentiment and design strategies such as dispelling rumors and post balanced messages to revert the public opinion [8]. A strategy of building statistical models from the social media dynamics to predict collective sentiment dynamics was developed and collective sentiment change was modelled without delving into micro analysis of individual tweets or users and their corresponding low level network structures.

3. Stock Price Prediction Approach

The method applied in this research involves first obtaining the twitter sentiments using twitter advanced search API. The criteria for collecting the sentiment was that the tweets mentioned the companies under study as well as their products under the specified research period. After classification, the tweets are scored using Wordnet and Sentiwordnet lexicons to determine the polarity of textual information and produce daily sentiment scores. Finally, neural network models are developed and accuracies compared when there is no sentiment versus when there is sentiment as inputs.

3.1. Sentiment Classification Process

The approach applied to classify the tweets corpus was as follows:

a) **Sentiment corpus collection**: A series of twitter sentiments were collected from Twitter Advanced search API and submitted to a Twitter sentiment model for optimal sentiment scoring. For Stock 01(Equity Bank), 3,950 tweets were collected while for Stock02 (Safaricom) 39,978 tweets were collected by use of Twitter Advanced Search API. Tweets that mentioned the related companies and their product were collected and manually labeled as either positive or negative.

b) **Tweets Corpus Labeling:** Tweets that were considered positive or negative were separated into two separate sets for each of the companies under research and labeled as either positive or negative. These sets were later subjected to a classification process using a hybrid process of machine learning and lexical-based approach to check on the accuracy of the models.

c) **Twitter Sentiment Model**: A sentiment analyzer model was applied to check on the accuracy. The steps that were used to build up the sentiment model:

1. **Tokenization**

   This involved the splitting up of the twitter corpus sentences into words that would be later be stored as word vectors for applying them in sentiment scoring process.

2. **Filter Tokens by content**

   In this process, tokens that were removed include the usernames i.e. twitter username accounts such as @safaricomltd, @safaricom_Care,

@Keequitykenya and # element(hashtag). These were eliminated as they had no sentiment value to be scored.

3. Filter tokens by length

The minimum character length for tokens was set as 3 and the longest token character length was set as 25 as some short text elements could not have any sentiment value.

4. Transform Case

All the twitter corpus content was transformed to lowercase.

5. Stem using Wordnet

Stemming using Wordnet dictionary was applied to get the adjectives, verbs, nouns and verbs as they would be key in ensuring the right sentiment values are obtained from the Sentiwordnet corpus.

6. Filter of stop words

Stopwords were eliminated from the twitter corpus as they were considered irrelevant in sentiment value. Some of the stopwords that were eliminated include 'a', 'an', 'the', 'it'.

d) Classification: After the above process was done, the processed twitter corpus was then classified using a machine learning process to obtain the accuracy measures. A series of algorithms were applied using cross validation model to the twitter data to check on the performance of the model. The model with highest accuracy was applied to progress with the sentiment scoring process.

3.2. Sentiment Scoring Process

After the Twitter Sentiment Analysis model achieved the desired accuracy measures with good accuracy percentages and better precision and recall values, the model was applied to extract the daily twitter sentiment scores which would be applied to the Neural Network for comparative analysis.

Here were the steps taken to develop and deploy the required model:

1. Retrieving of the word vectors

The word vectors that were obtained from the sentiment classification model were retrieved to be used for scoring the sentiments within the daily sentiments.

2. Uploading the day tweets

Each of the daily tweets was uploaded so that they could be cleansed and analyzed for their tweets.

3. Tokenization

This involved the splitting up of the twitter corpus sentences into words that would be later added to the words vector for applying the model to get proper sentiment scores from the model.

4. Filter Tokens by content

In this process, tokens that were removed include the usernames i.e. twitter username accounts such as @safaricomltd, @safaricom_Care, @Keequitykenya. These were eliminated as they had no sentiment value to be scored.

5. Filter tokens by length

The minimum character length for tokens was set as 3 and the longest token character length was set as 25 as some short text elements could not have any sentiment value.

6. Transform Case

All the twitter corpus content was transformed to lowercase.

7. Filter of stop words

Stopwords were eliminated from the twitter corpus as they were considered irrelevant in sentiment value. Some of the stopwords that were eliminated include 'a', 'an', 'the', 'it'.

8. Stem using Wordnet

The Wordnet corpus dictionary was applied to get the adjectives, verbs, nouns and verbs as they would be key in ensuring the right sentiment values are obtained from the Sentiwordnet corpus. The stemming process also helped to reduce derivationally related words with similar meanings.

9. Extraction of Day sentiment scores using Sentiwordnet.

This process applied the use of Sentiwordnet 3.0.0 to extracts sentiment scores. Wordnet 3.0 and a SentiWordNet 3.0.0 database was used to extract sentiment of the tweets inputs document. The sentiment value is in range [-1.0, 1.0] where -1.0 meant very negative and 1.0 meant very positive. The sentiment value was added to the metadata of the tweets for daily sentiment scoring. Wordnet and Sentiwordnet are connected by Synset IDs. To calculate sentiment of a document the sentiment of each word was calculated, where the first meaning of a word was considered to have the most influence on a sentiment and each next meaning has less influence on a sentiment. Document sentiment was then calculated as the average value of all word sentiments.

The processed twitter sentiments were then run through the base Twitter sentiment classification model to extract the classifications accuracy and extract the final computations of sentiment scores from the unlabeled twitter corpus in order to get the daily twitter sentiment score computations.

For calculation of daily sentiment, the following process was followed:

a) For a given lemma with **n** senses (lemma#n), the formula applied to all the n posScores and negScores of the lemma.

$$positive\ Score\,(Pos\ Score) = pos\ Score\,1$$

$$\text{and } negative\ Score\,(Negscore) = neg\ Score\,1$$

b) Calculation of the tweet score was then performed by the lexicon using the following formula:

$$Tweet\ Score = \frac{Total\ sum\ of\ scores\ within\ the\ Tweet\,(s)}{Total\ Number\ of\ words\ in\ the\ tweet\,(n)}.$$

Tweet Score was the positive or negative scores of sentences; score(s) is the positive or negative scores of the word in the Tweet; n is the number of words in the tweet.

c) The calculation of the daily tweet score was calculated to decide about positivity or negativity of overall tweets of the day by averaging the positive and negative scores for the company tweets.The formula for this computation was as follows:

$$Day\ Tweet\ score = \frac{Total\ sum\ of\ scores\ for\ sentences\,(x)}{Total\ number\ of\ day\ tweets\,(y)}.$$

The results were then fed into the Neural Network system as additional inputs to check for any improvement in accuracy of predictions.

3.3. NARX Neural Network Creation.

3.3.1. NARX Networks

An important and useful class of discrete time nonlinear systems for dynamic time series prediction is the "Nonlinear Autoregressive models with exogenous input (NARX model)". It is described as a dynamic neural network that contains feedback connections that enclose several layers for the network. This is a powerful class of models which have been demonstrated to be well suited for modeling nonlinear systems and especially time series. These networks converge much faster and generalize better than other networks hence their ability to discover long time dependences than conventional recurrent neural networks [15].

NARX is a dynamical recurrent neural network based on the linear autoregressive models with exogenous input (ARX) model. The next value of the dependent output signal $x(t)$ is regressed over the latest n_x values of the independent input signal and n_u values of the dependent output signal. n_x and n_u respectively represent the dynamical order of the inputs and outputs of the NARX. A mathematical description of the NARX model is summarized in (1) in which f is a nonlinear function.

$$x(t) = f\begin{pmatrix} x(t-1), x(t-2), \dots, x(t-n_x), \\ u(t-1), u(t-2), \dots, u(t-n_u) \end{pmatrix}. \quad (1)$$

3.3.2. Training of NARX Neural Networks

Each time a neural network is trained results in a different solution due to different initial weight and bias values and different divisions of data into training, validation, and test sets. As a result, different neural networks trained on the same problem can give different outputs for the same input. To ensure that a neural network of good accuracy has been found, it was observed that one has to retrain several times until the optimal results have been obtained.

There are three sets of data that are key in processing and evaluating a NARX network model [21]:

a) **Training Set**: this data set is normally applied to adjust the weights on the neural network.
b) **Validation Set**: this data set is used to minimize over fitting. This dataset helps in verifying that any increase that occurs in the training set of the data improves the accuracy of any data that hasn't been trained on. Ideally, the validation set has to be on a higher accuracy level than the training set otherwise the neural network will be deemed to be over fitting.
c) **Testing Set**: this data set is used only for testing the final solution in order to confirm the actual predictive power of the network.

3.3.3. NARX Network Processes

a) Variable selection

The model applied in the research used the stock market indicator values obtained from the Stock market data vendors. They included: day's stock highest price, the lowest stock price of the day, the highest stock price in the year and the lowest stock price in the year as well as the volumes of stock sold per day. The closing price was used as the target for testing the effectiveness of the network.

b) Data Collection and dealing with missing observations.

The cost and availability for data was considered for conducting a successful research process. Technical data was readily available from many vendors at an affordable cost. For the missing observations as well as for days that trading didn't occur, the research process adopted a concave function solution [1] to approximate the missing values. If the company stock value on a given day is m and the next available data is n with q days missing in between, an approximation for the missing data was done by estimating the first day after m to be $(n+m)/2$ and then following the same method recursively till all gaps were filled. This approximation was justified as the stock market data usually follows a concave function, unless of course at anomaly points of sudden rise and fall.

c) Data partitioning into training, testing and validation sets.

The baseline network architecture applied for the research process was 70:15:15 whereby 70% of the input data was used for training, 15% was used for testing the network and then 15% used for testing the performance of the network.
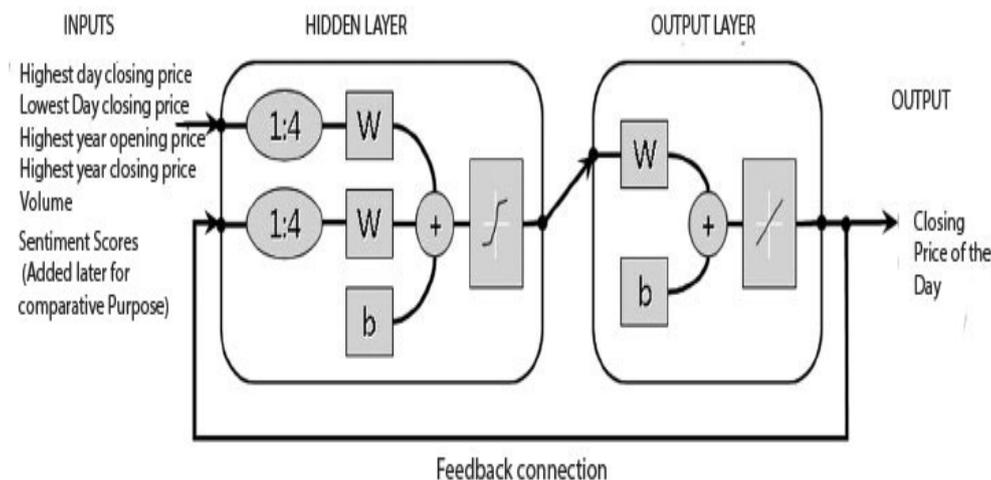

Figure 1. The representation of a NARX neural network

d) Neural Network Architecture

The Neural Network design was selected as follows:

Inputs: The inputs nodes to the Neural Network were computed based on the independent variables of stock data provide for the research process. The research process involved a comparative approach of determining if sentiment scores are added as additional inputs to a neural network will cause the network to perform better.

Hidden Layers: There are some empirically-derived rules-of-thumb that can be applied in improving the performance of a neural network; and of these; the most commonly relied on is "the optimal size of the hidden layer is usually between the size of the input and size of the output layers [16]. This was the applied method in determining the optimal hidden layer size for prediction analysis for the research analysis. For the input delays and feedback delays, they were both maintained as 2. This was to ensure a formidable baseline strategy for better comparative prediction process for the varying inputs.

e) Training of the Neural Network

The main objective of training in a neural network was to obtain the optimal set of weights and bias that provide optimal result for computations by the network and global minimum of the error function.

The back propagation algorithm applied for the training the neural network was the Levenberg - Marquardt optimization algorithm. This robust algorithm and performs best on minimal amounts of data for computational analysis. The algorithm is described as a standard technique for computations of nonlinear least squares method usually thought of as a combination of steep descent and Gauss-Newton method [17]. A technique that locates local minimum of a function is utilized and then expressed as the mean square errors of the non-linear functions [20].

The equation to approximate a function on the above algorithm is solved by the following equation:

$$\left(\mathbf{J^t J} \ + \ \lambda \mathbf{I}\right)\delta = \mathbf{J^t E} \tag{2}$$

Where   J = Jacobian matrix of the system
λ = Levenberg's damping factor,
δ = weight update vector and
E = error vector containing the output errors for each input vector used on training the neural network.

The weight update vector (δ) suggests by how much the network weights need to be changed inorder to achieve a better solution. The $\mathbf{J^t J}$ matrix can also be known as the approximated Hessian. The λ (damping factor) is normally adjusted at each iteration, and helps to guide the optimization process [22].
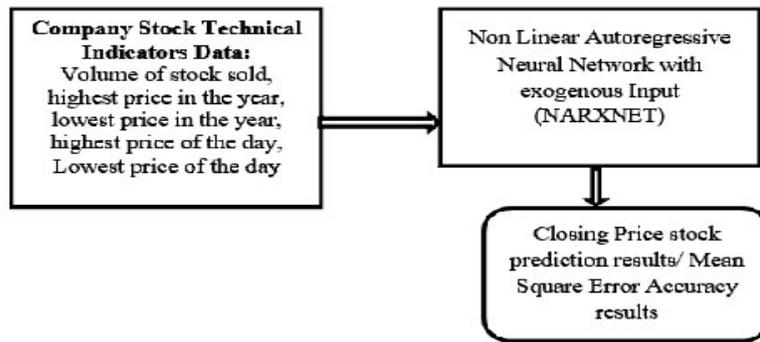
Figure 2. Prediction without Twitter Sentiwordnet scores. This describes the flow and features used to perform the prediction accuracy measurement before sentiment data is added
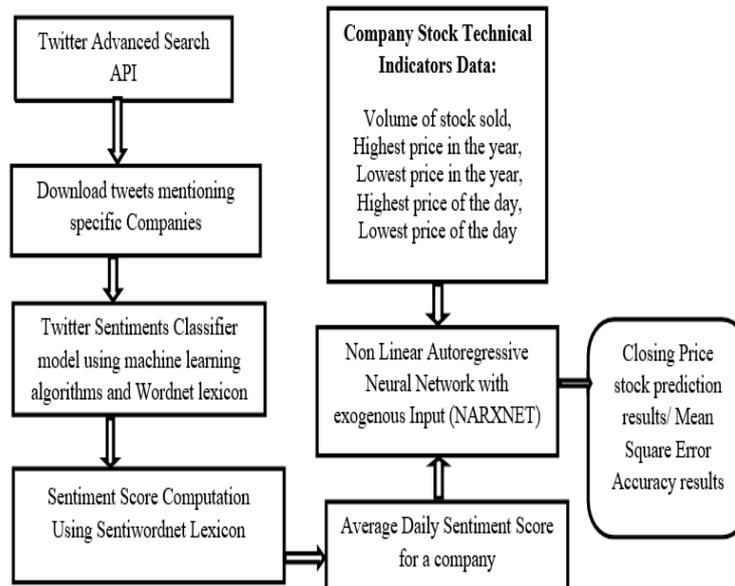
Figure 3. Prediction inclusive of Twitter Sentiwordnet scores. This describes the flow and features used to perform the prediction accuracy measurement after sentiment scores data is added

f) Neural Network Output.

The NARX Neural Network outputs were then achieved by compiling the results on the test set for the regression and training, validation and test set MSE values which was the accuracy measure of the network. Retraining the network model several times was applied in order to ensure better accuracy values were obtained. The smaller the Mean Squared Error, the closer the fit is to the observations data. Lower MSE values are usually preferred as it is an indication of a better performing model.

4. Experiment and Results

4.1. Experimental Set-up

For both sentiment classification and sentiment scoring processes, RapidMiner version 7.0 software was applied. It has modules and operators that made it possible to analyze huge sets of sentiment corpus for classification and scoring processes. The neural network model applied the use of Matlab neural network tool to analyze and develop neural network models to test the accuracy of the models using Mean Squared Error. The data was collected between September 2015 and December 2015.

4.1.1. Sentiment Classification Experiment Set up in Rapid Miner

A series of steps were applied with the processes that were picked from the Rapid Miner software.

Step 1: (Process Documents from files): The tweets under study are loaded in two sets.(Positive set and negative set) for the classification process to occur.

Step 2: (Store) The classification model wordlist is stored for use in the scoring process.

Step 3: (Validation) Under Validation process, the processed documents are classified to determine the accuracy of the model. First, the process model is chosen, eg Naïve Bayes, Neural Network, and KNN, on the training set then under the testing section, the modelling process chosen is applied to get the performance accuracy results.(see Figure 5).

Step 4: (Store) The classification accuracy results are then stored for further analysis.
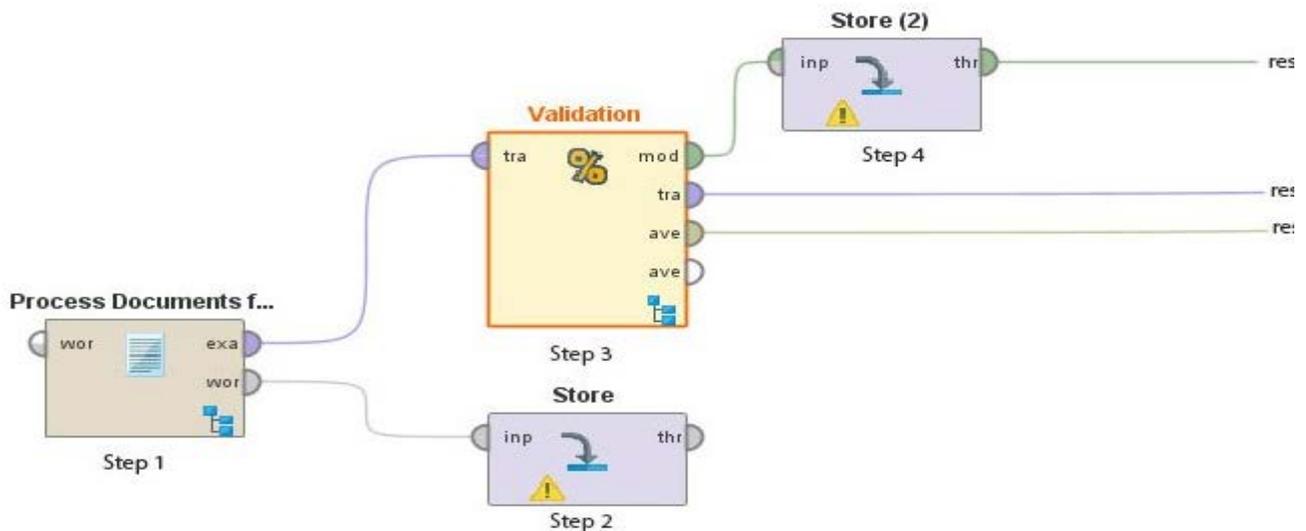

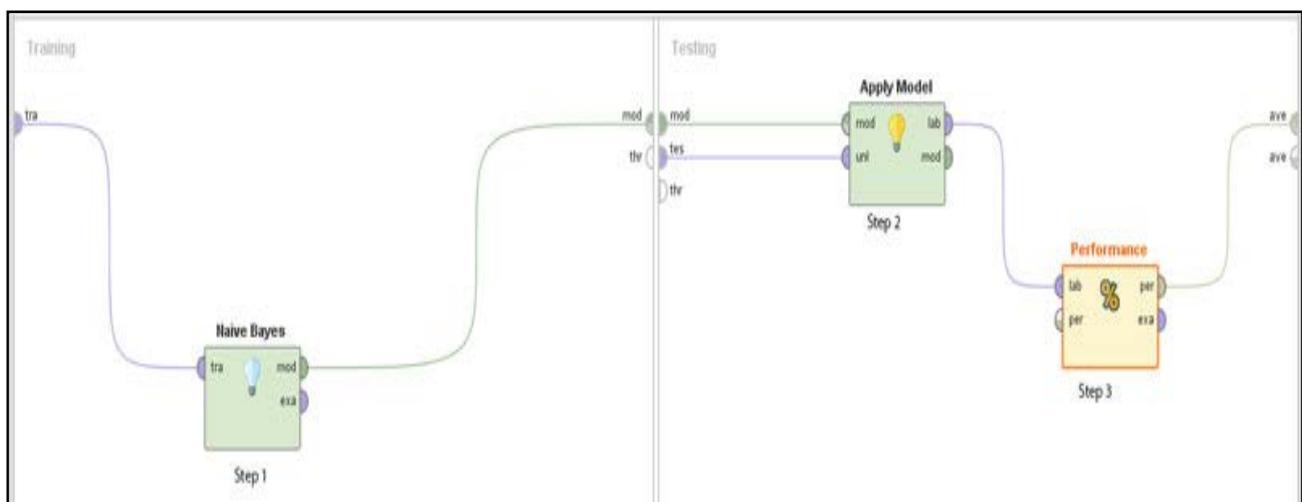**Figure 4.** The Sentiment classification process as carried out in RapidMiner Software.

Figure 5. The validation sub-processes. They include: Adding of the classification model (step 1); Applying the model after the training process under the test section (step 2); and extraction of the performance values (step 3)
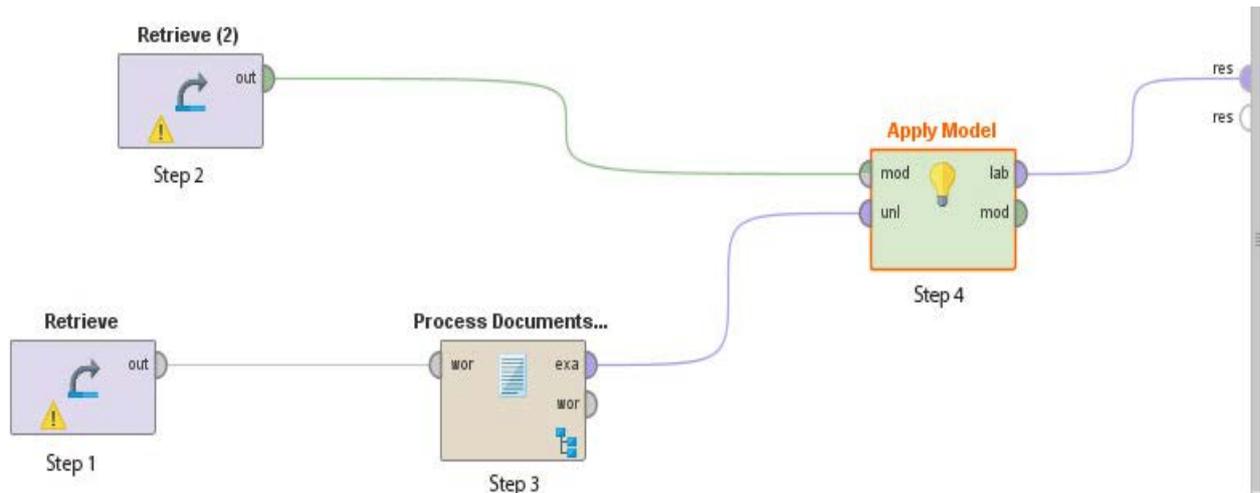
Figure 6. The sentiment scoring processes. This is the processes that take place after the sentiment classification has been done
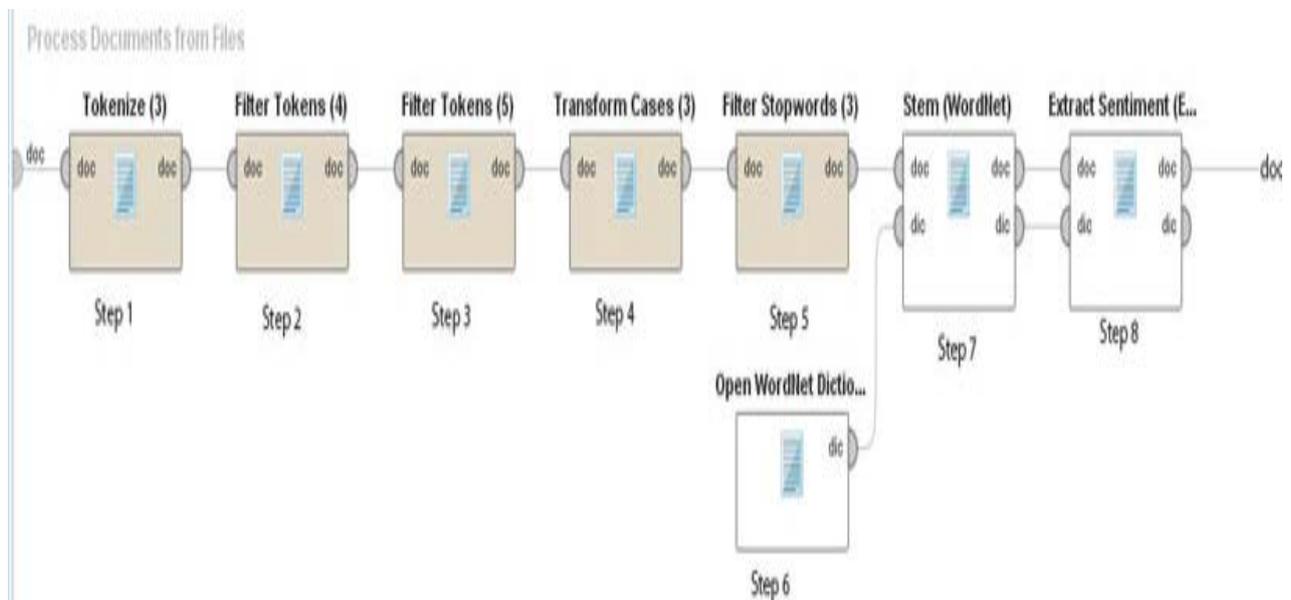

Figure 7. Illustration of the Sub-processes under the Process Documents from file Process in the sentiment scoring process (Step 3)

4.1.2. Sentiment Scoring Set up

For Figure 6, here is a description of the steps undertaken.

Step 1: The processed word list using Wordnet lexical database is retrieved and used at the process documents sub-process.

Step 2: The classification model results is retrieved to be later applied at the Apply Model process (step 4).

Step 3: The Process Documents sub process uses the retrieved word vectors from step 1, and processes the daily tweets to get the daily sentiment scores. Sub processes include Tokenization (removing non- letters from the daily tweets corpus, filter the word tokens based by character length, remove unnecessary symbols eg hashtags(#), and at (@), transforming all the text to lowercase, remove stop words (a, an, at) as they don't add any value to text processing and scoring, stem the processed tokens into nouns, adverbs, verbs and adjective using the Wordnet Lexical database, then apply the Sentiwordnet process to compute the daily sentiment scores for the text.(See Figure 7)

Step 4: Apply the classification model and the sentiment scoring process to obtain the final sentiment scores for the day.

4.1.3. NARX Neural Network Experimental set-up.

For the NARX Network process, the MATLAB Neural Network toolbox was used.

The dynamic time series application was loaded and the data under study was loaded for processing. They were loaded in two sets: Training set: (for training the data) and Target set (To compare the outputs with the processed results.)

Later on the other variables such as the hidden layers, the input delays were set up. The training method for the network was then chosen and the process run to display the Mean Square Error results for the training set, validation set and test sets.

The process was done in two steps for each of the data sets under study. One did not contain the sentiment scores while the other contained the sentiment scores. This was done to compare the results of the network when subjected to additional data (sentiment scores) to check on the improvement in accuracy.
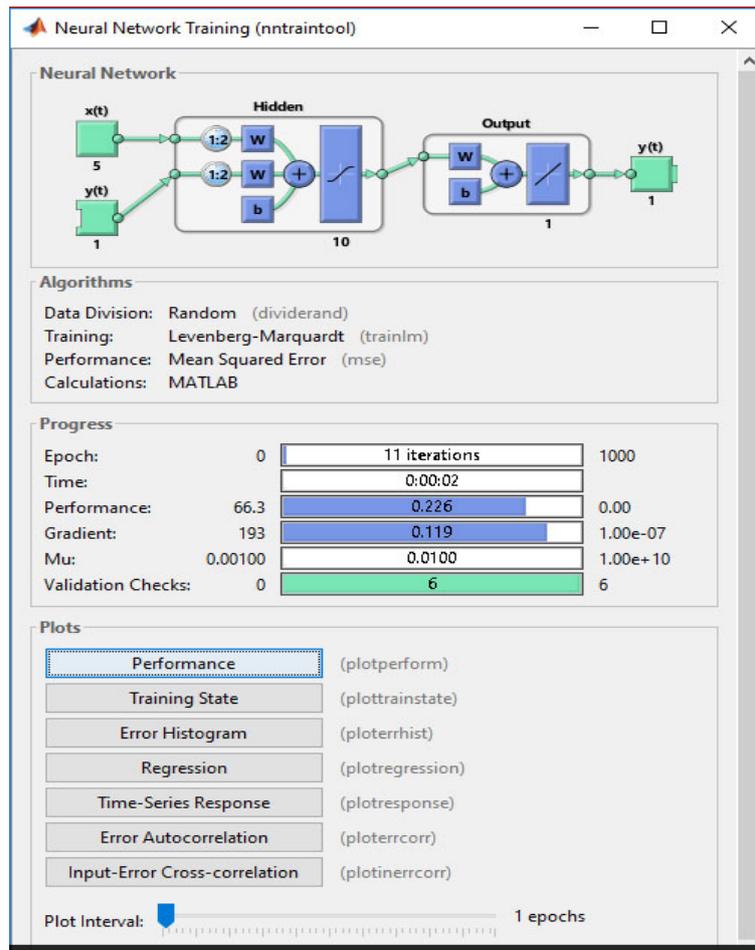
Figure 8. MATLAB neural network training tool. On the top, it shows the structure of the neural network. The structure displays x(t) showing the number of inputs into the network and y(t) representing the processed output that is regressed back into the network as additional input. It also shows the number of hidden layers and the input and feedback delays as well weights(w) and bias(b).On second subsection, the algorithms applied are displayed. The third subsection displays the progress of the training. The plots section shows the various plots and graphs that can be displayed on the completion of the training process

4.2. Results

4.2.1. Accuracies of the Sentiment Classification Model

For evaluation of the sentiment corpus classification model, accuracy values were considered. A series of learning algorithms were applied to test their efficiency and the one with the best values was considered. For purposes of comparative analysis, three algorithms were used: Naïve Bayes, K-NN and Neural Network. Accuracy was chosen as an evaluation metric and the resulting average accuracy has demonstrated the feasibility of the approach.

K-NN is described as "an instance based machine learning algorithm that classifies feature space based on the closest training cases [25]. K-NN finds the k closest instances to a predefined instance and decides its class label by identifying the most frequent class label among the training data that have the minimum distance between the query instance and training instances."

Naïve Bayes algorithm treats all features independently and how they make a prediction with no feature depending on other features values [25]. It's easy to construct, it doesn't need parameter estimation, interpretation is easy and can be performed by both expert and novice data mining developers. It performs well in comparison with other data mining methods.

Neural Networks also work well with classification problems as they can generalize well and get the best classification results based on the amount of data fed to it. Neural Nets also tend to perform well with large datasets as they are able to learn and generalize better when exposed to large amounts of data.

The table below shows the results after the tweets have been subjected to the classification process.

Table 1. Classification accuracy results for stock 01(Equity Bank) and Stock 02(Safaricom)

Scoring Algorithm	Accuracy Results	
	Safaricom Tweets (Stock 02)	Equity Bank (Stock 01)
Naïve Bayes	57.69%	52.14%
KNN	57.71%	63.04%
Neural Network	61.69%	61.63%

For both the companies under study, the considerations for modelling were based on the following three modelling samples. It was observed that the classification algorithms differed in terms of their accuracies. For stock 01, K-NN algorithm performed much better than neural network and Naïve Bayes. This showed that the classification process was much better when subjected to the K-NN algorithm thus facilitating a better approach to

obtaining better word vectors and also facilitating production of much accurate sentiment scores. For Stock 02, Neural Network produced much better and accurate scores than the other two algorithms under comparison. This would be interpreted as the ability of the algorithm to produce better results when subjected to much larger data sets hence the result being different from stock 01.Neural networks tend to over fit in the classification process when subjected to small amounts of data hence lower accuracy.

4.2.2. Sentiment Scores Model Results

The sentiment scores for the companies under study were successively computed and the results were compiled in one table. The daily sentiment score was arrived by averaging the total sum of sentiment score of each tweet in one day. The sentiment scores obtained were then added later as additional inputs to the NARX neural network for the comparative analysis function.

On observation, it was found that most of the sentiment scores were close towards the 0 value with some above 0 and other slightly below the 0 value. (Table 2)

4.2.3. Results of the NARX Neural Network Predictions

The neural network applied for the comparison was created and trained in open loop form. Open loop (single-step) is perceived to be more efficient than closed loop (multi-step) training due to its ability to allow supply to the neural network with correct past outputs as we train it to produce the correct current outputs. The Mean Squared Error computations were computed using Matlab Neural Network Toolbox [10] using the NARX Neural Network that is under the Dynamic Time Series Neural Network Box. For each of the Stock company under study, the lowest Mean Squared Error on the test set was computed. The baseline comparative ratio for training, validation and test set was set as 70:15:15 respectively. Different numbers of hidden layers were experimented to determine the best number for optimal results. The results were compiled and results presented in form of bar graphs as shown below.

Table 2. Sample table of sentiment scores results

Date	Stock01 (Equity Bank)	Stock02 (Safaricom)
9/3/2015	0.080180825	0.039540097
9/4/2015	0.04295992	0.020876819
10/17/2015	0.057992204	0.046042574
10/18/2015	0.183905909	0.003063902
10/19/2015	0.096326947	0.027075861
10/20/2015	0.160838729	0.052229502
10/22/2015	0.091332901	0.074488674
12/16/2015	0.029963832	0.083033785
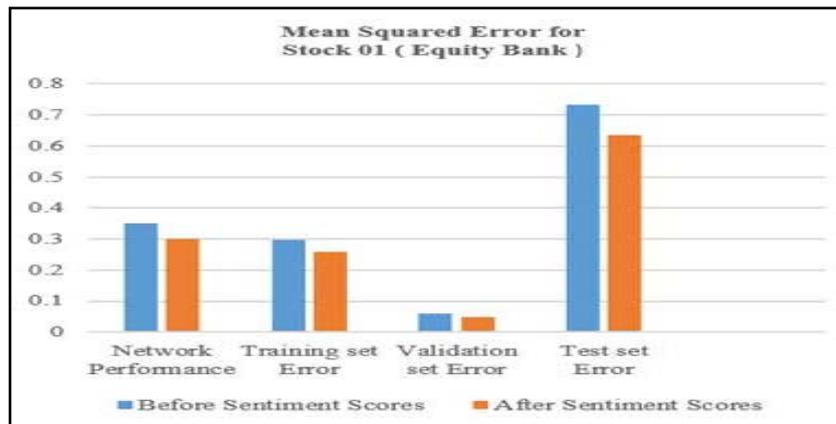
Figure 9. Mean Square Errors values for Stock 01 (Equity Bank) before and after sentiment scores are added as inputs to the NARX neural network

Figure 10. Mean Square Errors values for Stock 02 (Safaricom) before and after sentiment scores are added as inputs to the NARX neural network.

4.3. Observation and Analysis

4.3.1. Results before Sentiment Scores Addition to the Neural Network

The Mean Square Error for each of the stocks on the test set data was computed comparatively on the differentiated basis of inclusion of sentiment scores as inputs to other technical indicators data for prediction purposes. Prior to adding of sentiment scores as additional inputs to the Neural Network, Stock 01(Equity Bank) obtained the most optimal Mean Squared Error Performance Accuracy on the training set of 0.2984, 0.0609 on validation set and 0.7316 on test set. The network performance obtained was 0.3498 and had four hidden layers. Stock 02(Safaricom) obtained the most optimal Mean Squared Error Performance Accuracy on the training set of 0.0280, 0.0430 on validation set and 0.0427 on test set. The network performance obtained was 0.0324 and had one hidden layer.

4.3.2. Results after Addition of Sentiment Scores to the Neural Network

After addition of Sentiwordnet Twitter sentiment score as additional inputs to the neural network, Stock 01(Equity Bank) obtained the most optimal Mean Squared Error Performance Accuracy on the training set of 0.2576, 0.0493 on validation set and 0.6361 on test set. The network performance obtained was 0.2993 and had four hidden layers. Stock 02(Safaricom) obtained the most optimal Mean Squared Error Performance Accuracy on the training set of 0.0279, 0.0236 on validation set and 0.0309 on test set. The network performance obtained was 0.0277 and had four hidden layers.

According to the obtained results, they clearly demonstrated a reduction in the Mean Squared Error values after the Sentiment scores were included as additional inputs to the neural network. For Stock 01, the training set had its MSE lowered by 0.0408, 0.0116 on the validation set and 0.0955 on the test set. The network performance was lowered by 0.0505. For Stock 01, the training set had its MSE lowered by 0.0001, 0.0194 on the validation set and 0.0118 on the test set. The network performance was lowered by 0.0047.

4.3.3. Discussion of the Results

Lower MSE values for both Stock 01(Equity Bank) and Stock 02 (Safaricom) were obtained when the Sentiwordnet twitter sentiment scores were added to the NARX network for prediction. Since Mean Squared Error is a measure of performance accuracy of the network used in the prediction analysis, lower values were desirable. Our results for the research process show that for both of the stocks data, the lowest MSE predicted was the ones that had Sentiwordnet-based twitter sentiment scores as additional inputs to the Neural Network. This meant that the NARX neural network model was quite reliable in determining if addition of sentiment scores as inputs to a neural network, there is an improvement or reduction in prediction performance. NARX networks are usually suitable for their task based on their accuracy.

During the research process, research limitations that were experienced include limited time in obtaining a substantial amount of data for analysis. It would have been more desirable to obtain more quantities of research data to do prediction over a longer timeframe. Other sources of input data like fundamental indicators as well as daily sentiments from other sources such as newspapers and stock-exchange related blogs would have been considered as more additional inputs for prediction purposes.

5. Conclusion and Future Works

In this paper we provided a way and means of providing sentiment scores by use of classification and scoring approaches on company specific tweets to obtain numerical polarity values on daily basis. The scores are then used in neural network models in combination with other values to determine its impact in accuracy improvement. The research experiments show that adding sentiment data using Sentiwordnet lexicon-based approach as inputs to the neural network would be viable approach to improving the prediction accuracy of a neural network.

A major limitation was the large corpus of texts that were not in English language. They were largely ignored in the tweets processing as they were is Swahili as well as corrupted English-swahili lingo commonly known as "sheng". There needs to be a future way to adapt the local languages in parsing such dictionaries inorder to apply them in solving natural language processing problems. Furthermore, research can be done to explicitly show the most causative sentiment elements that trigger market reactions by use of deep learning methods. Deep learning is relatively a new field but has proved to be an excellent tool in improving the accuracy of prediction and also plays a huge role in pattern recognition. It is currently being applied in many technology companies in different contexts hence forming a huge part in developing methods to classify and analyze text.

As future work, we will extend the analysis by evaluating the accuracy of more lexical resources as well as additional sentiment datasets from a larger corpus of text such as other forms of social media such as blogs, facebook posts, and news articles. We also hope to see whether there can be a combination of several lexicons and models for better accuracy and efficiency of classification.

References

[1] Anshul M., Arpit G. (2013) "*Stock Prediction Using Twitter Sentiment Analysis.*" Stanford University Computer Science Department.

[2] Asif Perwej, Yusuf Perwej (2012). "*Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm.*" Journal of Intelligent Learning Systems and Applications, pages 108-119.

[3] Bar-Haim R., Dinur E., Feldman R., Fresko M., and Goldstein, G. (2011). *Identifying and following expert investors in stock microblogs.* In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11, pages 1310-1319, Stroudsburg, PA, USA. Association for Computational Linguistics.

[4] Bollen, J., Mao H., and Zeng, X. (2011b). *Twitter mood predicts the stock market.* Journal of Computational Science, pp. 1-8.

[5] Jared D. (2014) "*Big Data, Data Mining and Machine Learning*". John Wiley & Sons, Inc.

[6]   Jiuzhen L., Wei S., and Mei W. (2011), *"Stock Price Prediction Based on Procedural Neural Networks.*" Department of Electrical Engineering, Jiangnan University, Wuxi 214122, China Advances in Artificial Neural Systems Volume 2011, Article ID 814769.

[7]   Kara Y., Acar B. & Baykan K. (2011). *"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.

[8]   Le T., Pang W., William C., Wei P., Ying Z. (2012). *"Predicting Collective Sentiment Dynamics from Time-series Social Media."*

[9]   Lipo W. & Shekhar G. (2013). *"Neural Networks and Wavelet De-Noising for Stock Trading and Prediction"*. Time Series Analysis, Modeling and Applications Volume 47 of the series Intelligent Systems Reference Library pp 229-247.

[10] MathWorks (2015) *"Neural Network Toolbox."* Retrieved from http://www.mathworks.com/.

[11] Oxford Business Group (2013) *"Kenya looks to woo investors amid global stock market slowdown."* Retrieved from http://www.oxfordbusinessgroup.com/news/kenya-looks-woo-investors-amid-global-stock-market-slowdown.

[12] Rather, A.M. (2011). *"A prediction based approach for stock returns using autoregressive neural networks."* World Congress on Information and Communication Technologies (WICT), pp. 1271-1275.

[13] Olatunji S.O., Mohammad S., Moustafa E. & Yaser A. (2013). *"Forecasting the Saudi Arabia Stock Prices Based on Artificial Neural Networks Model."* International Journal of Intelligent Information Systems. Vol. 2, No. 5, pp. 77-86.

[14] Senti Word Net (2011) *"About SentiWordNet"*. Retrieved from http://sentiwordnet.isti.cnr.it/.

[15] Songbian Z. (2013) *"Africa Economic Growth Forecasting Research Based on Artificial Neural Network Model: Case Study of Benin."* International Journal of Engineering Research & Technology (IJERT)ISSN: 2278-0181 Vol. 3 Issue 11

[16] Stack Exchange (2013) *"Choosing the layers of hidden nodes in a neural network"* Retrieved from

http://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw.

[17] Ryan Sangjun Lee, Gregery T. Buzzard and Peter H. Meckl (2012). "Optimal Parameter Estimation for Long- Term Prediction in the Presence of Model Mismatch".

[18] Suchira C. (2013). *"An investigation into the use of neural networks for the prediction of the stock exchange of Thailand."*

[19] Tien Thanh Vu, Shu Chang, Quang Thuy Ha, Nigel Collier (2012) *"An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter"*. Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data, pp 23-38.

[20] Uma B., Sundar D., Alli P. (2013). An Optimized Approach to Predict the Stock Market Behavior and Investment Decision Making using Benchmark Algorithms for Naive Investors.

[21] Stack Exchange (2016). *"What is the difference between training, validation and test set"*. Retrieved from http://stackoverflow.com/questions/2976452/whats-is-the-difference-between-train-validation-and-test-set-in-neural-networ.

[22] César Souza (2009): Neural Network Learning by the Levenberg-Marquardt Algorithm with Bayesian Regularization. Retrieved from http://crsouza.com/2009/11/18/neural-network-learning-by-the-levenberg-marquardt-algorithm-with-bayesian-regularization-part-1/.

[23] Anais Collomb, Crina Costea, Damien Joyeux, Omar Hasan, Lionel Brunie (2015)*A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation.* University of Lyon.

[24] Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, Tiziana Guzzo (2015). "Approaches, tools and Applications for Sentiment Analysis Implementation". International Journal of Computer Applications. (0975-8887) Volume 125-No.3, September 2015.

[25] Benard Nyangena Kiage (2015). A data mining approach for forecasting cancer threats. Published Masters thesis. Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya.