Agrekon

Agricultural Economics Research, Policy and Practice in Southern Africa

ISSN: 0303-1853 (Print) 2078-0400 (Online) Journal homepage: www.tandfonline.com/journals/ragr20

Real-Time Grain Commodities Price Predictions in South Africa: A Big Data and Neural Networks Approach

Kayode Ayankoya, Andre P. Calitz & Jean H. Greyling

REAL-TIME GRAIN COMMODITIES PRICE PREDICTIONS IN SOUTH AFRICA: A BIG DATA AND NEURAL NETWORKS APPROACH

Kayode Ayankoya
Department of Computing Sciences
Nelson Mandela Metropolitan University
s212400096@nmmu.ac.za

Andre P. Calitz
Department of Computing Sciences
Nelson Mandela Metropolitan University
andre.calitz@nmmu.ac.za

Jean H. Greyling
Department of Computing Sciences
Nelson Mandela Metropolitan University
jean.greyling@nmmu.ac.za

ABSTRACT

The prices of agricultural grain commodities are known to be volatile due to several factors that influence these prices. Moreover, different combinations of these factors, such as demand, supply and macroeconomic indicators are responsible for the price volatility at different times. Big Data presents opportunities to collect and integrate datasets from several sources for the purpose of discovering useful patterns and extracting actionable insights that can be used to gain competitive advantage or improve decision making. Neural Networks presents research opportunities for training computer algorithms to model linear and non-linear patterns that might exist in datasets for the purpose of extracting actionable insights such as making predictions. This article proposes a Big Data and Neural Networks approach for predicting prices of grain commodities in South Africa. It was identified that disparate data that influence the grain commodities market can be acquired, integrated and analysed in real-time to predict future prices of grain commodities. By utilising SAP HANA as the enabling Big Data technology, data acquired from several sources was used to create an integrated dataset, and a predictive model was developed using Backpropagation Neural Network algorithms. This model was used to predict the daily spot prices of white maize on the Johannesburg Stock Exchange (JSE) at the end of each trading day. The initial results indicate that the approach can be scientifically used to predict future prices of grain commodities in a real-time environment.

Keywords: Big Data, Neural Networks, grain commodities prices, predictions, white maize
JEL Codes: Q11, Q13, C63, C89

Agrekon
Volume 55 | Number 4 | 2016
pp. 483–508

DOI: 10.1080/03031853.2016.1243060
Print ISSN 0303-1853 | Online 2078-0400

1. INTRODUCTION

The trading of grain commodities is coordinated in South Africa by the Johannesburg Stock Exchange (JSE). The implication of trading grain commodities on the stock exchange is that the grain commodities market in South Africa is *laissez faire* in nature. In essence, this indicates that the market, and effectively the prices of grain commodities are influenced by several local and international economic, political and social factors that are rapidly changing. Therefore, stakeholders in the grain commodities market are constantly exposed to price-related risks due to the volatility of prices of grain commodities (Venter, Strydom and Grové, 2013). The volatility of prices of grain commodities and the associated price-related risks suggest that stakeholders, specifically grain farmers, will be confronted with important decisions when marketing their products.

The volatility in the prices of grain commodities and other agricultural products has been a source of concern for academic researchers as well as governmental and non-governmental organisations for many decades (Trostle, 2008; Wright, 2011). Previous studies have shown that many South African grain commodities farmers might be disadvantaged in the market because they do not have the required skills, knowledge and time to monitor and interpret several market indicators (Jordaan and Grové, 2010; Venter *et al*., 2013). This has been attributed to the complexities associated with determining the grain commodities market intelligence and future outlook (Jordaan, Grové, Jooste and Alemu, 2007; Venter *et al*., 2013).

In order to optimise income and reduce price risks, it is required that stakeholders in the industry sift through volumes of economic, political and social data that has to be sourced from various places (Wright, 2011; Trostle, 2008). Moreover, they are required to make sense out of the changes in this data as it relates to the grain commodities price on a regular basis (Mofokeng and Vink, 2013; Venter *et al*., 2013). This is essential for them to devise strategies for selling their produce in order to manage price-related risks and increase profitability (Venter *et al*., 2013).

Contextually, this could be described as the dilemma of the average grain commodities farmer who enjoys farming activities but is unable to get the best price for his/her produce. Within the value chain of the grain commodities production and trade in South Africa, the grain commodities farmer who is unable to get the best value for his/her produce seems to be absolute price-takers. In the long run, this can be seen as a threat to the sustainability of the operation of such farms, due to price-related risks faced annually.

The grain commodities farming sector contributes significantly to the South African economy through job creation, foreign exchange earnings and supply of raw materials to other industries. The income from the production of maize, soya beans and sunflower seed during the year 2014 was in excess of R36 billion, with maize production responsible for more than 75% of that income (DAFF, 2015). Perhaps even more important is the role grain commodities production plays in ensuring food security in South Africa. This is because maize meals are considered to be staple food for millions of people across the country. Thus, enabling the stakeholders, especially the grain commodities farmers, to make the right decisions when selling their commodities could make a significant socio-economic impact in the industry and the country. Therefore, a system that helps stakeholders, with limited skill and experience, in forecasting grain commodities prices so that they can make better decisions in managing their price risks and increase profitability will be beneficial to all stakeholders.

The factors that influence the grain commodities industry include several variables that affect grain prices (Trostle, 2008; Abbott, Hurt and Tyner, 2011; Wright, 2011; Venter et al., 2013; Khamis, Nabilah and Binti, 2014). These factors can be categorised as:

- Historical and recent market data;
- Domestic demand and supply;
- International demand and supply;
- Macroeconomics; and
- Political factors.

Recent developments regarding the concept of Big Data make it easier to gain access to datasets on several subjects or areas. Big Data has been described as a concept with the potential to influence all aspects of life, including work and play (McAfee and Brynjolfsson, 2012).

Big Data is based on the ability that now exists to collect a large volume of datasets compared with what was possible previously. Other characteristics that define Big Data include the wide variety of datasets that complement one another, the velocity at which data is created and the associated veracity (complexity, uncleanness and inaccuracy) of Big Data as a result of the heterogeneity and rate at which the data is created (Davenport and Patil, 2012; Mayer-Schonberger and Cukier, 2013). However, it could be erroneous to consider large datasets as Big Data just because of their volume (Goes, 2014). It is the combination of some of these characteristics that makes Big Data different, hence, requiring new thinking and approach for storing and processing data (Chen and Zhang, 2014). This uniqueness, compared with traditional data, is also what defines the opportunities to generate dynamic and real-time insights, support decision making, predict the future and facilitate organisational learning from Big Data.

The ability to collect and integrate datasets from several sources, open new opportunities in different fields of interest for the purpose of discovering useful patterns and extracting actionable insights. Big Data also enables new research opportunities to investigate relevant concepts and provide solutions to difficult challenges in different fields (Ayankoya, Calitz and Greyling, 2014). This has led to the evolution of several tools, techniques and technologies that make it possible to leverage large datasets for innovations both in research and practice (Chen and Zhang, 2014).

Predicting the prices of grain commodities will require the collection of the market data and data on the external factors that influence grain commodities prices. Large volumes of historical data are available, and can be used to understand historical relationships. However, there is also the need to collect data as events take place in order to be able to provide real-time intelligence and insight. This can be achieved by taking advantage of the availability of large datasets, together with new technologies, tools and the ability to incorporate all these into a real-time solution that provides a platform for better support for decision makers (Power, 2014). Having more data available in real-time or near real-time together with sufficient tools, techniques and technologies that can be used to extract insight from such data could provide valuable support to improved decision making. The financial markets have been a generator of large datasets for many years through millions of transactions processed daily. However, the availability of relevant datasets in real-time creates new opportunities to analyse data from such transactions as they take place, which offers improved decision making (Ruta, 2014). The implementation of Big Data concepts, tools and technologies makes it possible to capture, store and use such torrents of data in a stream as they are created (Chen and Zhang, 2014).

Neural Networks can be implemented together with Big Data and enabling environments for extracting actionable insights from large datasets. It is a branch of Artificial Intelligence that is able to learn complex patterns from data for the purpose of solving difficult problems and making decisions. The implementation of Neural Networks is founded on the biological research into the ability of the neural system of the human/animal brain to learn, recognise, store information, generalise and make decisions based on prior knowledge. Research on the application of Neural Networks for understanding complex time series data indicates that it is suitable for making predictions from patterns that can be found in historical time series data (Qi and Zhang, 2008; Crone and Kourentzes, 2010) there has been no general consensus on how to model the trends in time-series data.

It has been found that using Neural Networks for modelling and forecasting future time series observations is not limited by the constraints of statistical approaches such as seasonal trends and stationarity (Qi and Zhang, 2008). Moreover, Neural Networks are able to deal with complex patterns and significant changes in patterns that might occur in the time series because of the ability to use non-linear learning to detect changes and relationships that might exist in the data (Zhang, 2003; Qi and Zhang, 2008; Bukharov and Bogolyubov, 2015). Neural Networks are also considered to be better than statistical techniques in time series analysis because they are able to analyse and forecast qualitative and discrete data types (Bukharov and Bogolyubov, 2015). Therefore, comparative studies from different areas of application have found Neural Networks to be more efficient than time series analysis, which is based on statistical techniques (Co and Boosarawongse, 2007; Zou, Xia, Yang and Wang, 2007; Bennett, Stewart and Lu, 2014).

The main research problem identified in this study is that grain farmers do not utilise the large volume of datasets available for grain price future prediction in order to sell at an optimum price. Based on this problem, the objective of this paper is to develop an application that utilises diverse large datasets and neural networks to predict future grain commodity prices.

This study explores the use of Neural Network for predicting prices of grain commodities in South Africa. The spot prices of white maize will be used as an experimental case study. The article is structured as follows: Section 2 will discuss the factors that influence prices of grain commodities in South Africa. Section 3 will provide an overview of scientific grounding, tools, techniques and methodology used in this study. This is followed by the experimental results of an implementation of the suggestions in this paper in Section 4 and concluding remarks in Section 5.

2. FACTORS AFFECTING GRAIN COMMODITIES PRICES IN SOUTH AFRICA

Past trading activities on the grain commodities market are known to influence future trading and the prices of grain commodities (Jordaan *et al*., 2007; Wright, 2011). It is, therefore, important to consider local market transactions of the grain commodity of interest, as well as international trade, for the same commodity in countries where factors such as prices, demand and supply of grain commodities in such countries influence the prices of grain commodities in South Africa. Variables to consider should include trade data such as price, volume traded, bidding prices and so on, as provided by the stock exchanges. Several of the grain commodities can be used for the same purpose, therefore they are considered as substitutes and their economics are considered to be interdependent (Wright, 2011). Hence it will be important to include the effect of substitutes in studying the factors that affect grain commodities prices in South Africa. The following discussion provides an overview about the factors that affect grain commodities prices in South Africa.

2.1 Demand, supply and storage

Economic theories suggest that prices will go up when there is an increase in demand for any commodity, especially when the supply of such a commodity does not increase with demand (Burda and Wyplosz, 2009). In reverse, the price of commodities is forced down when there is over-production, reduced demand or a huge stockpile of commodities. This summarises the impact that the local and international utilisation of grain commodities for domestic and industrial purposes has on the grain commodities price.

Variables under this theme include factors that influence the ability of farmers to supply or those factors that cause over-supply and the calming or panic effect that the level of the grain stockpile has on the volatility of grain prices (Wright, 2011; Abbott *et al*., 2011; DAFF, 2014). The demand for grain commodities as an important source of calories for human consumption and industrial demand for animal feeds and biofuel also play a role (Wright, 2011; Trostle, 2008). The prominent variables under the demand, supply and storage theme that influence grain prices are as follows:

- Domestic utilisation;
- Industrial utilisation;
- Utilisation by major importing countries;
- Production level in major exporting countries;
- Influence of weather on production;
- Input costs;
- Local stockpile;
- International stockpile;
- Price, demand, supply and storage of substitutes; and
- Level of utilisation compared to stockpile (stock-to-use-ratio).

2.2 Macroeconomics

Macroeconomic factors have been identified as influencing the changes that occur in the price of grain commodities. Similar to the previous theme, there are several variables that influence grain prices falling under this theme. Studies show that some of the macroeconomic variables influence the prices because they are linked directly to the factors of production (Trostle, 2008). On the other hand, the influence of the other macroeconomic factors is simply a reflection of the state of the local or global economy (Abbott *et al*., 2011). Although there are suggestions that the use of macroeconomic variables for understanding grain commodities prices requires further research (Wright, 2011), it remains an important part of the discourse on the price of grain commodities (Abbott *et al*., 2011; DAFF, 2014; Wright, 2014; Trostle, 2008). The macroeconomic variables that influence the price of grain commodities, identified from the literature that has been cited above, include:

- Currency exchange rates (especially US dollar to other currencies);
- Price of crude oil; and

- Local interest rates.

The factors explored in Sections 2.1 and 2.2 provide insights into the factors that influence the grain commodities prices. These include variables for which data is generated and stored monthly, daily, hourly and by the minute in some cases.

3. METHODOLOGY

3.1 Backpropagation Neural Network for forecasting

Several modelling algorithms exist for the different Neural Network architectures for making predictions or classifications. Wilamowski (2009) alluded that the choice of algorithm should be based on the type and complexity of the problem for which a model is being trained. The Backpropagation Neural Network (BPNN), which is based on the feed forward Neural Network, has been found to be widely suitable for problems requiring prediction from data such as a time series (Khashei and Bijari, 2010; Evans, Pappas and Xhafa, 2013; Khamis *et al*., 2014).

BPNN follows the multi-layer learning networks system as shown in Figure 1, where there is an input layer composed of neurons (1 to n) representing the independent variable. A typical BPNN comprises one or more hidden layers with neurons (1 to j) that are weighted and they determine the degree of influence during the learning process; these hidden layers enable the network to use a non-linear function to model complex patterns (Alpaydin, 2010). Finally, the BPNN also contains an output layer with neurons (1 to m) representing the estimated variables (dependent variables) as shown in Figure 1. During the learning process, the BPNN sends a signal about errors from the output layer back to the hidden layer. The connections that exist between the neurons from each layer facilitate the learning process through the use of mathematical functions depending on the type of Neural Network and its set-up (Engelbrecht, 2007). This ensures that subsequent learning produces an output with lesser error value until an optimal output is discovered (Alpaydin, 2010).

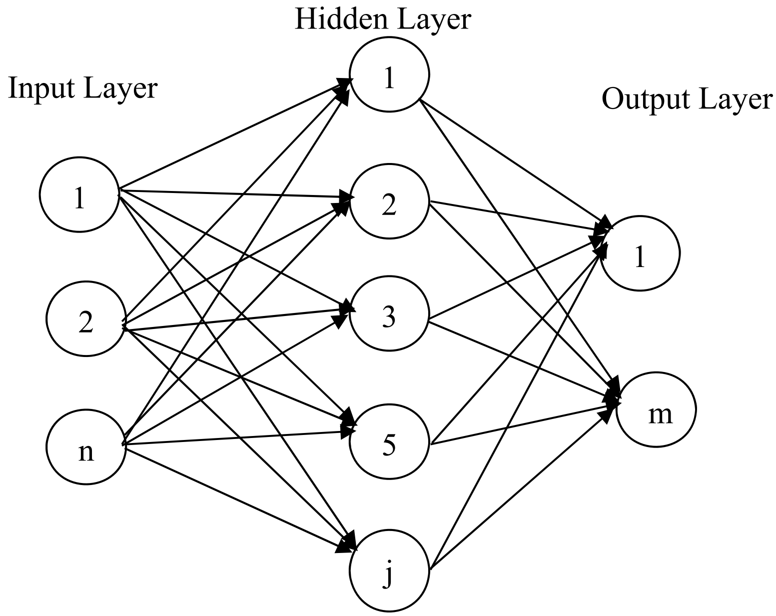

Figure 1: Simple Backpropagation Neural Networks

Previous studies have shown that BPNN is suitable for making predictions that are based on historical data even when they involve complex patterns (Ghwanmeh, Mohammad and Al-Ibrahim, 2013; Tsadiras, Papadopoulos and O'Kelly, 2013) especially in the rural areas where less support and care, due to lack of advanced heart diagnosis equipment. Also, physician intuition and experience are not always sufficient to achieve high quality medical procedures results. Therefore, medical errors and undesirable results are reasons for a need for unconventional computer-based diagnosis systems, which in turns reduce medical fatal errors, increasing the patient safety and save lives. The proposed solution, which is based on an Artificial Neural Networks (ANNs). Hence, many time series-related problems in areas such as financial forecasting, engineering and medical research have successfully implemented BPNN. Thus, this study adopts BPNN for the implementation of the Neural Network modelling and predicting prices of grain commodities.

Zhang (2003) and Qi and Zhang (2008) suggested that the relationship that exists between the input variables and the output variables in a feed-forward Neural Network can be represented mathematically as:

$$y_t = \propto_0 + \sum_{j=1}^{q} \propto_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}\right) + \varepsilon_t \quad (1)$$

where y_t is the expected output, α_j (j = 1,2, …, q) and β_i(i = 1,2. …, p) represent the weights for the connections between the neurons in the hidden layer and the output nodes; p represents the number of input nodes and the number of hidden nodes is represented by q in the equation. During the learning process, the transmission of information between the layers in the network is determined by the activation function (Lantz, 2013). Depending on the architecture of the network, transfer of signal from the neurons across the different layers could be weighted or not. Available activation functions include the linear function, sigmoid function, hyperbolic tangent function and Gaussian function. Generally, the selection of the appropriate activation function should depend on the type of problem at hand. The sigmoid function is often used in financial time-series and business-related problems (Wiles and Enke, 2014). For the purpose of this study, in-sample experiments were carried out to decide the choice of activation function. Using the available data, the resulting predictions using the hyperbolic and tangent and Gaussian function had large variances that were greater than R100/ton of white maize in most cases. On the other hand, when the same experiments were carried out using the sigmoid function, the recorded variances were as low as R1/ton in some cases. Therefore, the sigmoid function was adopted for this study. The sigmoid function is expressed as:

$$g(x) = \frac{1}{1 + \exp(-x)} \tag{1a}$$

The mathematical representation of the Neural Network denotes a non-linear autoregressive relationship that exists between the future value y_t and past observation (y_{t-1}, y_{t-2}, …, y_{t-p}) (Khashei and Bijari, 2011). Hence, the Neural Network model in equation (1) can be presented mathematically as:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \emptyset) + \varepsilon_t \tag{2}$$

Where f (.) denotes the Neural Network model and Ø is a vector of the parameter in equation (1). However, for a time series model where external variables are considered besides the internal autocorrelation function, the past observations of the external variables can be included in the model as:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, x_{1t-1}, x_{1t-2}, \dots, x_{1t-p}, \dots, x_{1t-1}, x_{1t-2}, \dots, x_{rt-p}, \emptyset) + \varepsilon_t \tag{3}$$

where x_{rt-1} denotes the observation for the external variable r collected during period t - 1. This has been included because the predicted outcome, y_t, is influenced by past observations of the same series, as well as by the past observations of the external variables. Hence, this implementation will consider the prices of previous trading days as input, as well as observations of the past trading days for the factors that influence the price of white maize.

By using the model represented in equation (3) to make predictions for a period $t + n$ in the time series, where is the current time and n is a positive integer, there is a need to make provision for the fact that data from the period between time t and $t + n$ will not exist. Hence, the Neural Networks model can be built to find the pattern between the independent variables at the current time t and the associated past observations, for predicting the future time for time $t + n$. Therefore the equation (3) can be written as:

$$y_{t+n} = f(y_t, y_{t-1}, \dots, y_{t-p}, x_{1t}, x_{1t-1}, \dots, x_{1t-p}, \dots, x_{1t}, x_{1t-1}, \dots, x_{rt-p}, \emptyset) + \varepsilon_t \quad (4)$$

This model can be used as the foundation for predicting future prices while taking into consideration the changes in the market dynamics. It also provides a basis for the retraining of the model to ensure that changes in the market dynamics are captured continuously. Thus, technological advancements that come with a Big Data environment such as in-memory, cloud and parallel computing can be leveraged by developing and retraining different models on how the combination of historical and real-time data influences different periods in the future. Hence, at the close of a business day, different models can be retrained based on the historical patterns that include the day's transactions to determine what will happen in the next 1, 2, 3 days and so forth.

3.1.1 Features selection for model

Selection of the right input variable that optimally captures and explains the patterns in a time series model is considered very crucial for the degree of accuracy of the model resulting from a Neural Network (Crone and Kourentzes, 2010; Qi and Zhang, 2008). The existing literature shows that deciding on the input variable for time series modelling, using Neural Networks, might be an art as much as a scientific expedition. Several authors conclude that there is generally no accepted theoretical background to follow in deciding the input variables in a Neural Network based time series modelling (Zou *et al*., 2007; Khashei and Bijari, 2011; Jabjone and Wannasang, 2014).

In developing a multivariate model, there is a need to consider not only the past observations of the variable being modelled but also to examine the influence of external variables as denoted in equation (3). Thus, in a multivariate analysis, the choice of external variables that will lead to an optimised model is crucial. Several studies on multivariate time series modelling used the analysis of correlation to support the choice of external variables (Yu and Ou, 2009; Khamis *et al*., 2014;

Jabjone and Wannasang, 2014). However, it is important to highlight that correlation analysis does not imply that these variables are, of a certainty, responsible for the patterns that exist in the price data (Irwin, Sanders and Merrin, 2009; Bukharov and Bogolyubov, 2015). Hence, the choice of external variables can also be supported by previous knowledge in the field of interest (Wiles and Enke, 2014).

In the case of univariate analysis, as well as multivariate time series analysis, after the external variables have been selected, there is still a need to decide how far back to go in including the effect of past observations in predicting future values. One of the major reasons for using the Neural Networks for time series analysis is to identify and capture non-linear relationships that might be in the dataset (Qi and Zhang, 2008; Bukharov and Bogolyubov, 2015). However, there is empirical and theoretical evidence that complex time series data with non-linearity patterns can also possess some linear characteristics (Khashei and Bijari, 2011). Qi and Zhang (2008) supported the use of techniques such as lagging to include the linear effect of past observations in the Neural Network-based time series data. There are no generally acceptable foundations for selecting the lag length for a Neural Network based time series modelling. However, similar work has made use of random experiments to determine the lag length that produced the best model (Zou *et al*., 2007; Khashei and Bijari, 2011).

The total number of input variables for the Neural Networks will consist of the dependent and independent variables that have been selected. These will also include the lagged variables for each of the selected variables.

3.1.2 Additional Neural Network parameters

In addition to other parameters that are required in setting up a BPNN the number of hidden layers, learning rate, momentum factor and the activation function are included. The learning rate, η, determines the number of steps that is taken in the search for the output. If the chosen learning rate is too large, the optimum can be missed, and when it is too small, the network can take too long to train (Engelbrecht, 2007). The momentum factor, α, determines the degree of influence that weights of previous learning will have on the current learning. It allows the training process to use the identified weights of the previous learning iteration so that the weights of the past iteration are introduced as inertia into the current learning iteration (Larose, 2005). The momentum factor ranges from 0 to 1, meaning that when the momentum term is close to or equal to one, the weight of the current iteration will be essentially the same as the previous one.

The learning process in Neural Networks is iterative; therefore, the number of iterations for the learning process should be set from the beginning as an exit criterion for the network. However, depending on the selected learning rate or the momentum rate, it could take much longer to achieve the set number of iterations. In such cases, a target error level can also be set for which the learning process will be terminated when it is achieved (Larose, 2005). The selection of the optimum learning rate, momentum term and exit criterion is a balancing act considering the implication that each of the parameters has on performance of the network. The use of experiments is also suggested as the approach for choosing the other network topology parameters such as the hidden layer, learning rates and momentum factor (Tsadiras *et al*., 2013; Ghwanmeh *et al*., 2013; Khamis *et al*., 2014)but this can be difficult, especially for large production lines, because the task is currently highly time consuming. Designers would be interested in a tool that would rapidly provide the solution to the BAP, even if only a near optimal solution is found, especially when they have to make their decisions at an operational level (e.g. hours).

Besides the mentioned network parameters, it is suggested that the input data for Neural Network modelling be pre-processed into a normalised format ranging between -1 and 1 or 0 and 1 (Engelbrecht, 2007; Khamis *et al*., 2014). Transforming the input data into a range of 0 to 1 can be achieved by using the equation:

$$y_t = \frac{y_t - y_{min}}{y_{max} - y_{min}} \tag{5}$$

where y_t is an observation for time t, y_{min} and y_{max} are the minimum and the maximum observed values of all the observations of a given variable.

It is further essential to determine how the network will initialise the weights for the neurons in the network during each learning process. Engelbrecht (2007) suggested the use of random numbers close to zero as weights for each neuron to ensure that learning takes place in the network.

3.2 Model evaluation

The purpose of identifying the correct parameters and topology for training an optimal Neural Network for different problems is so that the resulting model can adequately be used to estimate future occurrences based on historical data. However, care needs to be taken to ensure that the output from a model is not a result of just memorising the historical data (Provost and Fawcett, 2013a). In such cases, the model will have high accuracy when used to forecast a subset of the data used in training the model, but will not produce a reasonable result when used to predict a dataset not seen by the model during training. Hence the model has not learned the patterns in the data, but it has only memorised the observations in the data. This problem is regarded as overfitting (O'Neil and Schutt, 2014).

Contrary to just memorising the observations in a training dataset, the desired model from a modelling exercise is one that is able to accurately estimate future outcomes based on input data that has not been seen by the training model at all. This is regarded as generalisation (Provost and Fawcett, 2013b). The level of accuracy of a model can be measured by its ability to generalise even when there is a significant change in the input data. There is a risk of overfitting a model to the training data when the model becomes too complex, such as having too many hidden nodes or too few observations compared with the number of input nodes (Alpaydin, 2010). However, the ability of the model is reduced greatly with an overly simple network (Co and Boosarawongse, 2007). Hence, there is a need to strike a balance between generalisation and overfitting.

A common practice for avoiding overfitting is to split the available dataset into a training and a test set (O'Neil and Schutt, 2014; Provost and Fawcett, 2013b; Alpaydin, 2010), where the test set is kept completely separate and not used in the training process. The performance of the model is then checked by using the model to forecast the series in the test set and to compare the results with the actual data. Statistical measures such as the Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) provide quantitative measures for comparing the results of predictions from the training set and the test set. MSE is a modelling evaluation statistic that gives an indication of how much a set of values that has been predicted by using a model varies from the actual observations (O'Neil and Schutt, 2014). It represents the loss function between the result of a trained model when compared with the actual value, hence it is regarded as the training error for Neural Networks (Wilamowski, 2009; O'Neil and Schutt, 2014). The MSE is defined as:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2$$

where F_t represents the predicted values, Y_t the observed actual values and n the total number of values. However, a more popular measure of the accuracy of a model is the Root Mean Squared Error that is obtained by taking the square root of MSE (Khashei and Bijari, 2011; Bennett *et al*., 2014). RMSE is represented as:

$$RMSE = \sqrt{MSE}$$

The Mean Absolute Percentage Error (MAPE) is another measurement of accuracy of a predictive model which presents the predictions error as a percentage of the actual observed values. It calculates the absolute value of the ratio of the error to actual values (Tofallis, 2015), and calculates it as a percentage by multiplying it by 100. MAPE is obtained as:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t - F_t}{Y_t}\right|$$

These statistics are generally used for measuring model accuracy in time series forecasting (Enders, 2010; Tsay, 2010) and have also been adopted in measuring the accuracy of a Neural Network-based time series model as well (Zou *et al*., 2007; Crone and Kourentzes, 2010; Khashei and Bijari, 2011; Khamis *et al*., 2014) artificial neural network and the linear combination models for forecasting wheat price in Chinese market. Empirical results show that the combined model can improve the forecasting performance significantly in contrast with its counterparts in terms of the error evaluation measurements. However, as far as turning points and profit criterions are concerned, the ANN model is best as well as at capturing a significant number of turning points. The results are conflicting when implementing dissimilar forecasting criteria (the quantitative and the turning points measurements).

4. EXPERIMENTAL RESULTS

The proposed approach was implemented by setting up experiments for predicting the spot price of white maize traded on the Johannesburg Stock Exchange. SAP HANA was adopted as the technology of choice to demonstrate the proposed approach because of its ability to support Big Data and advance analytics solutions (Chen and Zhang, 2014). It provides the required technology to handle real-time acquisition, pre-processing and predictive analytics of large datasets.

4.1 Implementation

For the purpose of this implementation, historical data on spot transactions on grain commodities was obtained from the website of the JSE with permission to use the data for research purposes. End-of-day data for spot prices was captured directly from the newsfeed provided on the website of the JSE while end-of-day data was captured from the website of a major grain commodities storage company (www.senwes.co.za). To include the influence of other markets outside South Africa, this implementation included the effect of the grain commodities market in the USA as a major producer of corn. Data on the corn trade in the USA was collected from the Chicago Board of Trade (CBOT) through third party data subscription. Moreover, the demand and supply data was collected from a service made available by the South African Grain Information Services (SAGIS) website. Data on the production and consumption in the USA was also collected through free services offered by the Economic Research Services (ERS) of the United States Department of Agriculture (USDA) on its website. Other sources of data include the websites of the Reserve Bank, where historical and the current interest rate in South Africa, as well as the daily South African Rand-US Dollar currency exchange rates data was collected. Finally, historical and current data on the prices of Brent crude oil was accessed through open data services on www.quandl.com.

The data collected was largely unstructured and in various formats. The Big Data tools and techniques, however, made it possible to clean and structure the datasets into an integrated time series data, by using the date as the integrating factor. Data for this experimental analysis was then extracted from the integrated repository of data. Historical data of the end-of-day spot price of white maize was taken from 2 January 2007 till 31 July 2015, resulting in a total of 2 149 observations together with the data on factors influencing the prices as independent variables.

For the purpose of this study, datasets from 1 January 2010 was considered for the modelling of spot prices of white maize. This is primarily to ensure that the knowledge base of the resulting model is based on trends from a reasonable past (Ruta, 2014). It is expected that the predicted spot prices of white maize might be different using the same BPNN topology but a different subset of the available data. However, the trend of the predicted prices should be the same, irrespective of the subset of the dataset used, if the selected dataset is a reasonable representation of the current trend in the market.

4.2 Model training

Correlation analysis was carried out between each of the dependent variables (spot prices of white maize) and the independent variables in order to decide the input variables for the training of the networks. This was carried out with the support of extracts from literature on the factors that influence grain commodities prices in South Africa. Table 1 presents a list of the variables that were selected for modelling the spot prices of white maize.

Table 1: Input variables for Neural Network model for WMAZ spot price

No	Variables	Correlation with spot price of WMAZ
1	Spot price of WMAZ (lagged)	
2	Spot price of Wheat	0.6280 (n=2149)
3	USD-Rand exchange rate	0.5885 (n=2149)
4	Spot price of Brent Crude oil	0.3191 (n=2149)
5	Prime interest rate in SA	-0.3428 (n=2149)
6	Price of Corn in USA	0.2860 (n=2149)
7	Volume of Corn Trade in USA	0.2848 (n=2149)
8	Demand for WMAZ in SA	0.2474 (n=2149)
9	Demand for Wheat in SA	0.3347 (n=2149)

Initial experiments were also carried out to determine the other network parameters that would produce the best model. The optimal model was identified to have a lag length of 5, signifying the inclusion of the effect of the previous 5 trading days in the model. Considering the number of variables and lag length, the model with 7 hidden layers were also found to perform better than the rest. The model was found to perform optimally with the learning rate set to 0.4 and the momentum factor set to 0.001. SAP HANA provides an option to initialise all the weight for all the input neurons to zeros, to use normally distributed weights or a randomly selected weights between 0 and 1. The latter option was selected for this study because of the expected non-linearity characteristics of the data and it is widely used for business-related applications (Engelbrecht, 2007; Wiles and Enke, 2014).

4.3 Cross-validation of model

A BPNN model for spot prices of white maize in South Africa was created. Subsequently, a validation process was also carried out to ensure that the model was able to generalise and not just to memorise the input data. The model was used to make predictions by using subsets of the training dataset and the testing dataset. Using a subset of the training data to make predictions is known as the in-sample evaluation while predicting with a dataset that is totally separate from the one used in training the model is known as the out-sample evaluation (O'Neil and Schutt, 2014). The BPNN model for the spot price of white maize in South Africa was trained by using historical data of transactions that happened between 1 January 2010 and 31 December 2014. In-sample evaluations were carried out with subsets of the training data while out-of-sample evaluations were carried out with the testing data. For both categories, the created model was used to make predictions for 1-month and 3-month periods.

The dataset for the trading days in the last month of the training data from 1 December 2014 to 31 December 2014 was predicted and compared with the actual prices. In-sample predictions were also made and compared with actual prices over a period of 3 months using data from 1 October 2014 till 31 December 2014. A comparison of the predicted and actual spot prices of white maize for in-sample as well as out-sample predictions (from 01 January 2015 to 31 January 2015) within a 1-month period are presented in Figure 2 and Figure 3, respectively. The graph in Figure 2 shows that the in-sample predictions are very close to the actual values and it also indicates that the in-sample predictions followed the trend of the actual prices quite closely. On the other hand, Figure 3 presents the result of the out-sample predictions. The predicted prices were also close to the actual prices, however, not as much as they are with the in-sample predictions. A comparison of the results of the in-sample and the out-sample predictions suggests that the model is able to generalise and the predictions are not a result of the model memorising the data.

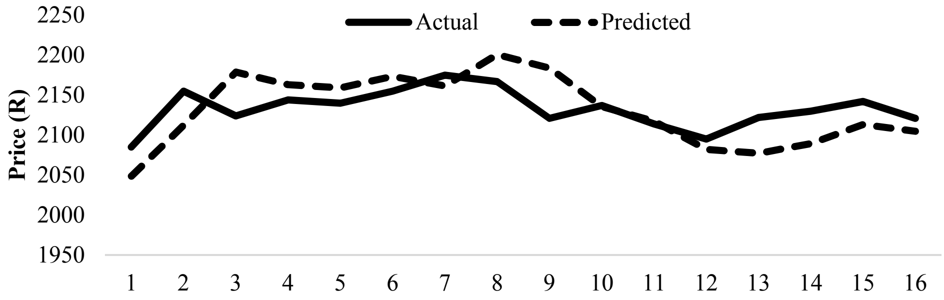

Figure 2: Comparison of actual vs predicted spot prices of white maize (1 month in-sample).

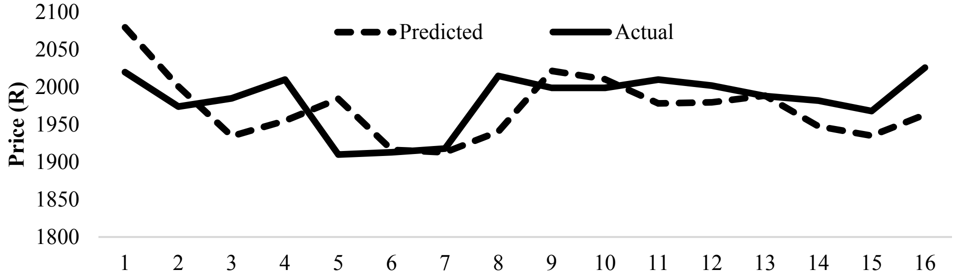

Figure 3: Comparison of actual vs predicted spot prices of white maize (1 month out-sample)

By using the same model, the spot prices of white maize over a period of 3 months were also predicted. Figures 4 and 5 show a comparison of the actual spot prices of white maize and the predicted prices over a 3-month period for the in-sample and out-sample predictions, respectively. The 3-month in-sample predictions were done over the period of 1 October 2014 to 31 December 2014 and the data for the spot prices of white maize over the period of 1 January 2015 to 31 March 2015 was separated for the out-sample comparison. The results show that the in-sample predictions over the 3-month period were very close to the actual prices of white maize. However, the accuracy of the out-sample predictions depreciated significantly from about the prediction for the 25th trading day as shown in Figure 5. One of the factors that might be responsible for this could be the need to identify and include data on other factors that influence the price of white maize in the medium to longer term.

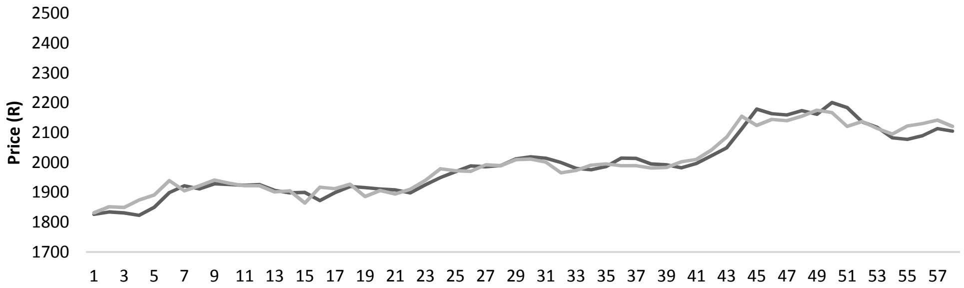

Figure 4: Comparison of actual vs predicted spot prices of white maize (3 months in-sample).

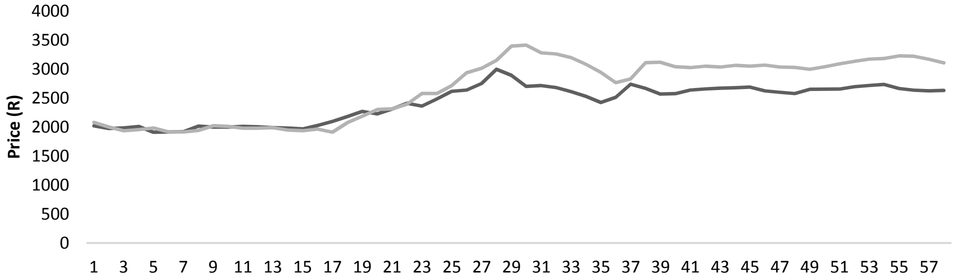

Figure 5: Comparison of actual vs predicted spot prices of white maize (3 months out-sample).

To measure the prediction accuracy of the model, the Mean Absolute Percentage Error (MAPE) statistic for the in-sample and out-sample predictions for the two different periods were compared. The results, as shown in Table 2, indicate that the MAPE of the in-sample predictions (1.31%) and that of the out-sample predictions (2.26%) over a period of a single month were relatively close. However, for the predictions over a 3-month period, the MAPE for the in-sample predictions was 0.97%, while that of the out-sample predictions was 9.20%. This signifies a noticeable difference when compared with the result obtained for predictions over a single month. However, the correlation between the predicted prices and the actual prices for the 3 months was 0.9709 and 0.9598 for the in-sample and out-sample predictions, respectively. This suggests that both the in-sample and the out-sample predictions over the 3-month period followed the trend of the actual spot price better than the predictions over 1-month period with 0.6568 and 0.1412 correlation for the in-sample and out-sample predictions, respectively. These results suggest that the model is able to generalise and make predictions for unseen data, although there is room for further research into improving the model.

Table 2: Summary of verification of BPNN model for spot prices

Period	In-sample			Out-sample		
	MAPE(%)	RMSE	R2	MAPE(%)	RMSE	R2
1 month	1.31	32.97	0.6568	2.26	61.02	0.1412
3 month	0.97	24.61	0.9709	9.20	348.64	0.9598

The predictions depicted by the graphs in Figures 2 to 5 show that the model is more accurate with in-sample predictions as expected, especially for predictions over 3 months. When the same model is applied for making out-sample predictions using the input dataset that was not used for the training process, the model was less accurate. However, the results of the out-sample predictions suggest that the model was intelligent enough to recognise the market trend, although, the deviation between the actual and the predicted price increased significantly with time. This result suggests that the identified BPNN topology and architecture could be used for predicting spot prices of white maize in South Africa. However, there is a need to implement strategies that will improve the accuracy of the predictions. This could include the use of other modelling techniques, supplementing the input datasets to include other contributing factors or additional pre-processing of the input datasets.

4.4 Real-time predictions

The cross-validation in Section 4.3 is based on the assumptions that the external data is available for the period for which the price of white maize is being predicted. However, as proposed in Section 2.2 with the model denoted as equation (4), a model can be built based on all the available data for predicting the spot and futures contract prices for different days into the future. This model can then be retrained periodically as new data becomes available to ensure that new market dynamics are captured in the Neural Network.

Based on the suggestions of Ruta (2014) on the use of Big Data for real-time learning for financial assets trading, new BPNN algorithms for building models for 14 trading days ahead were written. Each of the models was run continuously until 10 different predictions were recorded for each day. Thereafter, the mean value of the 10 predictions captured for each day was taken as the final prediction.

Table 3: Tables showing the input datasets used for modelling

Training		**Prediction**		
Start	**End**	**Start**	**End**	**Results for**
2010-01-01	2015-07-15	2015-07-16	2015-07-31	2015-08-03
2010-01-01	2015-07-18	2015-07-19	2015-08-03	2015-08-04
2010-01-01	2015-07-19	2015-07-20	2015-08-04	2015-08-05
2010-01-01	2015-07-20	2015-07-21	2015-08-05	2015-08-06
2010-01-01	2015-07-21	2015-07-22	2015-08-06	2015-08-07
2010-01-01	2015-07-22	2015-07-23	2015-08-07	2015-08-10
2010-01-01	2015-07-25	2015-07-26	2015-08-10	2015-08-11
2010-01-01	2015-07-26	2015-07-27	2015-08-11	2015-08-12
2010-01-01	2015-07-27	2015-07-28	2015-08-12	2015-08-13
2010-01-01	2015-07-28	2015-07-29	2015-08-13	2015-08-14
2010-01-01	2015-07-29	2015-07-30	2015-08-14	2015-08-17
2010-01-01	2015-08-01	2015-08-02	2015-08-17	2015-08-18
2010-01-01	2015-08-02	2015-08-03	2015-08-18	2015-08-19
2010-01-01	2015-08-03	2015-08-04	2015-08-19	2015-08-20

The experiments were set up to use a rolling subset of data as the input for the training and the predictions as shown in Table 3. The experiments made use of datasets between 01 January 2010 and 15 July 2015 as the training set for building the model for the first trading day in the month of August. New daily data was included in the input data for retraining the model at the end of each day. This was also applied to the input data for the predictions, by adding data from the previous trading day as shown in Table 3.

Besides the use of the measurement of accuracy statistics to measure the technical abilities of the models, 8 expert grain commodities traders (referred to experts A – H) agreed to voluntarily participate in the evaluation exercise. The experts who agreed to participate are from 3 different companies listed on the Johannesburg Stock Exchange's website as registered to trade grain commodities in South Africa. Moreover, some of these trading companies also buy and sell grain commodities as financial assets on the Johannesburg Stock Exchange. The experts were asked to predict the future spot prices of white maize on the Johannesburg Stock Exchange for the month of August 2015 before the beginning of the month of August 2015.

The results in Table 4 indicate that the predictions from the BPNN model had lesser deviation from the actual spot prices than the predictions from all the experts. The measurement of accuracy statistics shows that the predictions by the BPNN model had the minimum error with the Mean Absolute Percentage Error (MAPE) = 1.44% and Root Mean Square Error (RMSE) = 49.91 when compared with the predictions of the experts. This is only followed by the predictions of Expert C with MAPE = 2.16% and RMSE = 85.78. Figure 6 provides a graphical representation of the results, showing that the price predicted by the BPNN model is about the closest to the actual prices recorded, although there is more room for improvements.

Table 4: Comparison between predictions from experts and implemented DSS for spot prices of white maize.

Day	Expert A	Expert B	Expert C	Expert D	Expert E	Expert F	Expert G	Expert H	BPNN	Actual
1	3045	2950	3165	3250	3250	3200	3150	3190	3161	3131
2	3058	2930	3140	3265	3200	3225	3148	3220	3162	3142
3	3035	2900	3120	3280	3150	3195	3155	3260	3094	3138
4	3021	2930	3000	3350	3080	3196	3160	3230	3093	3125
5	2985	2900	3130	3280	3060	3190	3170	3230	3114	3073
6	2985	2850	2980	3240	3040	3210	3190	3280	3075	3124
7	2912	2820	2982	3190	2980	3200	3200	3330	3075	3074
8	2875	2860	2985	3240	3000	3180	3250	3350	3043	3011
9	2901	2890	2960	3260	2970	3150	3240	3320	3059	2987
10	2915	2850	2940	3295	2940	3153	3230	3330	3013	2969
11	2874	2800	2950	3330	2910	3180	3200	3350	3008	2941
12	2877	2790	2940	3290	2870	3190	3205	3380	3014	2960
13	2908	2750	2900	3210	2890	3185	3200	3400	3000	3024
14	2945	2720	2880	3250	2860	3187	3190	3400	2966	3068
MAPE	3.46%	7.11%	2.16%	6.46%	2.26%	4.20%	4.27%	7.50%	1.44%	
RMSE	106.22	212.67	85.78	228.51	87.54	145.40	167.71	280.49	49.91	
R-squared	0.9099	0.5241	0.6454	-0.1554	0.7771	0.7457	-0.7851	-0.7141	0.7153	
	(n=14)	(n=14)	(n=14)	(n=14)	(n=14)	(n=14)	(n=14)	(n=14)	(n=14)	

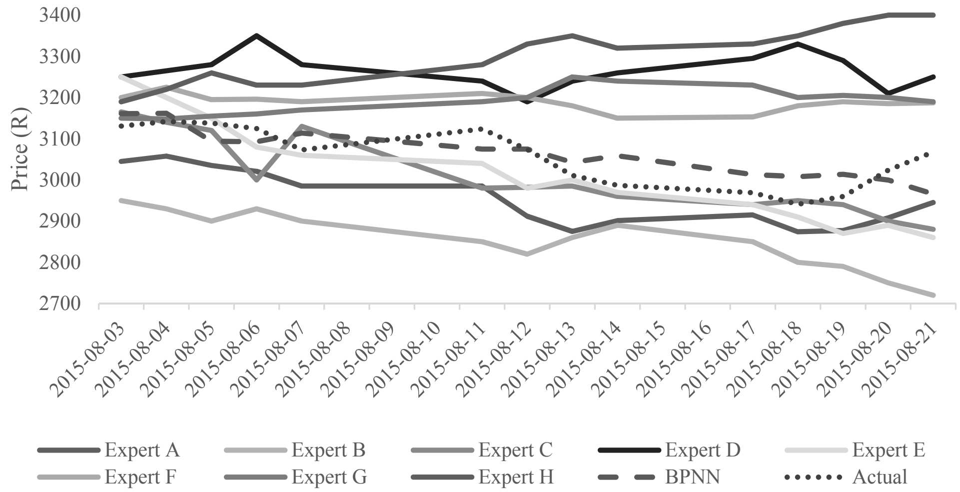

Figure 6: Prediction of spot prices of white maize by experts and BPNN model.

The calculation of MAPE on Table 4 indicates the difference between each predicted value and the actual value that was recorded as the percentage, showing the size of error between the predicted value and the actual value. On the other hand, the RMSE shows a measurement of how much the predicted prices deviates from the actual prices. Figures 7 presents a graphical view of the error statistics, which compares the performance of the BPNN model with predictions by experts. The graph shows that the BPNN model implemented performed relatively better with minimum deviation from the actual prices in terms of value.

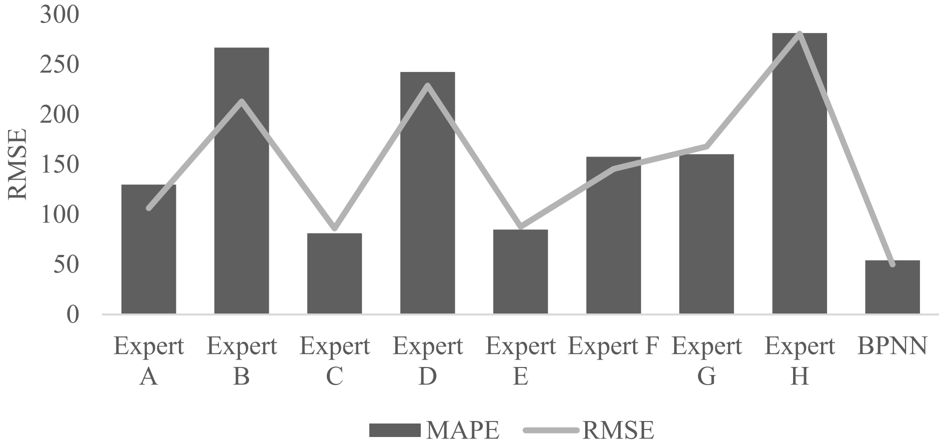

Figure 7: Error measurements of experts and BPNN model predictions for spot prices.

Both measurements of accuracy suggest that the predictions by the BPNN model performed better than the predictions made by the experts. The practical implication of this result is that the acquisition and analysis of data on factors that influence the grain commodities market in real-time present opportunities to create Decision Support Systems (DSS) for trading in grain commodities in South Africa. Such DSS can be used to assist stakeholders, such as the farmers, with limited skills and resources, in making decisions about trading their grain commodities.

5. CONCLUSIONS

This article sets out to demonstrate that grain commodities prices in South Africa can be predicted in real-time or near real-time by using Neural Networks and by taking advantage of the evolution in the concept of Big Data. It was identified that the grain commodities market data and data on the factors that influence the markets are available from different sources. Although the data is scattered in different locations and is often available in different formats, the tools and techniques of Big Data make it possible to source, acquire and integrate this data, even in real-time. Local demand and supply of grain commodities, international grain commodities markets, and macroeconomic indicators were identified as some of the factors that influence the grain commodities market in South Africa, that is, besides the influence of past grain commodities market transactions. However, it should be acknowledged that there might be other variables that could influence the price of grain commodities not identified by this study. Such variables can be added in the future to improve the outcome of the propositions in this study.

A Backpropagation Neural Network makes it possible to explore patterns in datasets regardless of the fact that these patterns might be linear or non-linear. It has also found its application in modelling time series problems in fields such as medical, econometrics and engineering. It was demonstrated in this article that the BPNN can be used to model and predict grain commodities prices. Furthermore, by using SAP HANA as a Big Data platform, it was demonstrated that with the acquisition of data in real-time or near real-time, a BPNN model can be retrained periodically as new data becomes available. This will ensure that changes in the market data are captured early enough and used in making predictions about future grain commodities prices. Empirical results in this study revealed that this approach could provide better predictions than those made by experienced grain commodities traders.

This study is part of a bigger research into how support for a decision can be provided for grain commodities trading in South Africa, especially for farmers with limited skills and resources for predicting grain commodities prices. The results presented provide improved support for short-term prediction of grain commodities prices using the Big Data and Neural Networks approach. Further studies will explore creating decision support, such as medium/long-term predictions, recommendations and discoveries that can be extracted from relevant datasets in real-time for trading grain commodities. Studies will also be carried out to explore how such market intelligence can be made available by using mobile technologies for easy access.

REFERENCES

Abbott, P.C., Hurt, C. and Tyner, W.E. 2011. What's driving food prices in 2011. *Farm Foundation Issue Report*.

Alpaydin, E. 2010. *Introduction to machine learning*. Cambridge, MA: The MIT Press.

Ayankoya, K., Calitz, A.P. and Greyling, J.H. 2014. Intrinsic relations between Data Science, Big Data, Business Analytics and Datafication. In *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference*, 192–198. Centurion: ACM Digital Library.

Bennett, C., Stewart, R.A. and Lu, J. 2014. Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies* 7(5): 2938–2960.

Bukharov, O.E. and Bogolyubov, D.P. 2015. Development of a decision support system based on neural networks and a genetic algorithm. *Expert Systems with Applications* 42: 6177–6183.

Burda, M. and Wyplosz, C. 2009. *Macroeconomics*. 5th ed. New York: Oxford University Press.

Chen, P.C.L. and Zhang, C.Y. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences* 275: 314–347.

Co, H.C. and Boosarawongse, R. 2007. Forecasting Thailand's rice export: Statistical techniques vs. artificial neural networks. *Computers and Industrial Engineering* 53(4): 610–627.

Crone, S.F. and Kourentzes, N. 2010. Feature selection for time series prediction: A combined filter and wrapper approach for neural networks. *Neurocomputing* 73: 1923–1936.

DAFF, 2014. *Trends in the agricultural sector 2013*. Department of Agriculture, Forestry and Fisheries.

DAFF, 2015. *Economic review of the South African agriculture*. Department of Agriculture, Forestry and Fisheries.

Davenport, T.H. and Patil, D.J. 2012. Data scientist: The sexiest job of the 21st century. *Harvard Business Review*: 70–77.

Enders, W. 2010. *Applied econometric time series*. Hoboken, NJ: John Wiley.

Engelbrecht, A.P. 2007. *Computational intelligence: An introduction*. West Sussex: John Wiley.

Evans, C., Pappas, K. and Xhafa, F. 2013. Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation. *Mathematical and Computer Modelling* 58: 1249–1266.

Ghwanmeh, S., Mohammad, A. and Al-Ibrahim, A. 2013. Innovative artificial neural networks-based decision support system for heart diseases diagnosis. *Journal of Intelligent Learning Systems and Applications* 5: 176–183.

Goes, P.B. 2014. Big Data and IS research. *MIS Quarterly* 38(3): 3–8.

Irwin, S.H., Sanders, D.R. and Merrin, R.P. 2009. Devil or angel? The role of speculation in the recent commodity price boom (and bust). *Journal of Agricultural and Applied Economics* 41(2): 377–391.

Jabjone, S. and Wannasang, S. 2014. Decision support system using artificial neural network to predict rice production in Phimai district, Thailand. *International Journal of Computer and Electrical Engineering* 6(2): 162–166.

Jordaan, H., Grové, B., Jooste, A. and Alemu, Z.G. 2007. Measuring the price volatility of certain field crops in South Africa using the ARCH/GARCH approach. *Agrekon* 46(3): 306–322.

Jordaan, H. and Grové, B. 2010. Factors affecting forward pricing behaviour: Implications of alternative regression model specifications. *South African Journal of Economic and Management Sciences* 13(2): 113–122.

Khamis, A., Nabilah, S. and Binti, S. 2014. Forecasting wheat price using Backpropagation and NARX Neural Network. *The International Journal of Engineering and Science* 3(11): 19–26.

Khashei, M. and Bijari, M. 2010. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications* 37(1): 479–489.

Khashei, M. and Bijari, M. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing Journal* 11(2): 2664–2675.

Lantz, B. 2013. *Machine Learning with R*. Birmingham: Packt Publishing.

Larose, D.T. 2005. *Discovering knowledge in data: An introduction to data mining*. 2nd ed. Hoboken, NJ: John Wiley.

Mayer-Schonberger, V. and Cukier, K. 2013. *Big Data*. London: John Murray.

McAfee, A. and Brynjolfsson, E. 2012. Big Data: The management revolution. *Harvard Business Review* 90(10): 61–68.

Mofokeng, M. and Vink, N. 2013. Factors affecting the hedging decision of maize farmers in Gauteng province. In *Fouth International Conference of the African Association of Agricultural Economists*, Hammemet.

O'Neil, C. and Schutt, R. 2014. *Doing Data Science*. Sebastopol: O'Reilly Media.

Power, D.J. 2014. Using 'Big Data' for analytics and decision support. *Journal of Decision Systems* 23(2): 222–228.

Provost, F. and Fawcett, T. 2013a. Data Science and its relationship to Big Data and data-driven decision making. *Big Data* 1(1): 51–59.

Provost, F. and Fawcett, T. 2013b. *Data Science for business*. Sebastopol: O'Reilly Media.

Qi, M. and Zhang, G.P. 2008. Trend time series modeling and forecasting with neural networks. *IEEE Transactions on Neural Networks* 19(5): 808–816.

Ruta, D. 2014. Automated trading with machine learning on Big Data. In: *2014 IEEE International Congress on Big Data*. Anchorage: IEEE.

Tofallis, C. 2015. A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operation Research Society* 66(8): 1352–1362.

Trostle, R. 2008. Global agricultural supply and demand: Factors contributing to the recent increase in food commodity prices. Available at http://www.ers.usda.gov/publications/wrs0801/ (accessed 15 March 2015).

Tsadiras, A.K. Papadopoulos, C.T. and O'Kelly, M.E.J. 2013. An artificial neural network based decision support system for solving the buffer allocation problem in reliable production lines. *Computers & Industrial Engineering* 66(4): 1150–1162.

Tsay, R.S. 2010. *Analysis of financial time series*. Hoboken, NJ: John Wiley.

Venter, M.M., Strydom, D.B. and Grové, B. 2013. Stochastic efficiency analysis of alternative basic grain marketing strategies. *Agrekon* 52: 46–63.

Wilamowski, B.M. 2009. Neural Network architectures and learning algorithms. *IEEE Industrial Electronic Magazine*: 56–63.

Wiles, P.S. and Enke, D. 2014. Nonlinear modeling using Neural Networks for trading the soybean complex. *Procedia Computer Science* 36: 234–239.

Wright, B.D. 2011. The economics of grain price volatility. *Applied Economic Perspectives and Policy* 33(1): 32–58.

Wright, B.D. 2014. Data at our fingertips, myths in our minds: Recent grain price jumps as the perfect storm. *Australian Journal of Agricultural and Resources Economics* 58: 538–553.

Yu, S. and Ou, J. 2009. Forecasting model of agricultural products prices in wholesale markets based on combined BP neural network-time series model. In: *International Conference on Information Management, Innovation Management and Industrial Engineering*, 558–561, Xian.

Zhang, G.P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50: 159–175.

Zou, H.F., Xia, G.P., Yang, F.T. and Wang, H.Y. 2007. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing* 70: 2913–2923.