Restricted Boltzmann Machines for the Prediction of Trends in Financial Time Series

Conference Paper · July 2018

DOI: 10.1109/IJCNN.2018.8489163

Carlos Assis
Centro Federal de Educação Tecnológica de Minas Gerais

Adriano C. M. Pereira
Federal University of Minas Gerais

Eduardo G. Carrano
Federal University of Minas Gerais

Wanderson Ferreira Dias
Federal University of Minas Gerais

Restricted Boltzmann Machines for the Prediction of Trends in Financial Time Series

Carlos A. S. Assis
PPGMMC
Centro Federal de Educação Tecnológica de Minas Gerais
Belo Horizonte – MG, Brazil
Email: carlosassis@dppg.cefetmg.br

Adriano C. M. Pereira
DCC
Universidade Federal de Minas Gerais
Belo Horizonte – MG, Brazil
Email: adrianoc@dcc.ufmg.br

Eduardo G. Carrano
DEE
Universidade Federal de Minas Gerais
Belo Horizonte – MG, Brazil
Email: egcarrano@ufmg.br

Rafael Ramos
DEE
Universidade Federal de Minas Gerais
Belo Horizonte – MG, Brazil
Email: rramosc@lakeheadu.ca

Wanderson Dias
DCC
Universidade Federal de Minas Gerais
Belo Horizonte – MG, Brazil
Email: wandersonf.dias@gmail.com

***Abstract*—Nowadays, it is possible to note many machine learning techniques being applied to predict financial time series. However, recent studies indicate that the performance of such techniques can be strongly affected by data representation. In this manuscript we propose a combination of two machine learning algorithms to detect trends in stock market prices. In this approach, Boltzmann Restricted Machines are used as the latent feature extractor and Support Vector Machines work as the classifier. We performed tests with real data of five assets from the Brazilian Stock Market, BM&FBOVESPA. The results obtained with the proposed combination were better when compared to those ones reached by Support Vector Machines only. This suggests that the proposed approach can be suitable for the considered application.**

***Index Terms*—Restricted Boltzmann Machines. Machine Learning. Stock Market.**

I. Introduction

Predicting the future is certainly one of the greatest ambitions of human beings. There is no system that manages to accomplish such a task with perfection, but it is possible to find in the literature many attempts of forecasting some given quantity [1]. Considering that financial markets are environments that provide great opportunities, many studies have been developed in order to carry out predictions for this kind of application [2], [3]. Most studies focus on minimizing the prediction, which can support the task of identifying the best moment for buying or selling an asset. The training of prediction models is performed based on historical data.

Many researchers [4]–[8] have worked towards the improvement of existing predictors. These studies often rely on recent machine learning techniques, such as Support Vector Machines (SVM) [9], Artificial Neural Networks (ANNs) [10] and Genetic Programming (GP) [11]. However, it is well known by the machine learning community that the high dimensionality of inputs usually leads to the reduction of computational performance and precision on classification and regression applications [12]. This difficulty resulting from high-dimensional spaces is usually called curse of dimensionality [13].

The precision of classification algorithms can be improved by a prior selection of dataset features. Nevertheless, this step usually depends on an expert in the considered application. This manuscript proposes the use of a recent artificial neural network, called Restricted Boltzmann Machine (RBM) [14], to carry out the selection of features in financial time series. This approach of deep neural is used to perform dimension reduction and, consequently, improve data classification [15]. Finally, once the features are selected by RBM, a SVM is employed to predict trends for supporting investment decisions.

In face the numerous researches used in the modeling of time series over the last few years, it shows that the subject is of great importance and relevance, which motivates the development of this work.

The main contribution of this paper is to present the power of RBM focused to predict stock market asset trends in scenarios with high dimensional features (more than 150). Five real data sets of the BM&FBOVESPA were used to validate the study. Comparisons were also made between combined techniques and SVM by itself, with the proposed approach generally presenting better results.

The remaining of this text is organized as follows: Section II describes the problem and some correlated studies; Section III presents a description of the RBM technique used; Section IV describes the methodology that was applied; Section V presents the results obtained in five real financial time series. Finally, Section VI presents our conclusion.

II. Problem Definition & Related Work

The main goal of this study is to predict the variations of stock market asset prices. More specifically, historical data on price and volume are assessed, using technical analysis [16], with the purpose of predicting price changes.

This section presents some basic concepts of financial market and correlated studies.

A. Financial Market

The financial market is the place where people can negotiate (buy or sell) assets. The purpose of the market is to gather many sellers in a single place, which makes them easily reachable by interested buyers. The stock exchange is the negotiation environment in which investors may buy or sell titles through direct negotiation, with or without the support of negotiation correspondents. In the case of the Brazilian stock exchange, the negotiation is done through brokers [17][1]. In Brazil, the role of the stock exchange is represented by BM&FBOVESPA [18], which is the owner of two stock exchanges: BM&F, which is focused on the negotiation of agriculture/livestock products and financial instruments; and BOVESPA, which is focused on the negotiation of stocks and stock options.

In the stock market, the investor earns money by buying undervalued stocks and selling them at a higher value. The profit of the investment is determined by the difference between buying and selling prices, adding benefits and discounting transaction fees [17].

Usually, in order to predict if the value of a stock will increase or not, analysts employ one of two methods [19]:

- **Fundamental Analysis**: in this kind of analysis, the references for the investor are parameters that define the financial situation of the company, such as net profit, level of indebtedness and distribution of dividends, among others. In summary, the fundamental analysis assumes that stocks have an intrinsic value which would correspond to its fair price. This price, in turn, would be determined by the income stream measured for the stock and effectively distributed throughout a given period of time, discounting the present value.
- **Technical Analysis**: this analysis focuses on information regarding stock price and buying/selling movement in a given period of time. Only this data is used to estimate the trajectory or value of stock prices in a given future. Since this approach is applied in this study, it will receive major attention along the remaining of this section.

Investors who use the technical analysis seek to identify possible trends, since this technique assumes these trends follow a cyclical pattern [20]. These patterns have been translated into numerical or logical indicators to facilitate the automatic processing of time series. These technical analysis indicators are usually categorized as [17]:

- **Momentum Indicators**: they usually indicate the moments of buying and selling in the market (e.g., Relative Strength Index, Williams %R and Rate of Change).
- **Trend Indicators**: they indicate market direction – upward or downward (e.g., Moving Averages, Average Directional Index, Moving Average Convergence / Divergence).
- **Volatility Indicators**: they show whether prices are too volatile, i.e. without a defined trend (e.g., Bollinger Bands, Price Channel, Average True Range).
- **Volume Indicators**: they are based on the fact that volume usually precedes price movements (e.g., On Balance Volume, Volume Oscillator, Chaikin Oscillator).

Considering the innate complexity and dynamism of financial market predictions, there is a constant debate regarding the possibility of predicting stock price changes. According to [21], the traditional analysis methods (technical and fundamental) are not capable of identifying the non-linear relations between the many variables that comprise the price of a stock and its movement upwards or downwards, leading to the need of using more advanced techniques.

B. Related Studies

Considering the scenario of uncertainties of the stock market, many studies have been developed to help on trend predictions. Nelson and Pereira [22] applied Long Short-Term Memory (LSTM) neural networks for predicting price trends. A prediction model was created and a series of experiments were carried out using assets of BM&FBOVESPA. The results found were considered satisfactory, with an average accuracy of up to 55.9% in predicting upward trends in the immediate future. The model was also assessed under financial perspectives, showing promising results regarding return.

Liang et al. [23] proposed to incorporate RBM and several classifiers to predict short-term stock market trend. Eleven technical indicators was firstly inferred by using trading data, e.g., close price, lowest price, open price and highest price. Afterwards, these technical indicators was conveyed to binary values by using a trend deterministic preparation layer and applyed a RBM to extract features from binary valued features from the last step. The experimental study demonstrated this model's effectiveness compared with several traditional methods.

Soto et al. [24] constructed of intelligent hybrid architectures and the optimization of the fuzzy integrators for time series prediction; interval type-2 fuzzy neural networks (IT2FNN). IT2FNN used hybrid learning algorithm techniques (gradient descent backpropagation and gradient descent with adaptive learning rate backpropagation). Was used interval type-2 and type-1 fuzzy systems to integrate the output (forecast) of each Ensemble of ANFIS models. Particle Swarm Optimization (PSO) was used for the optimization of membership functions (MFs) parameters of the fuzzy integrators. The Mackey-Glass time series was used to test of performance of the proposed architecture. Simulation results showed the effectiveness of the proposed approach.

[1]Financial institutions that intermediate the negotiations.

Franco and Steiner [25] compared Multi Layer Perceptron (MLP), Radial Basis Function (RBF), and Layer Recurrent Network (LRN) neural networks to predict the future value of certain stocks in BM&FBOVESPA. A total of 496 closing prices in reverse auction for four assets were used in the data set in the period ranging from February 27th, 2012, to February 25th, 2014. The accuracy measure used for validation was the Mean Squared Error (MSE) between the values predicted by the neural networks and the real values. The best prediction technique was that of LRN, with error values of the order of 10^{-11}.

Zhu et al. [26] implemented a decision-making support system for buying and selling assets. This system uses Deep Belief Networks (DBN) to predict stock prices. The experiment covered a set of 400 stocks from the *S&P 500*. The data set comprised 12 financial indicators. The authors stated that the proposed system was capable of predicting stock prices and attaining high financial performance. However, they showed that DBNs require a lot of time to be trained with historical data. For that reason, speed was an obstacle to the system.

Lawrence and Lee [27] developed an algorithm based on RBM to extract latent features of Nasdaq assets. The authors used a data set of the period between 1990 and 2009. The results showed that the use of RBMs enabled the reduction of input data dimensionality and the extraction of important features to support future price prediction.

Based on survey [28], it is possible to observe that many studies in the literature address the theme of predicting trends in financial series but few of them use RBM. In addition, we could not find studies combining SVM, RBM, and technical analysis. Finally, none of the RBM related studies were applied to BM&FBOVESPA.

The next section presents a theoretical background to this study.

III. Theoretical Background

This section addresses the main concepts required for understanding the proposed tool.

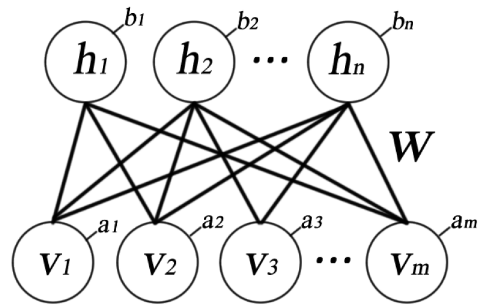
A. Restricted Boltzmann Machines (RBM)

The Restricted Boltzmann Machines [14] are unsupervised learning neural networks. They are mainly characterized by their ability to learn internal representations and to solve complex combinatorial problems [29].

In general, the RBM is a stochastic network whose structure is composed of a visible layer and a hidden layer. The visible layer represents the observed data and is connected to the hidden layer. This second one is responsible to learn how to extract relevant features from the input data. Originally, the RBM was proposed for handling binary data. Further, Hinton and Salakhutdinov [14] proposed a Gaussian-Bernoulli RBM (CRBM) for dealing with other data types. This study describes the basic concepts of the CRBM approach, considering we dealing with continuous inputs.

In RBM, the connections between neurons are bidirectional and symmetrical, which means that there is information traffic in both directions of the network. Besides, in order to simplify inference procedures, neurons of the same layer are not connected between themselves. Figure 1 shows a RBM with m neurons in the visible layer $(v_1, \ldots, v_m)$, n neurons in the hidden layer $(h_1, \ldots, h_n)$, being $(a_1, \ldots, a_m)$ and $(b_1, \ldots, b_n)$ the bias vectors (b), and $\mathbf{W}$ the connection weight matrix. The tuple $(\mathbf{W}, \mathbf{a}, \mathbf{b})$ will be referred as $\boldsymbol{\theta}$.

Figure 1. RBM with 4 inputs and 3 hidden neurons.



The RBM is based on an energy-based probabilistic model. The joint probability distribution of the configuration (v,h) is achieved using Equations 1 and 2:

$$p(\mathbf{v},\mathbf{h};\boldsymbol{\theta}) = \frac{e^{-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta})}}{\sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta})}} \tag{1}$$

$$E(\mathbf{v},\mathbf{h};\boldsymbol{\theta}) = -\sum_{i=1}^{m} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n} b_j h_j - \sum_{i,j=1}^{m,n} \frac{v_i}{\sigma^2} h_j w_{ij} \tag{2}$$

The probability that is assigned to a visible vector $\mathbf{v}$ is given by the sum of all the probabilities of the hidden vectors $\mathbf{h}$, calculated by Equation 3:

$$p(\mathbf{v};\boldsymbol{\theta}) = \frac{\sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta})}}{\sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta})}} \tag{3}$$

Since the RBM is restricted, the probability distributions of $\mathbf{h}$ given $\mathbf{v}$ and of $\mathbf{v}$ given $\mathbf{h}$ are evaluated as shown in Equations 4 and 5:

$$p(\mathbf{h}|\mathbf{v};\boldsymbol{\theta}) = \prod_{j=1,\ldots,n} p(h_j|\mathbf{v}) \tag{4}$$

$$p(\mathbf{v}|\mathbf{h};\boldsymbol{\theta}) = \prod_{i=1,\ldots,m} p(v_i|\mathbf{h}) \tag{5}$$

In the CRBM [14] the visible layer is continuous and the hidden layer is binary. In this direction, the conditional distributions are described by Equations 6 and 7:

$$p(h_j = 1|\mathbf{v};\boldsymbol{\theta}) = \phi(b_j + \sum_{i=1}^{m} v_i w_{ij}) \tag{6}$$

$$p(v_i = v|\mathbf{h};\boldsymbol{\theta}) = N(v|a_i + \sum_{j=1}^{n} h_j w_{ij}, \sigma^2) \tag{7}$$

in which $\phi(x) = \frac{1}{1+e^{-x}}$ (logistic function) and N is a normal distribution, with mean v and standard deviation σ^2.

The purpose of the RBM is to estimate the values of the components of vector $\boldsymbol{\theta}$ that decrease the energy level of the network. Since $p(\mathbf{v};\boldsymbol{\theta})$ is the input data distribution, $\boldsymbol{\theta}$ can be estimated by the maximization of $p(\mathbf{v},\boldsymbol{\theta})$ or, in an equivalent manner, log $p(\mathbf{v},\boldsymbol{\theta})$. Therefore, the descending gradient of log $p(\mathbf{v},\boldsymbol{\theta})$ regarding $\boldsymbol{\theta}$ is calculated as shown in Equation 8.

$$\frac{\partial p(\mathbf{v},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \langle v_i h_j \rangle_d - \langle v_i h_j \rangle_m \tag{8}$$

in which the components $\langle v_i h_j \rangle_d$ and $\langle v_i h_j \rangle_m$ are used to represent the computed expectations about the data and the model, respectively.

The term $\langle v_i h_j \rangle_d$ is estimated very simply, by means of the conditional probabilities $p(h_j = 1|\mathbf{v};\boldsymbol{\theta})$ and $p(v_i = v|\mathbf{h};\boldsymbol{\theta})$. However, estimating of $\langle v_i h_j \rangle_m$ is much harder. This can be done by means of Gibbs sampling [30], in which random data is used to feed the visible layer. However, this procedure may take a long time to reach an adequate result. Fortunately, a quicker procedure, so-called Contrastive Divergence (CD), was proposed in [31]. The idea behind this method is to feed the visible layer with training data and execute Gibbs sampling only once, in a phase the authors call reconstruction.

For the application of the CD algorithm, the first step is to match the visible layer $\mathbf{v_0}$ to the input data and, hereafter, estimate the hidden layer $\mathbf{h_0}$ using the conditional probability $p(h_j = 1|\mathbf{v};\boldsymbol{\theta})$, being $\langle \mathbf{v}\mathbf{h}^{\mathbf{T}} \rangle_d = \mathbf{v_0}\mathbf{h_0^T}$. Then, based on $\mathbf{h_0}$, $\mathbf{v_1}$ should be estimated using the conditional probability $p(v_i = v|\mathbf{h};\boldsymbol{\theta})$. Similarly, based on $\mathbf{v_1}$, $\mathbf{h_1}$ is estimated, again by $p(h_j = 1|\mathbf{v};\boldsymbol{\theta})$, being $\langle \mathbf{v}\mathbf{h}^{\mathbf{T}} \rangle_m = \mathbf{v_1}\mathbf{h_1^T}$. Finally, the set of parameters $\boldsymbol{\theta}$ are updated as follows:

$$\begin{gathered}
\mathbf{W^{t+1}} = \mathbf{W^t} + \Delta\mathbf{W^t} \rightarrow \Delta\mathbf{W^t} \\
= \eta(\mathbf{v_0}\mathbf{h_0^T}\mathbf{v_1}\mathbf{h_1^T}) - \rho\mathbf{W^t} + \alpha\Delta\mathbf{W^{t-1}} \\
\mathbf{a^{t+1}} = \mathbf{a^t} + \Delta\mathbf{a^t} \rightarrow \Delta\mathbf{a^t} = \eta(\mathbf{v_0} - \mathbf{v_1}) + \alpha\Delta\mathbf{a^{t-1}} \\
\mathbf{b^{t+1}} = \mathbf{b^t} + \Delta\mathbf{b^t} \rightarrow \Delta\mathbf{b^t} = \eta(\mathbf{h_0} - \mathbf{h_1}) + \alpha\Delta\mathbf{b^{t-1}}
\end{gathered}$$

considering that $(\mathbf{W},\mathbf{a},\mathbf{b})$ are randomly initialized. The pseudocode of the CD algorithm is presented in Algorithm 1:

The parameters η, ρ, and α are known as learning rate, weight decay, and momentum, respectively. In [32], the authors suggest the values $\eta = 0.01$ and $\rho = [0.01, 0.0001]$. For α, they suggest $\alpha = 0.5$ for less than 5 iterations or $\alpha = 0.9$ otherwise.

The RBM has four hyper-parameters: the amount of neurons in the visible layer (v), the amount of neurons in the hidden layer (h), the learning rate (lr), and the number of cycles (ep). In one hand, if the learning rate is excessively low, network learning is also low. On the other hand, if it is excessively high, it cause oscillations in the training and can preclude the convergence of the learning process. Usually, lr value varies from 0.1 to 1.0. The number of cycles is the number of times in which the training set is presented to the network. An excessive number of cycles can cause the network to lose its generalization capacity (overfitting). However, if ep is small, the network may not be capable of modeling the general behavior of the system (underfitting) [33].

Algorithm 1 Pseudocode of the algorithm: Contrastive Divergence

procedure CD
    Prepare set of input data
    Inform the number of neurons for hidden layer $\mathbf{h}$
    Initialize parameters η, ρ and α
    Randomly initialize $\boldsymbol{\theta}$
    while number of iterations or minimum error satisfied **do**
        Match visible layer v_0 to input data;
        Estimate hidden layer h_0 using the condit. probability $p(\mathbf{h}|\mathbf{v})$)
        Estimate, based on h_0, the visible layer v_1 using the equation $p(\mathbf{v}|\mathbf{h})$
        Estimate, based on v_1, the hidden layer h_1 using the equation $p(\mathbf{h}|\mathbf{v})$
        Update $\boldsymbol{\theta}$ using the updating equations described above
    Return: $\boldsymbol{\theta}$ training

B. Support Vector Machines (SVM)

Support Vector Machines (SVM) are based on the theory of statistical learning, developed by Vapnik [9] based on studies initiated in [34]. This study establishes a series of principles that should be followed in order to obtain classifiers with a good generalization.The SVM machine learning algorithms have the purpose of determining decision limits that produce an optimal separation between classes through the minimization of errors. The SVMs stand out due to at least two characteristics: solid theoretical foundation and high performance in practical applications [35].

In its original proposal, SVMs are linear classifiers that separate data in two classes by means of a separating hyperplane. An optimal hyperplane separates data with the maximum margin possible, which is defined by the sum of the distances between the positive points and the negative points that are closer in the hyperplane. These points, called support vectors [36], are circled in Figure 2. The hyperplane is constructed based on prior training using a finite data set.

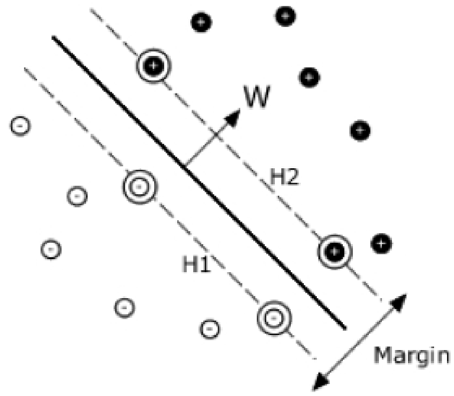
Assuming the training set $\{\mathbf{x}_i, y_i\},\ y_i \in \{-1, 1\}$, $x_i \in \mathbb{R}^n$, in which x_i is the ith input element and y_i is its respective class value for x_i, $i = 1,\ldots,l$. The evaluation of the hyperplane with optimal margin is given by the minimization of $\|w\|^2$, considering the following constraints:

$$\begin{aligned}
x_i w + b &\geq +1, \quad y_i = +1 \\
x_i w + b &\leq -1, \quad y_i = -1
\end{aligned}$$

in which w is the normal to the hyperplane. This is a quadratic optimization problem and may be converted to a dual problem, which depends only on the Lagrange multipliers α_i:

$$u = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \mathbf{x}_j)$$

Figure 2. Classification of a data set using a linear SVM.



according to the equality linear constraints:

$$\sum_{i}^{N} \alpha_i y_i = 0,$$

and the bound constraints:

$$\alpha_i \geq 0, \forall i$$

with the solution given by:

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

in which N is the number of training examples.

For most real problems, the data set is not separable through a linear hyperplane and the calculation of support vectors using the formulations described above would not be applicable [37].

This drawback can be surpassed by the introduction of margin expansion variables ξ_i, which relax the constraints of the linear SVM, allowing for some margin failures but also penalizing failures through the control variable C. The transformation of this optimization problem into its dual form only rely on changing the constraints to:

$$x_i w + b \geq +1 - \xi_i, \quad y_i = +1$$
$$x_i w + b \leq -1 + \xi_i, \quad y_i = -1$$

The SVM has some hyper-parameters to be chosen: kernel function, gamma, and cost. The kernel functions are responsible to provide a simple bridge between linear and non-linear algorithms. The cost parameter determines the balance between training errors and separation margins. Finally, the RBF kernel gamma parameter also controls the flexibility of the classifier [38].

IV. Proposed Approach

The approach proposed in this work comprises five steps: **[a]** extraction of historical data; **[b]** transformation; **[c]** dimensionality reduction and feature extraction; **[d]** classification; and **[e]** analysis of results. Figure 3 describes the proposed method. These steps are detailed in the following subsections.

A. Data Extraction

A historical data set of all assets of BM&FBOVESPA was extracted for the period between August 2014 and August 2015. This data was composed of daily candles[2]. Figure 4 shows the time series of the five assets at the period.

B. Transformation

Based on the values on candles, it is possible to assess the technical indicators. These indicators aim to support in the prediction of future market movements [16]. The assessment was made using a Java code, built by the authors, that communicates with the *API TA-Lib (Technical Analysis Library)*. This API is capable of generating more than 100 technical indicators for a candle set provided.

Although the technical indicators are essential for the proposed approach, it was possible to observe that the amount of indicators generated by the *API TA-Lib* expressively increases the dimensionality of the data. Therefore, it was necessary to apply an approach to reduce inputs.

C. Dimensionality Reduction

Two dimensionality reduction approaches for learning problems stand out in the literature: selection of features and extraction of features [40]. Selection, as the name implies, selects, according to a given criterion, the best subset within the original set of features. Extraction, in general terms, creates new features through transformations or combinations within the original set of features [40].

In this step, the RBM is used to perform feature selection and, consequently, reduce data dimensionality. The implementation used in this study was the CRBM from the library *Deep learning library for node.js*[3]. The tool was adapted in order to provide, in addition to the reduced set of indicators, a *label* indicating if the price of the asset increased or not. The label assignment is performed as follows:

$$label = \begin{cases} 1 & \text{if } closing_{i+1} > closing_i \\ 0 & \text{if } closing_{i+1} \leq closing_i \end{cases}$$

D. Classification

The classification is employed to identify the class of previously not inspected objects. Such a task is accomplished by a classification model, which is often built based on known data (the training set).

In this study, the Support Vector Machines (SVM) were used in this step. The implementation adopted was the *LibSVM*, from the library *Support Vector Machine for node.js*[4].

The case study used to evaluate this approach is presented in the next section.

[2]A candle represents the variation in the prices of a given asset in a given time unit (e.g., daily, weekly, monthly) [39]

[3]https://www.npmjs.com/package/dnn/

[4]https://www.npmjs.com/package/node-svm/

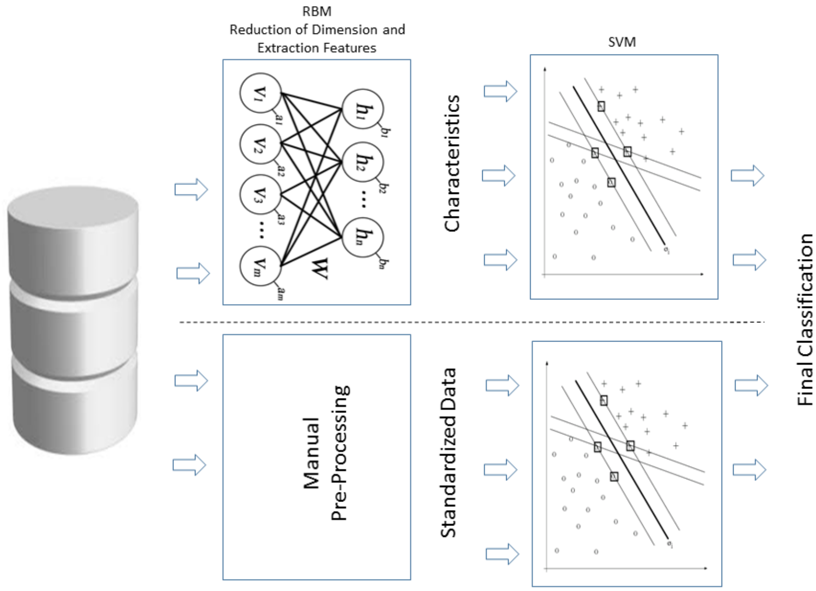


Figure 3. Proposed methodology

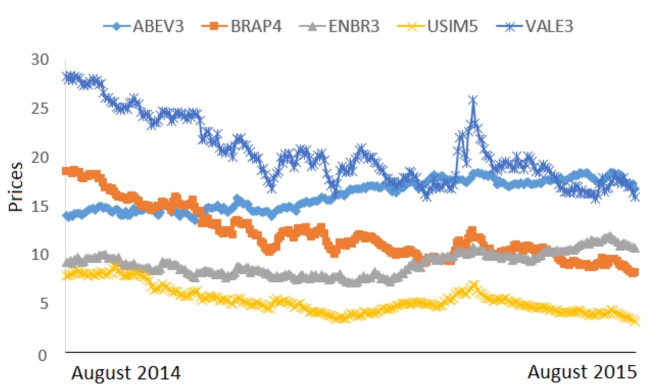


Figure 4. Evolution of prices of daily assets from August 2014 to August 2015.

V. Results

In order to validate our approach, we applied it to five real time series from BM&F BOVESPA[5]:

- **VALE3** is the ticker symbol of the common stocks of Vale S/A, world leader in the production of iron ore, pellets, and nickel.[6]
- **ENBR3** is the ticker symbol of the common stocks of EDP - Energias do Brasil S/A, a Brazilian company of the energy sector. EDP was assessed by a German consultancy company as one of the 20 best companies of the energy sector in the world in terms of performance.[7]
- **BRAP4** is the ticker symbol of the preferred stocks of Bradespar S/A, an investments company with relevant participation in many leading companies in its areas of operation.[8]
- **USIM5** is the ticker symbol of the Class A preferred stocks of Usinas Siderrgicas de Minas Gerais - Usiminas S/A, one of the largest siderurgy industries in Brazil.[9]
- **ABEV3** is the ticker symbol at Bovespa of the common stocks of Ambev S/A, the world largest beer manufacturer.[10]

The class distribution of the assets used can be found in Table I. The column NI presents the distribution of the class ”did not increase”, the column I presents the distribution of the class ”increased” and, finally, the column Dim presents the dimension of the input data set. It is possible to observe that there is not a great unbalance between the classes. Thus, accuracy is an adequate metric for performance in this case [41]. Besides, it can be seen that the dimension of each set is high, showing that solving this classification problem can become computationally complex and expensive.

[5] http://www.bmfbovespa.com.br/
[6] http://www.vale.com/
[7] http://www.edp.com.br/
[8] http://www.bradespar.com.br/
[9] http://www.usiminas.com.br/
[10] http://www.ambev.com.br/

Table I
DATA SETS OF BM&FBOVESPA

	Classes		
Asset	**I. (%)**	**NI. (%)**	**Dim.**
VALE3	49.5%	50.5%	180
ENBR3	50%	50%	180
BRAP4	48.5%	51.5%	180
USIM5	48.5%	51.5%	180
ABEV3	47%	53%	180

A. *Configuration of Algorithm Parameters*

The hyperparameters of the algorithms were defined using empirical tests. For training, tests were performed with the following combinations to RBM and SVM, respectively:

$$\begin{gathered}(\mathbf{v}, \mathbf{h}, \mathbf{ep}, \mathbf{lr}) \in \\ (\{180\} \times \{20, 15, 10, 5\} \times \{1000, 1500, 2000\} \\ \times\{0.9, 0.6, 0.3\})\end{gathered} \tag{9}$$

$$\begin{gathered}(\mathbf{kernel}, \mathbf{gamma}, \mathbf{cost}) \in \\ (\{RBF\} \times \{0.01, 0.25, 0.5\} \times \{0.01, 0.5, 1\})\end{gathered} \tag{10}$$

The hyperparameters of RBM and SVM are presented in tables II and III, respectively. The values are those that produced the best accuracy results among the various configurations considered.

Table II
RBM CONFIGURATIONS

RBM	
v	180
h	10
ep	1500
lr	0.6

Table III
SVM CONFIGURATIONS

SVM	
kernel	RBF
gamma	0.25
cost	1

Since the data is temporal, the result validation was performed in a training set/test set strategy over a sliding window. Therefore, at each step, the tool was trained with 20 candles and tested for the next 10 candles, until the end of the time series. These window sizes were also obtained based on experiments.

B. *Results Analysis*

The accuracy was adopted in this work for performance assessment. It is evaluated as the quantity of positive and negative samples correctly classified divided by the total quantity of samples, such as shown in Equation 11.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

in which: TP is the proportion of positive cases that were correctly identified. FP is the proportion of negatives cases that were incorrectly classified as positive. TN is defined as the proportion of negatives cases that were classified correctly. FN is the proportion of positives cases that were incorrectly classified as negative.

The results found for the combination of RBM + SVM were compared to those achieved with SVM only, such as presented in Table IV. The proposed method (RBM + SVM) led to results ranging from 0.54 (VALE3) to 0.66 (USIM5), which is higher than the results obtained with SVM isolated, (0.51 to 0.61).

Table IV
RESULTS OF THE ACCURACY OF THE EXPERIMENTS

	Accuracy	
Asset	**SVM**	**RBM + SVM**
VALE3	0.51	**0.59**
ENBR3	0.55	**0.61**
BRAP4	0.53	**0.59**
USIM5	0.61	**0.66**
ABEV3	**0.54**	**0.54**

The RBM+SVM association was able to outperform SVM only in four of the five assets, being equivalent in the other one (ABEV3). These results support the assumption that RBM could improve classification results through adequate selection of problem features.

VI. CONCLUSION

This study aimed to explore the capability of a deep neural network, specifically a Restricted Boltzmann Machine (RBM), to improve prediction of trends in the Brazilian Stock Market, BM&FBOVESPA. The results showed that this machine learning approach has the potential to reduce the dimensionality of input data and extract latent features to be considered by the main classifier. This enables the generation of additional information and, consequently, supports the process of data classification.

The proposed RBM + SVM obtained better results in four of the five data sets (it was equivalent in the remaining), when compared to the SVM classifier alone. Therefore, the proposed approach seems to be promising and may contribute to future studies on this type of application.

An immediate future study recommendation would be to incorporate the developed solution in a negotiation model to financial impact assessment and evaluate other datasets, including international financial markets (e.g NYSE, NASDAQ, Tokyo).

Acknowledgments

The authors would like to thank the AeC Contact Center for research support, the CEFET-MG, Brazilian agencies CAPES, CNPq, and FAPEMIG for the financial support.

References

[1] Barrymore, John , "Como funcionam as tendncias do mercado de aes." http://empresasefinancas.hsw.uol.com.br/tendencias-mercado-de-acoes.htm, 2017.

[2] M. Tkac and R. Verner, "Artificial neural networks in business: Two decades of research," *Applied Soft Computing*, vol. 38, pp. 788 – 804, 2016.

[3] R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega, and A. L. Oliveira, "Computational Intelligence and Financial Markets: A Survey and Future Directions," *Expert Systems with Applications*, vol. 55, pp. 194 – 211, 2016.

[4] P. Meesad and R. I. Rasel, "Predicting stock market price using support vector regression," in *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 1–6, May 2013.

[5] C. A. Nametala, A. Pimenta, A. C. Pereira, and E. G. Carrano, "An Automated Investment Strategy Using Artificial Neural Networks and Econometric Predictors," in *Proceedings of the XII Brazilian Symposium on Information Systems on Brazilian Symposium on Information Systems: Information Systems in the Cloud Computing Era - Volume 1*, SBSI 2016, (Porto Alegre, Brazil), pp. 152–159, Brazilian Computer Society, 2016.

[6] L. D. Persio and O. Honchar, "Artificial Neural Networks Approach to the Forecast of Stock Market Price Movements," *International Journal of Economics and Management Systems*, vol. 1, pp. 158–162, 2016.

[7] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162–2172, 2015.

[8] A. Pimenta, F. G. Guimaraes, E. G. Carrano, C. A. L. Nametala, and R. H. C. Takahashi, "GoldMiner: A genetic programming based algorithm applied to Brazilian Stock Market," *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pp. 397–402, 2014.

[9] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[10] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology*, vol. 5, pp. 115–133, Dec. 1943.

[11] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992.

[12] A. L. S and M. K. Albertini, "Anlise e visualizao de dados utilizando redes neurais artificiais auto-organizveis," in *Proceedings of The II Symposium on Knowledge Discovery, Mining And Learning*, KDMiLe 2014, (So Carlos, Brazil), Brazilian Computer Society, 2014.

[13] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2007.

[14] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, July 2006.

[15] R. Hrasko, A. G. Pacheco, and R. A. Krohling, "Time Series Prediction Using Restricted Boltzmann Machines and Backpropagation," *Procedia Computer Science*, vol. 55, pp. 990 – 999, 2015.

[16] C. Kirkpatrick and J. Dahlquist, *Technical Analysis: The Complete Resource for Financial Market Technicians*. FT Press, first ed., 2006.

[17] Bssola do Investidor, "Bssola do Investidor." https://www.bussoladoinvestidor.com.br/, 2017.

[18] BM&FBOVESPA , "BM&FBOVESPA." http://www.bmfbovespa.com.br, 2017.

[19] M. G. Anghel, "Technical Analysis versus Fundamental Analysis of Securities," *Romanian Statistical Review Supplement*, vol. 61, no. 2, pp. 257–262, 2013.

[20] M. Noronha, *Analise Tcnica: Teorias Ferramentas Estratgias*. Editec, 2003.

[21] F. A. d. Moura, "O uso de redes neurais artificiais na previso de tendncias no mercado de aes," Master's thesis, Universidade Federal de Pernambuco, 2006.

[22] D. M. Q. Nelson and A. C. M. Pereira, "Using LSTM and Technical Indicators to predict price movements," in *Proceedings of The IV Symposium on Knowledge Discovery, Mining And Learning*, KDMiLe 2016, (Recife, Brazil), Brazilian Computer Society, 2016.

[23] Q. Liang, W. Rong, J. Zhang, J. Liu, and Z. Xiong, "Restricted Boltzmann machine based stock market trend prediction," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1380–1387, May 2017.

[24] J. Soto, P. Melin, and O. Castillo, *Particle Swarm Optimization of the Fuzzy Integrators for Time Series Prediction Using Ensemble of IT2FNN Architectures*, pp. 141–158. Springer International Publishing, 2017.

[25] D. Franco and M. Steiner, "Comparative study of artificial neural network for financial time series prediction," *Blucher Marine Engineering Proceedings*, vol. 1, no. 1, pp. 303–313, 2014.

[26] C. Zhu, J. Yin, and Q. Li, "A stock decision support system based on DBNs," *Journal of Computational Information Systems*, vol. 10, no. 2, pp. 883–893, 2014.

[27] L. Takeuchi and Y.-Y. A. Lee, "Applying Deep Learning to Enhance Momentum Trading Strategies in Stocks," *Stanford University*, 2014.

[28] Y. Li and W. Ma, "Applications of Artificial Neural Networks in Financial Economics: A Survey," in *2010 International Symposium on Computational Intelligence and Design*, vol. 1, pp. 211–214, Oct 2010.

[29] X. Cai, S. Hu, and X. Lin, "Feature extraction using Restricted Boltzmann Machine for stock price prediction," in *Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on*, vol. 3, pp. 80–83, May 2012.

[30] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, Nov 1984.

[31] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, 2006.

[32] G. Hinton, "A practical guide to training restricted boltzmann machines," *Momentum*, vol. 9, no. 1, p. 926, 2010.

[33] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2nd ed., 1998.

[34] V. N. Vapnik and A. Y. Chervonenkis, "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities," *Theory of Probability & Its Applications*, vol. 16, no. 2, pp. 264–280, 1971.

[35] E. M. dos Santos, "Teoria e Aplicao de Support Vector Machines Aprendizagem e Reconhecimento de Objetos Baseado na Aparncia," Master's thesis, Universidade Federal da Paraba, 2002.

[36] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, pp. 121–167, Jun 1998.

[37] J. C. Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," in *Advances in Kernel Methods*, pp. 185–208, Cambridge, MA, USA: MIT Press, 1999.

[38] A. Ben-Hur and J. Weston, *A User's Guide to Support Vector Machines*, pp. 223–239. Totowa, NJ: Humana Press, 2010.

[39] GrafBolsa, "Indicadores Tcnicos." http://www.grafbolsa.com/help/indtec.html, 2017.

[40] T. E. de Campos, *Tcnicas de seleo de caractersticas com aplicaes em reconhecimento de faces*. PhD thesis, Universidade de So Paul0, 2001.

[41] M. d. A. Pereira, *Classificao de Dados Hbridos Atravs de Algoritmos Evolucionrios*. PhD thesis, Universidade Federal de Minas Gerais, 2012.