MC Medical Sciences
Volume 1 Issue 5 December 2021
Research Article

Deep LSTM-based Automated Learning Environment using Smart Data to Improve Awareness and Education in Time Series Forecasting

Zair Bouzidi[1,2]*, Abdelmalek Boudries[2] and Mourad Amad[1]

[1]LIMPAF Laboratory, Computer Science department, Faculty of Science and Applied Science, Bouira University, Algeria

[2]LMA Laboratory, Commercial Science department, Faculty of Economy, Business and Management, Bejaia University, Algeria

***Corresponding Author:** Zair Bouzidi, Commercial Science Department, Faculty of Economy, Business and Management, Bejaia University, Campus Abudaw, Bejaia, Algeria.

Received: October 30, 2021; **Published:** November 11, 2021

Abstract

This proposed commercial model is based on deep Long Short-Term Memory Network (LSTM), used because of its ability to learn long term dependencies, in taking the concatenated function and market data as input, while integrating Smart Data encapsulations, retrieving information by combining various search results (all the Web). It provides some good ideas for that we have extended to improve Corporate Marketing and Business Strategies. We show that the proposed model learns to both localize and recognize their different aspects. We evaluate it on the challenging task of detecting Fraud in Financial Services and Financial Time Series Forecasting and show that it is both more accurate than the state-of-the-art of other neural networks and uses fewer parameters and less computation.

Keywords: Awareness; Business; Education; LSTM; Smart Data

Abbreviations

AI : Artificial Intelligence
ANN: Automated Neural Network
API : Application Programming Interface
ConvNets : Convolutional NN ou CNN
DL : Deep Learning
DOSN : Decentralized Online Social Network
EANN: Elman Automated Neural Network
EMH : Efficient Market Hypothesis
FFNN : FeedForward Neural Network
FFNN : feed-backward neural network
IOT : Internet Of Things
LSTM : Long Short-Term Memory
MFF : Multi-Layer Feed-forward
NN : Neural Network
ONA: Our New Approach
OSN : Online social networks
P2P: Peer to Peer
RL : reinforcement learning

Citation: Zair Bouzidi., et al. "Deep LSTM-based Automated Learning Environment using Smart Data to Improve Awareness and Education in Time Series Forecasting". MC Medical Sciences 1.5 (2021): 03-17.

RNN: Recurrent NN

RW : Random Walk

TID : identifiers of the transactions

Introduction

The prime goal of a financial time series model is to provide reliable future forecasts which are crucial for investment planning, fiscal risk hedging, governmental policy making, etc. These time series often exhibit notoriously haphazard movements which make the task of modeling and forecasting extremely difficult. As per the research evidence, the random walk (RW) [1] is so far the best linear model for forecasting financial data. Artificial neural networks (ANN) are another promising alternative with the unique capability of nonlinear self-adaptive modeling. Numerous comparisons of the performances of RW and ANN models have also been carried out in the literature with mixed conclusions [2].

We propose a new real-time automated learning model based on Long Short-Term Memory Network (LSTM) that integrates encapsulations using Smart Data to retrieve information by combining various search results from various sources (all the Web) using Smart Data. Thus, we provide, not only, a solution to this challenge, but also, we propose better performances.

In this paper, we try to identify relevant content dealing with financial time series. Once this information is retrieved (distinguished, of course, from large amounts of other content and also distinguished from abusive information), it can be used to improve Marketing, Business, Fraud Detection in Financial Services and Financial Time Series. Our main contributions are listed below.

1. We develop a LSTM-based model that uses low-level content learning capabilities to automatically separate relevant information from redundant or abusive.
2. We develop the model that uses content learning capabilities of various sources (all online channels from social media and/or connected objects to websites) to automatically and efficiently capture real-time dynamics of financial data. Using a set of knowledge related to financial market, this model collects, using Smart Data, according to their lexical similarity, accurate forecasting of volatility from financial time series.
3. We have adapted some algorithms to streaming to get Smart Data. Smart Data, a different concept of Big Data, even in opposition to it, is based primarily on real-time data analysis. This term refers to an approach to data analysis that directly analyzes the data at the source, without the need to transmit it to a centralized system. Big Data is the mass of information circulating via the web, connected objects or smart phones, while Smart Data can be defined as the intelligent and relevant way of processing data.
4. Keeping in mind the limitations of the previous work, we develop an event-independent model that can be used directly to filter content on various source at a time in future events. Experiments on multiple financial market-related contents flows with diverse characteristics show that our proposed model outperforms forecasting-based approaches. While our approach filters content issued from all online channels, including connected objects.
5. Once we have developed this real-time LSTM-based model, and annotated manually the first information deducted by the LSTM, using various source content learning capabilities to automatically and efficiently capture real-time accurate forecasting of volatility from financial time series. Using a set of knowledge related to financial market and a set of tagged contents, it collects, reliable future forecasts which are crucial for investment planning, fiscal risk hedging, governmental policy making, etc.

The rest of this paper is organized as follows: section 2 presents the background and related works. Section 3 describes our proposed model. We provide details on accurate forecasting of volatility from financial time series, and preliminary results, followed a discussion of the results obtained. Finally, we conclude, giving some future works.

Background & Related Works

As background, we study some concepts of market basket analysis, followed by Social Networking and Smart Data with, notably streaming, and finally automated learning in general.

Market Basket Analysis

With a wide variety of products and buying behaviors, the shelves, on which products are presented, are the most important resources in the retail environment. Retailers can not only increase profits, but also reduce costs by properly managing shelf allocation and product display [3]. Using learning models in the organization of shelves in supermarkets by grouping products that are usually bought together; we can extract the following relation: customers who buy the product X at the end of the week, during the summer, generally also buy the product Y.

There is also the credit institution that permits to decide whether or not to grant credit based on the applicant profile, his / her request, and past loan experiences: it is used in data mining. The overbooking (optimization of the number of seats in planes, hotels, etc.), the targeting of offers (organization of advertising campaigns, promotions) and the analysis of business practices, strategies and their impact on sales are in Data Mining. This knowledge, unknown at first, may be correlations, patterns, or general trends in that data. Table 1shows a comparison of works with various economic tasks, including ours.

Economic Tasks	***References***	***Our New Approach (ONA)***
Detect novel frauds	[4]	ONA
Trading performance	[5]	
Exchange rate prediction	[6]	
Stock prediction	[7]	
Trade on the stock market	[8]	
Company stock prices	[9]	
Forecasting of financial time series	[10, 11]	

Table 1: Comparative table of all economic tasks used.

Experimental data are necessary to verify the correction of the system or the estimation of some difficult parameters to mathematical modeling. Data Mining is a field that has emerged with the explosion of the amount of information stored, with significant progress notably in processing speeds and storage media. The purpose of data mining is to discover, in large amounts of data, valuable information that can help understand the data or predict the behavior of future data. Since its inception, data mining has used several tools for statistics and artificial intelligence to achieve its objectives. It is an essential component of Big Data technologies, large data analysis techniques and recently data smart streaming. It is often defined as the process of discovering new knowledge. By examining large amounts of data (stored in warehouses or in streaming) using pattern recognition technologies as well as statistical and mathematical techniques used for the study, it describes the method used for the data collections.

There are also other specific tools as Extraction Element Set, extraction of sequential models, market basket analysis, Web Mining, Text Mining.

Extraction of element set is the model where some sequence exploration problems lend themselves to the discovery of frequent item sets and their order. Traditionally, Extraction of element set is used in marketing applications to detect patterns among competing elements in large transactions, for example, by analyzing customer shopping basket transactions in a supermarket, according to Han et al. [12].

The extraction of sequential models has many real applications because the data is encoded as sequences in many areas such as bio-informatics such as genomics and Proteomics [13].

We also see the development of market basket analysis which consists of studying sales (Sales receipt analysis) [14].

Thanks to the huge amount of information available online, the World Wide Web is a fertile ground for research in data mining: Web mining. Its research is at the crossroads of research conducted by several research communities, such as databases, information retrieval and within Artificial Intelligence (AI), particularly the sub-domains of learning and natural language processing [15].

While Text Mining is a branch of Data mining that specializes in the processing of text corpora to analyze the content and extract knowledge. The main tasks to be accomplished are the recognition of the information presented in the document and its interpretation. In unsupervised learning category, class labels either unknown or assumed to be unknown and clustering techniques are employed to figure out (i) distinct clusters containing fraudulent samples or (ii) far off fraudulent samples that do not belong to any cluster, where all clusters contained genuine samples, in which case, it is treated as an outlier detection problem. In supervised learning category, class labels are known and a binary classifier is built in order to classify fraudulent samples.

Fraud (including cyber fraud) detection is increasingly becoming menacing and fraudsters always appear to be few notches ahead of organizations in terms of finding new loopholes in the system and circumventing them effortlessly. On the other hand, organizations make huge investments in money, time and resources to predict fraud in near real-time, if not real time and try to mitigate the consequences of fraud. Financial fraud manifests itself in various areas such as banking, insurance and investments (stock markets). It can be both offline as well as online. Online fraud includes credit/debit card fraud, transaction fraud, cyber fraud involving security, while offline fraud includes accounting fraud, forgeries etc.

Advances in technology and breakthrough in deep learning models have seen an increase in intelligent automated trading and decision support systems in financial markets, especially in the stock and foreign exchange markets.

However, time series problems are difficult to predict especially financial time series. On the other hand, Neural and Deep learning models have shown great success in forecasting financial time series despite the contradictory report by efficient market hypothesis (EMH) [1], that the stock and foreign exchange follows a Random Walk (RW) and any profit made is by chance. This can be attributed to the ability of Neural Networks to self-adapt to any nonlinear data set without any statically assumption and prior knowledge of the data set. Deep leaning used both fundamental and technical analysis data, which is the two, most commonly, used techniques for financial time series forecasting, to trained and build deep leaning models.

Fundamental analysis is the use or mining of textual information like financial news, company financial reports and other economic factors like government policies, to predict price movement.

Social Networks

Social networking forms an important part of online activities of Web users. Web sites such as Facebook, MySpace and Orkut have millions of users using them every day. A decentralized online social network (DOSN) is a distributed system for social networking with no or limited dependency on any dedicated central infrastructure, while being a solution to the violation of privacy, especially thanks to P2P architecture.

Recent trends in the use of social networking highlight the fact that there is not only an increasing number of users of social networking applications, but also a significant increase in the number of such applications. In a short time social networks have invaded the daily lives of Internet users and Web professionals. Table 2 shows a Comparative table of all techniques and methods used in Models including our approach.

References	***Identification Methods***	***Used OSN & Smart Data***
[17]	Flood Disaster Game-based Learning	Twitter
[18]	Educational Purposes among the Faculty of Higher Education with Special Reference	
[19]	Summarization with social-temporal context	
[20]	Capitalizing on a TREC Track to Build a Tweet Summarization Dataset	
[21]	semi-automated artificial intelligence-based classifier for Disaster Response	
[22]	Summarizing situational tweets in crisis scenarios: An extractive-abstractive approach	
[23], [24]	Based on Artificial Neural Network (ANN)	Twitter & Facebook
[25]	Based on Feed Forward Neural Network (FFNN)	All the Web
[26]	Based on Recurrent Neural Network (RNN)	All the Web & Smart Data
[27]	Based on Long Short-Term Memory (LSTM)	
Our New approach	Deep Learning from Social Media and Big Data to improve Marketing, Business Strategies, Fraud Detection and Financial Time Series Forecasting	

Table 2: Comparative table of all techniques and methods used in Models including our approach.

Among the existing research studies, a group of studies identifies useful social networking information, using machine learning, to successfully extract structured information from unstructured textual social media contents.

People use social networking to post situational updates in various forms such as text messages, images and videos. Numerous studies have shown that this online information is useful for a quick response to a particular situation. Communication via social networking is direct, easy and instant and can simplify quick responses. Custom sites like Facebook, Twitter, Instagram, YouTube and Xing can subjectively offload the first contact of authorities and service providers. These analyzes of the use of social networking in events have identified a distinct role for users, who are more likely to generate useful information to improve situational awareness. Social networking can be considered as a practical and reliable emergency communication tool. While the predominant function of social networking remains social interaction, social networking sites are also considered the fourth most popular source of information. Different social networks have different characteristics and are therefore more or less suitable for use during a given situation. Thank to Web 2.0, a growing number of studies have examined the use of social networking data to gain knowledge of areas of human activity as diverse as the detection of disease as epidemics and predicting the stock market. However, understanding these voluminous and high velocity data is a difficult task.

Contents were collected from all online channels tracked automatically by the Online Listening Tool, namely Radian6 [28] or any its competitor, such as Awario, Brand24.com, Brandwatch, Mention, Keyhole, Socialert.net, SocialPilot.co, Simplify360, etc. from websites to social media, such as Twitter, Facebook, LinkedIn, Instagram, Google+, Youtube and so on. Actually, many networking platforms allow access to their data via Application Programming Interface (API) [29]. Table 3 shows a Comparative table of all techniques and methods with Big Data used in Models including our approach.

Models	Identification Usage of Big Data
[30]	Human Computing and Machine Learning to Make Sense of Big Data
[31]	Decision-Making and emerging Big Data
[32]	Quality of Social Media Data in Big Data
[33]	Big Data privacy in public Social Media
Our Previous approaches [26, 27]	Smart Data privacy in public Social Media
Our New approach	Automated Learning using Big Data to improve Marketing, Business Strategies, Fraud Detection and Financial Time Series Forecasting

Table 3: Comparative table of all techniques and methods with Big Data used in Models including ours.

Online listening tools provide the model, which reasonably represents the essentials, namely, as shown in figure 1:

- Harvesting contents: (such as conversations at the social media, news or any information in the Web);
- Cleaning the data of duplication and replication content: eliminating, from the content, any dubbed information like retweet, and any information harmful or redundant;

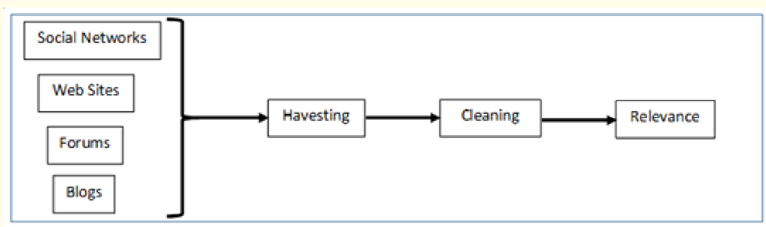

Figure 1*:* Online Methodology Reflecting our Listening and Monitoring Approach.

- Enabling relevance: thanks to the neural learning, obtaining of relevant information by using the machine learning with the learning corpus obtained thanks to the tagged messages. These tagged messages are realized by volunteers;
- Analyzing the results: During this stage, the verification and analysis of the results is carried out in order to ensure adequacy so that it is obtained to build disaster information such as situational awareness and/or education.

Automated Learning

Learning is a set of mechanisms leading to the acquisition of know-how and knowledge. While Automated learning is a branch of Artificial Intelligence (AI) that deals with the development of algorithms that make capable of accomplishing complex tasks without having been explicitly programmed for that purpose, making extensive use tools and concepts of AI, of mathematics, others cognitive sciences and so on. It can relying on statistical approaches to give the ability to "learn" from data using two phases. The first one, namely Model Design Phase (Training) consists in estimating a model from data, called observations. The second one is Production Phase where the model being determined, new data can then be submitted to obtain the result corresponding to the desired task. Handwriting recognition is a good and complex example because two similar characters are never exactly equal. An automatic learning system can be designed to learn to recognize characters by observing "examples", that is, known characters. Table 4 shows a Comparative table of all AI Concepts used including our approach.

AI Concepts	*References*	*Our New Approach (ONA)*
Elman ANN (EANN)	[2], [23], [24]	ONA
Multi-Layer Feed-forward	[2], [27], [25]	
ConvNets / Autoencoder	[9]	
RNN	[26], [11]	
LSTM	[2], [5], [6], [8], [27]	
Memory Networks	[2], [5], [6], [8], [27]	
Social Media	[2], [5], [6], [8], [27]	
Analysis	[5], [11], [27]	

Table 4: Comparative table of all AI Concepts used including our approach.

According to the information available during the learning phase, it is qualified in different ways. If the data is tagged, it is a supervised learning. We are talking about "*classification*" if the labels are discrete or "*regression*" if they are continuous. If the model is learned incrementally based on a reward, it is called "*reinforcement learning*". When data (or "*tags*") are missing, the model must use untagged examples that can still provide information. In medicine, for example, it can be an aid to diagnosis. It is said that learning is "semi-supervised". While the labeling of data is partial when a model states that a data does not belong to a class A, but perhaps to a class B or C, (A, B and C being 3 diseases for example evoked as part of a differential diagnosis). This is called *partially supervised learning*. In the most general, unlabeled case, we try to determine the underlying structure of the data: this is *unsupervised learning* [16].

Automated learning is used for a wide range of applications, such as diagnostic aid [16], outlier detection, missing data detection, relevant information retrieval from multiple sources (social media) [23-26], fraud detection, financial market analysis [26] & [14] and so on.

It is not just about a set of algorithms, but a list of steps to take into account and execute in order to reach an optimal result. Data Acquisition is the first step of this list where the algorithm feeds on input data and where the success of the project is collecting relevant data and in sufficient quantity. Preparation and cleaning of the data is the second step. The third is creation of the model. The fourth step is Evaluation which consists of evaluating trained model on the other (second) set of data. The fifth is Deployment where the model will be deployed in production to make predictions, and potentially use new input data to re-train and improve its model.

However, care must be taken to use an adequate number of neurons and hidden layers, to detect and thus avoid over-learning. Thus, the data is divided into two subsets: the learning set which allows changing the weight of the neural network. The validation set allows verifying the relevance of the network, while avoiding over-learning.

Automated Learning Environment to improve Fraud & Time Series Forecasting

Deep learning has become an important research method for forecasting or recognition in recent years. In this section, we describe the LSTM-based model.

Long Short-Term Memory (LSTM)

To overcome the deficiency of RNN, Hochreiter and Schmidhuber (1997) [34-37] proposed Long Short-Term Memory, one of the most successful RNN architectures for sequence learning. Compared with the Elman RNN, LSTM introduces the memory cell, a computation unit replacing conventional artificial neurons in the hidden layer.

Long Term Memory Networks (LSTMs) [34-37] are a kind of RNN that uses specific a "*memory cell*" with standard units. This *cell* is a component of LSTM units that can hold information in memory for a long time. LSTMs are often referred to as Sophisticated RNNs: LSTM improves the lack of long-term memory of RNN and prevents the problem of gradient disappearance. LSTM can dynamically learn and determine the next recursive input: it can retain important content providing a good reference in building a model of prediction.

The memory cell LSTM contains main components: input gate, Forget gate, Output gate. Neurons, via these gates, decide what store information, when read, write and forget information as shown in Figure 1 where:

x_t is the input vector in time t,
h_t is the output vector,
c_t contains the state of the union,
i_t is the vector of input gate,
f_t is the vector of forgotten gate,
o_t is the vector of output gate, and
σ_i, σ_f, σ_o, σ_c, σ_h are activation functions.
The mathematical definition of the computation of the LSTM model can be described as follows:

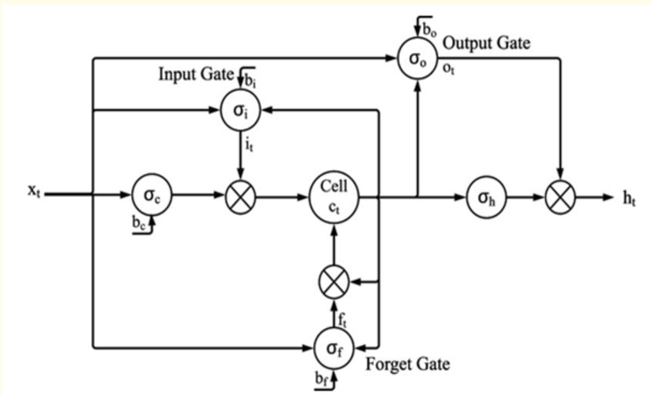

Figure 2: LSTM with gates and activations functions.

$$i_t = \gamma\,(\sigma_{ix} \cdot x_t + \sigma_{ih} \cdot h_{t-1} + b_i) \qquad (1)$$

$$f_t = \gamma\,(\sigma_{fx} \cdot x_t + \sigma_{fh} \cdot h_{t-1} + b_f) \qquad (2)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(\sigma_{cx}).\, x_t + \sigma_{ch} \cdot h_{t-1} + b_c) \qquad (3)$$

$$o_t = \gamma\,(\sigma_{ox} \cdot x_t + \sigma_{oh} \cdot h_{t-1} + b_o \qquad (4)$$

$$h_t = o_t \odot \tanh(c_t) \qquad (5)$$

Where
ʘ denotes element-wise multiplication.
σ is the logistic sigmoid function.

i, f, o, and c are respectively the input gate, forget gate, output gate, and cell activation vectors, all of which are in the same size as the hidden vector h at the level t.

LSTM-based Automated Learning Environment

Contents were collected from all online channels tracked automatically by the Online Listening Tool, namely Radian6 from websites to social media, and objects connected thanks to Smart Data with streaming intelligent.

We use LSTM with a hidden layer that takes, as input to the network, a content e, as:

$$e = (w_1, \dots, w_i, \dots, w_n) \qquad (6)$$

Containing words w each coming from a finite vocabulary \Upsilon. $\mathbf{C}^n$ is the set of contents issued from the social media.

LSTM Modeling

Let:

$$\forall\ i \in [1, N]\ e_i \in \mathbf{C_n} = E \text{ with } e_i = (w_{i1}, w_{i2}, \dots, w_{in}) \qquad (7)$$

Containing words each coming from the set of words **W** where each word comes from a finite vocabulary **V**, the incorporation of a content of the source message i relevant for, at least, a keyword or a hashtag such as : $\mathbf{E_K}$

$$\exists\ j \in [1,M] \ / \ hj \in H \qquad (10)$$

We want the learning of a generic space with the neural network, as:

$$\mathbf{E_K} = \max_{1 \le k \le K} \{ek\} \qquad (11)$$

Which normalizes the differences:

$$\mathbf{E_K} = [\mathbf{E} - \mathbf{RDF}] \text{ where } \mathbf{RDF} = [\mathbf{R} + \mathbf{D} + \mathbf{F}] \qquad (12)$$

Thanks to the neural network, the transformation of $\{e_i\}$ into $\{e_k\}$ can be explained by:

$$\exists\ j \in [1, M] \ / \ h_j \in \mathbf{H} \ \&\ \exists l \in [1, L] \ / \ w_l \in \mathbf{W}\ /$$

$$\{ei \rightarrow e_k = \{e_i \mid e_i \text{ is relevant for } h_j \ \&\ w_l\} \text{ with I } [1, N]\ \} \qquad \mathbf{(13)}$$

$$\&\ e_i \notin [\mathbf{R} + \mathbf{D} + \mathbf{F}]$$

Where

R, **D** and **F** denote respectively the set of duplicate re-tweets, duplicate contents and false alerts.
The objective is then to maximize the size K of the set $\mathbf{E_K}$.

The model structure of our previous work [26] is a state-of-the-art natural language ***Sequential Pattern Mining*** tool, called SPM-Tool, used to recognize the sentiments of text.

The long short term memory neural network (LSTM), useful at analyzing time series data, is applied to time financial series forecasting with information of Market Basket Analysis and text sentiments issued from SPM-Tool.

Figure 2 is the diagram of the SPM-Tool model with (e_1, e_2, ..., e_n) as input and (Yes, Percent) / (No, Percent) as output.

The SPM-Tool model pre-trains a deep linguistic representation model based on the semantics of all layers.

Functioning of LSTM-SPM-based Model

Figure 3 shows the functioning the LSTM-SPM-based Automated Learning Environment to Enhance Time Series Forecasting. Once the messages have been collected via the LSTM-based Model, we apply the following algorithms to these contents.

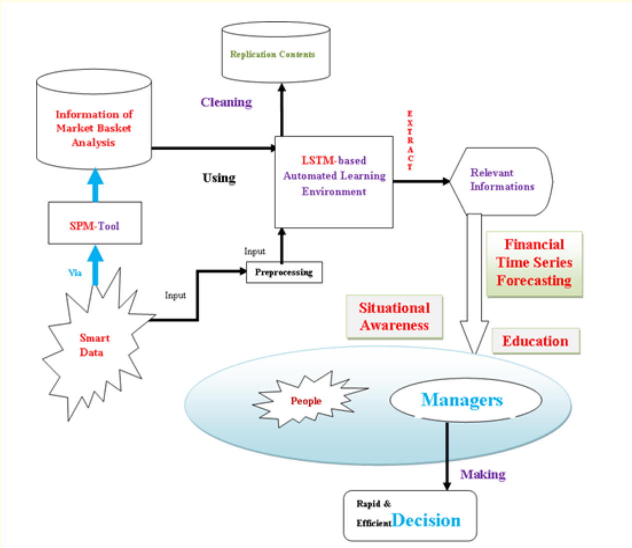

Figure 3: LSTM-SPM-based Automated Learning Environment to enhance Time Series Forecasting.

Automated Learning Environment

A new ad hoc real-time automated learning environment with accurate forecasting of volatility from financial time series is presented here. It is based on a new multi-view recovery model from multiple sources using Smart data. Such an approach is really useful for helping to make strategic appropriate decisions (see Figure 3).

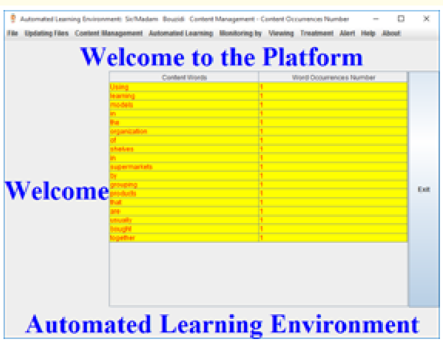

Figure 4: Example of Content Occurrences Number in the Automated Learning Environment.

Among the goals of the optimizations this algorithm is to facilitate streaming playback with a single read with sufficient computation and storage. The threshold σ is set by the analyst. This can follow an iterative approach by setting a threshold at the start and, depending on the result, change the threshold value. The algorithm proposed by Savasere [34] solves the memory space problem of the previous algorithm. The advantage of this algorithm is that it requires only one reading at most.

Figure 5: Example of Content Size in the Automated Learning Environment.

Figure 6: Example of Content Occurrences Number in the Automated Learning Environment.

Social media and Smart Data

With tools for statistics and artificial intelligence that are used with Big Data technologies, large data analysis techniques and recently data smart streaming (from social media and/or connected objects), as the discovering process of new knowledge by analyzing large quantities of data, in streaming, using Deep Learning as well as statistical and mathematical techniques.

Social media data help respond to needs and assess needs. Side by side, citizens easily turn to social networks to confide in, quickly disseminate information and learn useful insights. Social media improves people's knowledge of the situation, facilitates the dissemination of information and express people needs, and enables to learn useful insights and early warning. People and organizations are increasingly using social media to report and take action on high profile events. Researchers show a correlation between per capita social media activity and event, making it easy to quickly assess [10]. Countries are faced with both an increasing frequency and an increasing intensity of events. Reliable connectivity and data security allow us to offer seamless, efficient services for remote control and surveillance of fire detection and alarm systems. Remote Services transform our aid into a state-of-the-art IoT solution. It permits remote access to connected objects tools for manipulating, maintenance and live monitoring of any feature, and trouble transmissions to smart devices.

Situational Awareness

An important aspect of situation awareness using social media and/or Smart Data consists to detect and characterize how a supermarket works smartly. Thus, we will be better equipped to take all the precautions and the luck on our side. Knowing that a supermarket generates a lot of money, it allows several families to live decently, prompts us to fight all the evils that persist in disrupting this wonderful life, such as everything relating to the smart management of a large surface without parasites nor fraud.

Education

Education plays a leading role in human evolution in the pursuit of sustainable development or outright in its emancipation. Previous experiences have shown the positive effects of education in human societies, especially children. It turned out that those who have been trained or sensitized, thus warning others and protecting themselves in random destruction, do not live. This Model is designed to support an introductory traineeship in emancipation for citizens, trainees and future managers.

Thus, the trainee can use this tool in three modes. Novice Mode permits him to use a complete set of automated design and learning tools, such as observing of various programs at work, experimenting them and gradually learning from his experience, observations and mistakes. Beginners Mode permits him, at any point, to ask this tool to generate (move on) the next step. This tool analyzes knowledge and provides both the optimal stage and a list of all relevant operations. Not satisfied with proposed operation, he can choose any appropriate operation using adaptive hierarchical menus.

In the Online manual rehearsal Mode, at any time during the work, the trainee has a menu to access all previous courses: presentation of any previously learned concept, demonstration of all the examples learned and analysis of any problem explained or resolved. This mode provides access to the material learned from the course as a reference, thereby supporting example-based online help.

Educational messages play a role in enhancing situational awareness at any time. But above all, this education consists of constantly rehashing this advice on all information channels, websites and all social and networking media to have as much situational awareness as possible.

Experimental Results

In this study, two different models, including this model were established and compared based on the input sample. The details of input sample and models are shown in Table 2. The input sample is captured by LSTM-based Automated Learning Environment, according to Figure 3. The results are evaluated with the RMSE, MSE and R^2 calculated with the formulas (15, 16 and 17). Final results are shown in Table 3.

In this section, we present the experiments carried out to compare the performance of models, including our proposed LSTM model, tested with the dataset, introduced in the following subsection, which have been preprocessed. The mean squared error (RMSE), the mean absolute error (MAE) and the (R-Square) were the measures used to assess model performance across all experiments.

$$RMSE = \frac{1}{N} * \sum_{i=1}^{N} \sqrt{(y_i - y_i^*)^2} \qquad (15)$$

$$MSE = \frac{1}{N} * \sum_{i=1}^{N} | y_i - y_i^* | \qquad (16)$$

$$R^2 = \frac{1}{N} * \frac{\sum_{i=1}^{N} (y_i - y_i^*)^2}{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \qquad (17)$$

Where

N represents the number of content flow, y_i is the real content in flow i, and y_i* is the relevant content flow. $y_i^{\wedge}$ is the mean value of the relevant content number.

Data Description

We have divided the data sets into a training set and a verification set. The learning set is applied to train different deep learning models, while updating the weights and bias of the neural cell. And then the verification set checks the skill of these models.

Results

LSTM is an important part of the CNN-LSTM framework and provides vector characteristics based on historical information. The final experimental results are presented in Table 8.

In this section, we have checked the effectiveness of the proposed ConvLSTM model against the benchmarks: the RNN and LSTM prediction method are the widely used deep learning models. In the experiment, these deep learning / machine learning models must learn (finding best hyper-parameters), including find the number of neurons, the number of layers of neural networks and the activation function of the neural network. After a complete experiment, we obtained the final configuration results of this model through the evaluation of the verification set.

Model	***RMSE***	***MAE***	***R^2***
Neural Network [23, 24]	17,088.3797	18,471	0.3284
Feed Forward NN [25]	16,100.9272	19,461.5	0.4645
RNN [26]	13,359.4722	19,962.5	0.4805
LSTM [27]	16,704.4894	19,557	0.5064
Our New Approach	21,070.1960	12,809.5	0.6998

Table 5: Examples of Relevant Content of Covid-19 for a Hashtags and Keywords Set from social media.

To be fair, the number of relevant contents is taken as the historical information for NN, FFNN, RNN, LSTM and our new approach. Further, from the RMSE and MAE, it is obviously that our new approach is more accurate than other models since combining the advantages of both. This result indicates that the model is more suitable to retrieve relevant content than the Neural network, the Feed forward Neural Network, the original RNN and LSTM models.

Conclusion and Perspectives

A new ad hoc real-time automated learning environment with accurate forecasting of volatility from financial time series is presented here. It is based on a new multi-view recovery model from multiple sources using Smart data. Such an approach is really useful for helping to make strategic appropriate decisions.

As the contents are known to be written informally, the contents follow no syntax, no logic, noisy, may contain spelling mistakes, abbreviations, etc. and there is only collected English content posted, so, this work may have some limitations.

1. As a result, domain-specific biases may exist in the dataset.
2. Side by side, content published in other languages may contain different types of reasons in relation to English content.
3. The features of the automated learning environment have been developed based on the analysis of specific content.

As for future works, this study can have many potential applications for the future. The proposed model can be completed with a series of new research questions and perspectives. Pure improvements can start from:

1. Enhancing the validation of the information before launching this update information in order to avoid errors in accurate forecasting of volatility from time series, with abusive information.
2. Extending it to process images and videos of social media.
3. Integrating multiple languages, notably local languages in its use.

References

1. Fama EF. “Random walks in stock market prices”. Financial analysts journal 51.1 (1995): 75-80.
2. Adhikari R and Agrawal RK. “A combination of artificial neural network and random walk models for financial time series forecasting”. Neural Computing and Applications 24 (2014): 1441-1449.
3. Aloysius G and Binu D. “An approach to products placement in supermarkets using Prefix Span algorithm”. Journal of King Saud University - Computer and Information Sciences 25.1 (2013): 77-87.
4. Fiore U., et al. “Using generative adversarial networks for improving classification effectiveness in credit card fraud detection”. Information Sciences 479 (2019): 448-455.
5. Sermpinis G., et al. “Neural networks in financial trading”. Annals of Operations Research (2021): 293-308.
6. Calvez A (le) and Cliff D. “Deep learning can replicate adaptive traders in a limit-order-book financial market”. 2018 IEEE Symposium Series on Computational Intelligence, (SSCI 2018): 1876-1883.
7. Kodogiannis V and Lolis A. “Forecasting Financial Time Series using Neural Network and Fuzzy System-based Techniques”. Neural Computing and Applications 11 (2002): 90-102.
8. Fischer J., et al. “Atomic Orchid: A mixed reality game to investigate coordination in disaster response”. Proceedings of the 11th International Conference on Entertainment Computing (ICEC 2012): 572-577.
9. Pinheiro., et al. “Stock market prediction with deep learning: A character-based neural language model for event based trading”. Proceedings of the Australasian Language Technology Association Workshop (2017): 6-15.
10. Kryvasheyeu Y., et al. “Rapid assessment of disaster damage with social media activity”. Science Advances 2.3 (2016): e1500779.
11. Mohammad., et al. “Hybrid deep learning model for stock price prediction”. IEEE Symposium Symposium Series on Computational Intelligence SSCI. IEEE (2018): 1837-1844.
12. Han J-H., et al. “Neuronal Competition and Selection During Memory Formation”. Science 316.5823 (2007): 457-460.
13. Fournier-Viger P., et al. “A Survey of Sequential Pattern Mining”. Data Science and Pattern Recognition, Ubiquitous International 1.1 (2017): 54-78.
14. Masseglia F., et al. “Sequential pattern mining: A survey on issues and approaches”. Encyclopedia of Data Warehousing and Mining (2005).
15. Kosala R and Blockeel H. “Web Mining Research: A Survey”. ACM SIGKDD Explorations Newsletter, ACM 2.1 (2000): 1-15.
16. Abouelhoda M and Ghanem M. “String Mining in Bioinformatics”. Scientific Data Mining and Knowledge Discovery, (2009): 207-247.
17. Zaini NA., et al. “Design and Development of Flood Disaster Game-based Learning based on Learning Domain”. In International Journal of Engineering and Advanced Technology (IJEAT) 9.4 (2020): 679-685.
18. Vivakaran MV and Neelamalar M. “Utilization of Social Media Platforms for Educational Purposes among the Faculty of Higher Education with Special Reference to Tamil Nadu”. In Higher Education for the Future 5.1 (2018): 4-19.
19. He R., et al. “Twitter summarization with social-temporal context”. In World Wide Web 20.2 (2016): 267-290.
20. Dussart A., et al. “Capitalizing on a TREC Track to Build a Tweet Summarization Dataset”. In Text REtrieval Conference, (TREC’2020).
21. Lamsal R and Kumar TVV. “Classifying Emergency Tweets for Disaster Response”. In International Journal of Disaster Response and Emergency Management (IJDREM) 3.1 (2020): 14-29.
22. Rudra K., et al. “Summarizing situational tweets in crisis scenarios: An extractive-abstractive approach”. IEEE Transactions on

Computational Social Systems 6.5 (2019): 981-993.

23. Bouzidi Z., et al. "A New Efficient Alert Model for Disaster Management". Proceedings of Conference AIAP'2018: Artificial Intelligence and Its Applications, El-Oued, Algeria (2018).
24. Bouzidi Z., et al. "Intelligent and Real-time Alert Model for Disaster Management based on Information retrieval from Multiple Sources". International Journal of Advanced Media and Communication 7.4 (2019): 309-330.
25. Bouzidi Z., et al. "Towards a Smart Interface-based Automated Learning Environment Through Social Media for Disaster Management and Smart Disaster Education". Advances in Intelligent Systems and Computing. SAI 2020, Springer, Cham 1228 (2020): 443-468.
26. Bouzidi Z., et al. "Deep Learning-based Automated Learning Environment using Smart Data to improve Corporate Marketing, Business Strategies, Fraud Detection in Financial Services and Financial Time Series Forecasting". International Conference on "Managing Business through Web Analytics - (ICMBWA2020)", Khemis Miliana University, Algeria (2020).
27. Bouzidi Z., et al. "LSTM-based automated learning with smart data to improve marketing fraud detection and financial forecasting". 5th International Scientific Conference on Economics and Management, (EMAN 2021), Serbia (2021).
28. Rappaport SD. "Listening Solutions, A Marketer's Guide to Software and Services". Journal of Advertising Research 50.2 (2010): 197-213.
29. Imran M., et al. "Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions". Information Processing & Management 57.5 (2020): 1-9.
30. Ofli F., et al. "Combining Human Computing and Machine Learning to Make Sense of Big (Aerial) Data for Disaster Response". Big Data 4.1 (2016): 47-59.
31. Horita Flavio EA., et al. "Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil". Decision Support Systems 97 (2017): 12-22.
32. Immonen A., et al. "Evaluating the Quality of Social Media Data in Big Data Architecture". IEEE Access 3 (2015): 2028-2043.
33. Smith M., et al. "Big data privacy issues in public social media". 6th IEEE International Conference on Digital Ecosystems Technologies (2012): 1-6.
34. Hochreiter S and Schmidhuber J. "Long short-term memory". Neural Computation 9.8 (1997): 1735-1780.
35. Bouzidi Z., et al. "Deep Learning and Social Media for Managing Disaster: Survey". Intelligent Systems and Applications (2021): 12-30.
36. Bouzidi Z., et al. "Enhancing Crisis Management because of Deep Learning, Big Data and Parallel Computing Environment: Survey". Proceedings of the 3rd International Conference on Electrical, Communication and Computer Engineering (ICECCE), Kuala Lumpur, Malaysia (2021).
37. Bouzidi Z., Am et al. "A Survey on Deep Learning in Big Data and its Applications". International Conference on Innovations in Energy Engineering & Cleaner Production IEECP'21, Silicon Valey, California, USA (2021).

Volume 1 Issue 5 December 2021