ANALYSIS AND FORECASTING OF BUDGET ACCOUNTING SYSTEMS USING MACHINE LEARNING ALGORITHMS

Bogdan DRAGULIN PhD Student
IOSUD-SDSE Valahia University of Targoviste, Romania, dragulinbogdan@gmail.com

Veronica STEFAN, PhD Professor
IOSUD-SDSE Valahia University of Targoviste, Romania, veronica.stefan@ats.com.ro

ABSTRACT: *The paper debates and amplifies a topic of great interest, related to the current socio-economic context of the field, trying to establish new relationships of interdependence between social-economic factors and the evolution of income and expenditure. Through this paper we will address and analyze aspects of the evolution of income and expenditure, finding certain measurable links between socio-economic factors and the evolution of income and expenditure, finding and expressing strong or weak relationships between cause (socio-economic factors) and effect (amount of income and expenses). The present research was motivated by the use of modern analysis and prediction elements for modeling economic phenomena and for achieving a more accurate decision support. An important objective is to analyze the applicable regression methods in relation to the implementation of Machine Learning algorithms. The public budget is a perfect source of Big data for the implementation of a Machine Learning algorithm, because it allows us to define multiple dimensions for the same information contained. The conclusions and proposals resulting from the analysis of the causality and the interdependence of the analyzed factors are intended to represent a decisional support for the state institutions and at the same time an element of understanding and forecasting of the economic phenomena.*

Keywords: Prediction methods; Simulation models; Budget forecasting; Machine Learning algorithms; Big Data tools

JEL Classification: M41, O21, H61, H68, H83

INTRODUCTION

At this moment, both the budget execution and planning represent an activity with a high impact on society. The transparency and efficiency of budget execution directly results in an increase of society's trust in the state's mechanisms.

This work aims to define the methods for perfecting the forecasting of the public budget so that it will become an efficient instrument for the implementation of fiscal policies.

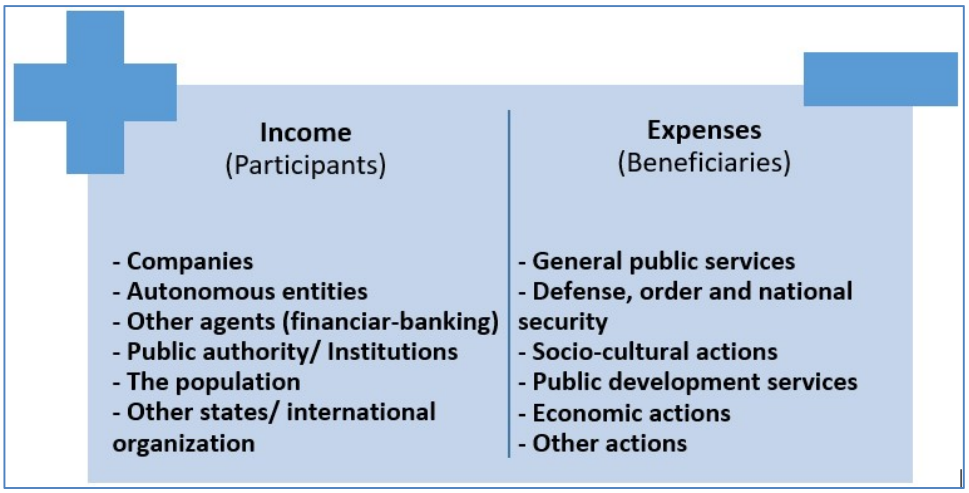
For highlighting the public necessities and for allocating the financial resources needed for achieving them, public authorities need an adequate organizational structure of the financial relations, this is known as "public budget" (Mosteanu T., 1997).

From a juridical standpoint, the public budget represents the use of public authorities of an instrument called "right" for structuring, organizing and highlighting social relations and also for protecting the fundamental social values involved.

In this regard, Leon Duguit remarked, talking about the juridical nature of legal budget dispositions that "*the budget is in part law when creating dispositions and in part administrative act when creating individual and concrete dispositions*" (Duguit L., 1923).

From another perspective, the notion of budget is often used also with the meaning of budget fund, meaning the ensemble of monetary resources mobilized for public authorities and allocated for different destinations.

Figure 1. The budget in terms of participants and relations between participants



Source: Own source

The appearance of the public budget in its legal acceptance laid the foundations for its consecration as a primary instrument for the economic and financial management, maintaining this quality to this day. It appears as a document with mandatory legal value that fixes in a certain structure the public income and expenditure indicators for a period of one year, following default rules, authorizes collecting revenues and making expenses and ensures the premises of control over them.

In this respect, staging the works specific for materializing the public budgets (*projection, approval, execution and control*), is fixed by law, including the meaning of a precise calendar.

The legal acceptance of the budget highlights the implication of public authorities in the process of the budget, income and expenditure indicators projected by the executive being subjected to the authorization by the legislature, which is entitled to a political control over the financial effort of the nation and the directions to which it is channeled.

The consistency with the government programme is ensured, based on which the government has been invested with the trust of the representatives of the nation.

This typical collaboration also expands for the approved revenue and expenditure indicators, respectively the conclusion of the budget execution.

In our country, the legal reglementation of public budget (The law of public finances) genetically defines the notion of budget as "*the document through which there are provided and approved every year the income and expenses or, as the case, only the expenses, depending on the financing system of the public institutes"*.

It is mandatory to specify in this context that in the current language it is also encountered, for the case of different public entities, the notion of "*budget of income and expenses*", seemingly partially debatable by reference to the content of the notion of budget, which would imply income and expenses. It justifies the need to differentiate this budget compared to other categories: *exploitation budget, investment budget, treasury budget* etc. or to highlight or separately treat the two parts of the budget (***income budget and expenses budget***).

The present research was motivated by the use of modern analysis and prediction elements for modeling economic phenomena and for achieving a more accurate decision support.

The interdisciplinary character (*mathematical elements, IT tools, new technologies, factors and social needs are involved, as well as their economic modeling*), makes the approached topic have an original, current character, necessary in order to interconnect the information from the approached fields.

The budget project in terms of revenues and expenditures is carried out in several stages:

- Elaboration of the draft budget for revenues and expenditures
- Approval of the draft budget
- Budget execution
- End the execution of the budget
- Control of budget execution
- Approval of budget execution

Budget planning and execution are organized, at the level of budgetary institution, by *functional classifications* and within the functional classifications, by *economic classifications*. The planning and execution are detailed at the levels subordinated to the institution, and are centralized at the level of the superior units, allowing a follow-up of the activity.

Revenues and expenditures are grouped in the budget based on budget classification.

Revenue is structured by chapters and subchapters of structure revenue; ***expenditure*** is structured by parts, chapters, subchapters, titles, articles, and paragraphs, as appropriate.

The context of this scientific research is related to the ability to forecast as accurately as possible the evolution of the public budget. New technologies and the evolution of algorithms and computing systems in the contemporary period allow determining the interdependence of causal factors of an event at an unprecedented level, allowing their analysis without having an explicit direct relationship between them.

The following research hypotheses are considered:

• The social structure directly influences the realization and execution of the public budget;

• Identifying the factors and their share in the influence is a necessity for understanding budgetary phenomena;

• The use of advanced algorithms and technologies increases the level of understanding and representation of the interaction between social factors and the public budget;

• Making predictions as close to reality as possible allows an efficiency of the activity of public authorities; the correct predictions allow the realization of budgets close to the needs of the society and achieve an increase of the social trust in the state mechanisms and at the same time a transparency of the decisions of the executive factors.

1. LITERATURE REVIEW

Predicting financial and economic time series is a complicated operation primarily due to unprecedented changes in economic trends and changes, as well as incomplete information. Frequent changes in legislation as well as rapid developments in recent years make it very difficult to make a rigorous financial forecast. Therefore, the assessment of forecast accuracy is necessary when using different forms of forecasting methods and, more specifically, forecasting using regression analysis, as they have many limitations in applications.

The main objective of this article is to investigate which forecasting methods provide the best predictions in terms of lower forecast errors and higher forecast accuracy. The best known method for univariate series is *Auto-Regressive Moving Average* (ARMA). In this method, the *Auto-Regressive* (AR) and *Moving Average* (MA) models are combined.

The univariate *Auto-Regressive Integrated Moving Average* (ARIMA) method is a special type of regression in which differences are taken into account in the model. Multivariate ARIMA models and *Vector Auto-Regression* (VAR) models are other popular prediction models that generalize univariate ARIMA models and univariate *Auto-Regressive* (AR) model allowing more variables.

Machine learning techniques and, more importantly, ***Deep Learning algorithms*** have introduced new approaches to prediction problems in which the relationships between variables are modeled in deep and layered hierarchies (Levity, 2021; Great learning, 2021; Flatiron school, 2021). Techniques based on machine learning such as *Support Vector Machines* (SVM) and *Random Forests* (RF) as well as algorithms based on Deep Learning, such as *Recurrent Neural Network* (RNN) and *Long Short-Term Memory* (LSTM), have captured great interest in recent years given their multi disciplinary applicability (IBM, 2020).

Deep Learning methods are able to identify the structure and pattern of data sets, such as nonlinearity and complexity in time series forecasting.

LSTM has been used in many fields, such as: natural language processing (Tarwani & Edem, 2017); handwriting recognition (Graves et al., 2008); speech recognition [(Robinson, 2002), (Sak et al. 2014)]; time series prediction [(Hochreiter & Schmidhuber, 1997), (Gers et all, 2002), (Zhang et all, 2019), (Brownlee, 2016), (Gamboa, 2017), (Roondiwala, 2017)]; modeling economic and financial data series [(Heaton et all, 2016), (Kohzadi, 1996), (Giles et al. 2001), (Huck, 2009), (Xiong, 2015)].

An interesting and important research question is the comparison of the accuracy and precision of traditional forecasting techniques in relation to forecasting algorithms based on

Deep Learning. To our knowledge, there is no specific empirical evidence for the use of the NRM / LSTM method in forecasting economic and financial time series to be able to assess performance and compare them with traditional econometric forecasting methods, such as ARMA.

This paper aims to make a comparison between the classic *Auto-Regressive* (AR), *Moving Average* (MA), *Moving Average Auto Regressive* (ARMA) forecast models compared to Deep Learn (RRN and LSTM models).

The models were chosen in order to be able to compare the result in terms of errors and accuracy.

2. DATA AND METHODOLOGY

2.1 CHRONOLOGICAL DATA SERIES

Timelines are everywhere. We can see them in weather forecasts, price fluctuations, the evolution of expenses or incomes, evolution trends (Moore's law), audio samples, power consumption, etc.

Chronological data series are an ordered sequence of values, usually at equal intervals in time, distributed per year, month, day, hour, minute or even microseconds. If for each unit of time there is only one value associated with each step, the chronological data series are called ***univariate time series.***

Figure 2. An univariate chronological data series

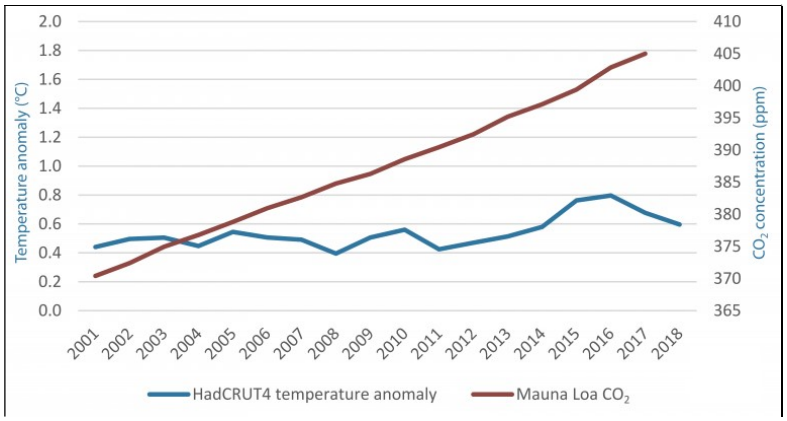
Source: Web

Time series that have two or more values associated with each moment are called ***multivariate time series***. Examples of multivariate chronological data series can be found in

the evolution over time of births in relation to deaths, in the evolution over time of heating in relation to the increase of CO2 emissions.

These time series can be treated as two or more univariate time series, but they can also be treated together in order to highlight the relationships between them.

Figure 3. A multivariate chronological data series



Source: Web

The analysis of time series has many ***applications***, the most obvious is *the forecast of the evolution of the series based on previous experiences.*

For example, analyzing the incomes for the last four years, we could predict what the income will be for the next year. This forecast can be made at different scales of the temporary step (annual, quarterly or monthly forecast can be made).

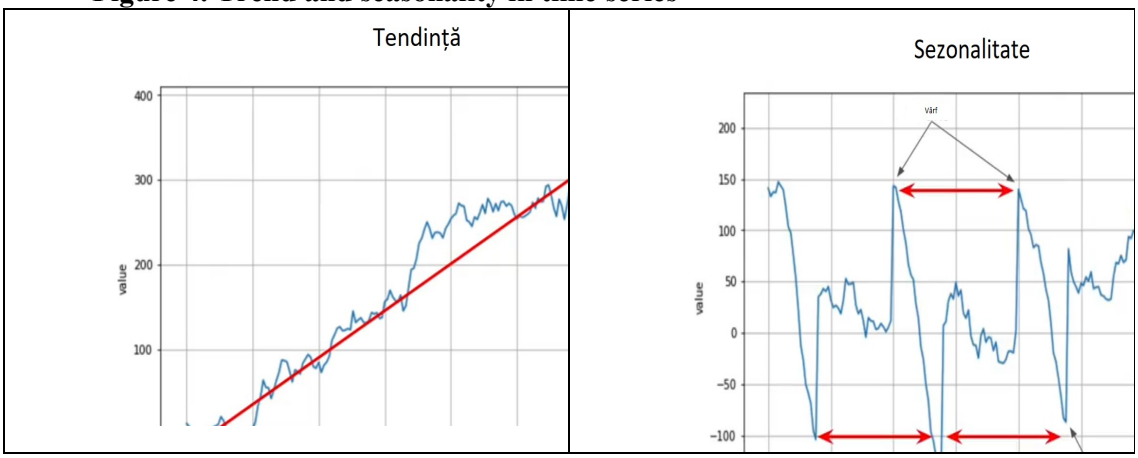
The chronological data series can be also used to analyze the processes or factors that generated the evolution of values (we can analyze the socio-economic factors that influence the evolution of the income of a Territorial Administrative Unit).

Some series evolve gradually increasing or decreasing over time. The characteristic of ascending or descending evolution is called ***tendency*** (ascending tendency or decreasing tendency).

Other series have repetitive time periods in which evolution retains the same characteristics. The evolution of these series is defined by peaks and depressions that are repeated at intervals and are called ***seasonal*** series (near the payment deadlines associated with taxes and duties there is a significant increase (peak) of budget revenues, at the same time, during the holidays and vacations there is a decrease (depression) of payments to the budget.

There are very common series that have a tendency to increase or decrease but also have seasonality.

Figure 4. Trend and seasonality in time series



Source: Web

2.2 METHODOLOGY

The main objective is to identify the biunivocal influences between causal socio-economic factors and the public budget (directly or indirectly through its elements: public revenues and expenditures), in a framework that starts from the premise that both the state with its institutions and civil society play an important role.

The main methodological objectives are the following:

- Identifying the potential factors that influence the evolution of the public budget;
- Identifying the quantitative and qualitative properties of the identified factors, their dependence or interdependence; we will analyze potential determining factors in the evolution of incomes and expenses, we will achieve their normalization and structuring, analyzing both from the point of view of the temporal evolution and from the point of view of the spatial structure;
- Analysis of the regression methods applicable in relation to the implementation of Machine Learning algorithms. Within this objective we analyze the efficiency of different regression methods in terms of effectiveness. At the same time we eliminate the possible variables that are linearly dependent or that do not have a direct influence on the generated models. At the same time we implement artificial neural networks (simple and convolutional).

The research process uses *both empirical and theoretical research methods*, with both a qualitative and quantitative character, in which thematic analyzes are performed to highlight the characteristics and attributes of the analyzed data.

Quantitative research is performed using data from official statistical records as the primary source, but also data from forecast situations as a secondary source.

The data on the budget execution of revenues and expenditures of Targoviste municipality were extracted from the website "NATIONAL REPORTING SYSTEM" of the Ministry of Finance published within the "BUDGET TRANSPARENCY" program. The website can be accessed through the browser by navigating to the address:

https://extranet.anaf.mfinante.gov.ro/anaf/extranet/EXECUTIEBUGETARA/Rapoarte_Forexe

Multiple criteria search for export. The search parameters used were:

- Type of Report: Detailed budget execution
- Budget sector: 02 - Local budget (local government)
- CIF Main Authorizing Officer: 4279944
- CIF Public Entity: 4279944
- County: Dâmbovița

Figure 5. The source of data: Website ForExeBug

Source: own source

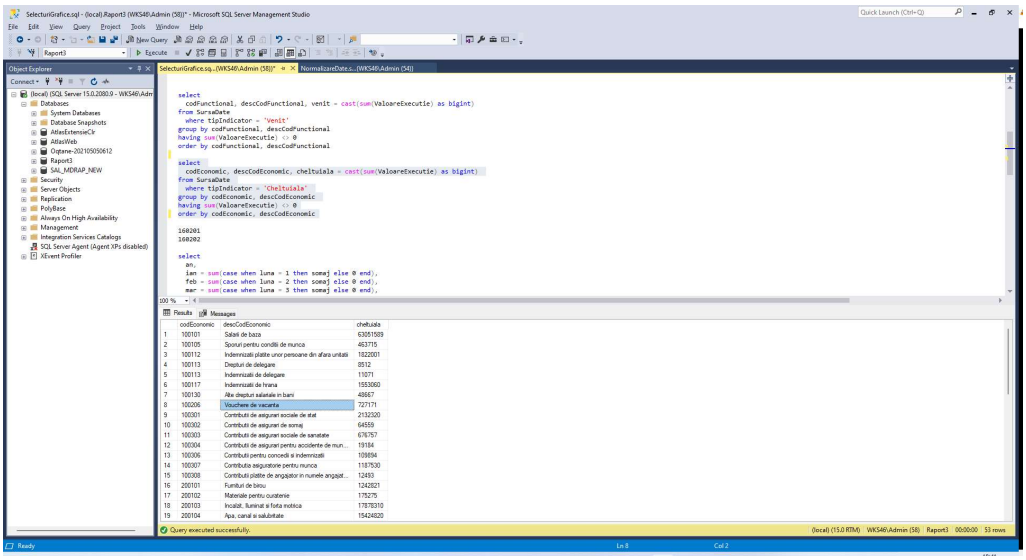
The data were extracted *for the period 2016 - 2021* in the form of *xml*. In order to be able to use the data sets throughout the period, we are analyzing, transfer, deserialization and normalization operations were necessary. A SQL Server 2019 Express Edition was used to process the xml and to organize, normalize, serialize, reserialize, and structure the data.

Normalization of data sets involved:

- Uploading data to the database
- Data deserialization
- The actual normalization of the data
- Uniformization of the field of values
- Export in csv format

It was used to export data in csv format the native function of SQL Server Management Studio (SSMS).

Figure 6. The result of running the SQL query



Source: own source

Based on the data analyzed by inductive methods, ***a qualitative analysis*** of the information is performed.

The conclusions of the interpretations, through ***inductive and deductive methods***, allow the establishment of relationships between primary data, financial statements and financial developments.

We aim to identify some distinctive elements, defining in order to make connections with activities of public institutions that are directly influenced by future budget provisions. Thus, it can be seen and validated the possibility of using as feedback information for modern analysis processes accounting data, to optimize the allocation of resources for future activities.

The Machin-Learning algorithms fit perfectly in a research activity, being necessary the structuring of the representative data, the realization of the successive analysis of the input data as well as of the obtained results. It is also necessary to optimize the indicators by identifying the linearly dependent elements and eliminating them from the analyzed data sets.

The public budget is a perfect Big Data source system for implementing a Machine Learning algorithm, because it allows us to define multiple dimensions for the same information contained.

For the realization of an assisted learning algorithm to make a budget forecast we have:

- ***Task S***: making a budget forecast
- ***Performance measurement P***: shift from previous budget executions
- ***Experience E***: making predictions for different sets of input data.

In the process of modeling algorithms using Machine Learning, we go through the following ***steps:***

1. *Defining the areas of interest*, the indicators for which the forecast is made ex (Expenditures: investments, personnel expenses, auxiliary materials, services, etc.; Revenues: direct, from transfer, attracted)
2. *Filtering the indicators* that are relevant in the monitored areas; Iterative procedures determine the impact of each indicator on the result. In this step we modify the parameters to perform local validations (validation set)
3. *Eliminates indicators that are linearly dependent*; By running successive micro predictions we eliminate data that are linearly dependent, that can be expressed by other data included in the learning set.
4. *Running distinct models for each measured indicator and determine the average errors*; In the optimization processes we calculate the average errors related to the reference set.
5. *Selecting the models with the optimal result* according to the minimum error.
6. *Validating the learned algorithm* by using the entire data set and compare it with the reference data.

Going through these steps we want to identify the intrinsic relationships of input data and use this data to analyze the current situation of public authorities and to make predictions based on data, learned based on Machine Learning algorithms.

3. RESULTS AND DISCUSSION

In the process we consider certain situations that can decisively influence the correctness of the resulting algorithm:

• We need to strike a balance between variation and movement and highlighting the variation by the fact that we have very different predictions for the analyzed data sets. We find the displacement when the prediction is systematically wrong.

• The number of parameters directly influences the results and the quality of the predictions. Through successive adjustments of the variation and displacement, we achieve an optimal framing.

• We need to refine the range of indicators for model stability within the input data space and eliminate the random events.

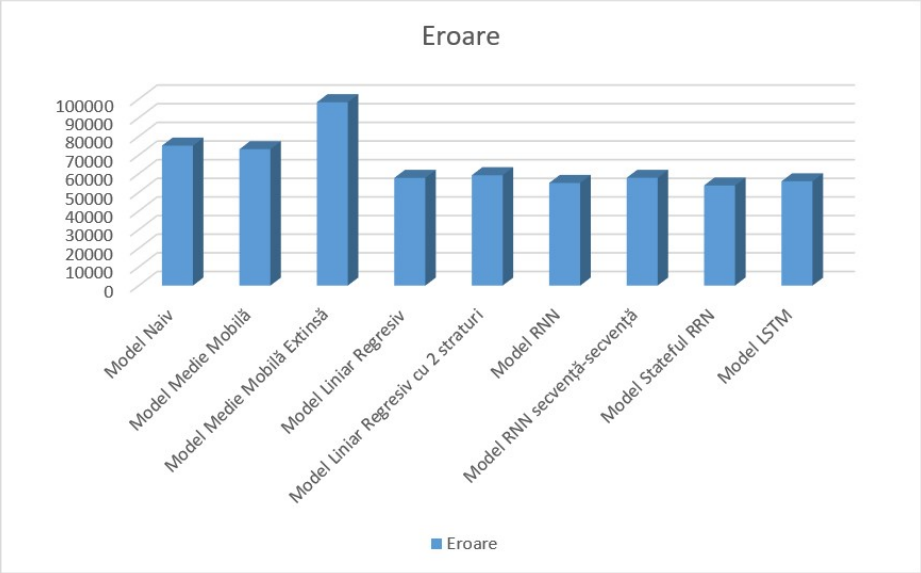
Table 1. The result obtained by using different forecasting methods

Model name	**Absolute Average Error**
Naïve Model	75049
Moving Average Model	73155
Extended Moving Average Model	98373
Regressive Linear Model	57767
Regressive Linear Model with 2 layers	59201
RNN Model	55041
Sequence-sequence RNN model	57795
Stateful RRN Model	53706
LSTM Model	55940

Source: own source

From the comparison of the results obtained by using different models, we obtain the table 1 and the following graph:

Figure 7. The result obtained by using different forecasting methods



Source: own source

From the data set and hyper parameters used for the prediction it results that *the classical RRN model has the smallest error.* The result is visibly better than the naive forecast.

CONCLUSIONS

The conclusions and proposals resulting from the analysis of the causalities and the interdependence of the analyzed factors, are intended to represent a decisional support for the state institutions and at the same time an element of understanding and transparency of the economic phenomena.

Financial planning is a process in which comprehensive strategies are developed and implemented in order to achieve the financial objectives of a public institution. These strategies are developed based on future projections, made for different time horizons.

Public institutions assess the likelihood of these projections and develop a response (financial plan) to mitigate adverse financial consequences. Although used mainly by firms, financial planning tools and techniques will eventually become part of current public administration practice.

After implementing the prediction models and calculating the absolute errors, it can be concluded that the phenomena that describe the chronological data series can be successfully modeled using Machine Learning algorithms.

Even if the data set was small and the analysis was limited to univariate data sets, the efficiency of recurrent neural networks proves to be quite good.

REFERENCES

1. CECCAR (2015), Practical guide for applying the accounting regulations regarding the individual annual financial statements and the consolidated annual financial statements approved by OMFP no. 1802/2014, Bucharest, CECCAR Publishing House
2. Coman, M.D., Denisa, C. M., & Florina, B. R. I. (2012), Managerial accounting–a source of information for an efficient management in SME, Procedia-Social and Behavioral Sciences, vol. 62, pp. 521-525
3. Duguit L., Traité de droit constitutionnel, Tome 2, La théorie générale de l'État, Boccard, Paris, 1923, pg. 144
4. Flatiron school, (2021), AI, https://flatironschool.com/blog/deep-learning-vs-machine-learning Accessed on May 2021
5. Gers F.A., Eck D., Schmidhuber J. (2002), Applying LSTM to Time Series Predictable Through Time-Window Approaches, Perspectives in Neural Computing, Springer, London, 2002
6. Graves A., Schmidhuber J., Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 2008
7. Great learning, (2021), https://www.mygreatlearning.com/blog/machine-learning-tutorial/
    - Accessed on August 2021
8. Heaton, J. B., Polson, N. G., and Witte, J. H., (2016), Deep learning in finance, February 2016, http://arxiv.org/abs/1602.06561.pdf
9. Hochreiter S., Schmidhuber J., (1997), Long short-term memory, Neural Computation Journal, 9(8):1735-1780, 1997
10. IBM (2020), Cloud, https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks Accessed on June 2021
11. Jason Brownlee J., (2016), Time Series Prediction with LSTM Recurrent Neural Networks in Python, Deep Learning for Time Series, 2016
12. Kanchan M. Tarwani, Swathi Edem, Survey on Recurrent Neural Network in Natural Language Processing, International Journal of Engineering Trends and Technology (IJETT), V48 (6), 301-304 June 2017, ISSN:2231-5381, www.ijettjournal.org
13. Kohzadi N., Boyd MS., Kermanshahi B., (1996), A comparison of artificial neural network and time series models for forecasting commodity prices, Neurocomputing. 10(2):169–181
14. Levity, (2021), AI, https://levity.ai/blog/difference-machine-learning-deep-learning
    - Accessed on July 2021
15. Moşteanu T., Buget şi trezorerie publică, Ed. Didactică Pedagogică, Bucureşti, 1997, p. 8
16. Robinson A. J., (2002), Connectionist speech recognition of Broadcast News, Speech Commun Journal, Volume 37, pg. 27-45
17. Roondiwala M., Patel H., Varma S., (2017), Predicting Stock Prices Using LSTM, International Journal of Science and Research (IJSR) 6(4)
18. Sak et al, (2014), Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition, Cornell University, 2014
19. Xuan Zhang et al, (2019), An Attention-based LSTM Model for Financial Time Series Prediction, Materials Science and Engineering, Volume 569, Issue 5