Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 Dec; 6 (10s2): 1649-1656

Developing An Interactive Web-Based Time Series Forecasting System With Deep Learning And LSTM For Student Enrollment Using Dash

Hassan Bousnguar[1*], Lotfi Najdi[2], Amal Battou[3]

[1*]IRF-SIC, Faculty of Science, Ibn Zohr University, Agadir, Morocco, h.bousnguar@uiz.ac.ma
[2]LIMA Laboratory, ENSA, Ibn Zohr University, Agadir, Morocco
[3]IRF-SIC, Faculty of Science, Ibn Zohr University, Agadir, Morocco

[*]**Corresponding Author:** Hassan Bousnguar
IRF-SIC, Faculty of Science, Ibn Zohr University, Agadir, Morocco, h.bousnguar@uiz.ac.ma

Abstract:

Forecasting methods are one of the most promising areas in the data analytics landscape. In this paper we demonstrate the added value of using the LSTM model to forecast the enrollment process in higher education context. We also proposed web-based systems that give the top university manager the ability to make correct decisions.
The system that we develop offers a dashboard to improve strategic decisions for the university to allow managers to make decisions for expansion to new sites or to create new courses.

Keywords: Deep Learning, LSTM, IDSS, Forecasting, Time series, Dash.

1. INTRODUCTION

Developing a system that provides tools for decision-makers to manipulate the historical data they have at their disposal, as well as allowing them to anticipate and forecast the number of new enrolments, has always been a challenge.
Converting data into valuable and useful knowledge for decision-making is one of the greatest challenges facing any strategic system. Institutions of higher education might solve that problem by using forecasting models to optimize resources and efforts. Forecasting is an important phase of the decision-making process, which is used in businesses and other organizations.
Forecasting has been a main domain within data analysis, which has caught the interest of many different researchers across different disciplines. Statistical researchers have been the first to lead research in this field. Two main methods are available in statistical approaches, qualitative methods and quantitative methods. Qualitative forecasting methods consist of the judgment of human experts in order to generate forecasts. Quantitative forecasting methods utilize past data to generate a forecasting model that extrapolates past and current behavior for the future forecast.
The objective of this work is to develop a web-application to plan future projects related to enrolment numbers of previous years in different higher education institutions.
This system implements the LSTM algorithm in an IDSS environment with the implementation of a dash library for graphs in interaction implementation.
The structure of the rest of this paper is as follows. In Section 2, basic forecasting techniques are introduced and forecasting work applied to the educational environment is summarized. Section 3 focuses on the methodology adopted. Section 4 presents the system and some results. This paper is concluded with a summary, discussion and an outlook for future work.

2. RELATED WORK

Forecasting is the method of making future predictions based on past and present data and most commonly through trend analyzes.
There are several types of methods in the literature and one cannot prefer each other because the choice is made on the basis of a number of parameters, including the area covered, the quality of the data used and also the quantity of information processed,
Another interesting method is that of the time series which used historical data as the basis for predicting future outcomes. They are based on the premise that history of past demand is a strong predictor of potential demand.
There are multiple algorithm of machine learning that are used in the forecasting process, RNN and LSTM Algorithms are networks that consist of standard recurrent cells such as sigma cells and tanh cells, which is a deep-learning NN,

which is explicitly designed to learn the long-term dependencies. It is able to remember information for long periods of time via the introduced gates (Elsheikh et al., 2021) ,
Other researchers are interested in new types of Forecasting methods based on machine learning algorithms. In the literature, some authors use nearest neighbors as the prediction value(Liu and Liu, 2017), multiple deep learning techniques are used like Recurrent Neural Network (RNN), long short-term memory (LSTM), or Bayesian Neural Network (Miao et al., 2020).
The design and development of models is important, but the exploitation of those models is much more important. The challenges faced when using models is the transition from the research environment to the decision-making environment.
Need to create an interactive decision support system, to facilitate the use of different forecasting models.
The concept of an interactive decision support system (IDSS) is very wide and their definitions depend on the perspective of the authors and the evolution of information technologies and studies on decision making (Ismail, 2008).
(Finlay, 1994) describes this type of system, in a brief manner, as "a computer system supporting decision making". Turban, on the other hand, proposes a more extensive definition, which also covers the composition of the SIAD: "A SIAD is an information system that is interactive, flexible, adaptable, and designed specifically to improve the quality of decision making. This system is based on both internal and external data, a set of models and user-friendly interfaces" (Turban, 2008)

3. METHODOLOGIES

The proposed system is built in order to provide a simple and efficient tool for the manager to have visibility on the evolution of higher education enrolment.
The system that we propose must be :
- Reliable: uses the most powerful models that give a high accuracy.
- Ergonomic: the system must have an appropriate graphic design with the use of the latest tools and libraries.
- Simple: the system must be simple, and consist of a minimum of windows and graphic objects.

The figure 1, illustrates the simplified architecture of our system.

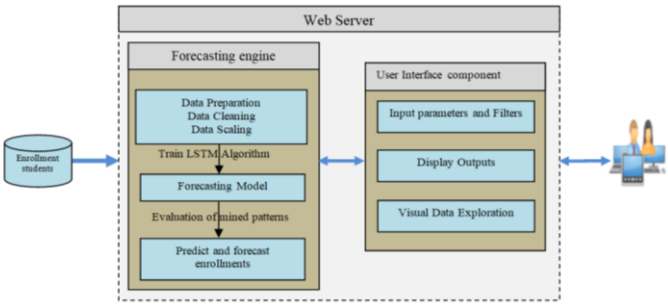


Figure 1: System Architecture .

The System that we are developing is composed of two main components, the Front-End Layer(user interface component) in which we manage the different interaction between the user and the system and also by using the appropriate libraries we provide exploitable dashboards for the user. The second component is the back-end layed (forecasting engine), in this part we define our forecasting model, and we prepare the results to transmit them to the user interface component.
The different components of the system are developed and managed using python language, both for the user interfaces, which are combined with HTML5 and JavaScript, and also for the business part are developed with Python.
Python is a high-level, object-oriented programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, largely used by scientists for data analysis, visualization, and AI modeling. Python provides a wide variety of data processing and graphical techniques, and is highly extensible.
For the user interface and data visualization we use Dash, which is an Open Source Python library for creating reactive, Web-based applications.
Python and dash was chosen for the advantage of the flexibility and strength of using python to run the data processing and machine learning algorithm, as well as the interactive and intuitive visualization of the data on the web.

1. SIAD

1.1 Definition

An interactive decision support system (IDSS) is a computer system that enables a company or organization to make better decisions. Interactive decision support systems serve management, operations, and planning within an organization and enable users to make decisions on problems that can change rapidly and are not easily defined in advance. IDSS is either fully automated, human-driven, or a combination of both.
The IDSS has a very particular position in the IT system. It is the end point for the IT system.

1.2 Constitution of a SIAD

According to (Marakas, 2003) the components of a SIAD (See Figure1) can be generally classified into five distinct parts:

- A database management system and associated database: which stores, organizes, sorts and retrieves data relevant to a particular decision context;
- A model base management system and its associated model base: which has a similar role to the database management system except that it organizes, sorts, and stores the organization's quantitative models;
- The knowledge engine: which performs tasks related to the recognition of problems and the generation of final or intermediate solutions as well as functions related to the management of the problem solving process.
- A user interface: which is a key element of the functionality of the overall system;
- A user: who is an integral part of the problem solving process

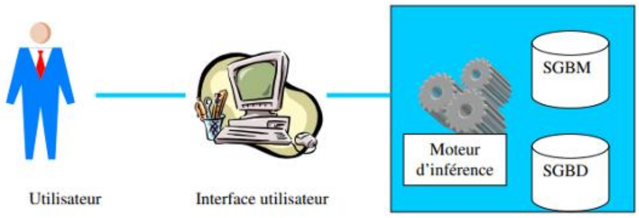


Figure 2: Architecture of interactive decision support system (Marakas, 2003)

2. User Interface

The web-based user interface helps users to load student's historical data, filter the courses , test the stationarity of our time series if we use a statistical approach and apply transformation if it isn't. Choose the parameters of our algorithm and visually inspect results of the forecasting algorithm. User preferences and inputs are then transferred to the python server for the execution of the main algorithm and the generation of data visualizations and summarizations before they are displayed afterward by this interface.

3. Forecasting engine

3.1 data preparation

Data preparation is considered as the main input of the interactive decision support system, the quality of the data depends on many factors such as the availability of data, the implication of human resource in data collection etc...
data in real word are incomplete, noisy, and inconsistent. An important task in data preprocessing is the task of completing the missing values in order to reduce the noise and correct the inconsistencies.
Some functions used in this phases :
- Missing values
Firstly, the function involves cleaning up any missing data values which confuse the whole data. Substitution with the mean value of the Time series, for this value in the list, has been used in our project.
- Detecting of outliers
This function is concerned with the treatment of outliers. An outlier is a data object that varies substantially from other objects as if a separate process was developed.
- Scaling Data
The third function is features scaling, which looks for limiting the variation range of independent variables or features of data. In data processing, it is also known as data normalization.

3.2 Forecasting module

In this paper we propose a web-application to forecast enrollments in higher education institutes. The user interface provides a rich IHM with multitudes of components which facilitates the navigation in different tabs of our web-application. The user can choose the dataset for training, and also can visualize the data in numerical and graphical representation. The implementation of LSTM is performed by a user interface that communicate with the forecasting engine.

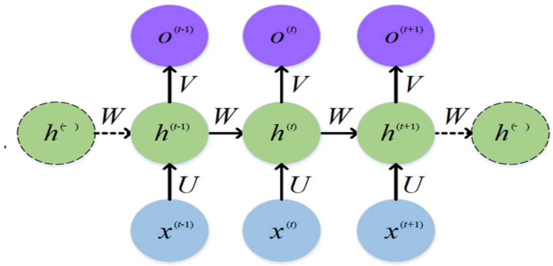


Figure 3: Standard RNN

Usually RNNs are networks that consist of standard recurrent cells such as sigma cells and tanh cells. Figure 1 shows a schematic of the standard recurrent sigma cell. The mathematical expressions of the standard recurrent sigma cell could be described as:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b),$$

$$y_t = h_t,$$

Where xt, ht, and yt denote the input, the recurrent information, and the output of the cell at time t, respectively; Wh and Wx are the weights;

The long short-term memory (LSTM) RNN, which is a deep-learning NN, Figure 2, is explicitly designed to learn the long-term dependencies. It is able to remember information for long periods of time via the introduced gates. There are three kinds of the gate and each one has a different task in the cell, the forget gate is able to discard redundant information, the input gate is able to select key information to be stored in the internal state, and the output gate is used to determine output information. In this case, the LSTM approach is able to store and update the key information efficiently over a long time without gradients vanishing.

4. A CASE STUDY

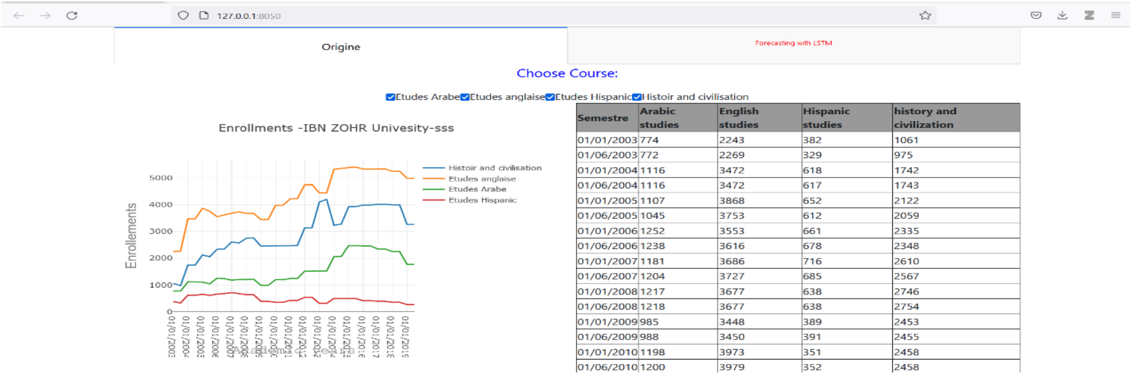
1- Data Presentation

In this section we demonstrate the relevant features of the proposed web-based system by simple examples. This analysis has been done on a dataset related to historical student's enrollments for four courses in college of law and economics studies.

"Table 1" shows the structure of the matrix used to train our algorithm, selected attributes of this dataset covering thirty four half-years of enrollment students. As shown in "Figure 2", the system displays and illustrates our matrix to give a graphical presentation, to assist the top manager in their analysis.

Table 1 : Matrix of time series

Years	Half-year	Institute_1	Institute_2	….	Institute_m
Y1	H1	A11	A12	…	A1m
	H2	A21	A22	…	A2m
Y2	H1	A31	A32	…	A3m
	H2	A41	A42	…	A4m
…	H1	…	…	…	…
	H2	…	…	…	…
Yn	H1	Y(2n-1)1	Y(2n-1)2	…	A(2n-1)m
	H2	Y(2n)1	Y(2n)2	…	A(2n)m



Semestre	Arabic studies	English studies	Hispanic studies	history and civilization
01/01/2003	774	2243	382	1061
01/06/2003	772	2269	329	975
01/01/2004	1116	3472	618	1742
01/06/2004	1116	3472	617	1743
01/01/2005	1107	3868	652	2122
01/06/2005	1045	3753	612	2059
01/01/2006	1252	3553	661	2335
01/06/2006	1238	3616	678	2348
01/01/2007	1181	3686	716	2610
01/06/2007	1204	3727	685	2567
01/01/2008	1217	3677	638	2746
01/06/2008	1218	3677	638	2754
01/01/2009	985	3448	389	2453
01/06/2009	988	3450	391	2455
01/01/2010	1198	3973	351	2458
01/06/2010	1200	3979	352	2458

Figure 3: plotting of time series

2- LSTM

LSTMs are sensitive to the scale of the input data. It can be a good practice to rescale the data to the range of 0-to-1, also called normalizing.

This default will create a dataset where X is the number of enrollments at a given period (t) and Y is the number of enrollment periods at the next time (t+1), Figure 4.

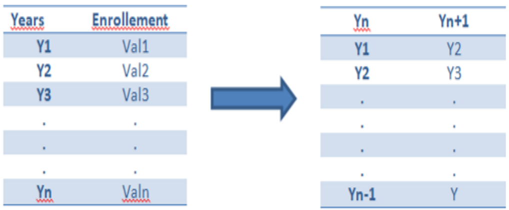


Figure 4: Transforming of time series to LSTM format

the system allows decision makers the possibility to choose the different parameters needed to build the LSTM models. We suppose that decision makers are not always comfortable with LSTM, and they may not be able to interpret the meaning of LSTM parameters, therefore we suggest optimal values that will be a starting block for our system.

We also offered the possibility to choose the courses over a list of courses in the faculty of humanities and social sciences.

the forecast result will differ according to different parameters of the model, as well as the time series, for some parameters will be good for a series and not for another one.

The following figures illustrate results for different values of the hyperparameters (number of epochs and batch size).

For example, for English studies course best accuracy is obtained with 200 epochs and a batch size of 8, while for history and civilization course best accuracy is obtained with a value of 100 epochs and a batch size of 16.

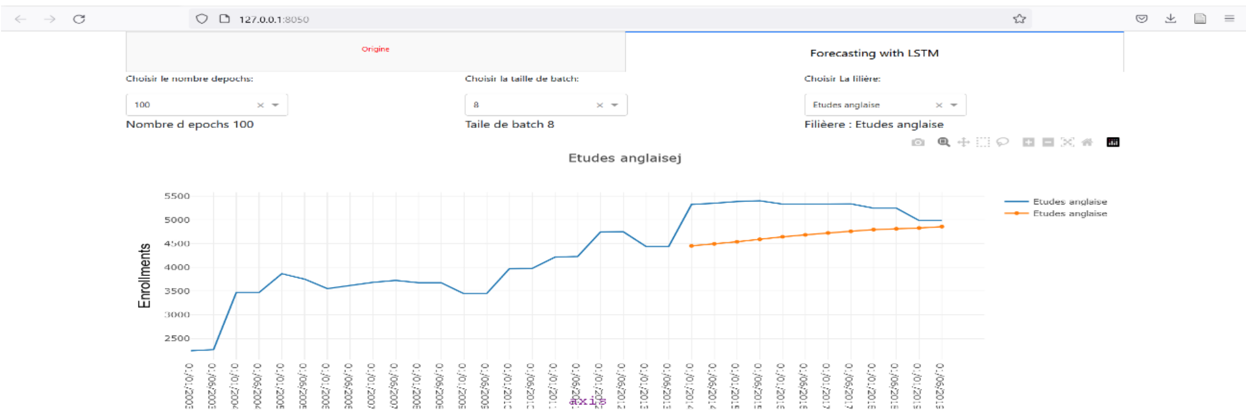


Figure 5: Forecasting for English studies with 100 epochs and batch size 8.

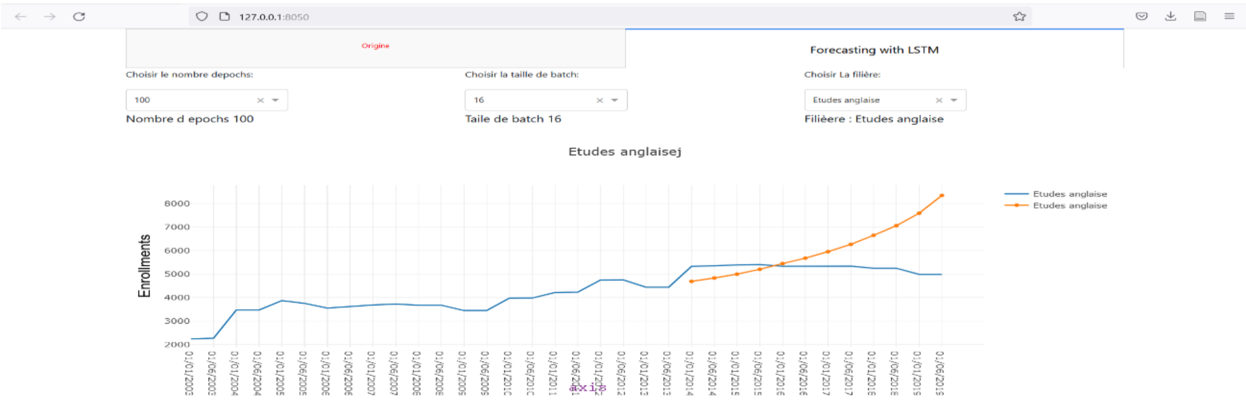


Figure 6: Forecasting for English studies with 100 epochs and batch size 16.

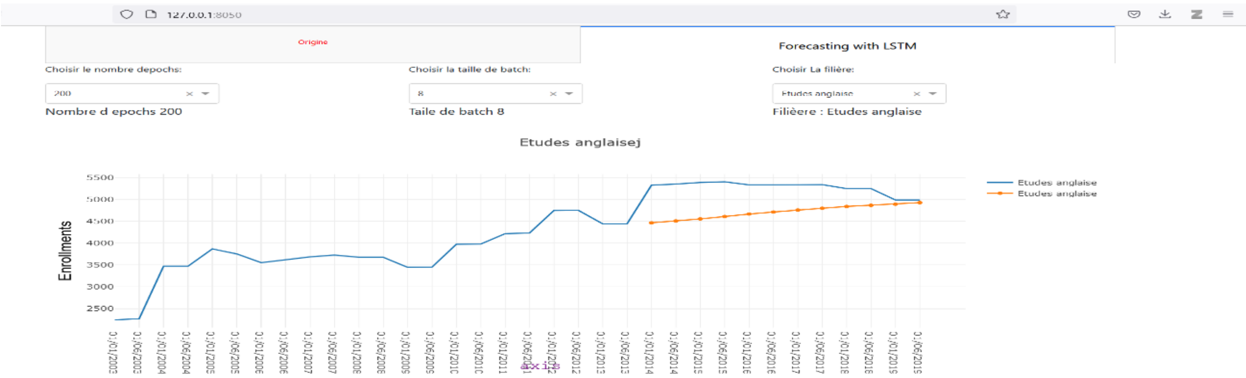


Figure 7: Forecasting for English studies with 200 epochs and batch size 8.

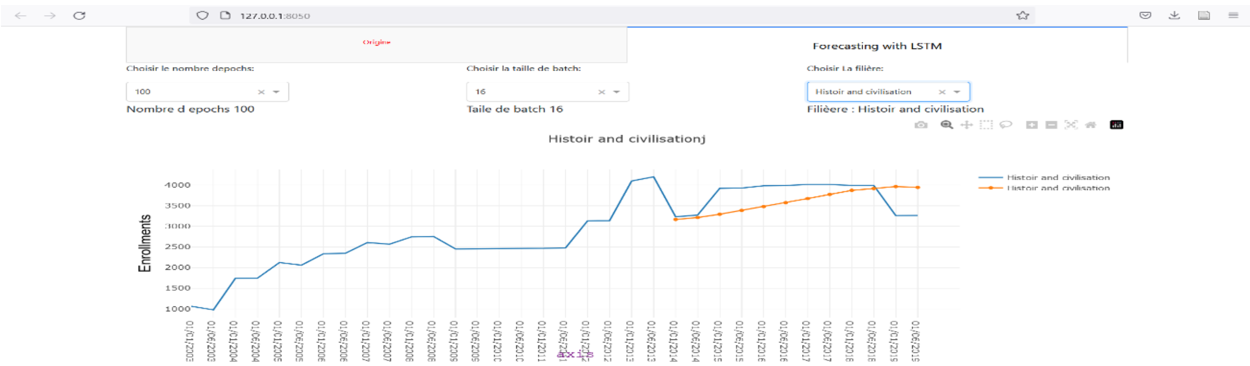


Figure 8: Forecasting for History and civilization with 100 epochs and batch size 16.

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 Dec; 6 (10s2): 1649-1656

Figure 9: Forecasting for History and civilization with 100 epochs and batch size 8.

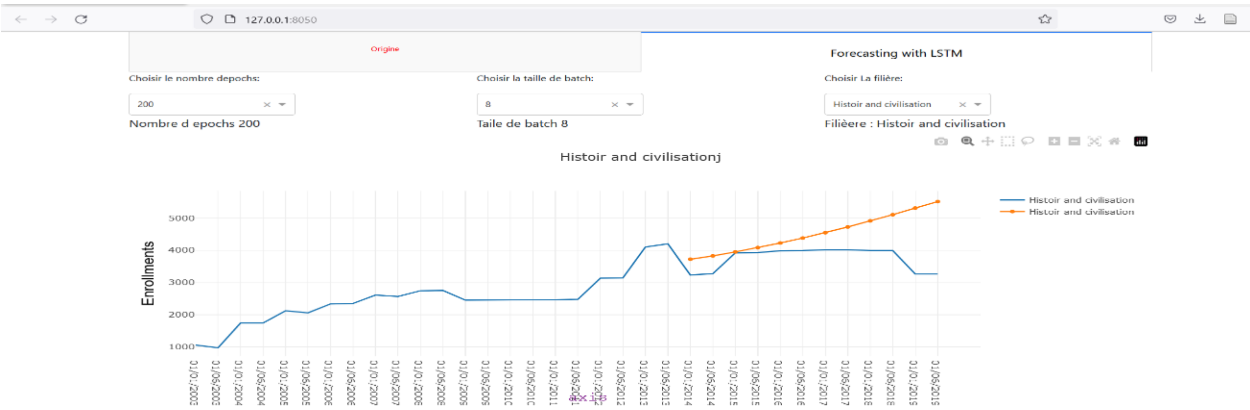


Figure 8: Forecasting for History and civilization with 200 epochs and batch size 8.

We followed a step secession to apply LSTM in our case, which will enable us to extract the useful information from our algorithm, in our user interface. The user can choose the number of epochs and the depth of our LSTM algorithms. The web interface allows decision makers to interactively achieve forecasting analysis using inputs in the side panel. For example he can choose the course from a range of checkList, choose hyper-parameters for dropdown components or let the system to affect tuning value of those hyper-parameters. The system performs necessary computing tasks and immediately populates the main panel with different tables, plots for analyzing and visualizing discovered behavior in our time series. These outputs are organized among tabs to enable the user to assess results within a simple and ergonomic design.

5. DISCUSSION AND CONCLUSION

This work discussed and implemented forecasting based models to explore the enrollment of students in several courses in the License program. The forecasting-based system was able to make an accurate forecast for future enrollments.

This web-based system has rendered forecasting tasks more usable for decision makers and academic administrators, by enabling them to select the studied courses, hyper-parameters or just let the system use tuning hyper-parameters.

Our system consists of two parts, the first one concerns the forecasting engine, where several steps are performed before the result can be displayed to the user, a first task of data collection and cleaning has been implemented and then the model is built with the optimized hyper-parameters, Then the second part of this system, which is visible to the user, is the web-based interface, which allows the decision maker to interact with the system and also control it by giving him the possibility to update the hyper-parameters

We have realized that the obtained results using different alternatives of hyperparameter values, depending on the time series we use, sometimes we can find that batch size 8 is more accurate than 16, for some series, but for others it is the opposite. The same thing for the number of epochs, sometimes increasing the number of epochs improves the accuracy, but in other cases the accuracy is better if you decrease the number of epochs.

REFERENCES

1. Elsheikh, A.H., Katekar, V.P., Muskens, O.L., Deshmukh, S.S., Elaziz, M.A., Dabour, S.M., 2021. Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate. Process Saf. Environ. Prot. 148, 273–282. https://doi.org/10.1016/j.psep.2020.09.068
2. Finlay, P., 1994. Introducing decision support systems. Blackwell Pub.
3. Ismail, Z., 2008. A decision support system for improving forecast using genetic algorithm and tabu search. ARPN J. Eng. Appl. Sci. 3, 13–16.
4. Liu, Q., Liu, C., 2017. A Novel Locally Linear KNN Method With Applications to Visual Recognition. IEEE Trans. Neural Netw. Learn. Syst. 28, 2010–2021. https://doi.org/10.1109/TNNLS.2016.2572204
5. Marakas, G.M., 2003. Amazon.com: Decision Support Systems in the 21st Century: 9780137441860: Marakas, George M.: Books [WWW Document]. URL https://www.amazon.com/Decision-Support-Systems-21st-Century/dp/013744186X (accessed 1.24.22).
6. Miao, K., Han, T., Yao, Y., Lu, H., Chen, P., Wang, B., Zhang, J., 2020. Application of LSTM for short term fog forecasting based on meteorological elements. Neurocomputing 408, 285–291. https://doi.org/10.1016/j.neucom.2019.12.129
7. Turban, E. (Ed.), 2008. Business intelligence: a managerial approach. Pearson Prentice Hall, Upper Saddle River, N.J.