The LSTM technique for demand forecasting of e-procurement in the hospitality industry in the UAE

Elezabeth Mathew, Sherief Abdulla
Faculty of Engineering and Informatics, British University in Dubai, United Arab Emirates

Article Info

Article history:

Received May 26, 2020
Revised Oct 5, 2020
Accepted Nov 11, 2020

Keywords:

Deep learning
Demand forecasting
E-procurement
Hospitality
Long- and short-term memory
Machine learning

ABSTRACT

The hospitality industry is growing at a faster pace across the world which has resulted in the accumulation of a huge amount of data in terms of employee details, property details, purchase details, vendor details, and so on. The industry is yet to fully benefit from these big data by applying ML and AI. The data has not been fully investigated for decision-making or revenue/budget forecasting. In this research data is collected from a chain hotel for advanced predictive analytics. Descriptive and diagnostic analytics is done to an extent across the hotel industry, whereas predictive and prescriptive analysis is done rarely. Demand forecasting for spend and quantity is done using the LSTM technique in e-procurement within the hospitality industry in the UAE. Five years of historical data from a chain hotel in the UAE is used for deep learning in this study. The results confirm the ability of LSTM model to predict e-procurement spend and order forecast for six months. LSTM time series analysis is considered the most suitable technique for demand forecasting to optimize e-procurement.

Corresponding Author:

Elezabeth Mathew
Faculty of Engineering and Informatics
British University in Dubai
United Arab Emirates
Email: 2016152110@student.buid.ac.ae

1. INTRODUCTION

Research and development (R&D) in the area of artificial intelligence have drastically advanced in the last few decades [1]. As a result, in many fields, R&D departments have been integrated for the same purpose, especially because a large amount of data is being accumulated in all industries [2]. The hospitality industry has only recently adopted artificial intelligence in a systematic manner [3]. The primary reason for late adoption is that there is limited collaboration between the hospitality industry and the practitioners of academic research [4]. A few early studies in the discipline were anecdotal and did not make any contribution to the industry, nor to academic research, because those studies focused on the individual operation or the locality only [5]. Another major reason for late adoption is that adoption requires a large amount of time and money, so the industry, especially small and medium enterprises, were reluctant to invest in this area [6]. Furthermore, senior management across the hospitality industry initially found it difficult to grasp the concept of artificial intelligence, as they could not relate it to any business benefits or profit [2]. Considering the impact of process effectiveness, the hospitality industry has to adopt artificial intelligence to continuously provide the customer with the best deals a quality experience. This is especially relevant for the hospitality industry, for which customers are everything; the requirements and expectations of customers keep on changing, so it is important to manage data more effectively to maximize business revenue and profitability [7].

The study shows that knowledge gathered online or offline should be used further for data analytics and predictions and, thus, to bring greater benefit to organizations efficiently and effectively. In the study, data has been collected for a chain of hotels in the UAE for the past five years for predictive and prescriptive analysis. The study is valuable for e-procurement in the hospitality industry as it will enhance cash flow management and foster improved profitability. Descriptive and diagnostic analytics is done to an extent across the hotel industry, whereas predictive and prescriptive analysis is done rarely [8-9].

The budget for the given financial year should be developed and approved beforehand in any organization. Most of the time, budgets are either over- or under-forecast due to the bullwhip effect ["The bullwhip effect is a distribution channel phenomenon in which forecasts yield supply chain inefficiencies" (Wiki)]. There exists no mechanism to predict a future budget accurately.

The purpose of the research is to evaluate the performance of several AI (Machine Learning) algorithms to optimize e-procurement based on data to enhance demand forecasting accuracy. In this research, the long short-term memory (LSTM) technique is used for the time series prediction for six months. From the time series graph plot, it is evident that the technique used for forecasting is done accurately for the six months, as the actual and the predicted data are accurate. The study done by [10] on forecasting economic and financial time series in 2018 depicts that the LSTM algorithm is superior and outperform traditional-based other time series algorithm as the error reduction rates have gotten is between 84% - 87% [11-12]. Moreover, the study also specified that "epoch" the number of training times has no effect on the performance of the trained forecast and it shows just an unsystematic behavior [10]. The study conducted by [13] states the LSTM techniques are more efficient than many other univariate forecasting methods in their study on multiple seasonal cycles. Often big data has a large amount of sequential time series as in sales demand for related products in retail with multiple related time series. The paper by [14] suggests the LSTM technique is popular to build a global model across multiple related time series.

2. THEORETICAL FRAMEWORK

As per [15], the three categories for data analytics are descriptive, predictive, and prescriptive. In the paper by [9] and [16], they suggest that there is four data analytics including diagnostic analysis. Furthermore, as per the latest study by [17], cognitive analytics is the fifth data analysis in his analytics maturity model.

- Descriptive analytics is conducted on existing data or processes to identify problems and opportunities. In most cases, descriptive analysis is undertaken in most organizations as part of report generation, which itself is part of online analytical processing (OLAP); it addresses the question "what happened?".
- Diagnostic analytics is another traditional analytics approach in which decisions are taken by assured delays. The delay is due to the necessity to gather and analyze data and then interpret them. Diagnostic analytics supports the finding of consistencies and measurable relations between variables via historical data analysis; it addresses the question "why did it happen?".
- Predictive analytics is mainly used to forecast and predict using carefully worked-out algorithms and programming to determine illustrative patterns within the data. Various techniques and programs can be used to do this, which include web/data mining using the Python data analytics tool; it addresses the question "what will happen in the future?".
- Prescriptive analytics is used to complete high-level decision-making and find alternatives to meet strategic goals, which are described by high dimensions and density to enhance business performances; it addresses the question "what action is to be taken?".
- Cognitive analytics are based on real-time analytics. Data is collected, organized, analyzed, and interpreted primarily to identify regularities and patterns. These models are kept in the data stream, which affects the collaboration between guests and the organization. That is the method of communication with the guests and the reception of a brand which involves real-time monitoring of the situation and guests' behavioral patterns, and finally selecting a behavioral pattern that is optimal; it addresses the question "what's the best next action?".

Prescriptive and predictive analytics plays a significant role in fostering the organization's importance in making effective decisions [18]. In this paper, the researcher is focused on integrating big data business analytics (BDBA) and supply chain analytics (SCA) to manage uncertainties in the organization by applying advanced predictive analytics.

"Big data collected from both internal and external services enable hospitality practitioners to make use of historical databases to forecast and predict business trends such as occupancy, rates and yield, labor costs and investment decisions" [19-20]. However, limited research has been conducted on the data gathered during procurement, although there are a lot of possibilities for useful research. Moreover, the available data does not follow any standardized format, so it is a challenge to retrieve and process to make a reliable sense of this large body of data. The hospitality industry involves a large number of stakeholders in the form of employees, suppliers, managers, dealers, customers, guests, etc. The data collected can be helpful to all of these stakeholders only if they can access and analyze it. The management relies on the historical and contextual data for prediction and forecasting of future trends in pricing to attract customers.

Previously, internal big data from previous years are used to support decision-making and forecasting on pricing, rate rules, distribution channel management, and inventory optimization. However, recently a small number of organizations have started to use a neural network to analyze the given input with the expected output to obtain a better multi-attribute decision or prediction regarding the result. Contextual information can be used to calculate the best price from vendors to gain long-term profit for those parties involved [21]. That is why the organization needs to combine both internal big data and contextual data to generate an efficient result. In this research, LSTM uses a peculiar version of recurrent neural network (RNN) for time series prediction. RNN can be considered as a feedforward neural network with feedback loops or backpropagation through time. Before the neurons get momentarily disabled, it fires for a limited time and enables other neurons to fire at a later stage. It has an additional time variable that is not in multi-layer perception (MLP). This additional feature will allow the model to not just use the current input but also the earlier input.

Long short-term memory (LSTM): S. Hochreiter and J. Schmidhuber introduced LSTM in 1997 [22-23]. LSTM is an improved version of RNN. RNN has an architecture that embeds the memory concept, yet this suffers from a certain problem called “the vanishing and exploding gradient problem”; LSTM is a solution that avoids this problem. The notion in LSTM is to pile the last output in memory and use that as input for the following step [21-24].

Effective supply chain management (SCM) is a critical success factor for any business. Managing cost and inventory in a multi-national hotel chain structure is a tedious task, as it is too multifaceted to predict the demand of the majority of the commodities [24]. Globally, travel and tourism have evolved massively due to social, political, and technological advancement in recent years [25]. As a result, cost and benefits may rise too due to unusual demands for resources [26]. Hence, accurate forecasts are vital for each stakeholder where they try to exploit the growth in market demand and balance local ecological and supply chain capacities [20]. The optimization of the supply chain is vital for any organization that is involved in buying and selling, as these procedures may openly affect customer service, inventory and cost, and reaction to the ever-changing situations. Therefore, decision-makers in SCM should think through basic uncertain events while combining the goals and objectives of the various processes involved [27].

3. RESEARCH METHOD

It is important to select a set of strategies that fit the research type. In this study, most of the questions are *how*, *what,* and why questions, so it is better to use case studies and experiments of different historical examples [27]. “A case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not evident; and in which multiple sources of evidence are used” [28].

The hotel chain company studied is referred to as *The X Hotels*. *The X Hotels* is a global brand having more than 2,500 properties across the world. In this study, we are considering only *The X Hotels'* properties located in the UAE. The name of the hotel company, its vendors, and suppliers have been disguised throughout the study for confidentiality. The data collected from *The X Hotel* is extracted, cleaned, analyzed for further reporting and prediction purposes, which could be beneficial for the key stakeholders in this study. Data analytics is conducted using Python 3.6. The methodology used for ML techniques is adapted from two books, namely ‘python for machine learning’ by [29], and ‘python data science cookbook’ by [30]. The proposed method is used for the construction of the demand forecast model for *The X hotels* proceeds as follows:

- Exploration of the dataset: The dataset is opened in the python environment and the data dictionary of the attributes is also included. The sample consists of five years of data for demand forecasting, which includes almost 4.5 million records.
- Data mining: Missing value concerns are not included in the data as only mandatory fields from the system are used. The issues in the dataset are mostly due to human error at the data entry stage. These errors will not be identified unless it is noticed during the data analysis stage. Making sure the CSV format is well-maintained was one of the issues that consumed time as the data set involved five years of data. The five-year dataset has precisely 4,503,218 unique entries. The data for demand forecasting is grouped into a monthly forecasting format.
- Baseline Forecast: A simple time-series graph is plotted to provide a baseline for comparison purposes. For quantity forecasting for a particular hotel and a particular product bought is found.

- Data Transformation: If there is an increasing or decreasing trend, the data is not stationary. It is necessary to first transform the data to make it stationary. Second, the time series must be converted to set features for the LSTM model. Third, the data must be scaled for optimum results.
- Feature Selection: Feature selection is especially important for any forecasting and predictive modelling. Features are selected carefully to avoid multi-collinearity and mitigate any redundant features that have a high correlation and, thus, to improve the overall performance. In this research, consumption, seasonal, room size, or other information were not available. The only available data was quantity, unit price, receiving date, property ID, product name, category name, and purchase type. Therefore, based on these variables all possible features are constructed. The adjusted R-squared is calculated after adding each feature to identify the optimum number of features.
- Test/Train Prediction: A comparison between the prediction accuracy was then conducted using the train/test split method. The test size for comparison was 0.2, which is 80% of the dataset, which was used for classifier training and the remaining 20% of the dataset for testing. Figure 1 summarizes the steps of the proposed method.

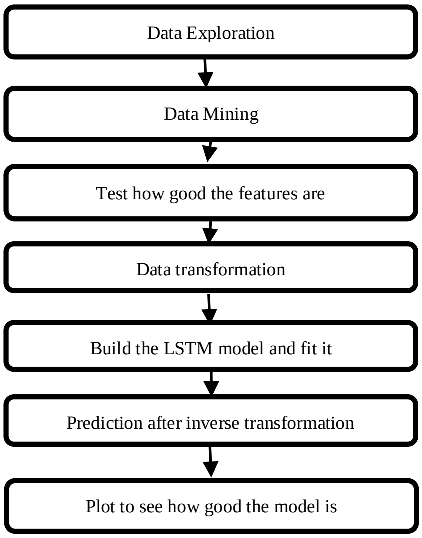

Figure 1. A proposed method for demand forecasting

4. RESULTS AND DISCUSSION

Demand forecasting gives an estimate for the number of commodities and services that each property will purchase in the foreseeable future, which will optimize top management decision-making [30]. The data considered are sequentially arranged in time. This dataset describes the monthly number of sales of all commodities over five years. First, a baseline of performance is developed for a forecast problem. Then data are organized, improved, and used to estimate an LSTM RNN for time series forecasting. The dataset has four columns showing the purchase dates, purchasing property, purchased items, quantity, and spend. The task is to forecast monthly total spend and total quantity. Data are aggregated at the monthly level and sum-up the spend and quantity column.

Step 1: As part of the transformation, a simple line plot is drawn as displayed in Figure 2 to see if there is an increasing or decreasing trend for monthly spending and monthly orders. Figure 2 clearly shows an increasing trend. Then the data is divided into training and test data. The experimental test setup data will be modelled on training data and predict for test data.

Step 2: Baseline forecast

A good baseline forecast for a time series with a linear increasing trend is a persistence forecast where the prior time is used to predict the current time. A rolling forecast scenario is made by shifting the training spend data once. An error score based on root mean square error (RMSE) is calculated based on the model developed to summarize the accuracy. In this case, the error is more than 9221.876 over the test dataset. A plot is drawn that displays the training set and the departing predictions from the test dataset is depicted in Figure 3. In the persistence model predictions, the predicted model is 1-step behind actuality. There are a rising trend and month-to-month noise in the spend figures, which shows the limitations of the persistence technique.

Step 3: Data preparation for LSTM

The LSTM RRN has the potential of educating an extensive series of datasets. The data preparation is conducted as follows. First, the data is compiled as a supervised learning model for machine learning, then it is ensured as stationary, and finally, the dataset is configured on a particular scale.

Step 4: Feature selection

Supervised ML requires input and output variables and uses a model to acquire the mapping function from the input to the output. The objective is to approximate the real underlying mapping even when the new dataset is used, it can complete predictions. For time-series observations this is implemented by using prior time steps as input variables and the current or next time step as the output variable. This method is called the sliding window method or lag method. In this case, the multistep sliding window forecasting is conducted. By applying lags 12 times there exist 12 lag long input sequences to predict the output. Transforming the structure of the data into a stationary format is undertaken by differencing. To find the difference, the prior data is subtracted from current data to remove the trend and to show the differences in the dataset. The calculated difference is shown in the Spend Diff column. Time series is stationary if they do not have trends or seasonal effects. The stationary time-series graph plotted is shown in Figure 4.

Before using it for modeling it is important to check if it is useful for prediction, which in this case means the adjusted R-squared is found. Adjusted R-squared is identified, if greater than 0.5 is moderately good, and more than 0.7 can be considered very good bonding. The adjusted R-squared value, which explains how much variation of difference variable is explained, is notable as the score is 79%; this is better than the persistence model that achieved an RMSE of 9221.876 over the test dataset. Feature variables are ready to build the model after scaling the data. Before scaling, the data should be split into train and test sets. For the test set, the last six months' spend is selected. LSTMs presume data to be within the rule of the activation function used by the network. The preferred range for a time series data is -1 to 1. In this case, MinMaxScaler is used for this transformation. To compile the RNN, the loss function and optimization algorithm must be given. The code block also prints how the model improves itself and reduces the error in each epoch. At the end of the run, a line plot of the custom RMSE metric is created is displayed in Figure 5. The model is now ready for prediction.

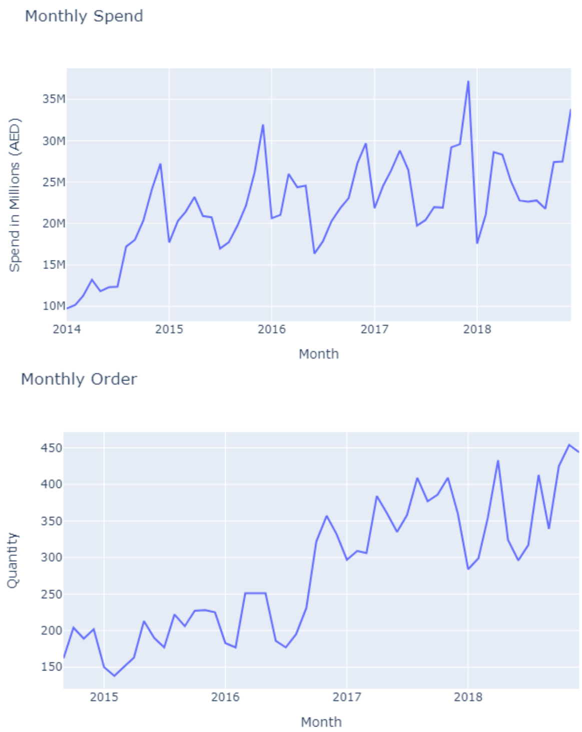

Figure 2. An increasing trend for spending. The increasing trend shows that there are behavioral changes over time and the graph is not stationary. As part of data transformation, the data need to be converted to stationary before converting to supervised learning with feature sets for the LSTM model

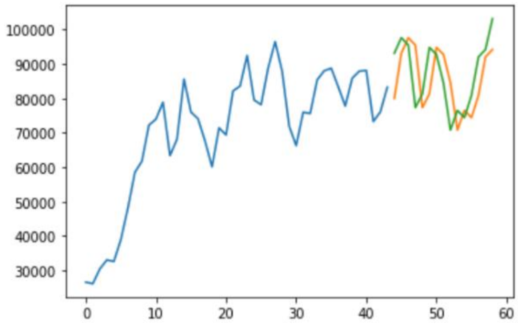

Figure 3. Baseline graph for spend forecasting. Notice that the predicted model (orange curve) is 1 step behind the actual values (green curve). This shows that the persistence technique has limitations in prediction as it did not avoid the noises in the spend values

Figure 4. Stationary time-series graph. Notice that there is no increasing or decreasing trend showing there are no behavioral changes over time. A stationary graph is a flat looking series, without trend, constant variance over time, and no periodic fluctuations

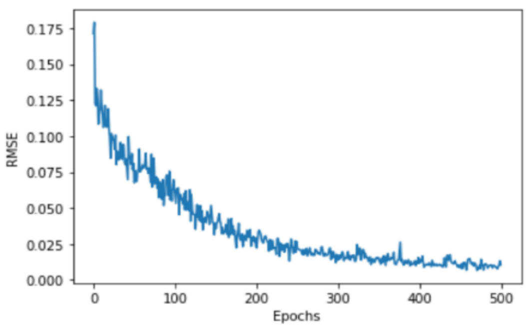

Figure 5. RMSE error graph. The RMSE value is 0.0137 at the end of the 500 epochs run with 8 layers

Step 5: Spend & quantity forecasting

The results of prediction look similar but they do not reveal much because it is the scaled data that demonstrate the difference. To see the actual, spend prediction first, the inverse transformation for scaling is conducted. Second, the data frame is built to show the dates and predictions. Transformed predictions display the difference. The actual values and predicted values for spend and quantity is displayed in the Table 1. Calculated predicted spend numbers should be plotted with the actual spending. Calculated predicted spend should also be shown in the same data frame as given in Figure 5 for quick comparison. This is an impressive prediction as an increase is shown before it happened, which makes the top management ready to manage cash flow more efficiently as seen in Figure 6a.

A similar study is done using quantity ordered for a particular product from the X Hotels. The initial steps in data transformation included converting the data to stationary, convert time series to a supervised model for having a feature set for the LSTM model, and scaling the data. As in the earlier study of spend forecasting, lag 1 to lag 12 is assigned values by using shift command. Calculated predicted quantity is shown in the same data frame with the actual quantity ordered as given in Figure 6b. This is indeed an impressive prediction as an increase is shown before it happened, which makes the top management ready to manage cash flow easier.

Table 1. Actual vs predicted for spend and quantity

Date	Actual Spend	Predicted Spend	Actual Quantity	Predicted Quantity
01/07/2018	22626740	18705284	317	306
01/08/2018	22781770	23740806	413	368
01/09/2018	21789480	25379430	339	427
01/10/2018	27425110	22904167	425	408
01/11/2018	27493800	32998242	454	450
01/12/2018	33823440	33063405	444	415

In this case, the forecast for spend and quantity is done for e-procurement in the hospitality industry. Six months' prediction is done with the least error. This information is important to all properties to expect the unexpected and be proactive. Furthermore, if there was a noticeable fluctuation in the budget allotted and money spent a further investigation would help the management to get more information on predicting next year's budget. One improvement that can be done for this model is to add holidays, breaks, and other seasonal effects. They can be done by merely adding these variables as a new feature.

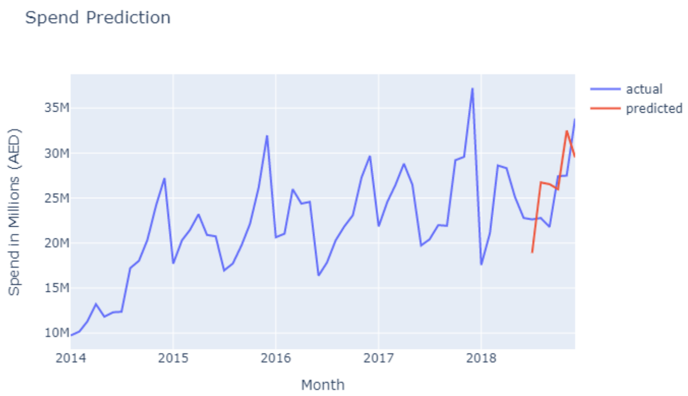

Figure 6a. Spend forecast for the last 6 months. From the plot, we can observe that the actual spending went up while our model also predicted that the spend will go up. This clearly shows how powerful LSTMs are for analysing time series and sequential data

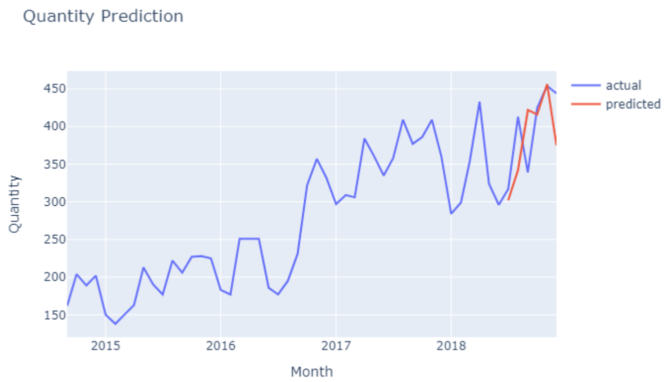

Figure 6b. Quantity forecasted for the last 6 months. From the plot, we can observe that the actual quantity bought went up while our model also predicted that the quantity bought will go up. This clearly shows how powerful LSTMs are for analyzing any variable in time series and sequential data.

In this research, the methodology was based on time series disintegration and LSTM RNN, to astound the problem in budget forecasting in the hospitality industry by training a combined model that adapts key structures, behaviors, and patterns common within a time-series set of data.

5. CONCLUSION

In this research, the forecast for spend and quantity is done for e-procurement in the hospitality industry for which a novel LSTM algorithm with 8 layers and 500 epochs is used. Six months' prediction is conducted with the least error of about 0.0137. The RMSE for the LSTM model was 0.0137 which is good improvement when compared to the baseline model's RMSE 9221.876. This information is important to all properties in anticipating developments and being proactive. Furthermore, if there was a noticeable fluctuation in the budget allotted and money spent, further investigation would assist the management in obtaining more information on predicting next year's budget. Additionally, it guides the top management to take strategic decisions to spend to acquire products and services most effectively. At the same time, it enhances cash flow management within the different properties across the hotel chain. With the latest spread on evolving refined ML-based techniques and, in particular, deep learning algorithms using LSTM, higher accuracy, and powerful results can be obtained for sequential time series data. Machine learning together with deep learning has good scope in e-procurement of the hospitality industry as the data has not undergone any exploratory data analytics study. LSTM has given good forecasting results and can be molded to forecast occupancy, wastage of food, consumption rates, etc. provided the organization is willing to share the data with skilled people for machine learning.

REFERENCES

[1] Bogetić, S., Đorđević, D., Ćoćkalo, D. & Bešić, C. The Analysis of Quality Aspects in the Development of Competitiveness of Domestic Hotel Enterprises. *International Journal "Advanced Quality"*, vol. 44 (2), p. 1, 2017.

[2] Mariani, M., Baggio, R., Fuchs, M. & Höepken, W. Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, vol. 30 (12), pp. 3514-3554, 2018.

[3] Edghiem, F. & Mouzughi, Y. Knowledge-advanced innovative behavior: a hospitality service perspective. *International Journal of Contemporary Hospitality Management*, vol. 30 (1), pp. 197-216, 2018.

[4] Gomezelj, D. A systematic review of research on innovation in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, vol. 28 (3), pp. 516-558, 2016.

[5] Brandon-Jones, A. & Kauppi, K. Examining the antecedents of the technology acceptance model within e-procurement. *International Journal of Operations & Production Management*, vol. 38 (1), pp. 22-42, 2018.

[6] Lamba, K. & Singh, S. Big data in operations and supply chain management: current trends and future perspectives. *Production Planning & Control*, vol. 28 (11-12), pp. 877-890, 2017.

[7] Massa, S. & Testa, S. A knowledge management approach to organizational competitive advantage: Evidence from the food sector. *European Management Journal*, vol. 27 (2), pp. 129-141, 2009.

[8] Greasley, A. (n.d.). *Simulating business processes for descriptive, predictive, and prescriptive analytics*. Frankfurt: Walter de Gruyter GmbH & Co.

[9] Mathew, E. Big Data Analytics in E-procurement of a Chain Hotel. *Advances in Internet, Data, and Web Technologies* [online]. Vol. 29, pp. 295-308, (2019). Available at: https://link.springer.com/chapter/

[10] Siami-Namini, S. & Namin, A. (2018). "Forecasting Economic and Financial Time Series: ARIMA vs. LSTM". arXiv.org [online]. Available at: https://arxiv.org/abs/1803.06386
[11] Yeom, H., Kim, J. & Chung, C. "LSTM Improves Accuracy of Reaching Trajectory Prediction From Magnetoencephalography Signals". *IEEE Access*, vol. 8, pp. 20146-20150, 2020.
[12] Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z. & Zhang, H. Deep Learning with Long Short-Term Memory for Time Series Prediction. IEEE Communications Magazine, vol. 57 (6), pp. 114-119, 2019.
[13] Bandara, K., Bergmeir, C. & Hewamalage, H. "LSTM-MSNet: Leveraging Forecasts on Sets of Related Time Series With Multiple Seasonal Patterns". *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-14, 2020.
[14] Lihong, D. & Qian, X. Short-term electricity price forecast based on long short-term memory neural network. *Journal of Physics: Conference Series*, vol. 1453, p. 012103, 2020.
[15] Wang, G., Gunasekaran, A., Ngai, E. & Papadopoulos, T. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, vol. 176, pp. 98-110, 2016.
[16] "Artificial Intelligence Trends: Decision Augmentation". (2020). Available at: https://www.gartner.com/en/documents/3979508/artificial-intelligence-trends-decision-augmentation
[17] Król, K. & Zdonek, D. Analytics Maturity Models: An Overview. *Information*, vol. 11 (3), p. 142, 2020.
[18] Delen, D. & Demirkan, H. Data, information, and analytics as services. *Decision Support Systems*, vol. 55 (1), pp. 359-363, 2013.
[19] Zhang, J., Yang, X. & Appelbaum, D. Toward Effective Big Data Analysis in Continuous Auditing. *Accounting Horizons*, vol. 29 (2), pp. 469-476, 2015.
[20] Buhalis, D. & Leung, R. Smart hospitality—Interconnectivity and interoperability towards an ecosystem. *International Journal of Hospitality Management*, vol. 71, pp. 41-50, 2018.
[21] Bendoly, E. Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management*, vol. 31 (1-2), pp. 62-71, 2012.
[22] Song, H., Qiu, R. & Park, J. A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, vol. 75, pp. 338-362, 2019.
[23] Wen, H., Li, S., Li, W., Li, J. & Yin, C. Comparison of four machine learning techniques for the prediction of prostate cancer survivability. *Institute of Information Science and Engineering*, (3), 2017.
[24] Ampazis, N. Forecasting Demand in Supply Chain Using Machine Learning Algorithms. *International Journal of Artificial Life Research*, vol. 5 (1), pp. 56-73, 2015.
[25] Wu, C., Patil, P. & Gunaseelan, S. Algorithms for Multiple Regression on Black Friday Sales Data. *IEEE explore*, vol. 14 (1), pp. 17-21, 2018.
[26] Alonzo, L., Chioson, F., Co, H., Bugtai, N. & Baldovino, R. (2018). A Machine Learning Approach for Coconut Sugar Quality Assessment and Prediction. *Engineering Research and Development for Technology (ERDT) of the Department of Science and Technology (DOST)*.
[27] Goudarzi, F. Travel Time Prediction: Comparison of Machine Learning Algorithms in a Case Study. *20th International Conference on High-Performance Computing and Communications*, vol. 4 (20), p. 1, 2019.
[28] Raschka, S., Julian, D. & Hearty, J. (2016). *Python deeper insights into machine learning*. 2nd edn. Birmingham B3 2PB, UK.: Packt Publishing Ltd.
[29] Subramanian, G. (2015). *Python Data Science Cookbook*. 1st edn. Birmingham B3 2PB, UK.: Packt Publishing.
[30] Pahwa, K. & Agarwal, N. (2019). Stock Market Analysis using Supervised Machine Learning. *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*.

BIOGRAPHIES OF AUTHORS

Dr. Elezabeth Mathew is the corresponding author of this paper who is an educator with over 12 years of teaching experience in the UAE. This paper is part of her thesis. Her interest remains in artificial intelligence, machine learning, big data, intelligent systems, etc. She has published several papers in the areas of interest.

Dr Sherief Abdallah is Professor at the British University in Dubai, and an Honorary Fellow at the University of Edinburgh, UK. He holds a PhD from the University of Massachusetts at Amherst, the United States. Dr Abdallah's research focuses on applying machine learning to real and novel problems, including mobile devices, network management, and information retrieval. He collaborated with world-class researchers in the United States, South America, and Europe.