

## Deep learning-based approach for forecasting intermittent online sales

Yashar Ahmadov<sup>1</sup> · Petri Helo<sup>1</sup>

Received: 10 April 2023 / Accepted: 7 June 2023

Published online: 19 December 2023

© The Author(s) 2023 [OPEN](#)

### Abstract

Deep Neural Networks (DNN's) present some of the leading applications of Artificial Intelligence (AI) which have proven suitability on various machine-learning use cases. Forecasting demand of intermittent on-line sales is a task which needs to be carried out automatically for a large number of Stock Keeping Units (SKU's). This paper discusses the intermittent online sales and proposes an AI-based model for forecasting demand. We provide empirical evidence by utilizing data from 17 different sellers with approximately 3000 orders in total. Our findings indicate that thanks to their multi-layered learning structure, the DNN's can provide up to 35% better accuracy than the classic models such as Moving Average, Exponential Smoothing, Croston's method and ARIMA. Also, it was revealed that the time between orders' arrivals follow Exponential distribution and the order sizes also generally follow Exponential distribution. Thus, most of the time, Poisson Exponential distribution can be used for modelling intermittent sales process through online platforms. The analyses show that Poisson Exponential distribution can generate values close to real sales with less than 7% error margin with real data.

**Keywords** Deep learning · Forecasting · Artificial intelligence · e-Commerce · Compound distributions

## 1 Introduction

The development of online marketplaces has encouraged many sellers to sell their goods online. This development is reflected in the number of sellers and amount of sales that have skyrocketed during the recent years. According to the European Union statistics (2016) [44], the proportion of internet users who have ordered goods through the internet has risen from 50 to 66% during the years 2007–2016. The importance of accuracy in forecasting such sales has been increasing respectively. Increased forecasting accuracy helps especially the commercial sellers to decrease the bullwhip effect in their supply chain. This became especially important after the COVID-19 pandemic hit and e-commerce platforms saw a two to threefold increase in the orders. According to Alfonso et al. [2], revenues increased in the first half of 2020 for Amazon (34% year on year), Alibaba (27%), JD (28%), Shopify (74%), Rakuten (16%) and Mercado Libre (50%). As a side note, the terms "e-business" and "e-commerce" have been used to express different meanings in different contexts, but in general it is accepted that e-business has more general meaning and it covers e-commerce transactions.

Fast-growing e-business combined with the developments in the Artificial Intelligence and Machine Learning models motivates us to find more advanced approaches to model and forecast online sales. With this pace of growth of the e-commerce and the number of Stock Keeping Units (SKU's), automated real-time working algorithms will

✉ Petri Helo, phelo@uwasa.fi | <sup>1</sup>Networked Value Systems, School of Technology and Innovation, University of Vaasa, P.O. Box 700, 65101 Vaasa, Finland.



have to be deployed in the future to handle the forecasting/inventory related tasks. Deep Neural Networks (DNN's) constitute the backbone of many modern technologies (e.g., face recognition, computer vision, self-driving cars, Natural Language Processing, etc.). One main motivation behind this paper is to investigate the applicability of Deep Neural Networks (DNN's) to forecasting the intermittent online sales. As of 2022 Q4, the share of the third-party sellers constitutes 59% of the total sales on Amazon. On special days (Black Friday, Cyber Monday), the sales of third-party sellers exceed one billion US dollars on Amazon, so, third-party sellers should not be underestimated [45]. Sales for many individual third-party sellers on online markets such as eBay, Amazon and Alibaba are intermittent and "nervous", especially for small and medium size stores. There are multiple complications with predicting such data: skewed distributions, high variability, different demand drivers, to name a few. On some days, sellers receive several orders and then receive no orders during several consecutive days. There are no papers currently attempting to model the described situation in this specific context. Our paper is inspired by a real-life situation that the authors have faced after starting selling goods online. The authors had tried to forecast their sales for the following year and noticed that they do not follow regular distributions. This led to further analyses of the data. It was found out that DNN's perform well for forecasting and Compound Distributions fit well for modeling purposes. As it will be discussed in the literature review section, DNN's' superiority have been proved in not only forecasting intermitted demand, but also in a wide variety of prediction tasks.

The origins of online selling date back to the mid-90 s; when eBay and Amazon were launched as pioneers in this field. Since that, electronic businesses have been the research topic of a significant number of articles. E-businesses have been analyzed from the aspect of Supply Chain integration by Lee and Whang [26]. They focus on information integration, synchronized planning, coordinated workflow and new business models and discuss how the integration of Supply Chains have been affected because of massive migration to online business. The article was written in the early 2000s, when the use of the internet started proliferating across the world. They had correctly predicted that the companies that successfully integrate their supply chains into e-business will gain significant competitive advantage [13]. This process is still ongoing, not all supply chains are connected from end to end and there is a long way to go in this field.

Swaminathan and Tayur [42] presented models used in supply chain and discussed ways to adapt the current models to e-business. They focus on visibility in supply chains, supplier relationships, distribution and pricing, real-time decision technologies and customization. The authors discuss the growing need for models that can give better insight into the effect of changing different parameters in the supply chain. Our paper tries to address this need and provide e-Supply Chains with better forecasting models. In quantitative research, testing and estimation require information about the population distribution. The traditional approaches do not work when usual assumptions about the population cannot be made. Bootstrapping method uses information based on the resamples from the sample to estimate the population distribution. It treats the sample as a population, draws  $X$  samples each containing  $n$  entries and computes the statistic that is of interest for the research. Then, the population statistics are estimated by considering all the bootstrap samples. A paper by Willemain, Smart and Schwarz [49] proposes using bootstrapping to forecast the demand for spare parts. They show that the proposed bootstrap method dominates the traditional Exponential Smoothing and Croston's method. But the performance of the bootstrap method deteriorates with the increasing lead time. Other methods have been used in the literature are Single Exponential Smoothing [3, 10], Croston's Method [6], Syntetos-Boylan approximation [43], Moving Average [36], Weighted Moving Average [38], Additive Winter [4], Grey Prediction Model [5], ARIMA [47] and Neural Networks (NN's) [5]. Gutierrez et al. [14] describe lumpy demand forecasting using NN's and compare their results with single exponential smoothing, Croston's method, and the Syntetos-Boylan approximation. NN's dominate the other traditional methods. The paper by Jiang et al. [18] discusses the application of Support Vector Machines (SVM) to modeling the spare part demand in the heavy-duty vehicle industry. In their case, the SVM approach results in better run times and better accuracy when compared to the Neural Networks (NN's). Turkmen et al. [48] have investigated the application of Deep Renewal Processes to the intermittent demand forecasting. In their case, the LSTM-based recurrent neural networks outperform the traditional methods.

First research task of this paper is (1) to study the applicability of Deep Learning (DL) models to forecasting the sales of online stores. Making reliable forecasts for the upcoming periods by treating the data as time series is included in this article. We employ DNN's to find the most suitable Neural Network (NN) architecture that would yield the least deviations from the actual values. DNN's have been successfully applied to forecasting problems from different domains. The structure of the data that we use is similar to the applications in spare parts demand forecasting. (2) The second research task is to find statistical distributions to model the total sales for online stores. Intermittent online sales do not follow regular distributions and we attempt to find statistically appropriate distributions for such data.

The remainder of the paper is structured as follows: literature review provides an overview of the most recent articles on the topic, then methodology is explained. Result explains the findings of the research and provides the statistical tests. Finally, conclusions are presented with discussion.

## 2 Literature review: demand forecasting for intermittent demand

This section reviews some prior works on forecasting intermittent demand and how neural networks and related techniques have been applied to solve similar kinds of problems.

For intermittent demand, a great share of literature until the early 2010's focuses on Croston's Method; after that many hybrid methods in forecasting the intermittent Supply Chain demand patterns have been proposed such as the methodology described by Fu et al. [9]. Initial version of the Croston's Method was found to have an error leading to the positive bias. Solis et al. [41] studied whether the two proposed corrections to the Croston's Method (SBA and SBJ) are superior to each other or not. The results showed no significant difference between the two approaches.

A hybrid model for forecasting was proposed by Khashei and Bijari [20]. According to this study, any forecasting problem can be divided into linear and non-linear components. Since ARIMA models assume that the generating process is stationary, and it does not depend on additional variables, they are used to model the linear portion. Then selected  $n$  last residuals ( $e_t$ 's), forecasted values ( $L_t$ ) and  $z_t$  values from the ARIMA model are used as inputs to the multilayer perceptron where

$$\left( z_t = (1 - B)^d (y_t - \mu) \right) \quad (1)$$

and  $B$  is the backward shift operator,  $d$  is the order of differencing,  $y_t$  is the actual value of the observation at period  $t$ , and  $\mu$  is the mean value. They use the sigmoid activation function in the NN's and structure of the best fitted network is as in Fig. 1.

The number of residuals and  $z$  values were decided as results of repeated trials, as such there is no single way to determine the number and constitution of the inputs. The proposed model is claimed to dominate other hybrid approaches in MAE (Mean Absolute Error) and MSE (Mean Square Error) values.

Another academic work worth discussion is about application of NN's to intermittent demand patterns by Kourentzes [22]. He discusses different ways to forecast the intermittent demand, especially comparing the variation of Croston's method to NN's. His findings indicate that NN's do not provide better forecast accuracy than other classical approaches, however, they dominate when taking the service levels into consideration. The article uses 1000 items for analysis and calculates the model parameters on the dataset of 3000 time series that were used by Syntetos and Boylan [43]. Inventory

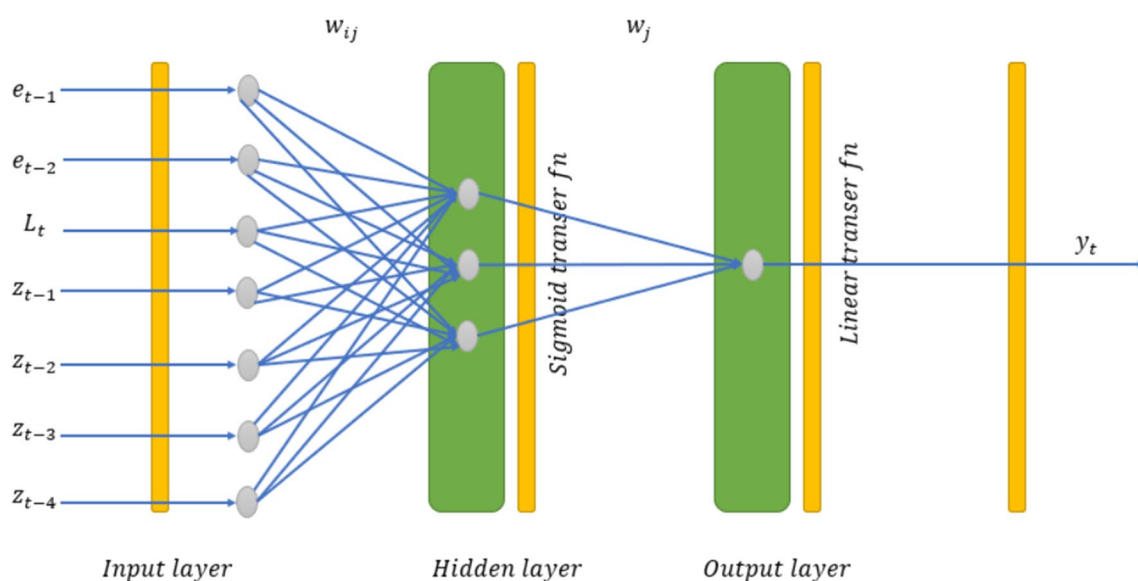


Fig. 1 The structure of NN used by Khashei and Bijari [20]

simulations were done on these items and the service level results have been openly in favor of Neural Network models. There are several articles in the literature discussing the applications of the Artificial Intelligence and Deep Learning models to the problems in the industry. Li et al. [28] make a review of the application of these models to intelligent manufacturing. Quantitative models for evaluating Supply Chains are discussed by Lima-Junior and Carpinetti [29].

The applications of the NN and DL models to the Operations Management (OM) and Supply Chain Management (SCM) problems has been studied throughout the last two decades. Such problems include production planning and scheduling, Material Requirements Planning (MRP) and inventory planning. Lot Sizing problem which is used to determine the optimal production quantity for each period is solved by using NN's in the article by Radzi et al. [34]. Their model inputs are holding cost, reorder cost and the demand values. Feed-forward multi-layered network with backpropagation learning algorithm was used. NN's input layer consists of 14 nodes; 12 for the demand values for the time periods  $t = 1, 2, \dots, 12$ , 1 for the holding cost, 1 for the reorder cost. These values are then passed into the hidden layer consisting of 13 nodes. The output layer consists of 12 nodes with binary values. If the result of any node is 1, it means that an order is to be placed in that period. The model in question has provided better and faster solutions when compared to the classical models.

The paper titled "Export Sales Forecasting using Artificial Intelligence" authored by Sohrabpour et al. [40], and published in *Technological Forecasting and Social Change*, dives into the domain of international trade and investigates the efficacy of AI methods in predicting export sales. Through empirical analysis, the study evaluates multiple AI techniques, such as Genetic Algorithm, against conventional time series forecasting methods, revealing that AI-based approaches outperform traditional models by achieving an average prediction accuracy improvement of around 15%. These results underscore the substantial potential of AI in revolutionizing export sales forecasting, offering businesses valuable insights into optimizing their strategies and resource allocation for enhanced competitiveness in the global market landscape.

In their 2020 paper titled "Data Analytics in Supply Chain Management: Review of Machine Learning Applications in Demand Forecasting," Aamer et al. [1] review the current state of the literature in the realm of the demand forecasting. They have found out that 77% of the articles used various ML techniques (such as NNs and SVM) for demand forecasting. Their second finding was that the ML algorithms can provide better accuracy and less computational cost.

In their 2019 article "A Comprehensive Literature Review of the Demand Forecasting Methods of Emergency Resources from the Perspective of Artificial Intelligence," Zhu et al. [51] conduct a comprehensive examination of demand forecasting techniques for emergency resources, with a focus on artificial intelligence (AI). The authors survey a range of AI-based methods used in this context, revealing the multifaceted landscape of forecasting approaches. Throughout the review, the authors emphasize the importance of AI-driven strategies in accurately predicting the demand for emergency resources, shedding light on innovative methodologies that contribute to more effective disaster preparedness and response efforts.

One of the classical problems in the OM is demand forecasting. An interesting approach has been proposed by Qiu et al. [33]. It is an ensemble method, i.e., a combination of different models working together. The authors justify their choice as ensemble methods being statistically, computationally and representationally effective. Their method works as follows: a number of Deep Belief Networks (DBN's) are trained with different numbers of epoch values. The outputs of DBN's are fed into the Support Vector Regression (SVR) model. The outputs of the SVR become the predictions for the coming periods. The proposed model outperforms the different ML algorithms when applied individually. The application of the DL to the Newsvendor problem has been studied by Oroojlooyjadid et al. [31]. Forecasting and inventory optimization have been incorporated into a single model rather than dealing with them separately. One advantage of such models is that they do not need assumptions about the demand distributions. The authors also mention data features included in the model, although they do not explain in detail what these features are. Their main finding is that when the demand is volatile, DL models work well. This coincided with our findings. Their other contribution is the extension of the DL models to the  $(r, Q)$  inventory policies. Another interesting paper by Jeble et al. [17] investigates the impact of big data and predictive analytics capability on supply chain sustainability. The fact that predictive analytics (including forecasting) can help firms to become more sustainable is another driving factor of this paper.

Another DL-based approach for time series forecasting was proposed by Torres et al. [46] with application to electricity load forecasting. They use the  $H_2O$  library in conjunction with Apache Spark to speed up the calculation process. Since  $H_2O$  does not allow multi-step ahead forecasting, the authors divide the problem into  $h$  subproblems ( $h$  denoting the number of time periods to be forecasted) and solve them separately. They turn this drawback into advantage by allowing a different number of neurons for each subproblem that would result in minimized errors. Their model can predict the real consumption values with less than 2% error margin. Larrain [24] takes the discussion to the macro-financial level by comparing NN's and Regression forecasts for the Purchasing Managers Index (PMI), T-Bill and Inventories. Their findings indicate that T-Bills can be used to forecast the PMI, PMI can be used to forecast the inventories. The forecasting of inventories is of paramount importance in Supply Chain Management.

Figueiredo [8] studied finding accurate forecasting models by employing the NN's for new products that are to be sold via online platforms. The main challenge in e-commerce is that any new product is offered to millions of potential customers at a point in time which makes forecasting and inventory control difficult. It is difficult to predict how the market will react to the new product, thus, the potential sales amount becomes a question. He uses the following input variables to the NN model: duration of the selling season, type of the season, product class, profile, demand in the first two weeks and demand in the weeks 3–4–5–6 and total demand. For some cases, he adds two more variables to the list: cumulative demands and demand moving averages. Figueiredo's model achieved 2.4–7.8% reduction in the average absolute error for 1000 lines in the test sample. This cannot be considered an important increase in the forecast accuracy. The author then proposes using combined forecast methods and this reduces the forecast error by 13.3% which can be considered as a significant change.

Ferreira, Lee and Simchi-Levi [7] discuss demand forecasting and price optimization for an online retailer called Rue La La. They achieve 9.7% increase in the revenues of the case company by applying more accurate forecasting methods and price optimization techniques. For the forecasting part, their findings show that regression trees with bagging outperform other regression methods in predicting the demand. For the pricing part, they propose a new formulation of the optimization problem and create an algorithm for the solution of this complex problem. Another importance of this paper is that they combine machine learning and optimization techniques into a pricing decision support tool that results in increase in the revenue of the online retailer.

The possibilities of using of artificial intelligence in supply chain management has been discussed recently [32]. Various technologies to support processes has been suggested in case studies [15] and literature reviews have been conducted to propose the application area [37]. However, the use of artificial intelligence technologies such as deep learning is far from maturity.

During the last decades, many versions of NN's have been developed to address different questions in the industry and Artificial Intelligence. Recurrent Neural Networks (RNN's) and its variants such as LSTM, Convolutional Neural Networks (CNN's), Generative Adversarial Networks (GAN's) and Kohonen Networks are a few examples. Most common applications of RNN's have been in the field of speech recognition. Graves et al. [12] discuss speech recognition with deep bidirectional RNN's. Their model works in both directions; conducts front and back propagation which outperformed the other approaches based on the TIMIT database at the time of publishing. CNN's have been successfully applied to the image classification and face recognition types of tasks such as the model described by Krizhevsky et al. [23] and Lawrence et al. [25]. Another important paradigm developed in the field of machine learning is reinforcement learning. Different from supervised and unsupervised learning, reinforcement learning deals with the actions that need to be taken by an agent to maximize the reward. The rules are usually stochastic and in this sense reinforcement learning is a Markov Decision Process, sometimes they are referred to as approximate dynamic programming models. Reinforcement learning concept is successfully applied to robotics as discussed in the survey by Kober et al. [21]. They are especially useful in complex cases where classical Dynamic Programming falls short [11].

Besides searching for the articles related to the e-business demand analysis, we also conducted a search to find papers discussing methodologies to model intermittent sales. We found out that similar types of demand patterns are observed in service logistics. There are many papers in the literature dealing with modeling spare parts sales. Spare parts show similarities to online sales; their demands are generally intermittent and lumpy. Lengu et al. [27] propose compound distributions to model the spare part demands. Depending on the coefficient of variation and mode of the data, they recommend using Poisson-Logarithmic, Poisson-Pascal, Poisson-Geometric and Poisson-Poisson distributions. Snyder, Ord and Beaumont [39] propose a new model for slow-moving inventories. They compare the performances of Poisson, Negative Binomial and hurdle shifted Poisson distribution. Based on the analysis of 1046 car parts SKUs, they conclude that using dynamic schemes in favor of static ones can provide substantial benefits to the companies.

### 3 Methodology

This section provides an explanation of the data, mathematical models, distributions and overall approach used in the analyses of the data.

**Table 1** The number of stores by categories

Store category	No. of stores
Collectibles	3
Home and Garden equipment	1
Sporting goods	3
Electronics	2
Books	1
Jewelry and Watches	2
Cosmetics	2
Fashion	2
Total	16

**Table 2** Summary statistics of the data sets

Dataset no	Coverage period	Avg. order size (\$)	Std. dev. of order sizes	Avg. interarrival time (days)	Total no. of orders
0	05.04.2015–21.12.2016	55.06	61.00	5.19	121
1	01.01.2015–05.01.2017	154.05	170.40	2.18	170
2	03.10.2015–27.02.2017	63.00	60.14	3.00	189
3	02.02.2014–12.11.2016	27.29	24.18	3.52	186
4	09.01.2014–05.02.2016	2.44	2.92	2.80	141
5	05.03.2015–23.03.2017	53.83	60.25	6.21	121
6	06.11.2015–06.02.2017	17.00	11.22	2.14	203
7	15.04.2015–01.03.2017	46.78	64.97	2.00	160
8	14.08.2015–08.03.2017	24.00	23.21	3.00	184
9	18.05.2015–26.06.2017	81.45	76.92	3.28	138
10	03.05.2015–28.03.2017	22.19	19.76	2.00	175
11	01.04.2015–05.10.2016	317.49	287.12	3.15	179
12	04.01.2014–29.01.2017	15.54	9.67	5.00	314
13	05.04.2014–06.06.2016	16.00	15.19	3.64	118
14	05.02.2016–20.02.2018	44.28	35.91	2.00	186
15	06.07.2014–13.07.2016	56.00	47.24	5.68	159
16	07.04.2015–04.08.2017	294.00	268.51	2.25	259

### 3.1 Data

The data collected covers a period of two years of online sales on eBay between April 2015 and January 2017. Bank notes, coins and other numismatics products were being sold; 121 orders were received. The data for early February were left out of consideration since it was the start up month of the sales and therefore, they do not realistically represent sales trends. We also acquired sales data of **16 different stores** from eBay and its Turkish subsidiary Git-tiGidiyor. The data span a period of 2 to 4 years. The information related to the sales on online platforms are available on the feedback pages of the sellers as such, the findings of this research can be verified and reproduced easily. The distribution of data across categories is seen in Tables 1 and 2. The SKU's can be held by the e-commerce platform owner, vendor, private seller depending on the configuration.

Summary statistics of the Data Sets:

These data were deliberately selected from stores with intermittent demand; they do not represent the whole e-business, only the individual sellers with non-continuous demand are present in this data set. The stores are from very different fields as depicted in the Table 1 to demonstrate that the proposed methods in this paper work for many domains. We also randomly picked **50 different items** to analyze and find out their demand patterns. These items included phone accessories, watches, home equipment, books, numismatics products, sports equipment, clothes and apparel. The data consists of the dates of orders for specific items and their respective amounts.

### 3.2 Classical forecasting models

In this paper we evaluate DNN's and compare their performance with different classical demand forecasting methodologies. In the literature there are following approaches proposed by various authors and we applied them to our data to see the respective results:

#### 1. Moving average (MA)

This method uses the average of  $n$  periods as a forecast for the next period. Simple moving average is the unweighted mean of the last  $n$  observations:

$$x_{t+1} = \frac{x_t + x_{t-1} + x_{t-2} \dots + x_{t-(n-1)}}{n} \quad (1)$$

#### 2. Single exponential smoothing (SES)

Single Exponential Smoothing is used when data does not show any patterns of trend or seasonality. It is a method of smoothing the time series data by assigning exponentially decreasing weights to the older observations, thus, giving higher priority to the most recent observations. The formula used is

$$x_{t+1} = \alpha * x_t + (1 - \alpha) * x_{t-1} \quad (2)$$

where  $\alpha$  is the smoothing factor between 0 and 1.

#### 3. Double exponential smoothing (DES)

DES is used when the underlying data has a trend. It contains the smoothing factors  $\alpha$  and  $\beta$  between 0 and 1 for data smoothing and trend smoothing, respectively. As in SES, the model assigns much higher weight to closer observations than to the older ones.  $x_1, x_2, x_3, \dots, x_t$  are the observations,  $s_t$  is the smoothed value for period  $t$  and  $b_t$  is the estimate of the trend for period  $t$ .

$$s_1 = x_1 \quad (3)$$

$$b_1 = x_1 - x_0 \quad (4)$$

$$s_t = \alpha * x_t + (1 - \alpha) * (s_{t-1} + b_{t-1}) \quad (5)$$

$$b_t = \beta * (s_t - s_{t-1}) + (1 - \beta) * b_{t-1} \quad (6)$$

To forecast  $m$  period ahead, the following formula is used:

$$F_{t+m} = s_t + m * b_t \quad (7)$$

#### 4. Croston's method

Croston's method separates two processes: the time between order arrivals and the magnitude of the individual orders. Demand process is described with the equation  $x_t = y_t * z_t$  where  $y_t$  is a binary variable representing whether an order will arrive or not and  $z_t$  showing the magnitude of the order if it arrives. As Simple Exponential Smoothing results in higher noise, Croston's Method is preferred when planning inventory levels.

#### 5. Grey prediction models

The superiority of Grey prediction models is that they can be used with limited amounts of data to estimate the values of unknown parameters [19]. They are represented as  $G(n, m)$  where  $n$  is the order of the difference equation and  $m$  is the number of the variables. Grey models are used to reveal the realistic governing rules of the data [19]. GM (1,1) type of models are the most frequently used type of Grey models in the literature.

### 6. ARIMA (p, d, q)

Auto Regressive Integrated Moving Average (ARIMA) models are used extensively to forecast data that has the characteristics of non-stationarity.  $p$  shows the number of lags,  $d$  the degree of differencing and  $q$  the order of the moving average model. ARIMA is suitable for stationary data; thus, if data's mean and/or variance changes over time, transformations need to be used. We used log transformation in this article. The general formula for ARIMA is:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \tag{8}$$

where;  $\hat{y}_t$ : forecast for period  $t$ .  $\phi$ 's and  $\theta$ 's: Coefficients of the model.

### 3.3 Neural networks and deep learning models

DNN is currently one of the most promising Artificial Intelligence technologies and we investigated their power in intermittent demand forecasting. In general, NN's are composed of algorithms that imitate the learning process of human brains. They are powerful tools to recognize patterns in a wide variety of data; they can be used for classification, regression and similar tasks. NN's take numerical data as inputs, map the relationship between inputs and outputs by assigning weights to the inputs. This weight assigning process is an iterative one and there are different methods to achieve close-to-optimum values. After a model is trained, it can be used for predicting future inputs. NN's are defined as "data-hungry" i.e., they need a lot of data for training to achieve the desired level of accuracy. There are plenty of papers discussing the application of NN's to the forecasting, classification and regression models.

NN's contain an input layer, hidden layers and an output layer. There are different activation functions used such as Sigmoid and RELU depending on the situation. The model fine tunes the weights to find the function that minimizes a certain loss function. When there is more than one hidden layer, the model can be classified as a Deep Learning (DL) model. More hidden layers allow to reveal a more complex relationship between the inputs and outputs. Development of the DL models have increased the power of working with unlabeled/unstructured data.

There are several properties of the NN's that make them attractive for forecasting tasks for researchers and practitioners. First important property is that they do not have any a priori assumptions regarding the distribution of the underlying data and the demand rate which becomes a problem when dealing with the traditional forecasting methods. Another important characteristic of NN is that they can generalize. As we will discuss in the Results section, DNN's can perform well even on the unseen part of the data. Thirdly, they are universal approximators and can estimate the complex nonlinear relationship between the inputs and outputs [50].

In addition to the traditional models, we tested two different types of NN's; one for one-step ahead forecasts and one for three-step ahead forecasts. We used the Keras library in Python with TensorFlow backend to do the calculations. The architecture of the NN's used for the LSTM model is depicted in Fig. 2.

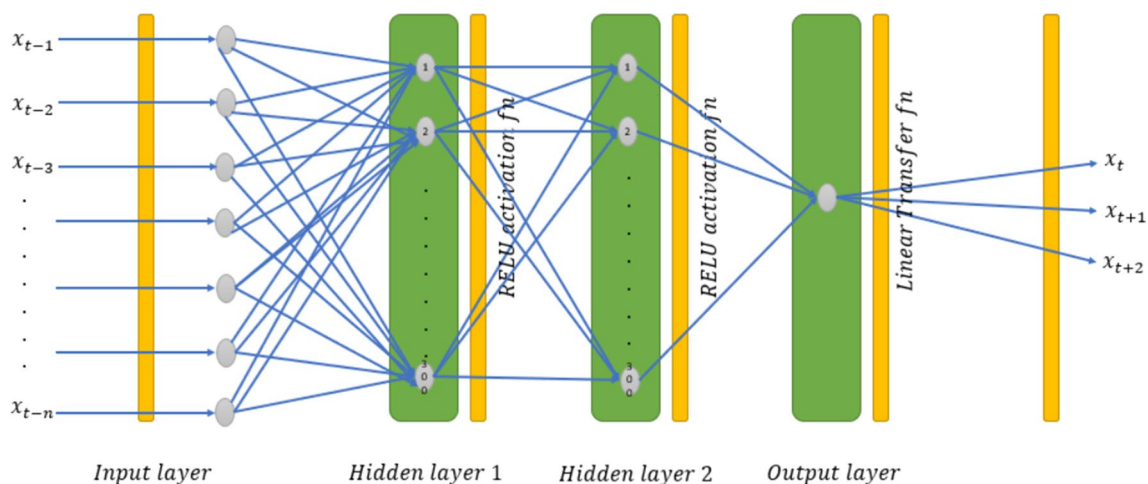


Fig. 2 The architecture of the LSTM Network

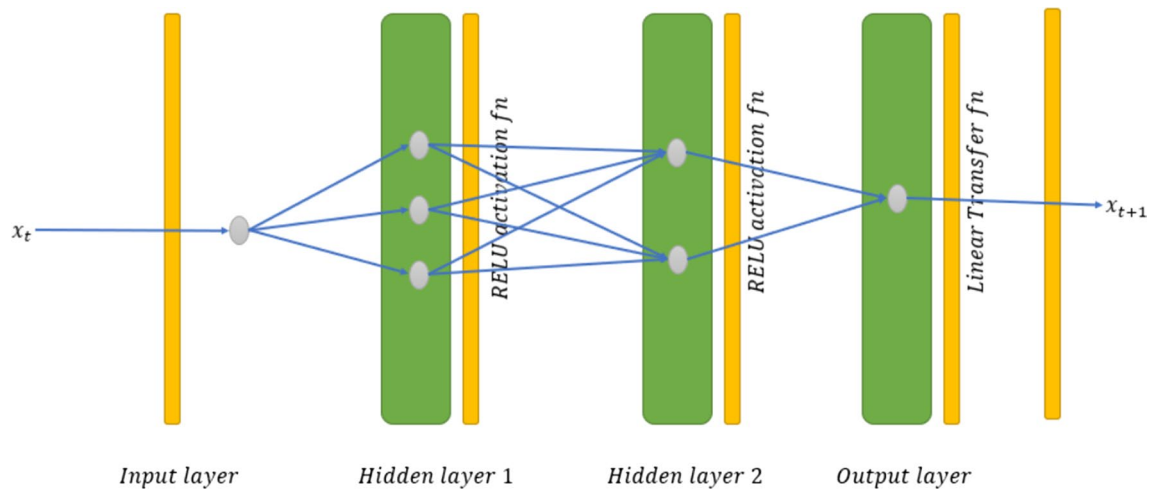
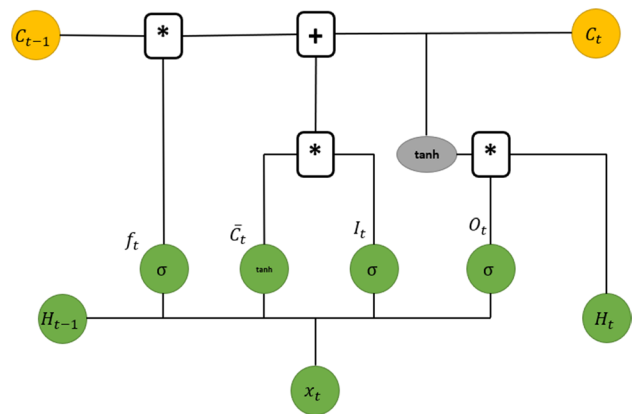


Fig. 3 The structure of the Ordinary Neural Network

Fig. 4 The structure of the LSTM Network



LSTM model uses the last 12 months’ data as input, passes through two hidden layers each with Rectified Linear Unit (ReLU) activation and 300 nodes. The outputs of the model are  $x_t, x_{t+1}, x_{t+2}$  which are the forecasts for the next 3 periods. The formula for ReLU is

$$f(x) = x^+ = \max(0, x) \tag{9}$$

Ordinary Neural Network structure is shown in Fig. 3.

Ordinary NN uses the last observation as input, passes it through 2 hidden layers with 3 and 2 nodes, respectively. The output is the forecast for the next period. As there is no single rule to decide on the parameters of the NN models, we also used trial-and-error method to find the most suitable parameter values.

For multi-step ahead forecasting, we used Long Short-Term Memory (LSTM) models which were developed by Hochreiter and Schmidhuber in 1997 [16]. They are a special type of Recurrent Neural Networks (RNN’s) which can handle long-term dependencies. NN’s use backpropagation algorithm to update the weights of the network. Vanishing and Exploding gradients are well-known problems associated with this algorithm which cause difficulties in training the data. Also, traditional NN’s had the shortcoming that they were not able to link the past information to the current context and LSTM’s were created to address this problem. They have a channel to convey this information. Gates control how much of past data will be used in the cell state which is regulated by Sigmoid function. Sigmoid generates values between 0 and 1, which acts as a portion of each data to be utilized in the cell state. A diagram of the LSTM cell is depicted in Fig. 4.

The notations used in the diagram are:

$$x_t : \text{InputVector}$$

$H_{t-1}$  : Outputofthepreviouscell

$C_{t-1}$  : Memoryofthepreviouscell

$H_t$  : Outputofthecurrentcell

$C_t$  : Memoryofthecurrentcell

The formulas 10–15 are associated with the LSTMs, where  $\sigma$  stands for Sigmoid activation function and  $\tanh$  for Hyperbolic Tangent activation function:

$$f_t = \sigma(x_t * U_f + H_{t-1} * W_f) \quad (10)$$

$$\bar{C}_t = \tanh(x_t * U_c + H_{t-1} * W_c) \quad (11)$$

$$I_t = \sigma(x_t * U_i + H_{t-1} * W_i) \quad (12)$$

$$O_t = \sigma(x_t * U_o + H_{t-1} * W_o) \quad (13)$$

$$C_t = f_t * C_{t-1} + I_t * \bar{C}_t \quad (14)$$

$$H_t = O_t * \tanh(C_t) \quad (15)$$

W, U: weights.

It must be noted that LSTM's have many variants and we used only one type for this study. To summarize, our selection criteria of LSTM over the other types of RNNs are the followings:

- handling long-term dependencies
- gating mechanisms
- avoiding gradient vanishing and exploding
- memory cells
- flexibility in sequence length
- efficient training

Nowadays, there are several tools on the market for applying Artificial Intelligence (AI) tools on the datasets. Some of them are open source and some are commercial. H<sub>2</sub>O and Keras libraries in Python and R are examples of open-source libraries that are very powerful in real life applications. RapidMiner is an example of commercial tools that are widely utilized in the industry and in the academic world. With RapidMiner one can apply more than ten data mining methods (including DL) to any set of data in a few minutes. Main challenge in using the AI tools is the hyperparameter tuning. For a Long Short-Term Memory (LSTM) network, the learning rate, the number of training epochs, the size of training batches and the number of neurons are the major hyperparameters that need to be tuned for optimized results. Reimers and Gurevych [35] test 50.000 network configurations to find the patterns for optimal hyperparameters. Nejedly et al. [30] propose a method based on differential evolution genetic algorithm to produce the hyperparameters.

External factors like weather, month, country, and other relevant features can be added into neural networks to improve the accuracy of predictions for intermittent online sales. This is done as a part of the feature engineering and can significantly enhance the predictive power of the model. RNN's, especially LSTM models are well-suited for capturing the temporal patterns. We did not do this since the articles we selected were random, however, if we focused on one industry, for example the fashion industry we would benefit from these additional features. Usually, the country names are one-hot encoded, weather-related factors (such as temperature, humidity, precipitation) are added as quantitative features.

### 3.4 Statistical distributions to model online sales

The methodology described in this part is similar to the one proposed by Thompson [45] and Lengu et al. [27]. Lengu et al. [27] used a similar approach to model the intermittent demand for the spare parts. The probability of having  $r$  orders in a two-week period is

$$p_r = \frac{\rho^r * e^{-\rho}}{r!} \quad (16)$$

where  $\rho$  is the mean rate of the orders.

The sum of  $r$  exponentially distributed independent orders with mean amount  $\mu$  has the density function

$$g_r = \frac{x^{r-1} * e^{-x/\mu}}{\mu^r * (r-1)!} \quad (17)$$

where  $x$  is the two-week order total. In other words, adding up Exponential random variables leads to Gamma distribution. That is what  $g_r$  corresponds to:  $r$  independent orders, each having order sizes which are exponentially distributed.

Thus, the probability density function of the order totals is

$$\sum_{r=1}^{\infty} p_r * g_r = e^{(-\rho - \frac{x}{\mu})} * \sum_{r=1}^{\infty} \frac{x^{r-1} * (\rho/\mu)^r}{r! * (r-1)!} \quad (18)$$

This probability distribution can be used to model intermittent online sales, for example in simulation or forecasting models.

#### 3.4.1 Parameter estimation

To estimate the population mean rate ( $\hat{\rho}$ ) and the mean amount ( $\hat{\mu}$ ) of orders, we can collect samples from actual data and do the following computations. In terms of  $\rho$  and  $\mu$ , the mean (1) and variance (2) of the total sales are

$$\bar{x} = \rho * \mu \quad (19)$$

$$s^2 = 2 * \rho * \mu^2 \quad (20)$$

Therefore, the moment estimates for the parameters  $\hat{\rho}$  and  $\hat{\mu}$  are

$$\hat{\mu} = \frac{s^2}{2 * \bar{x}} \quad (21)$$

$$\hat{\rho} = \frac{\bar{x}}{\hat{\mu}} = \frac{2 * \bar{x}^2}{s^2} \quad (22)$$

If the given data contains the sale dates and amounts, then it is more reasonable to calculate  $\rho$  and  $\mu$  separately from the data; it is easier and more reliable.

There are two components of the problem discussed in this article: order arrivals and magnitude of the orders. The orders' interarrival times follow exponential distribution, so, it is safe to assume that the order numbers follow the Poisson distribution. The magnitude of the orders follows Exponential distribution. To summarize, the sales data follows Exponential distribution and arrival of orders is a Poisson process.

Figure 5 shows the histograms of the two processes for the main data. We fail to reject the hypothesis that interarrival times are exponentially distributed with p-value of 0.717, so, it can be assumed that order arrivals follow Poisson distribution. The second graph depicts the order sizes histogram, which has p-value of 0.558. In both cases, at 5% significance level we fail to reject the null hypotheses that data come from exponential distributions. This makes sense; because usually many people buy cheap items and a few people tend to buy expensive items. The number of buyers decreases exponentially from the cheapest to the most expensive items.

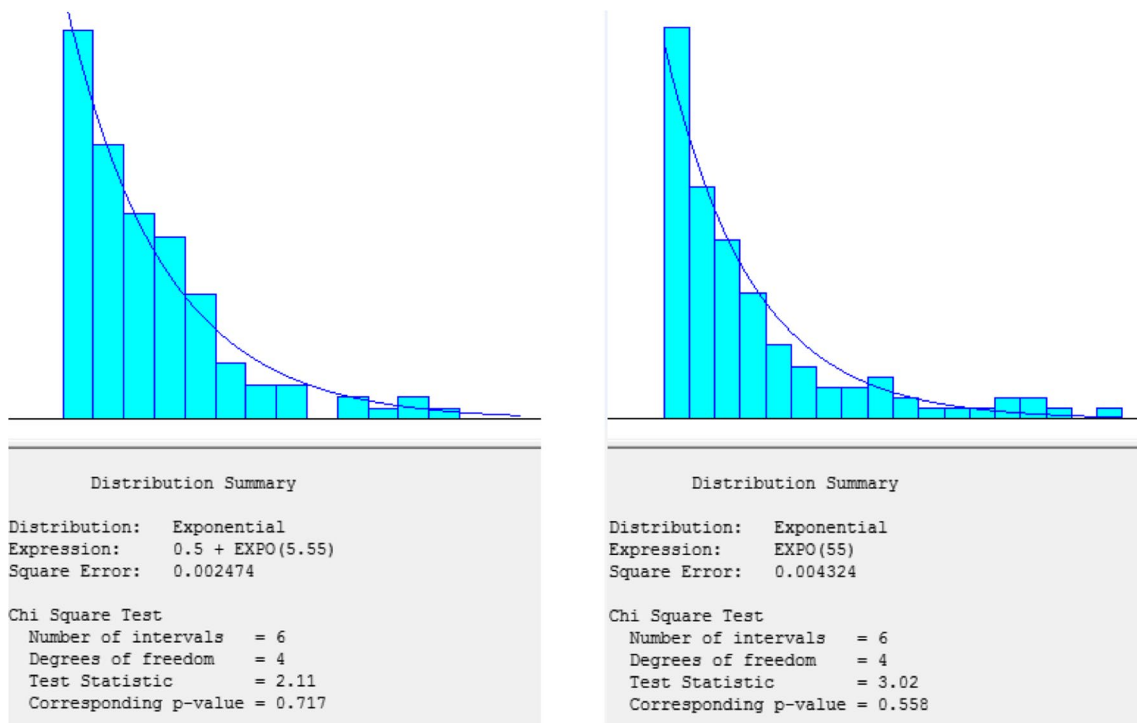


Fig. 5 Distribution summary for the main data (left: order arrivals, right: order sizes)

We propose to model such sales by using the Poisson Exponential compound distribution. In our case, the parameters of the distributions were calculated equal to 0.17 and 53.4 for  $\mu$  and  $\rho$ , respectively. The model was used to test how it fits the data and the results have shown a good fit.

The goodness-of-fit was tested by using the Kolmogorov–Smirnov test. The proposed distributions were tested against the empirical distribution. The tests were built with the hypotheses given as follows:

- $H_0$ : The data is distributed according to Poisson Exponential (PE) distribution
- $H_1$ : The data is not distributed according to Poisson Exponential (PE) distribution

The empirical distribution of data is defined as

$$F_n(x) = \frac{1}{n} * \sum_{i=1}^n I_{[-\infty, x]}(X_i) \tag{23}$$

where  $I_{[-\infty, x]}(X_i)$  is the indicator function.

Based on the empirical distribution given above, Kolmogorov–Smirnov test statistic is calculated by using the following formula:

$$D_n = \sup_x [|F_n(x) - F_0(x)|] \tag{24}$$

Then the  $D_n$  value is compared to the critical values taken from Kolmogorov–Smirnov tables. If  $D_n$  is less than the corresponding critical value, we fail to reject the null hypothesis that the provided data follows PE distribution. Otherwise, we reject the null hypothesis.

There are several assumptions behind the compound distributions. Firstly, the underlying distributions combined should be independent. Secondly, the compound distributions assume stationarity, i.e., properties of the combined distributions should stay constant over time. Thirdly, the combined distributions should be identically distributed. There are no “one size fits all” compound distributions for all the use cases, that is why, as discussed in the results section, we have tested a variety of them to find the most suitable ones.

## 4 Results and discussion

In this section, we discuss the findings of the paper. We applied the methodology discussed in Sect. 3 and recorded the results. Comparison was made between different approaches to find the level of effectiveness of applying DL to the problem in concern.

### 4.1 Forecasting online sales

We first discuss the results of NN's applied to our problem. As illustrated in the Methodology section, we applied two different types of feed-forward network structures with back-propagation learning algorithms and compared the outcomes with the classical forecasting methods. Before applying these methods, we need to normalize the data by transforming them to the scale of [0, 1]. The formula used for normalization in the pre-processing stage is:

$$X_t = \frac{X_n - X_{min}}{X_{max} - X_{min}} \quad (25)$$

where  $X_t$  is the normalized data,  $X_n$  is the observed data point,  $X_{min}$  is the minimum value of the dataset and  $X_{max}$  is the largest value of the dataset.

Based on the sales data collected from 16 different online stores with 3000 orders in total, we conducted a thorough hyperparameter optimization experiment on the cloud service of AWS called Amazon SageMaker. The objective was set to validation root mean square error and Bayesian approach was selected for reducing the total run time. We provided ranges for the number of hidden layers, hidden nodes, the number of epochs, batch sizes and the activation functions. Maximum total number of training jobs was set to 50 and the maximum number of parallel runs was set to 5. The experiment took 30 min to run. Based on the optimal hyperparameters, we created a script on Python Jupyter Notebook with Keras library installed.

Overfitting is a common problem in modeling with NN's. It happens when the accuracy is high on the training set, but low on the validation set; i.e., the model memorizes the characteristics of the training data. Common ways of handling this is adding regularization techniques such as L2 regularization and dropout, and/or early stopping and/or batch validation. If there is possibility of adding more input data either through data collection or data augmentation, it would yield better results. If the model architecture is too complex, one can try to simplify it as we did in our case. More complex LSTM models did not increase the validation accuracy thus, we went ahead with simpler model.

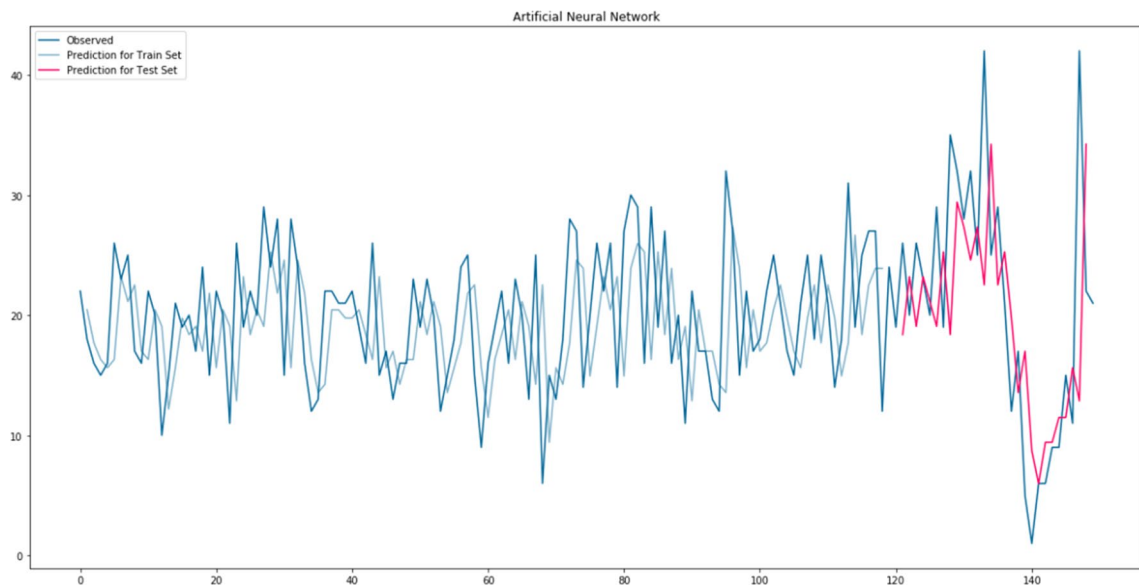
The optimal architecture and properties of the ordinary NN's were as follows:

1. Has two hidden layers each with three hidden nodes;
2. Each of the hidden nodes use RELU activation function;
3. There is one output layer as we are trying to forecast one step ahead;
4. Learning rates are automatically optimized by the model in Keras;
5. The training, validation and test parts were divided as 70%, 10% and 20%.

Then we applied DNN's of type LSTM to the datasets. The optimal properties of LSTM given are:

1. Has two hidden layers each with 300 hidden nodes;
2. Each of the hidden nodes use RELU activation function;
3. The number of output nodes are 3 as we are trying to forecast 3 steps ahead;
4. Learning rates are automatically optimized by the model in Keras;
5. The training validation and test parts were divided as 70%, 10% and 20%.

The training process aims to minimize the deviation between the desired and observed output values. Gradient Descent Algorithm was used to determine the direction of improvement during the training. The results show that DNN's forecast the data very close to the actual values; Fig. 6 presents an example of this. In this dataset, the behavior of the time series is very different (i.e. more volatile) towards the end. Even though the model was trained with the smooth part of the data, DNN's can perform well also in the test portion where the demand behavior is more volatile. The Root Mean



**Fig. 6** The graph of forecasted vs actual values

Square Error (RMSE) is 9.35 for this dataset. Our findings confirm the results obtained by Oroojlooyjadid et al. (2018) that DNN’s work well in forecasting the nervous demand.

Next, we discuss the results obtained from traditional models (Fig. 7). We applied log transformation to have suitable ACF and PACF for the ARIMA model. ARIMA (1,0,2) was the best fitting forecast among all ARIMA (p,d,q) for our data. All the p-values for AR and MA coefficients are zero and p-values for the Box-Pierce Chi-Square statistic are greater than 5%.

The problem with the ARIMA model is that the prediction intervals for the next periods are too large and thus, not much meaningful. Another point is that the AR coefficient is equal to 1 and the moving average component is not invertible. Croston’s method is one of the most famous models to forecast intermittent sales, however it contradicts with our main goal: we would like to see the periods with no order. Instead, Croston’s method distributes the total demand to all days and just smooths out intermittent demand. Grey prediction model was found to be not adequate in our case. The  $\alpha$  value was  $-0.98$  and this caused the model to generate extremely high values for the next periods. Basically, this means an exponential increase in the cumulative forecasts. For example, for period 28, the cumulative forecasted was expressed in terms of  $e^{28}$ . This confirms that Grey models work with smaller samples (for example, with 5 or 6 data).

Best-fitting forecasting methods among traditional models were Moving Average (MA), Single Exponential Smoothing (SES) and Double Exponential Smoothing (DES). Table 3 summarizes the results for all datasets. MSE values have been in the range of 6.70–23.69. In general, it is difficult to find accurate forecasting models among traditional models for such sporadic demand data. In the second column of the table, the best fitting traditional model is given; the next columns

**Fig. 7** ARIMA test results for the main data

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0,9995	0,0143	-69,91	0,000
MA 1	-1,2794	0,0000	-29563,27	0,000
MA 2	-0,2706	0,0687	-3,94	0,000

Number of observations: 53

Residuals: SS = 931468 (backforecasts excluded)  
MS = 18629 DF = 50

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	10,4	21,9	29,7	40,9
DF	9	21	33	45
P-Value	0,317	0,403	0,633	0,648

**Table 3** Forecasting methods results (MSE)

Dataset no	Best fitting traditional forecast model	Traditional—MSE	NN—MSE (% better than Trad)	LSTM—MSE (% better than Trad)
0	MA(3)	11.28	8.56 24%	9.35 17%
1	MA(3)	20.40	15.93 22%	16.85 17%
2	MA(3)	12.10	9.06 25%	10.22 16%
3	MA(4)	8.36	5.18 38%	5.47 35%
4	MA(3)	6.70	5.28 21%	5.42 19%
5	MA(3)	15.43	10.23 34%	11.54 25%
6	MA(3)	19.69	14.53 26%	15.28 22%
7	SES	8.10	4.99 38%	5.26 35%
8	MA(3)	15.52	10.98 29%	11.37 27%
9	MA(4)	16.16	13.46 17%	14.37 11%
10	DES	13.00	10.02 23%	10.48 19%
11	MA(3)	23.65	20.06 15%	20.45 14%
12	MA(3)	8.68	7.26 16%	7.91 9%
13	MA(3)	7.43	5.92 20%	6.35 15%
14	MA(3)	19.87	13.26 33%	14.93 25%
15	MA(3)	12.08	8.09 33%	8.15 33%
16	MA(3)	19.30	14.15 27%	15.26 21%

are the MSE's for the best fitting traditional model and the Deep Learning Model. NN's performed better than traditional methods and dominated them for all datasets. The MSE values for LSTM models were in the range of 5.26–20.45, meaning 9% to 35% better results than best fitting traditional model. As expected, ordinary NN's generated slightly better forecasts, thanks to the one-step-ahead outputs. LSTM's were required to generate three-step-ahead predictions which made their task difficult. Ordinary NN's generated 15% to 38% better forecasts than traditional models.

On top of the MSE, we calculated Mean Absolute Error (MAE) to have a better picture of the accuracy of the different approaches.

We tried adding more hidden layers but adding more layers is not always yielding the expected results. There are contradicting results in the literature: some researchers have found that there are small improvements by adding new hidden layers, some have found that no actual improvement is observed [50]. In our case also, no evident improvements were observed, on the other hand, training the data became much more difficult.

## 4.2 Modeling online sales

In this section, we discuss the proposed approach for modeling online sales. Figure 8 shows the distribution of order interarrival times for all 16 other stores. The p-values confirm that at 5% significance level, we fail to reject that the data come from Exponential distribution with respective parameters.

Then we conducted goodness-of-fit tests for the order sizes (Fig. 9). At 5% significance level, 11 stores appeared to follow Exponential distribution. The best fitting distribution for stores #6, #11 and #16 was Beta distribution. For stores #12 and #14, the best fitting distributions were Lognormal and Erlang, respectively. The reason for such deviations from Exponential distribution was that the sellers had fast-selling items in the intermediate intervals of the price range. For example, the store #12 had watches for sale which became favorite items among buyers, so we have more counts in the second bin of the histogram than in the first one.

Then, the Kolmogorov–Smirnov tests were conducted for the data by assuming  $\alpha=0.05$  (Table 4). In the results tables, we have omitted the datasets that do not follow the Exponential distribution for both order sizes and order interarrival times since they need slightly different formulation.

According to the table above, all input data have shown good fit with the proposed Poisson Exponential Distribution. In addition to the K-S test, we simulated order arrivals using the Poisson-Exponential distribution for 12 data sets that followed Exponential distribution for both the order interarrival times and order sizes. Simulations were done with 10 000 replications each, and the average values have been recorded. Two main parameters are total sales and

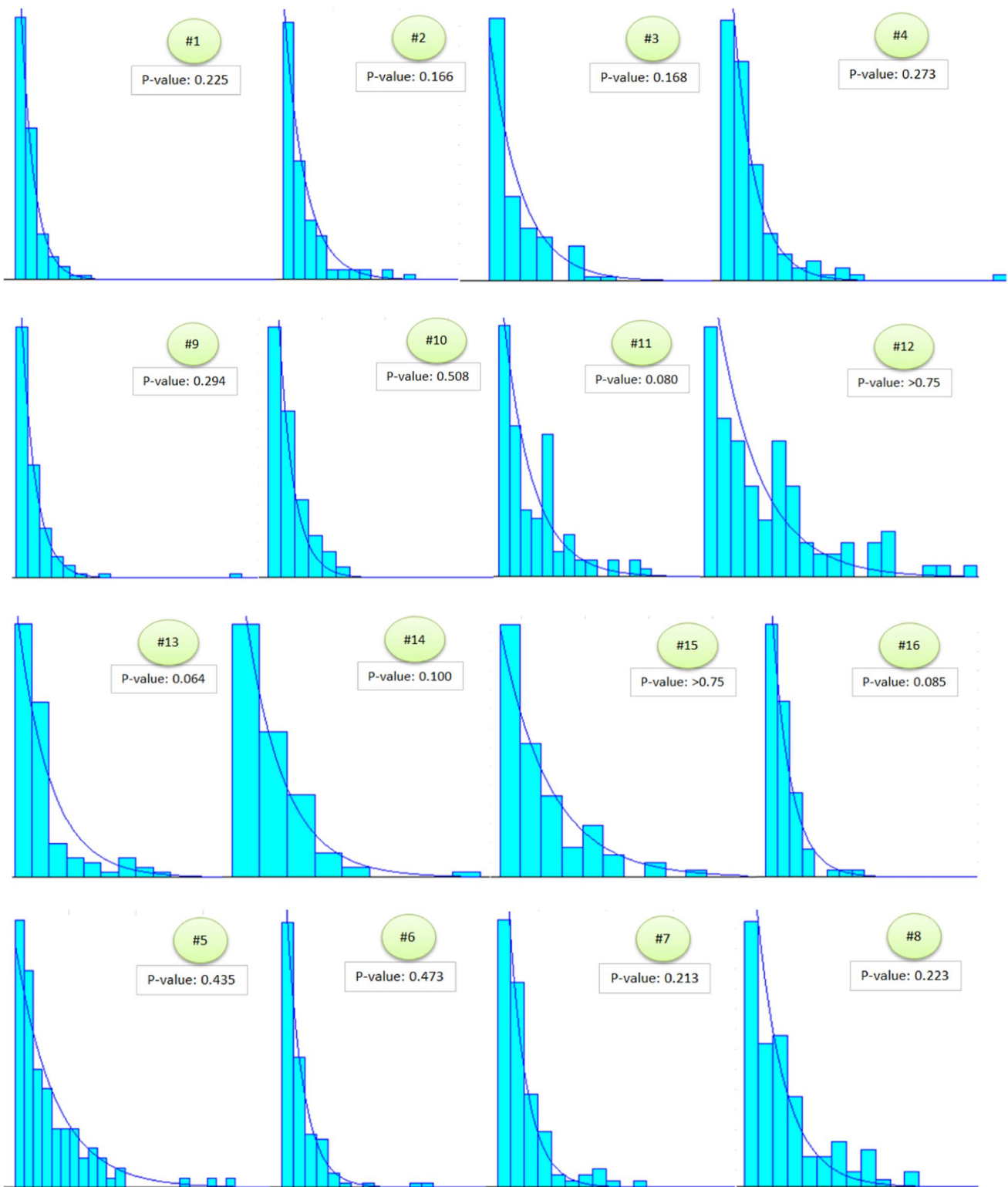


Fig. 8 Distribution test results for the order arrivals

days when order arrivals happen. Table 5 depicts the comparison between the real data and randomly generated data from the Poisson-Exponential distribution for the same period. We did not simulate the datasets 6, 11, 12, 14, and 16 since they do not follow the Poisson-Exponential distribution and need a different setup for the simulations.

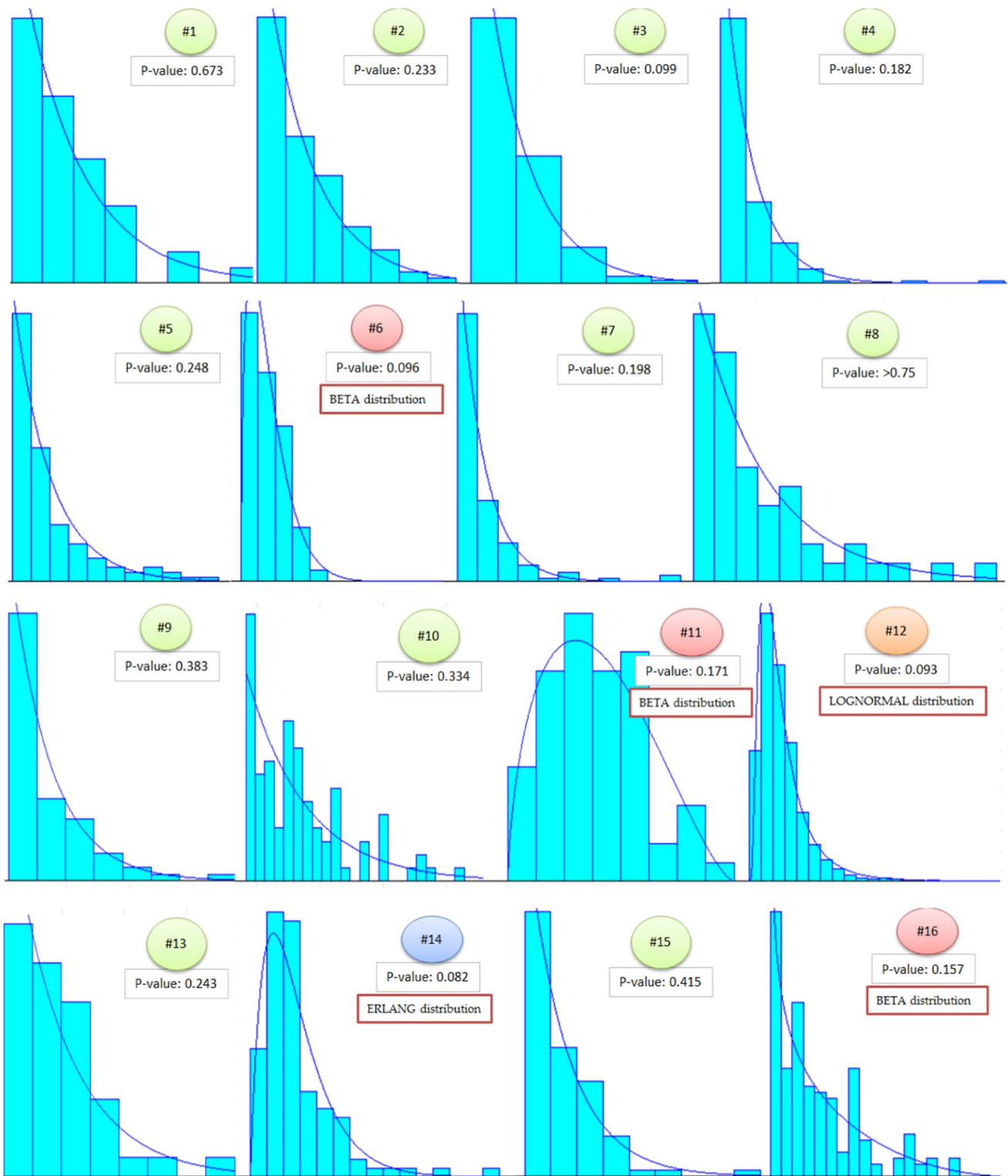


Fig. 9 Distribution test results for the order sizes

The accuracy in these parameters is in the range of 1–7%; showing the power of the Poisson Exponential distribution in modeling the intermittent online sales. The existing schemes have explored the application of the Neural Networks to other domains such as spare parts demand forecasting which is a well-known problem. In our case, we analyze both order interarrival times and order sizes, formulate the problem in terms of the Compound distributions, forecast the

**Table 4** Forecasting methods results (MAE)

Dataset no	Best fitting traditional forecast model	Traditional-MAE	NN-MAE (% better than Trad)	LSTM-MAE (% better than Trad)
0	MA(3)	3.16	2.9	8%
1	MA(3)	5.36	4.66	13%
2	MA(3)	3.96	3.93	1%
3	MA(4)	3.5	3.09	12%
4	MA(3)	2.37	1.48	38%
5	MA(3)	3.79	3.66	3%
6	MA(3)	4.9	3.98	19%
7	SES	2.14	1.52	29%
8	MA(3)	3.24	2.86	12%
9	MA(4)	3.13	2.98	5%
10	DES	3.36	3.03	10%
11	MA(3)	5.24	4.39	16%
12	MA(3)	3.06	2.75	10%
13	MA(3)	1.94	1.48	24%
14	MA(3)	3.84	3.46	10%
15	MA(3)	3.7	3.21	13%
16	MA(3)	4.57	4.27	7%

**Table 5** The K-S test results for the datasets

Dataset number	K-S test statistic	Critical value	Conclusion
0	0.13	0.23	We fail to reject the $H_0$
1	0.06	0.26	We fail to reject the $H_0$
2	0.13	0.23	We fail to reject the $H_0$
3	0.08	0.21	We fail to reject the $H_0$
4	0.11	0.26	We fail to reject the $H_0$
5	0.20	0.21	We fail to reject the $H_0$
7	0.08	0.28	We fail to reject the $H_0$
8	0.09	0.22	We fail to reject the $H_0$
9	0.13	0.25	We fail to reject the $H_0$
10	0.08	0.28	We fail to reject the $H_0$
13	0.13	0.26	We fail to reject the $H_0$
15	0.10	0.21	We fail to reject the $H_0$

future sales by using DNN's and provide their performance on real data from e-commerce platforms. The existing schemes have explored the application of the Neural Networks to other domains such as spare parts demand forecasting which is a well-known problem. The contribution of this paper is that we analyze both order interarrival times and order sizes, formulate the problem in terms of the Compound distributions, forecast the future sales by using DNN's and provide their performance on real data from e-commerce platforms. This is a more specific application which is based on the needs and requirements of e-business operators.

## 5 Conclusion

In this paper, we discussed novel ways to forecast and model intermittent online sales by applying DNN's and Compound distributions, respectively. The research task originated from a real-life situation faced by the online sellers. There conclusions for the two main problems are:

- (1) Forecasting online sales for different stores. The main finding is that Deep Learning models with two hidden layers dominate the traditional models in forecast accuracy.
- (2) Finding suitable distributions to model online sales. We found out that Poisson Exponential distribution can be used to model such sales in most of the cases.

We found out that LSTM models outperform traditional models for multi-step-ahead forecasting by up to 35%. For single step forecasts, ordinary NN's can be used which generate up to 38% better forecasts than traditional models. Another finding of this article is that if the order arrival process is following Poisson distribution, DNN's can be used for forecasting the future demand with higher accuracy than the traditional methods. In the modeling part, it was concluded that most of the time Poisson-Exponential distribution can be used to model intermittent online sales. We tested the distributions by using the Kolmogorov–Smirnov test and they showed good fit. These models can be used for many purposes such as simulating the sales process (Table 6).

Advanced methods of Artificial Intelligence, including Deep Learning will present new possibilities for forecasting complex demand patterns. Good forecasts are needed for automated order replenishment and inventory control decisions. There will be required powerful algorithms to perform these tasks; and Deep Learning Models provide us this power as discussed by many authors in the literature. The algorithms will make forecasts, find distributions of the underlying demand/lead time data and will make decisions on how much to keep in stock, when to order, how much to order etc. This paper has demonstrated how applying DL models for forecasting online sales, especially intermittent ones; DNN's can present a good potential in forecasting accuracy.

In this study, we used data mainly from individual sellers. In the future, the analysis should be extended to analyze the performance of high-volume selling items with complex patterns. Future research should attempt to use compound Poisson distribution with these to cover all the possible scenarios and to get a more accurate picture of the process. The data that we used from 17 online stores were selected to have low means and high standard deviations to analyze the intermittent sales patterns. As such they might not represent the whole e-business in general. It is evident that the stores with regular sales do not have intermittency in their demand and they can be modeled by using the Normal distribution most of the time. We have shown the use of DNN's in time series; but recently we have seen the rise of generative AI

**Table 6** Simulation results

	Real	Simulated	Error %		Real	Simulated	Error %
Dataset 0				Dataset 7			
Total amount of sales	6665	6213	− 6.6	Total amount of sales	6128	6236	1.8
Days with orders	121	114	− 5.8	Days with orders	161	169	5.0
Dataset 1				Dataset 8			
Total amount of sales	26,180	27,096	3.5	Total amount of sales	4272	4203	− 1.6
Days with orders	170	162	− 4.6	Days with orders	191	186	− 2.4
Dataset 2				Dataset 9			
Total amount of sales	11,907	11,203	− 5.9	Total amount of sales	11,892	11,203	− 5.8
Days with orders	171	178	4.1	Days with orders	123	115	− 6.9
Dataset 3				Dataset 10			
Total amount of sales	5076	4982	− 1.9	Total amount of sales	3883	3915	0.8
Days with orders	184	195	5.8	Days with orders	165	153	− 7.0
Dataset 4				Dataset 13			
Total amount of sales	344	326	− 5.2	Total amount of sales	1888	1796	− 4.9
Days with orders	140	132	− 5.7	Days with orders	118	112	− 5.1
Dataset 5				Dataset 15			
Total amount of sales	6513	6199	− 4.8	Total amount of sales	8904	8314	− 6.6
Days with orders	121	116	− 3.8	Days with orders	130	138	6.2

tools such as ChatGPT and Bard. It would be interesting to study the possible applications of such systems in forecasting tasks of e-commerce.

**Authors' contributions** All authors participated in conceptualization of the research and development of formal analysis. YA completed the methodology design and data analysis, and writing the manuscript. PH contributed in investigation, supervision and writing, review and editing. All authors read and approved the final manuscript.

**Data availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Competing interests** The authors declare that they have no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Aamer A, Eka Yani L, Alan Priyatna I. Data analytics in the supply chain management: review of machine learning applications in demand forecasting. *Oper Supply Chain Manag Int J*. 2020;14(1):1–13.
2. Alfonso V, Boar C, Frost J, Gambacorta L, Liu J. E-commerce in the pandemic and beyond. *BIS Bull*. 2021; 36.
3. Amin-Naseri MR, Tabar BR. Neural network approach to lumpy demand forecasting for spare parts in process industries. In *Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on computer and communication engineering*, (pp. 1378–1382). IEEE. 2008.
4. Archibald BC, Koehler AB. Normalization of seasonal factors in Winters' methods. *Int J Forecast*. 2003;19(1):143–8.
5. Chen FL, Chen YC. An investigation of forecasting critical spare parts requirement. In *Computer Science and Information Engineering, 2009 WRI World Congress on computer science and information engineering (Vol. 4, pp. 225–230)*. IEEE. 2009.
6. Croston JD. Forecasting and stock control for intermittent demands. *Oper Res Quart*. 1972; 289–303.
7. Ferreira KJ, Lee BHA, Simchi-Levi D. Analytics for an online retailer: demand forecasting and price optimization. *Manuf Serv Oper Manag*. 2015;18(1):69–88.
8. Figueiredo MCB. E-COMMERCE: FORECASTING DEMAND FOR NEW PRODUCTS, Proceedings of IADIS International Conference e-Commerce, 102–112. 2008.
9. Fu W, Chien CF, Lin ZH. A Hybrid Forecasting Framework with Neural Network and Time-Series Method for Intermittent Demand in Semiconductor Supply Chain. In *IFIP International Conference on Advances in Production Management Systems* (pp. 65–72). Springer, Cham. 2018.
10. Ghobbar AA, Friend CH. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Comput Oper Res*. 2003;30(14):2097–114.
11. Gosavi A. Reinforcement learning: a tutorial survey and recent advances. *INFORMS J Comput*. 2009;21(2):178–92.
12. Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645–6649). IEEE. 2013.
13. Gunasekaran A, et al. E-commerce and its impact on operations management. *Int J Prod Econ*. 2002;75(1):185–97.
14. Gutierrez RS, Solis AO, Mukhopadhyay S. Lumpy demand forecasting using neural networks. *Int J Prod Econ*. 2008;111(2):409–20.
15. Helo P, Hao Y. Artificial intelligence in operations management and supply chain management: an exploratory case study. *Prod Planning Control*. 2022;33(16):1573–90.
16. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–80.
17. Jeble S, Dubey R, Childe SJ, Papadopoulos T, Roubaud D, Prakash A. Impact of big data and predictive analytics capability on supply chain sustainability. *Int J Logist Manag*. 2018;29(2):513–38. <https://doi.org/10.1108/IJLM-05-2017-0134>.
18. Jiang P, Huang Y, Liu X. Intermittent demand forecasting for spare parts in the heavy-duty vehicle industry: a support vector machine model. *Int J Prod Res*. 2021;59(24):7423–40.
19. Kayacan E, Ulutas B, Kaynak O. Grey system theory-based models in time series prediction. *Expert Syst Appl*. 2010;37(2):1784–9.
20. Khashei M, Bijari M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl Soft Comput*. 2011;11(2):2664–75.
21. Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: a survey. *Int J Robot Res*. 2013;32(11):1238–74.
22. Kourentzes N. Intermittent demand forecasts with neural networks. *Int J Prod Econ*. 2013;143(1):198–206.

23. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105). 2012.
24. Larrain M. The PMI, the T-bill and inventories: a comparative analysis of neural network and regression forecasts. *J Supply Chain Manag.* 2007;43(2):39–51.
25. Lawrence S, Giles CL, Tsoi AC, Back AD. Face recognition: a convolutional neural-network approach. *IEEE Trans Neural Networks.* 1997;8(1):98–113.
26. Lee HL, Whang S. E-business and supply chain integration. In *Stanford Global Supply Chain Management Forum* (Vol. 2). 2001.
27. Lengu, et al. Spare parts management: linking distributional assumptions to demand classification. *Eur J Oper Res.* 2014;235(2014):624–35.
28. Li BH, Hou BC, Yu WT, Lu XB, Yang CW. Applications of artificial intelligence in intelligent manufacturing: a review. *Front Inf Technol Electron Eng.* 2017;18(1):86–96.
29. Lima-Junior FR, Carpinetti LCR. Quantitative models for supply chain performance evaluation: a literature review. *Comput Ind Eng.* 2017;113:333–46.
30. Nejedly P, Plesinger F, Viscor I, Halamek J, Jurak P. Prediction of Sepsis Using LSTM Neural Network With Hyperparameter Optimization With a Genetic Algorithm. In *2019 Computing in Cardiology (CinC)* (pp. Page-1). IEEE. 2019.
31. Oroojlooyjadid A, Snyder L, Takáč M. Applying deep learning to the newsvendor problem. 2016. arXiv preprint [arXiv:1607.02177](https://arxiv.org/abs/1607.02177).
32. Pournader M, Ghaderi H, Hassanzadegan A, Fahimnia B. Artificial intelligence applications in supply chain management. *Int J Prod Econ.* 2021;241: 108250.
33. Qiu X, Zhang L, Ren Y, Suganthan PN, Amaratunga G. Ensemble deep learning for regression and time series forecasting. In *2014 IEEE symposium on computational intelligence in ensemble learning (CIEL)* (pp. 1–6). IEEE. 2014.
34. Radzi NHM, Haron H, Johari TIT. Lot sizing using neural network approach. *Second IMT-GT Regional.* 2006.
35. Reimers N, Gurevych I. Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. 2017. arXiv preprint [arXiv:1707.06799](https://arxiv.org/abs/1707.06799).
36. Sani B, Kingsman BG. Selecting the best periodic inventory control and demand forecasting methods for low demand items. *J Oper Res Soc.* 1997;48:700–13.
37. Sharma R, Shishodia A, Gunasekaran A, Min H, Munim ZH. The role of artificial intelligence in supply chain management: mapping the territory. *Int J Prod Res.* 2022;60(24):7527–50.
38. Sheu SH, Tai SH, Hsieh YT, Lin TC. Monitoring process mean and variability with generally weighted moving average control charts. *Comput Ind Eng.* 2009;57(1):401–7.
39. Snyder RD, Ord JK, Beaumont A. Forecasting the intermittent demand for slow-moving inventories: a modelling approach. *Int J Forecast.* 2012;28(2):485–96.
40. Sohrabpour V, Oghazi P, Toorajipour R, Nazarpour A. Export sales forecasting using artificial intelligence. *Technol Forecast Soc Chang.* 2021;163: 120480.
41. Solis AO, Longo F, Mukhopadhyay S, Nicoletti L. An empirical investigation of comparative performance of approximate and exact corrections of the bias in Croston's method in forecasting lumpy demand. *Int J Simul Process Model.* 2017;12(6):535–50.
42. Swaminathan JM, Tayur SR. Models for supply chains in e-business. *Manage Sci.* 2003;49(10):1387–406.
43. Syntetos AA, Boylan JE. The accuracy of intermittent demand estimates. *Int J Forecast.* 2005;21(2):303–14.
44. The European Union, E-commerce statistics for individuals (2016). Retrieved from [http://ec.europa.eu/eurostat/statistics-explained/index.php/E-commerce\\_statistics\\_for\\_individuals](http://ec.europa.eu/eurostat/statistics-explained/index.php/E-commerce_statistics_for_individuals).
45. Thompson CS. Homogeneity analysis of a rainfall series: an application of the use of a realistic rainfall model. *J Climatol.* 1984;4:609–19.
46. Torres JF, Fernández AM, Troncoso A, Martínez-Álvarez F. Deep learning-based approach for time series forecasting with application to electricity load. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 203–212). Springer, Cham. 2017.
47. Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. *Technol Forecast Soc Chang.* 2002;69(1):71–87.
48. Turkmen AC, Wang Y, Januschowski T. Intermittent demand forecasting with deep renewal processes. 2019. arXiv preprint [arXiv:1911.10416](https://arxiv.org/abs/1911.10416).
49. Willemain TR, Smart CN, Schwarz HF. A new approach to forecasting intermittent demand for service parts inventories. *Int J Forecast.* 2004;20(3):375–87.
50. Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks: the state of the art. *Int J Forecast.* 1998;14(1):35–62.
51. Zhu X, Zhang G, Sun B. A comprehensive literature review of the demand forecasting methods of emergency resources from the perspective of artificial intelligence. *Nat Hazards.* 2019;97:65–82.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.