CHAPTER 6

DEEP LEARNING APPLICATIONS IN SALES FORECASTING

Erhan USTAOĞLU[1]

1. INTRODUCTION

Sales forecasting has always been of utmost importance to companies almost in all industries. Many forecasting techniques have been developed and applied by companies for various industries and time spans. Already in 1980s, most of the managers considered sales forecasting to be critical to companies' success since it provided an important input into planning process. Sales forecasting were used mostly for production planning but also for budget, strategic, marketing, inventory, and logistic planning. Regression, simulation, time series analysis and computer models integrated to companies' operational systems were the mostly deployed techniques for sales forecasting (Mentzer & Cox, 1984). Forecasting methods ranged from surveys of salesforce, customers, and industry to econometric models. Many companies used several forecasting models simultaneously in order to improve the forecast accuracy (West, 1994). Sales forecasting was considered to be a business process as well as choosing the right techniques and tools for the matter. Hence companies were supposed to employ dedicated units for sales forecasting (Mentzer et al.,1997). Sales organizations in companies generally used judgmental assessments based on past experiences and but operations are more based on quantitative data. However, when forecasting sales these quantitative and qualitative data were rarely interrelated. Sanders and Ritzman (2004) developed a framework to address this issue. They proposed four methodologies for combining judgmental and quantitative assessments for forecasting. They also pointed out that success of these methodologies was dependent on the structure of the organization although each methodology had its advantages.

Since the accuracy of sales forecasting had been becoming crucial with the growing competence, more sophisticated methods were developed and applied. Using time series and cross-sectional methods independently for sales forecasting

[1] Assis. Prof. Dr., Erhan Ustaoğlu, Marmara University, Faculty of Administration, Management Information Systems, erhan.ustaoglu@marmara.edu.tr

was not reflecting the relationships in data properly, then longitudinal data models were proposed. Frees and Miller (2004) introduced a longitudinal data mixed model and applied it using weekly lottery sales data together with postal codes. They achieved better results compared to using one dimensional data.

Various methods and techniques have been developed and applied for analyzing time series data for sales forecasting. Many of these methods are linear models which are easy to implement. If the performance of a simpler model is satisfactory for sales forecasting, it should be preferred over a more sophisticated model but they usually fail to represent the complex relationships in the data. Hence, these models are not always capable of predicting the future sales correctly. As a consequence, more complicated non-linear models have become popular in the field of time series analysis (Chu & Zhang, 2003). Although non-linear models have the capability of reflecting the non-linearity of the time-series, they are often appropriate for single dimensional data and every model is only suitable for certain trends in data based on certain assumptions. Moreover, it is not very clear before the implementation whether the model could reflect the non-linearity in the data or not, so a trial-and-error method should be conducted for choosing a suitable method. On the other hand, these models are unable to cope with multidimensional time series with non-linearities. In order to address these difficulties in dealing with multidimensional non-linearities, neural network (NN) models were proposed to implement for sales forecasting analysis. NNs had been already employed for forecasting in the field of economics and finance before its implementation in sales forecasting and the applications rapidly increased in this field too, due to the ease of implementation and promising results. Already in early 1990s, implementations of NNs in sales forecasting appeared. Chakraborty et al. (1992) applied NNs for predicting monthly flour prices using 100 monthly historical data and compared the results with that of classical autoregressive moving average (ARMA) model. The results were above their expectations outperforming the classical model. The root mean squared error values were better for NN by at least an order. They, then, concluded that the NN could be applied in other fields since no specific prior information is needed for the underlying phenomenon. Ansuj et al. (1996) compared NN with autoregressive integrated moving average (ARIMA) model in predicting the sales of a medium sized enterprise and reached better forecasting results for NN. Thiesing and Vornberger (1997) obtained better results for NN compared to naïve and moving average in predicting the sales of a German Supermarket. But not all of the researchers reached results that support the superiority of NNs over the classical time series analysis. Tang et al. (1991) compared NN and classical Box-Jenkins method for three time series of different complexity. They found out that although NN outperformed Box-Jenkins method, both methods yielded comparable results for time series of short memory. Hence, they concluded that although NN yielded good results with also long memory time series, there were difficulties choosing the parameters and the optimal structure for training.

Many different variations of NNs were also proposed in order to overcome the difficulties in sales forecasting. Kuo (2001) proposed a fuzzy NN for addressing sudden changes in sales. Initial weights for the training were produced by a generic algorithm taking into account the fuzzy rules obtained through the experiences of experts. The model was applied for the sales forecast of a convenience store and the results obtained were better compared to those of classical method and conventional NN. Chang et al (2005) proposed an evolving NN, replacing the backpropagation, which causes convergence problems, by a genetic algorithm. They applied their model to printed circuit board short- to mid-term sale forecasting and found out that the forecasting accuracy was better compared to classical forecasting methods and NN with backpropagation. They also pointed out that the implementation of the model would reduce the cost of forecasting for printed circuit board companies. Doganis et al. (2006) also proposed an adaptive NN with radial basis function and special genetic algorithm and applied it for the forecasting of a diary product. They compared the results with various linear and non-linear models and found out that the proposed algorithm outperformed the others. Another example of an extended NN model is an extreme learning machine presented by Sun et al. (2008). The extreme learning machine is a feedforward NN with one hidden layer. The model was applied for fashion retail sales forecasting and it was demonstrated that the results were superior to those NNs with backpropagation.

Although not as popular as NNs, other machine learning methods like support vector regression, K-nearest regression, Bayesian NN, generalized regression NN (kernel regression), Gaussian regression, have been deployed for time series analysis in general and for sales forecasting. A study, which tested NNs and support vector regression (SVR) on 36 artificial time series, showed that these two methods have comparable accuracies and outperform classical time series analysis methods (Crone, 2006). Accordingly, sales forecasting research based on SVR are not infrequent. Some of the examples of SVR for sales forecasting can be given as follows. Brühl et al. (2009) predicted German monthly, quarterly and yearly automobile sales with support vector machines with ε -Regression and Gaussian kernel and compared the results with those of multiple linear regression model.

They found out that support vector machine predictions were clearly more accurate than multiple linear regression. Guajardo et al. (2010) employed SVR for predicting the weekly sales of 5 different products. They used a dynamic SVR together with a model updating strategy for feature selection. They obtained better prediction accuracies for all products for three different error measures. Lu et al. (2010) applied SVR for predicting sales of information technology products. They first implemented a data mining technique called multivariate adaptive regression splines, for choosing relevant features for SVR. They reached better forecasting results with their combined model for three different information technology products compared to using SVR and multivariate adaptive regression splines independently. Yuan (2012) improved the accuracy of SVR by dynamically optimizing its parameters using a genetic algorithm. The resulting model then used for forecasting the monthly truck sales volumes in Taiwan. The proposed algorithm was compared to least-mean square algorithm, NNs, ordinary SVR and was found to be superior in predicting the sales. Karmy and Maldonado (2019) recently implemented SVR using a hierarchical time series approach for sales forecasting in retail business. They demonstrated that their SVR model outperformed traditional ARIMA and Holt-Winters models in predicting sales. Research of other machine learning techniques are relatively rare compared to NNs and SVR. One of the examples that is worth mentioning is a recent study with K-nearest neighbor regression for predicting retail drugstore sales. The results however indicated that the linear regression model yielded better prediction accuracy than K-nearest neighbor regression for the data set used (Kohli et al., 2021).

Deep learning techniques have been developed and used for machine learning tasks where it was thought that classical learning models, some of which were mentioned above, were insufficient or deep learning would provide better results. Deep learning models are applied successfully in many fields such as speech and pattern recognition or natural language processing. Recently the applications of deep learning techniques have become popular for time series forecasting too (Torres et al., 2021). Sales forecasting is a field, which is no exception, to this trend. In this survey deep learning models, which have been deployed for sales forecasting will be investigated. The research, which have been conducted will be classified according to industries, compared models, evaluation methods, techniques and features, which are used. First, the deep learning models, which are used for sales forecasting will be briefly described.

2. DEEP LEARNING MODELS IN SALES FORECASTING

Since sales forecasting is a time series analysis, the models used for general time series are also deployed for sales forecasting frequently. This study surveys the deep learning applications used in sales forecasting through Google Scholar. The survey encompasses the recent research from 2017. The most frequently used models and the hybrids of these models are included in the survey although there exist many other valuable novel deep learning model proposals. The deep learning models, which are included in this survey are briefly described in this section.

2.1. Deep Neural Network (DNN)

Deep neural networks (DNNs), are artificial neural networks (ANNs) with more than one layer. They can be considered as multilevel perceptron (MLP), which imitates the functioning of the neurons in human brain cells. DNNs are feed forward neural networks (FFNNs). The information flow in DNNs is always in the same direction, from the input to the output layer. When implementing a DNN for sales forecasting, like any other learning task, a set of parameters, i.e., features are fed to the input layer. The DNN weights are then computed through a backpropagation algorithm, which is the gradient decent, as to minimize the loss function. The loss function is based on the difference of the actual and the predicted values of the target value. That's why DNNs are, in some studies, called backpropagation neural networks (BPNNs). On the other hand, the calculated weights in input or any hidden layer are passed to the following layer through a non-linear activation function. A DNN architecture is depicted in Figure 1. The input vector contains the parameters for sales forecasting like historical sales, economic indicators, store, weather, sentimental data, etc., and would be the prediction for sales quantity or revenue.

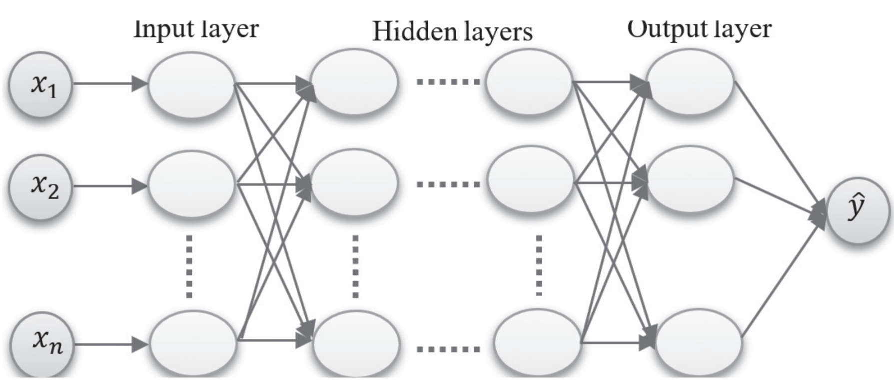

Figure 1. A DNN architecture

2.2. Convolutional Neural Network (CNN)

Sales forecasting can be a very hard task due to the nature of the data involved that comprises seasonality, high dimensionality, and non-linear interdependencies between attributes. Hence, it is generally difficult to determine the convenient set of features for the prediction. In this respect, CNNs are proposed for discovering and extracting the highly non-linear structures in sales data. In most of the studies CNNs are used in learning the relevant features, which may reflect the seasonality and non-linearity in data, and the other techniques are deployed for prediction using these extracted features (Zhao et al., 2017).

CCNs are usually preferred in retail and e-commerce sales forecasting due to the diverse features, which are common to these business fields. Different types of CCNs with different combinations of layer structures are implemented. Since CNNs have been successfully implemented for feature learning in the field of pattern recognition, semantic role labeling or sentence classification, Ma and Field (2020) proposed a double channel convolutional neural network (DCCNN) for meta-learning in retail sale forecasting. They, then used the proposed meta-learner in combination with several known base forecasters and showed that the forecasting was superior compared to that without the meta-learner. Xu et al. (2021) proposed a dilated CNN for distracting the diverse features for sales forecasting in sportswear retailing. They combined it with MLP and showed that the proposed method was superior to statistical methods AR, MA, and ARIMA. Yin and Tao (2021) used CNN in combination with a denoising automatic encoder for achieving the best performing model structure for e-commerce sales forecasting. They obtained better performance with their model compared to Adaboost.

CCNs were developed by imitating biological vision systems for pattern recognition especially for 2 dimensional tasks. However, they have found widespread applications in other field like time series analysis due to the properties of achieving automatic feature extracting and reducing computational loads. A classical CNN structure is made of three layers; convolutional, pooling and fully connected. Filters comprised of weights are applied to the input features in convolutional layer. These learnable filters, which are called kernels, are smaller in size compared to the input. Hence, the hidden structures in overall dataset are extracted decreasing the computational load in further training. This also helps reducing the risk of overfitting (O'Shea & Nash, 2015). A convolutional operation in 2-D is depicted in Figure 2, in which an input matrix size of 5x5 is convolved with a kernel of 3x3 resulting in a 3x3 feature map. The kernel is first flipped and slid over the entire input. Pooling, which is implemented by taking maximum or average of sliding a kernel over input matrix, is used to reduce the parameters further. This helps, on one hand, reducing computational load and chance of overfit, but on the other hand, it can be destructive since it results in data loss. Therefore, it should be handled with care.

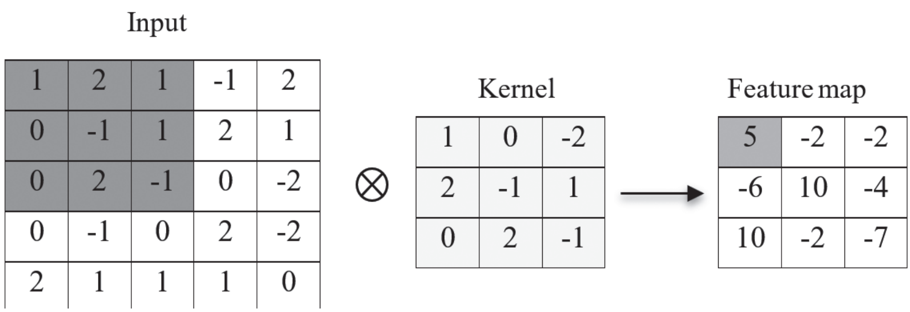

Figure 2. Convolutional layer operation for 2-D

Convolutional and pooling layers can be arranged in any order and can be optimized for the type of learning task. Fully connected layers are analogous to DNN layers and handle the similar learning tasks.

2.3. Recurrent Neural Network (RNN)

Recurrent neural networks (RNNs), contrary to FFNNs, are capable of handling temporal and sequential data. Unlike FFNNs, RNNs keep the information of prior set of inputs and this information is used for learning process together with the current input. Since sales forecasting requires complex temporal analysis, RNNs have recently been used more frequently in this respect. Pemathilake et al. (2018) combined ARIMA with RNN and forecasted sales of IT products using historical and sentimental data. The proposed model showed better performances compared to single ARIMA and RNN models. Lin et al. (2021) forecasted the sales of a service station on a freeway in Taiwan. They found that RNN outperformed ANN for holidays and vice versa for weekdays. Another study investigated the retail sales depending on the economic factors such as CPI (consumer price index), CCI (consumer confidence index), PMI (purchasing managers' index), WPI (wholesale price index), IPI (industrial production index), REP (retail employment population), real wage, unemployment rate, oil price, etc. The sales were forecasted using classical RNN, LSTM (long short-term memory), and GRU (gated recurrent unit). The performances were compared against RF (random forest) and GB (gradient boosting). It was found that RNNs performed slightly better compared to classical machine learning model and but concluded that RNNs have the capability to handle more complex learning tasks if sufficient data is available (Wang, 2022).

RNN nodes process data from both current input and underlying hidden layers in the model, which accounts for sequential and temporal dependencies. An RNN with one hidden layer is illustrated in Figure 3. The input vector contains the features and is the target to be predicted at any time .

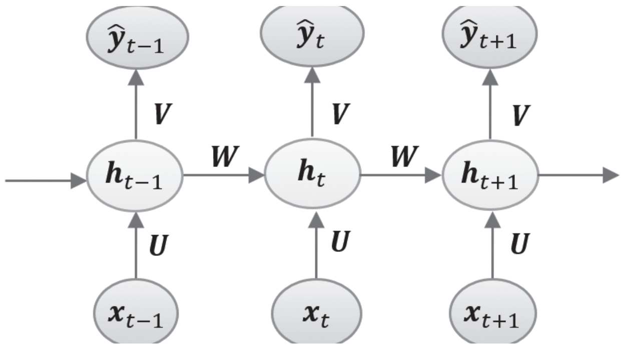

Figure 3. An RNN structure with one hidden layer

Hidden state value at any time *t* is given as follows for the RNN with one hidden layer in Figure 2

$$\boldsymbol{h}_t = z\,(\boldsymbol{U}\,\boldsymbol{x}_t + \boldsymbol{W}\,\boldsymbol{h}_{t-1} + \boldsymbol{b}) \qquad (1)$$

where z is an activation function such as sigmoid or tanh, $\boldsymbol{U}$ is the initial weight matrix for the input features, $\boldsymbol{W}$ is the weight matrix associated with hidden state h and $\boldsymbol{b}$ is a threshold vector. Weight matrices $\boldsymbol{W}$ and $\boldsymbol{U}$ matrices for multiple layer RRN structures are to be optimized through training process. The predicted target value is given as follows where z is an activation function, $\boldsymbol{V}$ is the weight matrix between hidden layer and the output layer to be optimized and $\boldsymbol{c}$ is a threshold vector.

$$\hat{\boldsymbol{y}}_t = z\,(\boldsymbol{V}\,\boldsymbol{h}_t + \boldsymbol{c}) \qquad (2)$$

2.3.1. Long Short-Term Memory (LSTM)

Time series analysis requires capturing long-term interdependencies intrinsic to historical data and most of the algorithms fail to deal with this problem. Sales forecasting, which usually incorporates multiple temporal non-linearly dependent features, is a field that this sort of analysis is definitely required. LSTM, a type of RNNs, belongs to machine learning models which can address this challenge. Hence, the number of research regarding implementation of LSTM neural networks has recently increased dramatically. Also, increase in the number of studies of sales forecasting has been striking (Yu et al., 2017). Helmini et al. (2019) used a peephole connected LSTM, which captures temporal patterns more accurately, for forecasting retail sales and found that LSTM outperformed XGB and random forest regression (RFR), although they reported overfitting issues for LSTM because of a small dataset. Hong (2021) also employed LSTM for forecasting online sales of Korean apparel retailer using historical and weather data and compared the forecasted sales with actual sales. He reported that forecasted sales figures very close to actual ones and that more training data would result in more accurate predictions. Shih and Lin (2019) applied LSTM for sales prediction in e-commerce. They used online historical sales data and comments to train an LSTM model and achieved accurate prediction results for short-term demands.

LSTM have also been used extensively in combination with other sales forecasting models. Han (2020) used ARIMA and LSTM for forecasting pharmaceutical sales. The linear part of the sales data was modeled by ARIMA whereas the non-linear part modeled by LSTM. The combined model achieved more accurate prediction compared to single ARIMA and LSTM models. Combination of LSTM with ARIMA have been a popular field of research with applications in real estate, e-commerce and automotive sectors too (Temur et al., 2019; Vavliakis et al., 2021; Cheng et al., 2018).

LSTM has also been implemented in combination with other machine learning techniques. He and Yu (2020) used an LSTM-LGBM (Light Gradient Boosting Machine) combined model for predicting short-term vegetable sales. The combined model outperformed single models. Zhao et al. (2021) added a feature selection function to LSTM using a denoising autoencoder (DAE), and applied the model for predicting online retail sales. They reported that the combined model outperformed multiple linear regression (MLR), SVR, ANN and single LSTM.

Number of recent LSTM implementations for sales forecasting as well as for time series and other application areas are far greater than that of classical RNN. The reason being that, classical RNN model suffers from a vanishing gradient problem that makes the effects of past information converge to zero. LSTM model was developed to address this issue. LSTM memory cells, unlike RNN, do not only transfer the previous information to upper layers through a sole activation function, but they control the information flow through three cell vectors, which are optimized through the training process and called gates. The forget gate in an LSTM cell controls how much of the previous information will be retained and how much of it will be discarded. The input gate controls the flow of new information into the LSTM cell and the output gate determines the cell state vector based upon the information coming from forget and input gates (Zhao et al., 2021). An LSTM cell structure is shown in Figure 4.

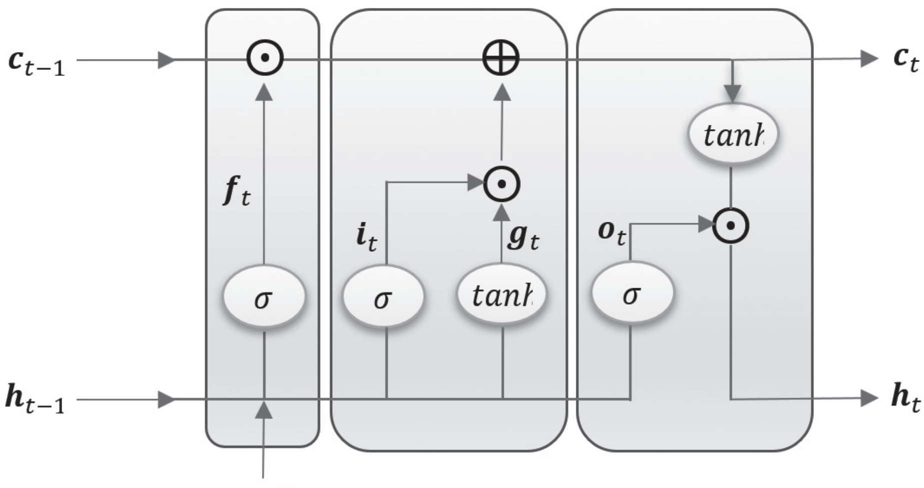

Figure 4. An LSTM cell structure

For any layer in the LSTM structure, the equations governing the gate functions are as follows

$$\boldsymbol{i}_t^l = \sigma\,(\boldsymbol{W}^{l-1}\,\boldsymbol{h}_t^{l-1} + \boldsymbol{W}^l\,\boldsymbol{h}_{t-1}^l + \boldsymbol{b}^i) \qquad (3)$$

$$\boldsymbol{o}_t^l = \sigma\,(\boldsymbol{W}^{l-1}\,\boldsymbol{h}_t^{l-1} + \boldsymbol{W}^l\,\boldsymbol{h}_{t-1}^l + \boldsymbol{b}^o) \qquad (4)$$

$$\boldsymbol{f}_t^l = \sigma\,(\boldsymbol{W}^{l-1}\,\boldsymbol{h}_t^{l-1} + \boldsymbol{W}^l\,\boldsymbol{h}_{t-1}^l + \boldsymbol{b}^f) \qquad (5)$$

$$\boldsymbol{g}_t^l = tanh\,(\boldsymbol{W}^{l-1}\,\boldsymbol{h}_t^{l-1} + \boldsymbol{W}^l\,\boldsymbol{h}_{t-1}^l + \boldsymbol{b}^g) \qquad (6)$$

where is the sigmoid function, and are the weight vectors of the previous and the current hidden layer respectively, and 's are bias vectors. Hence, cell state and hidden layer vectors are calculated as follows

$$\boldsymbol{c}_t^l = \boldsymbol{f}\,\odot\,\boldsymbol{c}_{t-1}^l + \boldsymbol{i}_t^l \odot \boldsymbol{g}_t^l \qquad (7)$$

$$\boldsymbol{h}_t^l = \boldsymbol{o} \odot \tanh\,(\boldsymbol{c}_t^l) \qquad (8)$$

where is the element-wise multiplication. It is to be noticed that LSTM calculates both layer state vector and cell state vector for every layer in the training process, contrary to RNN which keeps only layer state vector. This enables LSTM to make more accurate predictions while increasing computational load (Karpathy et al., 2015).

2.3.2. Gated Recurrent Unit (GRU)

Although not frequent like LSTM applications, GRU implementations have been seen in sales forecasting research. Ma et al. (2020) deployed GRU together with other deep learning algorithms in order to predict automotive spare parts demand. They observed similar prediction accuracies compared to LSTM and fully connected networks. Qi et al. (2019) proposed a deep neural framework for sales forecasting in e-commerce. They implemented a GRU encoder for determining the trend using 17 sales features. The proposed model outperformed classical time series analysis methods like AR, EMA, ARIMA and other deep neural network models like CNN and DNN.

GRU, proposed by Cho et al. (2014), has been becoming popular for time series analysis and sales forecasting as well as for other application areas, since it has a simpler architecture and computational advantages over LSTM.

2.4. Gradient Boosting Machines and XGboost

Gradient boosting machines (GBMs) and Xgboost algorithms belong to a class of algorithms called ensemble models, which optimize the predictions of two or more algorithms. GBMs are built by adding weak or base learners to the model sequentially by optimizing the whole model in order to minimize the total error (Natekin & Knoll, 2013). The loss function and the base learners can be chosen arbitrarily, but decision trees are more common in practice.

Examples of implementations of gradient boosting machines (GBMs) and XGboost have been seen more frequently in sales forecasting recently. Antipov and Pokryshevskaya (2020) predicted retail sales using standard POS data. They implemented a GBM built on regression trees and compared their results against RF and Elastic nets. They reported that GMB outperformed RF and Elastic nets. Weng et al. (2020) forecasted supply chain sales using a LightGBM algorithm, which is based on decision trees. They also combined it with LSTM, feeding the net features coming from LSTM together with input features to GBM. They achieved better prediction accuracy through the proposed combined model compared to single models.

XGboot is another gradient boosting algorithm similar to GMB but with a differentiable convex loss function. Due to its loss function structure XGboost converges quicker compared to GMB and has advantages when dealing learning task with large datasets (Shilong et al., 2021). Since it is a relatively fast algorithm, it has been more frequently applied in sales forecasting and shown to be superior to classical models like linear regression (LR), decision tree (DT), and ridge regression (RR) (Shilong et al., 2021; Behera & Nain, 2019). XGboost have also been combined with other machine learning techniques for sales forecasting (Massaro et al., 2021; Wang &Yang, 2021).

3. METRICS USED FOR PERFORMANCE EVALUATION

Almost all of the research surveyed in this study use evaluation metrics in order to assess and compare the performance of the used or the proposed models. Although there are a few others, the most used metrics are MAE (Mean Absolute Error), MAPE (mean absolute percentage error), MSE (Mean Square Error), and RMSE (Root Mean Square Error), which are given as follows where is the actual value, is the predicted value and is the number of predictions made for different time windows. Most of the surveyed studies use more than one metric for a better evaluation and comparison of the models.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i|} \tag{10}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

4. SURVEY OVERVIEW

The deep learning algorithm studies published in the last 5 years in sales forecasting have been reviewed in the survey, searched with keywords such as "deep learning sales forecasting", "deep learning sales forecasting DNN". Although there exist other models and techniques in the literature DNN, CNN, RNN, LSTM, GRU, GBM, and XGboost research, totally 56, are included. The majority of the research are in retail sales forecasting. These studies are classified under their sectors, even if the sales are online. Other online retail sales research (fashion, food, oil, etc.) are given under e-commerce. These studies usually include online customer behavior together with sales features. Real-estate, commodity, supply chains, agriculture, information technologies, and restaurant are classified under others. Number of studies according to sectors are given in Figure 5.

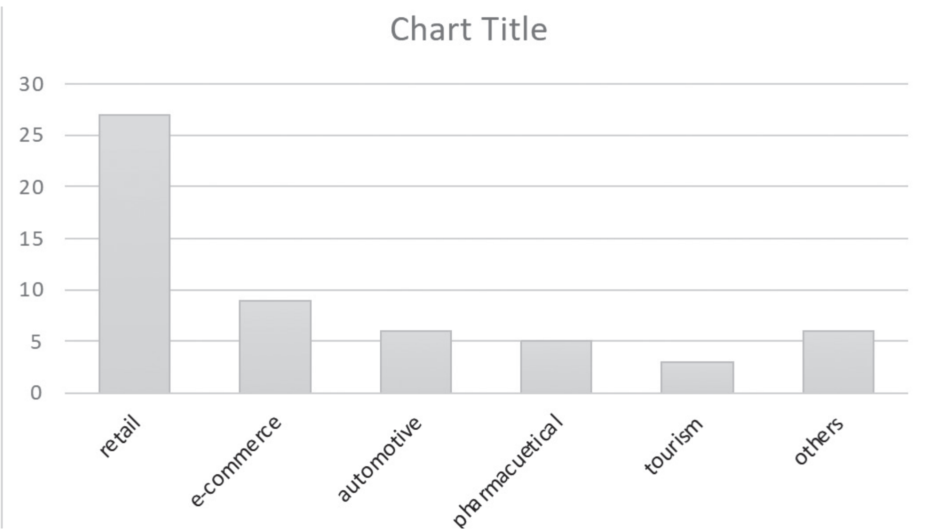

Figure 5. Number of research in sectors

All the sales forecasting applications using deep learning, which are reviewed in this survey, are tabulated in Table 1. The table includes the deep learning algorithms implemented, the baseline algorithms used for comparison, the sectors for which the study is conducted, features which are used for algorithm training, and the references. Hybrids are indicated with a plus sign between the models included, if the research includes more than one model, they are separated by a comma. The feature "historical sales" may contain features like items, dates, prices, shops or any others attributes related to them. When other features like economical indicators or weather data are included in features, they are additionally indicated. The abbreviations in Table 1, which are not given in the text, are found in the Glossary for Table 1.

Table 1: Sales Forecasting Applications Using Deep Learning

Model	Baseline	Sector	Features	Reference
DNN	DT, RF, SVM, ANN, LR	fashion retail	historical sales	Loureiro et al. (2018)
DNN+ LSSVR	ANN	automotive	historical sales data, economic indicators	Yuan & Lee (2020)
DNN+OLS+MLR	SVR, LR, SARIMAX	tourism	historical sales, weather data	Athapaththu et al. (2020)
DNN	MA	retail	historical sales data	Karb et al. (2018)
DNN, GBM	DNN, GBM	retail	historical sales data	Wanchoo (2019)
DNN (BPNN)	-	commodity	price, discounts, special days	Mia et al. (2021)
DNN	-	automotive	historical sales data, economic indicators	Kaya & Yildirim (2020)
DNN+ LightGBM	DNN, LightGBM	aviation	historical sales	Tang et al. (2019)
CNN+LSTM	MLR, online sentiment data	automotive	historical sales, sentiment data	Ou-Yang et al. (2022)
CNN	RNN	retail	historical sales	Kechyn et al. (2018)
CNN + LSTM + Attention Module	Näive, MLP, CNN, LSTM, GRU, XGboost	retail	historical sales	Kaneko (2019)
CNN+ MLP	AR, MA, ARIMA	sportwear retail	historical sales data	Xu et al. (2021)
CNN+ LSTM	LSTM, CNN, actual sales	retail	Historical sales data	Kaunchi et al. (2021)
CNN	AdaBoost	e-commerce	historical sales data, product, merchant, buyer attributes	Yin & Tao (2021)
SNN (twin- CNN)	näive, KmDt	fashion retail	Historical sales data	Craparotta et al. (2019)
CNN+ LSTM	MLP, CNN, LSTM	retail	Historical sales data	Uddin et al. (2021)
DCCNN	ETS, ADL, ARIMAX, SVR, ELM, ADLP, ELMP, RF, GBRT	retail	sales, price, feature advertisings and displays	Ma & Fildes (2021)

Dilated causal CNN	LSTM, GRU, ARIMA	retail petrol	historical sales data	Rizvi et al.
CNN+ Bi-directional LSTM	-	e-commerce	sales, items, comments, shop data	Zhu (2021)
CNN v3 + MLP	-	retail fashion	historical sales data and images	Giri et al. (2019)
CNN	MLP, näive, MA	pharmacy	historical sales data	Velastegui et al. (2020)
RNN	ANN	freeway retail	historical sales, weather, traffic data	Lin et al. (2021)
RNN, LSTM	SVM, ARIMA	grocery retail	historical sales data	Wang et al. (2019)
RNN+ARIMA	RNN, ARIMA	IT	historical sales and sentiment data	Pemathilake et al. (2018)
RNN, LSTM, GRU	ARIMA, RF, GB	retail	historical sales data, economic indicators	Wang (2022)
GRU+DSF	AR, EMA, ARIMA, GBRT, DNN, CNN-WD, DSF/SRN	e-commerce	static, date, user behavior, purchasing, and promotion features.	Qi et al. (2019)
LSTM	RFR, XGB	retail	historical sales data	Helmini et al. (2019)
LSTM	-	retail	historical sales	Yu et al. (2017)
LSTM	-	e-commerce (apparel)	historical sales, weather data	Hong (2021)
LSTM	-	e-commerce	historical sales, sentiment rating of comments	Shih & Lin (2019)
LSTM		tourism	active bookers, corona virus impact	Goel & Bajpai (2020)
LSTM	LR, MLP, BPNN	retail	historical sales	Lakshmanan et al. (2020)
LSTM	ARIMA, ETS, ANN, KNN, RNN, SVM	furniture	historical sales	Abbasimehr et al. (2020)
LSTM	ARIMA, SARIMA, DES, TES, CNN, Prophet	retail (furniture)	historical sales	Ensafi et al. (2022)
LSTM	SVM, LR	retail	historical sales	Li et al. (2020)
LSTM	ARIMA	pharmaceutical	historical sales	Ferrera et al. (2022)

Table 1: Sales Forecasting Applications Using Deep Learning ***(contunied)***

Model	Baseline	Sector	Features	Reference
LSTM, GRU, TFT	Previous instances	restaurant	historical sales	Schmidt et al. (2022)
LSTM+ LGMB	LSTM, LGMB	retail, e-commerce	historical sales	He & Yu (2020)
ARIMA+ LSTM	LSTM, ARIMA	pharmaceutical	historical sales	Han (2020)
ARIMA+ LSTM	LSTM, ARIMA	real estate	historical sales	Temur et al. (2019)
ARIMA+ LSTM	LSTM, ARIMA	pharmaceutical	historical sales	Vavliakis et al. (2020)
ARIMA+ LSTM	LSTM, ARIMA, Traditional Hybrid Model	automotive spare parts	historical sales	Cheng et al. (2018)
LSTM+ PSO	LSTM, LR, SVR, MLP, M5P, RF, KNN, ARIMA, TL, RNN	e-commerce	historical sales, page-views, unique visitors, promotions	He et al. (2022)
SA+LSTM	LSTM, BPNN	agriculture	historical sales	Wang et al. (2019)
DAE+LSTM	MLR, SVR, ANN, LSTM	online retail	historical sales	Zhao et al. (2019)
LSTM+ Hyperparame-ter search	KNN, RF, CatBoost, XGBoost, LightGBM	retail	historical sales	Dai & Huang (2021)
LSTM	Prophet	pharmaceutical	historical sales data	Meng et al. (2021)
Attention-LSTM	ARIMA, SVR, BPNN	automotive	historical sales, user attention data	Cuiqing et al. (2021)
FCN, LSTM, GRU, CNN, TN	LGBM	automotive spare parts	historical sales	Ma et al. (2021)
GBM	RF, Elastic nets	retail	historical sales	Antipov & Pokryshevskaya (2020)
LightGBM+ LSTM	LightGBM, LSTM	grocery supply chain	historical sales	Weng et al. (2020)
XGBoost	LR, DT, RR	retail	historical sales	Behera & Nain (2019)
XGBoost	LR, RR	retail	historical sales	Shilong (2021)
XGBoost, LightGBM, CatBoost	XGBoost, LightGBM, CatBoost	retail	historical sales	Smirnov & Sudakov (2021)

XGBoost with AD	XGBoost w/o AD	retail	historical sales, weather data	Massaro et al. (2021)
XGBoost+ GAN+ LSTM		e-commerce	historical sales, customer behavioral data	Wang & Yang (2021)

4. CONCLUSION

Sales forecasting has always been a very challenging task since the sales are usually affected by many variables, which are mostly non-linear and non-linearly correlated. Many models had been developed to overcome these difficulties. First, linear time series models were used accompanying judgmental forecasting. Then, since these models could not account for the non-linearity in datasets, non-linear models started appearing more frequently for sales forecasting. However, non-linear models could only be successful for certain forecasting tasks and they mostly failed when the data was multi-dimensional. Machine learning applications then started being applied for sales forecasting because they are capable of reflecting the non-linearity in sales data. Although classical machine learning models account for non-linearity, they fail to capture the temporal and sequential dependencies. Deep learning models have recently been implemented for sale forecasting to address these drawbacks that time series models and other classical machine learning techniques suffer.

As multi-layered versions of ANNs, DDNs have been implemented for sales forecasting tasks. Although research included in this survey reported some successful applications, their numbers are relatively low compared to other deep learning models. Determining the relevant features for any machine learning task is not straightforward but it is more crucial for sales forecasting where interdependencies among features are more complex. CNN models have been preferred for sales forecasting due to capability in extracting the relevant features data mitigating the task of data preparation. Therefore, in many studies CNNs are proposed or applied as a part of hybrid structure together with other machine learning models.

Due their superiority in reflecting temporal and sequential dependencies in data, RNNs have become popular for many machine learning tasks. They have also found widespread applications recently in sales forecasting as it is a field in which temporal dependencies are important. Although the number of applications of classical RNNs are not high, LSTM has turned out be the most common deep learning model captured by the survey, the reason being that LSTM overcomes the technical difficulties of classical RNNs. LSTM is implemented in many sales forecasting applications as a single model or in combination with other machine learning techniques. Another type of RNN, GRU, which has a simpler structure and computational advantage over LSTM, has been reported to have similar performances to LSTM. Relatively new ensemble algorithms like GBM and XGboost have become considerably popular and the number of their implementations are to be expected to rise.

Research reviewed in this survey report that deep learning models achieve superior performances against classical time series analysis and classical machine learning models in general. However, it should be noted that there is no model, which is appropriate for all kind of applications and the research for better machine learning models in sales forecasting will continue as it will in other fields.

REFERENCES

Ansuj, A. P., Camargo, M. E., Radharamanan, R., et al. (1996). Sales forecasting using time series and neural networks. Computers & Industrial Engineering, 31(1-2), 421-424. Doi: 10.1016/0360-8352(96)00166-0.

Antipov, E.A., Pokryshevskaya, E.B. (2020). Interpretable machine learning for demand modeling with high-dimensional data using Gradient Boosting Machines and Shapley values. *Journal of Revenue and Pricing Management,* 19**,** 355–364. Doi: 10.1057/s41272-020-00236-4.

Behera, G., Nain, N. (2019). A comparative study of big mart sales prediction. In International Conference on Computer Vision and Image Processing (pp. 421-432). Springer, Singapore. DOI: 10.1007/978-981-15-4015-8_37

Brühl, B., Hülsmann, M., Borscheid, D., et al. (2009). A sales forecast model for the German automobile market based on time series analysis and data mining methods. In *Industrial Conference on Data Mining* (pp. 146-160). Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-03067-3_13.

Chakraborty, K., Mehrotra, K., Mohan, C. K., et al. (1992). Forecasting the behavior of multivariate time series using neural networks. *Neural networks*, *5*(6), 961-970. Doi: 10.1016/S0893-6080(05)80092-9.

Chang, P. C., Wang, Y. W., Tsai, C. Y. (2005). Evolving neural network for printed circuit board sales forecasting. *Expert Systems with Applications*, *29*(1), 83-92. Doi: 10.1016/j.eswa.2005.01.012.

Cheng, J., Hong, T., Li, X., et al. (2018). Research on Sales Forecasting Method for Transmission Parts of Customized Production. *3rd International Conference on Communications, Information Management and Network Security* (CIMNS 2018) (pp. 220-224). Atlantis Press. Doi: 10.2991/cimns-18.2018.50

Cho, K., Van Merriënboer, B., Gulcehre, et al (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chu, C. W., Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of production economics*, *86*(3), 217-231. Doi:10.1016/S0925-5273(03)00068-9

Crone, S. F., Lessmann, S., Pietsch, S. (2006). Forecasting with computational intelligence-an evaluation of support vector regression and artificial neural networks for time series prediction. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings* (pp. 3159-

3166). DOI: 10.1109/IJCNN.2006.247299

Dai, Y., & Huang, J. (2021). A Sales Prediction Method Based on LSTM with Hyper-Parameter Search. In *Journal of Physics: Conference Series* (Vol. 1756, No. 1, p. 012015). IOP Publishing. DOI:10.1088/1742-6596/1756/1/012015

Doganis, P., Alexandridis, A., Patrinos, P., et al. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*, *75*(2), 196-204. Doi: 10.1016/j.jfoodeng. 2005.03.056

Ensafi, Y., Amin, S. H., Zhang, G., et al. (2022). Time-series forecasting of seasonal items sales using machine learning–A comparative analysis. *International Journal of Information Management Data Insights*, 2(1). Doi: 10.1016/j.jjimei.2022.100058

Frees, E. W., Miller, T. W. (2004). Sales forecasting using longitudinal data models. *International Journal of Forecasting,* 20(1), 99-114. Doi: 10.1016/S0169-2070(03)00005-0.

Guajardo, J. A., Weber, R., Miranda, J. (2010). A model updating strategy for predicting time series with seasonal patterns. *Applied Soft Computing*, 10(1), 276-283. Doi: 10.1016/j.asoc.2009.07.005

Han, Y. (2020). A forecasting method of pharmaceutical sales based on ARIMA-LSTM model. *In 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)* (pp. 336-339). Doi: 10.1109/ISCTT51595.2020.00064

He, Q. Q., Wu, C., Si, Y. W. (2022). LSTM with Particle Swam Optimization for Sales Forecasting. *Electronic Commerce Research and Applications*, 101118. Doi: 10.1016/j.elerap.2022.101118

He, Z., Yu, S. (2020). Application of LightGBM and LSTM combined model in vegetable sales forecast. *In Journal of Physics: Conference Series* (Vol. 1693, No. 1, p. 012110). IOP Publishing. Doi: 10.1088/1742-6596/1693/1/012110

Helmini, S., Jihan, N., Jayasinghe, et al. (2019). Sales forecasting using multivariate long short term memory network models. *PeerJ, PrePrints*, 7, e27712v1. Doi: 10.7287/ peerj.preprints.27712v1

Hong, J. K. (2021). LSTM-based Sales Forecasting Model. *KSII Transactions on Internet and Information Systems (TIIS)*, *15*(4), 1232-1245. Doi: 10.3837/tiis.2021.04.003

Karmy, J. P., Maldonado, S. (2019). Hierarchical time series forecasting via support vector regression in the European travel retail industry. *Expert Systems with Applications*, *137, 59-73.* Doi: 10.1016/j.eswa.2019.06.060

Karpathy, A., Johnson, J., Fei-Fei, L. (2015). Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*. Doi: 10.48550/arXiv.1506.02078

Kohli, S., Godwin, G. T., Urolagin, S. (2021). Sales Prediction Using Linear and KNN Regression. *Advances in Machine Learning and Computational Intelligence*, *321.* Doi: 10.1007/978-981-15-5243-4_29

Kuo, R. J. (2001). A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. *European Journal of Operational Research, 129(3), 496-517*. Doi: 10.1016/S0377-2217(99)00463-4

Lu, C. J., Lee, T. S., Lian, C. M. (2010). Sales forecasting of IT products using a hybrid MARS and SVR model. *In 2010 IEEE International Conference on Data Mining Workshops (pp. 593-599)*. Doi: 10.1109/ICDMW.2010.11

Ma, S., Fildes, R. (2021). Retail sales forecasting with meta-learning. *European Journal of Operational Research, 288(1), 111-128*. Doi: 10.1016/j.ejor.2020.05.038

Ma, Z., Wang, C., Zhang, Z. (2021). Deep Learning Algorithms for Automotive Spare Parts Demand Forecasting. *In 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI) (pp. 358-361). IEEE.* Doi: 10.1109/CISAI54367.2021.00075

Massaro, A., Panarese, A., Giannone, D., et al. (2021). Augmented Data and XGBoost Improvement for Sales Forecasting in the Large-Scale Retail Sector. *Applied Sciences, 11(17), 7793.* Doi: 10.3390/app11177793

Mentzer, J. T., Cox Jr, J. E. (1984). Familiarity, application, and performance of sales forecasting techniques. *Journal of forecasting, 3(1), 27-36.*

Mentzer, J. T., Moon, M. A., Kent, J. L., et al. (1997). The need for a forecasting champion. *The Jour-*

nal of Business Forecasting, 16(3), 3.

Natekin, A., Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics, 7, 21.* Doi: 10.3389/fnbot.2013.00021.

O'Shea, K., Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458.* Doi: 10.48550/arXiv.1511.08458.

Pemathilake, R. G. H., Karunathilake, S. P., Shamal, et al. (2018). Sales forecasting based on autoregressive integrated moving average and recurrent neural network hybrid model. *In 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 27-33). IEEE.* Doi: 10.1109/FSKD.2018.8686936

Qi, Y., Li, C., Deng, H., et al. (2019). A deep neural framework for sales forecasting in e-commerce. *In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp.299-308).* Doi: 10.1145/3357384.3357883

Sanders, N. R., Ritzman, L. P. (2004). Integrating judgmental and quantitative forecasts: methodologies for pooling marketing and operations information. *International Journal of Operations & Production Management, 24(5), 514-529.* Doi: 10.1108/01443570410532560.

Schmidt, A., Kabir, M. W. U., Hoque, M. T. (2022). Machine Learning Based Restaurant Sales Forecasting. *Machine Learning and Knowledge Extraction, 4(1), 105-130.* Doi: 10.3390/make4010006

Shilong, Z. (2021). Machine learning model for sales forecasting by using XGBoost. *In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 480-483).* IEEE. Doi: 10.1109/ICCECE51280.2021.9342304

Shih, Y. S., Lin, M. H. (2019). A LSTM approach for sales forecasting of goods with short-term demands in E-commerce. *In Asian Conference on Intelligent Information and Database Systems (pp. 244-256).* Springer, Cham. Doi: 10.1007/978-3-030-14799-0_21

Sun, Z. L., Choi, T. M., Au, K. F., et al. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems, 46(1), 411-419.* Doi: 10.1016/j.dss.2008.07.009

Tang, X., Gao, S., & Jiang, Z. (2019). A Blending Model Combined DNN and LightGBM for Forecasting the Sales of Airline Tickets. *In Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence (pp. 150-154).* Doi: 10.1145/3374587.3374652

Tang, Z., De Almeida, C., Fishwick, P. A. (1991). Time series forecasting using neural networks vs. Box-Jenkins methodology. *Simulation, 57(5), 303-310.* Doi: 10.1177/003754979105700508

Temur, A. S., Akgün, M., & Temur, G. (2019). Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models. Doi: 10.3846/jbem.2019.10190

Thiesing, F. M., Vornberger, O. (1997). Sales forecasting using neural networks. *In Proceedings of International Conference on Neural Networks (ICNN'97) (Vol. 4, pp. 2125-2128).* IEEE. Doi: 10.1109/ICNN.1997.614234

Torres, J. F., Hadjout, D., Sebaa, A., et al. (2021). Deep learning for time series forecasting: a survey. *Big Data, 9(1), 3-21.* Doi: 10.1089/big.2020.0159

Vavliakis, K. N., Siailis, A., Symeonidis, A. L. (2021). Optimizing Sales Forecasting in e-Commerce with ARIMA and LSTM Models. *In Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST 2021), pages 299-306.* Doi: 10.5220/0010659500003058

Wang, C. H. (2022). Considering economic indicators and dynamic channel interactions to conduct sales forecasting and for retail sectors. *Computers & Industrial Engineering, 107965.* Doi: 10.1016/j.cie.2022.107965

Wang, S. and Yang, Y. (2021), M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores. *Data Technologies and Applications, Vol. 55 No. 5, pp. 749-770.* Doi: 10.1108/DTA-11-2020-0286

Wang, Y., Chang, D., Zhou, C. (2019). The study of a sales forecast model based on SA-LSTM. *In Journal of Physics: Conference Series (Vol. 1314, No. 1, p. 012215).* IOP Publishing. Doi:10.1088/1742-6596/1314/1/012215

Weng, T., Liu, W., Xiao, J. (2020), Supply chain sales forecasting based on lightGBM and LSTM combination model. *Industrial Management & Data Systems, Vol. 120 No. 2, pp. 265-279.* Doi: 10.1108/IMDS-03-2019-0170

West, D. C. (1994). Number of sales forecast methods and marketing management. *Journal of Forecasting, 13(4), 395.* Doi: 10.1002/for.3980130405

Xu, J., Zhou, Y., Zhang, L., et al. (2021). Sportswear retailing forecast model based on the combination of multi-layer perceptron and convolutional neural network. *Textile Research Journal, 91(23-24):2980-2994.* Doi: 10.1177/00405175211020518

Yu, Q., Wang, K., Strandhagen, J. O., et al. (2017). Application of long short-term memory neural network to sales forecasting in retail—a case study. *In International Workshop of Advanced Manufacturing and Automation (pp. 11-17).* Springer, Singapore. Doi: 10.1007/978-981-10-5768-7_2

Yuan, F. C. (2012). Parameters Optimization Using Genetic Algorithms in Support Vector Regression for Sales Volume Forecasting. *Applied Mathematics, 3, 1480-1486.* Doi: 10.4236/am.2012.330207

Zhao, B., Lu, H., Chen, S., et al. (2017). Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics, 28(1), 162-169.* Doi: 10.21629/JSEE.2017.01.18

Zhao, Y., Ren, X., Zhang, X. (2021). Optimization of a Comprehensive Sequence Forecasting Framework Based on DAE-LSTM Algorithm. *In Journal of Physics: Conference Series (Vol. 1746, No. 1, p. 012087).* IOP Publishing. Doi:10.1088/1742-6596/1746/1/012087

Zheng, D., Zhang, W., Netsanet, S., et al. (2021). Short-term renewable generation and load forecasting in microgrids. Dehua Zheng (Ed)., *Microgrid Protection and Control, (pp. 57–96).* Academic Press. Doi: 10.1016/B978-0-12-821189-2.00005-X