International Journal of Information Management Data Insights 2 (2022) 100058

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Time-series forecasting of seasonal items sales using machine learning – A comparative analysis

Yasaman Ensafi[a], Saman Hassanzadeh Amin[a,*], Guoqing Zhang[b], Bharat Shah[c]

[a] Department of Mechanical and Industrial Engineering, Ryerson University, ON, Canada
[b] Department of Mechanical, Automotive and Materials Engineering, University of Windsor, ON, Canada
[c] Ted Rogers School of Management, Ryerson University, ON, Canada

ARTICLE INFO

Keywords:
Time-series forecasting
Sales forecasting
Seasonal items
Neural network
Big data

ABSTRACT

There has been a growing interest in the field of neural networks for prediction in recent years. In this research, a public dataset including the sales history of a retail store is investigated to forecast the sales of furniture. To this aim, several forecasting models are applied. First, some classical time-series forecasting techniques such as Seasonal Autoregressive Integrated Moving Average (SARIMA) and Triple Exponential Smoothing are utilized. Then, more advanced methods such as Prophet, Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) are applied. The performances of the models are compared using different accuracy measurement methods (e.g., Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE)). The results show the superiority of the Stacked LSTM method over the other methods. In addition, the results indicate the good performances of the Prophet and CNN models.

1. Introduction

In today's competitive world where sales are highly important for companies, accurate sales forecasting plays a key role in every successful retail business. It can be helpful in inventory management by avoiding overproduction and reducing overstock (Islam & Amin, 2020; Nguyen, Tran, Thomassey, & Hamad, 2021). In addition, it can be a great tool for increasing the profitability of a company by cost prediction (Huang, Fildes, & Soopramanien, 2019; Islam, Amin, & Wardley, 2021; Sohrabpour, Oghazi, Toorajipour, & Nazarpour, 2021; Verstraete, Aghezzaf, & Desmet, 2020).

There are some major factors that have impacts on the future sales. These factors can be identified by analyzing the sales patterns of total sales of a retail store or a specific product. It is worth mentioning that each product has a different difficulty level of forecasting. Some products like milk, have stable consumption over the year, and therefore, their sales can be predicted easily (Chopra & Meindl, 2016; Padilla, García, & Molina, 2021). Some other items such as fashion goods and furniture products contain trends and seasonality in their sales pattern that cause complexity in the forecasting process.

In this research, a dataset is investigated which has the information of a retail store sales from 2014 to the end of 2017. The products of this dataset belong to three different categories including office supplies, technology, and furniture. Among these three items, the products of furniture category are chosen to forecast their future sales. A public dataset with the occurrence of seasonal fluctuations cannot be found easily. However, the historical sales of items in the furniture category of the selected dataset hold this seasonality factor. Therefore, the furniture category has been chosen for predicting the upcoming sales which can help predict the future demand. Furniture is an inseparable part of every house and office across the world and people need furniture to make their houses and offices complete. In recent years, the development in the designing and manufacturing furniture, the rising demand for multi-functional furniture, and high-quality furnishings have made this industry a growing one. The value of the global furniture market was $564.17 billion in 2020, and it is predictable to grow and reach $671.07 billion in 2021. It should be mentioned that the outbreak of COVID-19 has had a major impact on furniture industry and has boosted the sales massively due to the remote working (Furniture Global Market Report, 2021).

The aim of this study is not to analyze the retail store's performance, but to use the past data to predict the future sales of furniture. As it has been mentioned before, seasonal times-series forecasting plays a key role in strategic decision-making and planning future activities. We also would like to question the ability of neural networks models for the task of seasonal sales forecasting which is widely encountered in numerous applications. In addition, we would like to compare the performance with various classical and advanced time-series forecasting methods, and to find the best method among them.

* Corresponding author.
E-mail addresses: yasaman.ensafi@ryerson.ca (Y. Ensafi), saman.amin@ryerson.ca (S.H. Amin), gzhang@uwindsor.ca (G. Zhang), bshah@ryerson.ca (B. Shah).

https://doi.org/10.1016/j.jjimei.2022.100058
Received 17 July 2021; Received in revised form 12 January 2022; Accepted 12 January 2022

In recent years, neural network methods have been known as a powerful tool and have been applied in different fields especially in forecasting. The questions that this research is going to answer are: (a) How well neural network methods will perform for seasonal items forecasting? (b) Which of these traditional and innovative time-series forecasting methods will obtain the most accurate results?

To the best of our knowledge, this is the first investigation that has compared the performances of more than ten different forecasting models including classical methods and advanced techniques such as different types of Long Short-Term Memory (LSTM) and Prophet methods. In this investigation, the capability of Convolutional Neural Network (CNN) which is mostly used for image recognition is also investigated for forecasting seasonal items sales.

The rest of this document is outlined as follows. In the next section, a literature review is provided which discusses different approaches for forecasting. With the goal of gaining more information about the dataset, the third section explains the results of Exploratory Data Analysis (EDA) process. The fourth section describes the data, the prepossessing steps, and the methodology. In Section 5, the performances of different approaches are reported, and the results are compared. Finally, Section 6 is dedicated to the conclusions.

2. Literature review

One important element of sales forecasting is the accuracy of the prediction. Therefore, a lot of efforts have been made to make this process more accurate. In the publications of this field, the accuracy is obtained using some error measurement methods like RMSE, MAPE, after comparing the actual sales with the predicted results. In this section, some related papers are reviewed.

2.1. Classical time-series forecasting methods

There is not a single best technique to solve time-series forecasting problems (Zhang & Kline, 2007). In order to deal with time-series forecasting, each problem might be solved with a different approach. Moving Average (MA) is one of the simplest prediction techniques for making projections about time-series without a noticeable seasonal pattern (Chopra & Meindl, 2016). In several papers, a more advanced version of MA which is called Autoregressive Integrated Moving Average (ARIMA) has been used. For instance, Ramos, Santos, and Rebelo (2015) used ARIMA and exponential smoothing method to compare the performances of these methods on forecasting the retail sales of women's footwear, which contain products with repeatable fluctuations in their patterns. The demand for purchasing boots in winter is an example of these fluctuations. Huber, Gossmann, and Stuckenschmidt (2017) applied multivariate ARIMA successfully to perform demand forecasting on perishable goods. Recently, Yang et al. (2021) applied combined ARIMA and neural network for network traffic forecasting. ARIMA model has been also used in a recent study by Tandon, Revankar, and Parihar (2021) for cryptocurrency price forecasting based on the impact of social media.

Seasonal ARIMA (SARIMA) is another classical forecasting method. This technique has been used successfully in various applications like forecasting tourism demand (Goh & Law, 2002) and predicting vehicular traffic flow (Williams & Hoel, 2003). However, Hamzaçebi (2008) indicated that SARIMA can cause limitations in prediction because of its linear form and inability to detect nonlinear and highly volatile patterns. Pai, Lin, Lin, and Chang (2010) introduced a Seasonal Support Vector Regression (SSVR) method for prediction. The outcomes proved the superiority of SSVR over SARIMA in forecasting seasonal items.

There are different types of exponential smoothing methods (Holt, 2004). Kalekar (2004) explained the difference between these methods and used the triple exponential smoothing method or Holt-Winters (HW) technique to deal with the seasonality in dataset. Exponential smoothing has been a powerful forecasting tool for prediction. For instance, Taylor (2011) applied a seasonal exponential smoothing method in forecasting the daily sales of a supermarket which showed high volatility.

Some studies have utilized more advanced techniques. Lv, Bai, Yin, and Dong (2008) proposed a new model for evaluating the new locations of retail stores based on sales forecasting. Li and Lim (2018) proposed a Greedy Aggregation Decomposition (GAD) method to overcome the limited space in a fashion retailer inventory related to Stock Keeping Units (SKU) daily demands. In addition, Giering (2008) implemented a system for online product recommendations based on large retail store sales prediction.

2.2. Prophet and Artificial Neural Networks time-series forecasting methods

The Prophet method is one of the recently introduced forecasting methods by the Facebook company (Taylor & Letham, 2018) and therefore, its capability has not been investigated comprehensively. In a study about monthly temperature predictability conducted by Papacharalampous, Tyralis, and Koutsoyiannis (2018), Prophet forecasting results were close to the results of traditional forecasting methods. However, in another research about hospital discharge volume prediction (McCoy, Pellegrini, & Perlis, 2018), the results were predicted more accurate than the results of SARIMA model. In addition, Subashini, Sandhiya, Saranya, and Harsha (2019) proposed a Prophet time series model to predict the traffic of a website which led to satisfying results. Satrio, Darmawan, Nadia, and Hanafiah (2021) applied ARIMA and PROPHET for forecasting COVID disease in Indonesia.

Artificial Neural Networks (ANN) are other possible methods for forecasting. Alon, Qi, and Sadowski (2001) investigated these alternative methods and compared them with traditional ways of predicting aggregate retail sales. Aggregate time-series are defined as large collections of time-series that are combined with different techniques, and they are usually more accurate than disaggregate forecasts (Chopra & Meindl, 2016; Hyndman & Athanasopoulos, 2018). Alon et al. (2001) indicated that ANN prediction outperformed winters' exponential smoothing and Box Jenkins techniques. Besides, multivariate regression had the lowest MAPE average among them. Verstraete, Aghezzaf, and Desmet (2019) developed a framework based on ANN to predict the impact of the short-term and long-term weather uncertainty on the retail products. Zekić-Sušac, Has, and Knežević (2021) combined ANN with some methods such as regression trees and applied it for forecasting energy cost of public buildings.

ANN models have been popular techniques for sales forecasting especially because of their flexibility for detecting patterns in data (Tkáč & Verner, 2016). Zhang and Qi (2005) investigated the ability of neural networks for modeling seasonal time-series, and used Multilayer Perceptrons (MLP), known as the universal approximators function (Sharma, Rana, & Kumar, 2021). MLP is a feedforward neural network method that has one layer or more of hidden units. Their results showed that to perform accurate prediction, prior data processing is vital. The data pre-processing techniques such as deseasonalizing and detrending can prevent high variance in the results and can improve the accuracy.

To improve ANN, lots of attempts have been made. Hamzaçebi (2008) introduced a novel model called Seasonal Artificial Neural Network (SANN) for determining the number of neurons (input and output) based on seasonal fluctuations in the pattern. The model let to accurate prediction compared to the other classical methods. Khashei and Bijari (2010) investigated model combining to find diverse patterns in the data. They proposed a new hybrid model of ANN using ARIMA model to improve the accuracy of prediction. The results confirmed the effectiveness of their model.

2.3. Deep learning

Machine learning and deep learning techniques are useful tools for accurate forecasting. Kotsiantis, Zaharakis, and Pintelas (2006) reviewed Swarm intelligence algorithms. They discussed insect-based and animal-based algorithms. Kar (2016) reviewed several bio inspired algorithms such as Neural Networks and Swarm intelligence. In an important research, Chakraborty and Kar (2017) focused on Swarm intelligence and reviewed the related papers. They discussed the potential applications and the future research. Kar and Dwivedi (2020) discussed the reasons of using big data. In addition, the gaps in this field of research and the future directions were provided in this article.

The idea behind deep neural networks is using Multilayer Perceptrons (MLP) with more hidden layers to be able to model more complicated functions (Alpaydin, 2009). This successful approach which is based on ANN has been used in several investigations, such as solar power forecasting (Gensler, Henze, Sick, & Raabe, 2016), stock price prediction (Zhuge, Xu, & Zhang, 2017), and short-term traffic forecasting (Zhao, Chen, Wu, Chen, & Liu, 2017). Kaneko and Yada (2016) presented a model for predicting sales of a retail store using deep learning approach. They trained and tested their model using three years of Point-Of-Sale (POS) data from supermarkets. Their results indicated the superiority of deep learning method over the logistic regression model. Lee, Rim, and Lee (2019) presented a novel idea that can rank the products based on their large number of online reviews with deep learning techniques.

Another method which has gained significant attention in recent years is Recurrent Neural Network (RNN). RNN is basically a network consists of loops which give them the ability to keep events from the past on their memory. Therefore, they can be very useful in time-series prediction (Gamboa, 2017). One of the special kinds of RNN which is a solution for short term memory is Long Short-Term Memory (LSTM) technique. Malhotra et al. (2016) presented a model for detecting anomalies in time-series with LSTM. Zhao et al. (2017) indicated the effectiveness of LSTM for forecasting. LSTM has been also applied in a recent study by Ghosh and Sanyal (2021) on predicting market fear and detecting a hidden pattern of influence during the COVID-19 timeline. However, machine learning models such as XGBoost achieved better results compared with the applied deep learning models.

A Convolutional Neural Network (CNN) is a kind of deep network which has been utilized in time-series forecasting recently. Zhao and Wang (2017) proposed a novel approach to learn effective features automatically from the data with the help of CNN and then used this method to perform sales forecasting. In another study, Selvin, Vinayakumar, Gopalakrishnan, Menon, and Soman (2017) proposed a deep learning-oriented model for stock price forecast. Their results indicated that CNN could detect changes in trends. Koprinska, Wu, and Wang (2018) explored the application of CNN for short-term forecasting of solar power and electricity. They showed the superiority of the CNN and multilayer perceptron neural networks. Momeny, Latif, Sarram, Sheikhpour, and Zhang (2021) developed a CNN based method considering the robustness of the problem.

2.4. Research contributions

The main research contributions of this study are as follow:

- A dataset including the sales history of furniture in a retail store with real data is investigated. In addition, a unique exploratory data analysis is applied to find more information about the dataset.
- Unlike the previous studies that have focused on applying a few forecasting methods, we apply several classical time-series forecasting techniques such as SARIMA to predict the sales of furniture.
- We apply some advanced forecasting methods based on Artificial Neural Networks such as Prophet, LSTM, and CNN.
- The results are compared using accuracy measurement methods such as RMSE and MAPE.
- We analyze the results, and we select the best methods for prediction.

3. Exploratory Data Analysis (EDA)

In this research, a dataset is investigated which has the information of the sales of a retail store from 2014 to the end of 2017 with the goal of sales forecasting.

This dataset contains 9994 data points, 21 features, and zero null values. The columns are attributes such as Order ID, Ship Date, Order Date, Ship Mode, Customer Name, Customer ID, Segment, City, Country, State, Region, Postal Code, Category, Product Name, Sub-Category, Sales, Discount, Quantity, and Profit. Each row of the dataset represents an order that has been shipped. To gain a better insight into this dataset and to be able to predict future sales more accurately, the dataset is divided to three parts based on product categories including furniture, technology products, and office supplies. In the next step, the data is resampled on a monthly frequency, and averages daily sales values are utilized. In addition, the start of the month is set as an index. Fig. 1 represents the time-series of each category and comparison of them.

Based on Fig. 1, the only category which displays seasonality in its pattern is Furniture sales. It is not surprising that the sales of this category reached their peak in the winter holiday season. During this time of year, retailers offer big sales to start the season. In addition to that, other post-Thanksgiving sales events such as Black Friday and Cyber Monday deals are very encouraging to the customers. As it can be seen in the Furniture sales graph of Fig. 1, the furniture sales tend to peak at the end of each year, and then decline after the holidays.

As can be seen in Fig. 1, the sales graphs of all three categories show fluctuations over time. However, unlike the furniture category, there is no strong seasonal characteristic available in the other two graphs. For instance, the peak in the sales amount at the end of each year which had occurred in the furniture category cannot be detected in two other items. To elaborate, for both technology products and office supply categories, the data does not experience any specific trend that recurs every year. Therefore, among these three available classes, furniture is chosen to be investigated regarding its seasonal pattern.

According to Brockwell, Davis, and Calder (2002), a time-series includes observations that each observation is recorded at a specific time. Each of time-series can be described by important components like trend, cyclicality, and seasonality. Trend refers to the overall upward or downward movement. Cyclicality is when a repeating pattern is available, but there is no fixed period. In addition, seasonality can be defined as periodic fluctuations in the time-series. One example is an observable increase in retail store sales during Christmas holidays.

Another important characteristic of time-series is stationarity. A time series is called stationary if its statistical features (e.g., mean, standard deviation) continue steadily over time, and this is highly important because if a time-series is stationary, there is a high probability that it will repeat its behavior in the future, and therefore it will be easier to forecast (Jain, 2016). The first approach to time-series modeling is removing the trend and seasonal factors to obtain stationary residuals (Brockwell et al., 2002). Thus, by applying the transformation, the higher values will be penalized more than smaller values. The stationarity can be checked with Dickey-Fuller test and plotting rolling statistics. Fig. 2 demonstrates the mean and standard deviation of the sales over time. Those values are almost constant over time.

The objective of Augmented Dickey Fuller (ADF) test is to decide that the time-series is stationary or non-stationary by checking the presence of unit root in a time-series. This method observes the difference between the value level and the mean. If it was higher than the mean, the next movement will be downward. Furthermore, if it was lower than the mean, the movement will be upward. Eq. (1) explains these value changes, where μ is a constant, β is the coefficient on a time trend, k is the lag order of the autoregressive process, and $\Delta y(t)$ can be defined as

$$\Delta y(t) \equiv y(t) - y(t-1), \Delta y(t-1) \equiv y(t-1) - y(t-2)$$

and so on (Corrius, 2018).

$$\Delta y(t) = \lambda y(t-1) + \mu + \beta t + \alpha_1 \Delta y(t-1) \pm \ldots + \alpha_k \Delta y(t-k) + \varepsilon_t \tag{1}$$

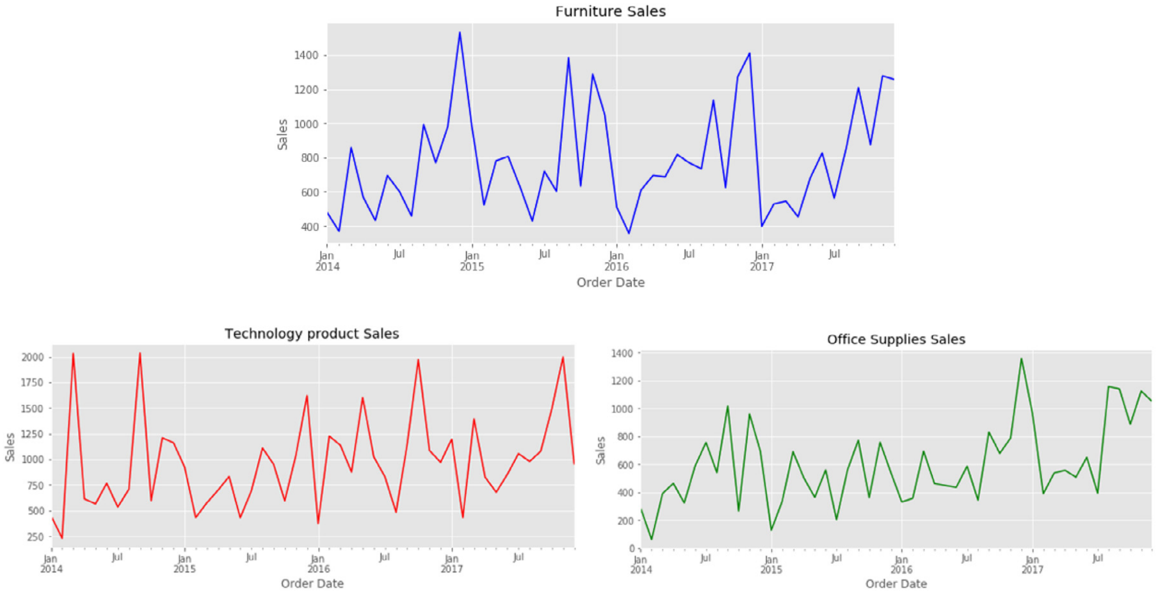

Fig. 1. Comparison of different sales categories.

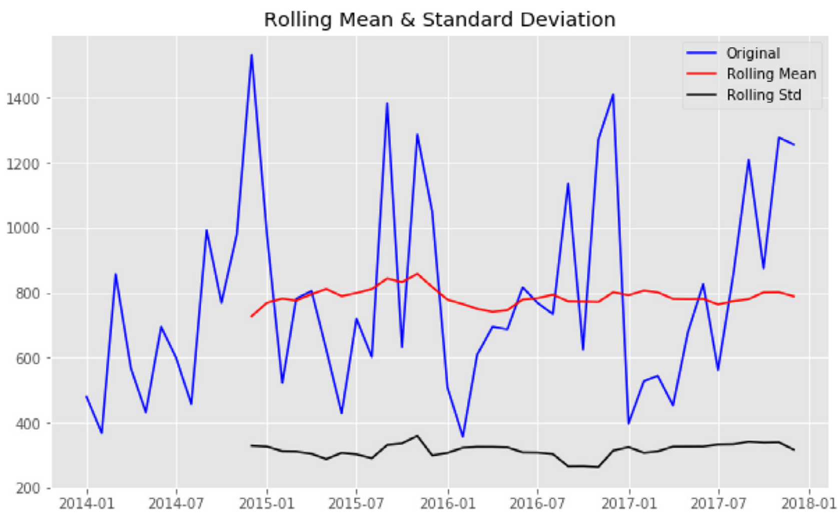

Fig. 2. Mean and standard deviation of the furniture sales.

The null hypothesis states that the time-series in non-stationary ($\lambda = 0$). If this hypothesis is rejected, it shows that the next movement ($\Delta y(t)$) is not just a random value, and it depends on the current level $y(t-1)$. Thus, the time-series is stationary. The results of this test in Python are test statistic and p-value (Singh, 2018). If the p-value is more than 0.05, the null hypothesis is not rejected, and it suggests that the time series has a unit root, and it is non-stationary. Otherwise, if the p-value is less than 0.05, it means that the data does not have a unit root, and it is stationary (Kang, 2017). The results of applying ADF test on the furniture dataset shows that the p-value is smaller than 0.05, and the time-series is stationary. It should be mentioned that ADF is the only test for trend stationarity. Therefore, it is a good idea to also check the stationary by visualization.

4. Methodology and experiments

4.1. Aim of this study

The aim of this study is to perform a time-series forecasting on a seasonal item with the help of different models such as SARIMA, Exponential Smoothing, Prophet, and Neural Networks. In this investigation, the model with the most accurate result is chosen as the best model. The process of forecasting is demonstrated in Fig. 3.

4.2. Dataset

There are several factors to consider for selecting the dataset for research, including availability, privacy, and scale of data (Zhang, Kuppannagari, Kannan, & Prasanna, 2018). Although there are not many public datasets with enough data points available for the task of forecasting seasonal time-series, we managed to find the Superstore sales dataset (Community.tableau.com, 2017) which displays seasonality in its sales pattern, and does not contain any missing values. This dataset which is used for this study describes this retail store sales from 2014 to the end of 2017, and it contains near 10,000 data points and 21 features and consists of sales information of three different categories including furniture, technology goods, and office supplies. In this investigation, the sale of furniture is the variable of interest which it contains seasonal patterns.

4.3. Data pre-processing

To increase the success rate of any project, data pre-processing plays a significant role. In this research about univariate time-series forecast-

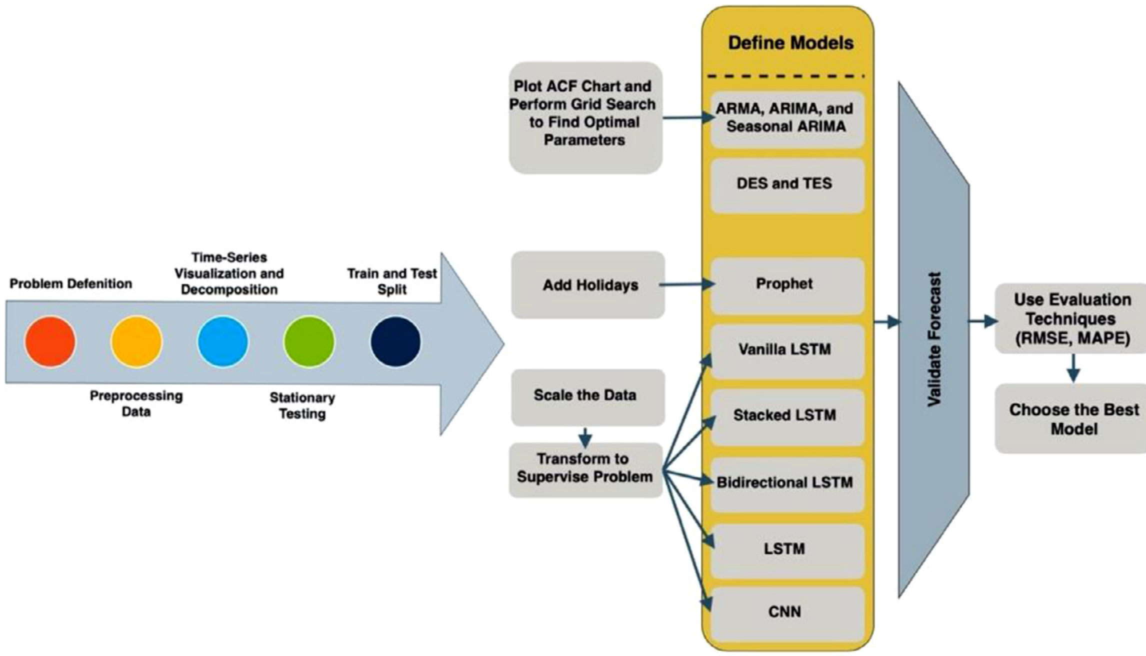

Fig. 3. Process of furniture forecasting.

ing, time-series aggregation is used which reduces computational resources significantly (Kotzur, Markewitz, Robinius, & Stolten, 2018), and enhances forecasting accuracy. Due to the high fluctuation in the numbers of daily sales of furniture prior to the forecasting, in this study the daily data is resampled into monthly frequency. Then, the averages daily sales values are utilized. The most important features for performing univariate forecasting are the sales amount and order date of each data point.

4.4. Measures of forecast error

Time-series forecasting performance measures indicate the capability of the models. There are many methods for evaluating a model. In this research, the model performance is assessed by three commonly used measures. Mean Squared Error (MSE) determines the average of the squared predicted error values. Root Mean Squared Error (RMSE) is the standard deviation of the prediction errors, and it is scale dependent. Therefore, it cannot be applied for comparing time-series with different units. Mean Absolute Percentage Error (MAPE) is obtained as the mean absolute percentage error function for the prediction and the eventual outcomes. This error measures expresses error as a percentage and can be used in evaluating models for different datasets (Hamzaçebi, 2008).

These estimators are calculated using Eqs. (2)–(4) where Y_t is the actual value and F_t is the forecasted value of period t (Hamzaçebi, 2008). In addition, the number of total observations is represented by n.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2 \tag{2}$$

$$RMSE = \sqrt{MSE} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{(Y_t - F_t)}{Y_t}\right| * 100 \tag{4}$$

4.5. Train-test split

The accuracy of every forecast can be determined by the model performance on new data which has not been used during the fitting of the model (Hyndman & Athanasopoulos, 2018). Therefore, one of the important steps in machine learning techniques is to split the data into train and test sets. The same thing needs to be done in time-series forecasting. However, the only difference is that they cannot be divided into groups randomly. In fact, observations in time-series are dependent on each other, and data should be split in time. Therefore, common techniques such as k-fold cross validation are not very useful for time-series.

In this study, a common split is used which indicates the proportion of 75% of the data to the training set, and 25% to the test set. This common split has been used in numerous machine learning studies including research by Arnoux et al. (2017), Kusiak and Zhang (2010), and Avuçlu and Elen (2020). Other than popularity of this common split percentage technique, this standard is chosen because the first 75% of the dataset is equal to the first three years of the sales history, and the final 25% is equal to the last year of the sales history which is aimed to be predicted. During items forecasting with LSTM method, Walk-forward validation method which is one of the suggested methods for back-test models for time-series forecasting (Brownlee, 2016) is applied on the test set. In this method, a model may be updated after each one-step forecast and therefore, provides opportunity to make better forecast at each step.

4.6. Forecasting models

4.6.1. ARMA, ARIMA, and SARIMA

Auto Regressive Moving Average (ARMA) has been produced by integrating the Autoregressive (AR) and Moving Average (MA) methods. This method can be shown as ARMA (p, q) where p represents the order of the AR part. In addition, q is the order of the MA part (Choi, 2012).

Autoregressive Integrated Moving Average (ARIMA) is a popular method for time-series forecasting which consists of the integration of autoregressive and moving average models. This model that was introduced by Box and Jenkins (1970) usually is applied on non-stationary time-series because of its ability called integration to make the sequence stationary by taking the difference of the sequences. In the ARIMA model, the future value is a linear combination of the previous values and the random errors, as follows:

$$\begin{aligned} y_t = {} & \theta_0 + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \ldots + \phi_p\, y_{t-p} \\ & + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} \end{aligned} \tag{5}$$

In the above formula, y_t represents the future value, and p and q are AR and MA polynomials, respectively. Besides, ε_t is the mentioned error. Moreover, θ_i and ϕ_i express the coefficients.

ARIMA model can be extended further. For predicting seasonal time-series, Seasonal ARIMA or SARIMA model is utilized which is denoted by ARIMA (*p, d, q*) (*P, D, Q*) *m*, where *p, d*, and *q* are non-seasonal, and *P, D*, and *Q* are seasonal factors. They present the order of the autoregressive part, degree of the first differencing, and order of the moving average part. In addition, the seasonal parameter *m* can be defined as the length of the cycle Hyndman & Athanasopoulos, 2018). The full model can be expressed by Eqs. (6-(7).

$$\Phi(B^m)\phi(B)\nabla_m^D\nabla^d X_t = \Theta(B^m)\theta(B)Z_t, \tag{6}$$

with

$$\nabla_m X_t = X_t - X_{t-m},\ \nabla X_t = X_t - X_{t-1},$$

$$\begin{aligned} \Phi(B^m) &= 1 - \Phi_1 B^m - \cdots - \Phi_P B^{PM} \\ \phi(B) &= 1 - \phi_1 B - \cdots - \phi B \\ \Theta(B^m) &= 1 + \Theta_1 B^m + \cdots + \Theta_Q B^{Qm} \\ \theta(B) &= 1 + \theta_1 B + \ldots + \theta_q B^q \end{aligned} \tag{7}$$

In these equations, B is the notation for the backshift ($B^n X_t = X_{t-n}$) which shifts the data back n periods, and it is a sequence of white noises with zero mean and constant variance (Zhang & Qi, 2005). Two essential steps in the ARIMA and SARIMA models are finding the differencing orders (*d, D*), and choosing the orders of AR and MA operators.

In order to identify the best parameters, one option is drawing Autocorrelation Function (ACF) and Partial Autocorrelations (PACF) diagrams. Autocorrelations are values in the range of -1 and 1 which can measure the correlation between the time-series and itself during different periods (lags). For instance, an autocorrelation at Lag 1 can show the relation between the values which are 1 period apart. The partial autocorrelation can be defined as the correlation among time-series and the lagged version of itself, after the impacts of correlation at smaller lags have been subtracted from it. By comparing the amplitude of these two plots, the order of the autoregressive part and moving average part (*p, q*) can be determined. Besides, the length of the cycle which is denoted by variable *m* can be found from the peaks of the ACF plots. In this case, the frequency is 12.

There are different approaches for finding the optimal hyper parameters of a model. In case of ARIMA and SARIMA models, in addition to drawing ACF and PACF diagrams that have been mentioned earlier, a Grid Search can be applied. Grid Search will find the best hyper-parameters of the model by searching over the set of possible parameters and comparing Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) which are the estimators that can indicate the quality of the model. After finding all the parameters, the model can be created and fit on the training data, and then it can be used to make a prediction. After applying Grid search method, the best hyperparameters for ARIMA and Seasonal ARIMA have been indicated as (6, 0, 0) and (0, 0, 0) x (1, 1, 0, 12), respectively.

4.6.2. Single, double, triple exponential smoothing

Another popular time-series forecasting method is Exponential smoothing. This model can detect trend and seasonality, and it is like ARIMA method. However, instead of using the weighted linear sum of past data points, it assigns an exponentially decreasing weight to each observation (Brownlee, 2018b). In the other words, it gives less importance to the observations over time.

Exponential Smoothing method can be divided into three main categories. Single Exponential Smoothing (SES) for forecasting univariate data without a trend or seasonality. Double Exponential Smoothing (DES) which supports trends in the time-series, and Triple Exponential Smoothing (TES) which can handle seasonality as well (Hyndman, Koehler, Ord, & Snyder, 2008). In this study, the last two methods which are DES and TES are utilized to predict the furniture sales of a retail.

DES also known as Holt's linear trend model has many hyperparameters which help to perform more accurate forecast. For instance, Alpha controls the smoothing factor for the level, and Beta is the smoothing factor for the trend which is between 0 and 1. Moreover, the different kinds of trends like additive or multiplicative can be set. Double exponential smoothing can be explained mathematically with Eqs. (8) and (9). In these equations, α and β are two smoothing constants with values between 0 and 1. Also, y and b_t represent an estimation for the level of the time-series and the growth of time-series at the time t, respectively.

$$y = \alpha x_t + (1-\alpha)\left(y_{t-1} + b_{t-1}\right) \tag{8}$$

$$b_t = \beta\left(y_t - y_{t-1}\right) + (1-\beta)b_{t-1} \tag{9}$$

Triple Exponential Smoothing method (or Holt-Winters method) supports detecting a seasonal pattern in the time-series. This model takes DES further and has additional hyperparameters like Gamma which control the impacts of seasonal components. In this method (which is the most advanced one among exponential smoothing models), the type of seasonality can be defined as linear seasonality or exponential seasonality. In this method, the length of the seasonal cycle or L can be set (as same as SARIMA). In this monthly dataset, the seasonal period which is repeated every year is 12. This model is expressed mathematically based on Eqs. (10)–(12), where y, b_t, and c_t represent the levels of the series, growth, and seasonal component. In addition, the parameters α, β, and γ are usually limited between 0 and 1.

$$y = \alpha \frac{x_t}{c_{t-L}} + (1-\alpha)\left(y_{t-1} + b_{t-1}\right) \tag{10}$$

$$b_t = \beta\left(y_t - y_{t-1}\right) + (1-\beta)b_{t-1} \tag{11}$$

$$c_t = \gamma \frac{x_t}{y_t} + (1-\gamma)c_{t-L} \tag{12}$$

4.6.3. Prophet

One of the most successful companies in the times series forecasting is Facebook. Recently, this company released its own open-sourced time-series forecasting model called Prophet. This model is based on an additive model which can make high-quality forecasts from hourly, daily, or weekly time-series that have at least a few months' history. It can also predict the future pattern of time-series which contains historical trend fluctuations, and many outliers or missing values. In addition, it can forecast the effect of holidays on the time-series. Taylor and Letham (2018) combined automatic forecasting with the analyst-in-the-loop approach to model a wide variety of business problems.

The first step in this approach is modelling the time-series with specified parameters. Then, the forecast is obtained, and the performance is evaluated. In the next step, if the results show poor performance, the problems will be informed to human analyst so they can adjust the model. The Prophet method uses the additive regression model $y(t)$, and consists of trend, seasonality, and holidays components which are denoted by $g(t)$, $s(t)$, $h(t)$, respectively. Eq. (13) includes the formula. In addition, the component ε_t represents the noise in the data (Polusmak, 2018).

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{13}$$

In the Prophet library, two trend models are implemented to model the non-periodic changes. In addition, the seasonal components give the model the ability to handle periodic changes. These components can be set differently and can improve the performances of prediction. For instance, the seasonality mode can be set as an additive or multiplicative depends on the time-series. Besides, a value between 10 and 25 can pass to seasonality prior scale and can allow the seasonality to be more flexible. The other parameter is the Fourier order. The seasonality in

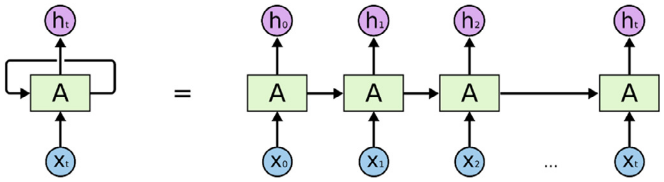

Fig. 4. Recurrent Neural Networks (RNN) structure (Olah, 2015).

Prophet can be modeled based on Fourier series. Therefore, changing the Fourier component can have a big impact on the results. One of the greatest advantages of Prophet is its ability to include the fluctuations caused by the holidays.

The last step before starting to fit the model is to change the name of the dataset's columns. To work with Prophet, the date column should be changed to "*ds*", and the variable of interest should be changed to "*y*". Then, the model can be generated and fit on the training data. It is good to mention that Prophet has a built-in cross-validation function. The goal of this function is to test the model during the training stage to overcome over-fitting or under-fitting. After fitting the model and making forecast, the results of forecast appear as a data frame which includes lots of columns like *yhat* which is the actual predicted value. In this data frame, *yhat_lower* and *yhat_upper* show the uncertainty levels.

4.6.4. Artificial Neural Network

In this section, Artificial Neural Networks (ANN) are discussed. ANN can learn from the data, identify the pattern, and predict complex relationships with high accuracy.

One of the most popular proposed methods for time-series prediction which is inspired by the human brain mechanism is Feed-Forward Network (Loureiro, Miguéis, & da Silva, 2018). Feed-Forward Network includes three layers. An input layer which consists of observations, hidden layers which process the received information from the first layer. The last layer is the output layer which is the actual prediction. In this method, the input of each layer is the output of the previous layer. The output of the first layer is generated by the weighted sum of the inputs and adding the specific bias like α_0 and β_{0j} to them. This model can be written as Eq. (14).

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-i} + \beta_{0j}\right) + \varepsilon_t \tag{14}$$

The number of nodes in the input layer and hidden layers are represented by *m* and *n,* respectively, and fit the activation function such as RELU or Sigmoid. It is good to mention that the most difficult job during the process of forecasting with ANN is choosing the architecture and its parameters like number of layers, number of units, and the assigned weights to the connections which lead the model to an accurate forecast. One of the problems with this model is that there is no memory for keeping the information from the past. To overcome this problem, Recurrent Neural Networks (RNN) approach which is based on ANN can be used.

The ability of RNN to keep events from the past on its memory can be very useful in time-series forecasting. The inputs of each layer go to the hidden layer which has a recurrent loop to the back. Therefore, the output is a function of the previous input, concatenated on the activation value of the past hidden layers. RNN is shown in Fig. 4.

4.6.4.1. Long Short-Term Memory Networks (LSTM). The issue about RNN is that after passing many hidden layers due to the multiplication, the results may vanish. In other words, vanishing gradient or exploding gradient may happen. Long Short-Term Memory Networks (LSTM) is a solution for short term memory of RNN. LSTM is trained to use back-propagation overtime.

LSTM consists of three gates including forget, input, and output gates. The structure of LSTM can be explained mathematically with the following formulas.

$$\begin{aligned} f_t &= \sigma_g\left(W_f x_t + U_f h_{t-1} + b_f\right) \\ i_t &= \sigma_g\left(W_i x_t + U_i h_{t-1} + b_i\right) \\ o_t &= \sigma_g\left(W_o x_t + U_o h_{t-1} + b_o\right) \\ c_t &= f_t \circ c_{t-1} + i_t \circ \sigma_c\left(W_c x_t + U_c h_{t-1} + b_c\right) \\ h_t &= o_t \circ \sigma_h\left(c_t\right) \end{aligned} \tag{15}$$

Input gate, forget gate, input and output gates activation vectors are represented by x_t, f_t, i_t, o_t, respectively. In addition, hidden state vector, cell state vector, weight matrices, and bias vector are denoted by h_t, c_t, *W, U, b,* respectively. In the first step of forecasting, because of the sensitivity of LSTM, it is better to rescale the dataset to the range of 0 to 1. Then, the model can be designed. After fitting, it can be used for predicting the future sales. To define LSTM model with Keras library, the time-series should be transformed to a supervised learning problem. It can be done using the previous observation (*t - 1*) as the input, and the actual observation at time Step *t* as the output. The process of performing this transformation is challenging. Therefore, TimeseriesGenerator library is used in the first three LSTM models to transform this univariate time-series to the supervised learning problem automatically.

In this research, four different LSTM models for univariate time-series prediction are utilized. The first one is Vanilla LSTM which consists of a single hidden layer of LSTM and 1 layer of output. The second LSTM model is called Stacked LSTM and includes of multiple LSTM on top of each other. In stacked LSTM, the output should be a sequence rather than a single value for each input. Therefore, the *return_sequences* parameter is set to True. The third LSTM model is called bidirectional LSTM which learns the input sequence forward and backward by wrapping the first hidden layer in a layer called Bidirectional (Brownlee, 2018a). The goal of using the fourth LSTM model in this study is to explore the effectiveness of a few changes in this method which are similar to the Vanilla LSTM model. First, instead of using TimeseriesGenerator, the data is transformed into supervised learning manually. In addition, by making differenced series, some efforts are made to make the time-series more stationary.

4.6.4.2. Convolutional Neural Network (CNN). Finally, a Convolutional Neural Network (CNN) is utilized in this research. CNN is a kind of deep neural network which consists of one or more convolutional layers, and it mainly used for image processing and text classification (Nasir, Khan, & Varlamis, 2021). However, because of its ability for identifying patterns, it can be utilized in forecasting as well. The results from previous CNN is fed into the next CNN layer. There is the max-pooling layer that takes the maximum number in the sliding window, and prevents the model to overfit (Ackermann, 2018). Between this layer and the dense layer, a flatten layer is used. Fig. 5 shows a typical CNN architecture using 1D convolutional layer (Hu, Huang, Wei, Zhang, & Li, 2015). In this study, a 1D CNN model is proposed which consists of 3 one-dimension convolution with 128 filters.

5. Results and discussions

The results are discussed in this section. Each model is trained on a trainset including the first three years of furniture sales history. The rest of the data which is equal to the last year of sales history is used for testing each model.

5.1. Classical time-series models

ARMA (1, 1) is the first model that is used for sales forecasting, and the second model is ARIMA. To find the optimal order for this ARIMA model, the grid search method is utilized that indicates its order as (6, 0, 0) (Brownlee, 2017a). Fig. 6 demonstrates the sales forecasting. None of these models could capture the seasonality pattern and can detect the growth of sales at the end of the year. By comparing the MAPE values of

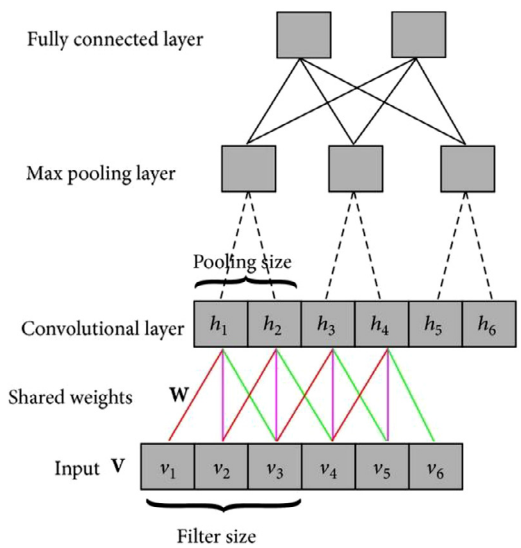

Fig. 5. Typical CNN architecture using 1D convolutions (Hu et al., 2015).

ARMA and ARIMA, it is observed that ARMA performs slightly better. However, RMSE value of ARIMA which is 282.50 is near 10 units lower.

Two different grid search methods are applied to Seasonal ARIMA (SARIMA) model (Vincent, 2017a). The first grid search calculates the optimal parameters with Auto-ARIMA from pmdarima Python library and reports back the best Akaike Information Criterion (AIC) value. It suggests $(p, d, q) = (0, 0, 0)$ and seasonal order = (1, 1, 0, 12) with AIC = 329.940 in the best model. Another grid search framework is developed manually to evaluate the performance of Auto-ARIMA grid search. It found that the optimal order is (1, 1, 0) x (1, 1, 0, 12). However, the value of RMSE for the first SARIMA model which has used Auto-ARIMA, has outperformed the second model by 35 units. Fig. 7 illustrates the results of sales forecasting of the first model, and shows the uncertainty level of forecast by calculating upper and lower sales predictions.

Among all the discussed models in the ARIMA family, the first SARIMA model had better performance with the MAPE value of 28.98% and RMSE value of 205.68. Fig. 8 shows that the SARIMA model can capture furniture sales seasonality in far future, although the prediction becomes less confident.

Fig. 9 displays a comparison of ARMA, ARIMA, and SARIMA sales forecasting. It indicates that SARIMA performed better, and it could capture the growth in the sales at the end of the year.

Double Exponential Smoothing (DES) and Triple Exponential Smoothing (TES) are two other classical time-series forecasting approaches that have been applied in this investigation. As expected, DES could not capture seasonality. The RMSE value is 636.08 which is the highest prediction error among the models. TES has performed better, but not as good as SARIMA. Fig. 10 illustrates the DES and the TES sales predictions.

5.2. Prophet method

The Prophet is one of the newest forecasting methods that has been released by Facebook company in 2017 (Vincent, 2017b). In this research, two Prophet models are utilized. In the first one, only the yearly seasonal component is set to True. In the second one, an attempt is made to improve the results of the first model by defining holidays argument which is one of the applicable tools of Prophet. After fitting the model and making a prediction, Prophet generates a table as a result which contains various data like yhat (the actual predicted value), and yhat_lower and yhat_upper which indicate the levels of uncertainty. Fig. 11 represents the classic prediction plot of the first Prophet model. In this graph, black dots are the actual time-series data points. In addition, the blue line shows the forecast, and the blue shade represents the uncertainty value by filling out between lower and upper levels of prediction.

Figs. 12 and 13 illustrate the results of sales forecasting of these models in another format to compare them easily according to the other graphs in this study. Based on these figures, both models have captured furniture sales seasonality. But the results of the first Prophet method have been improved by the second model and have reached RMSE of 167.29 from 194.93. According to the MAPE estimator, the error has decreased 4 percent, and has reached 22.62 percent. This method has outperformed the previous models such as SARIMA.

5.3. LSTM

Neural network algorithms initialize random weights during the training process and therefore, they produce different results even with

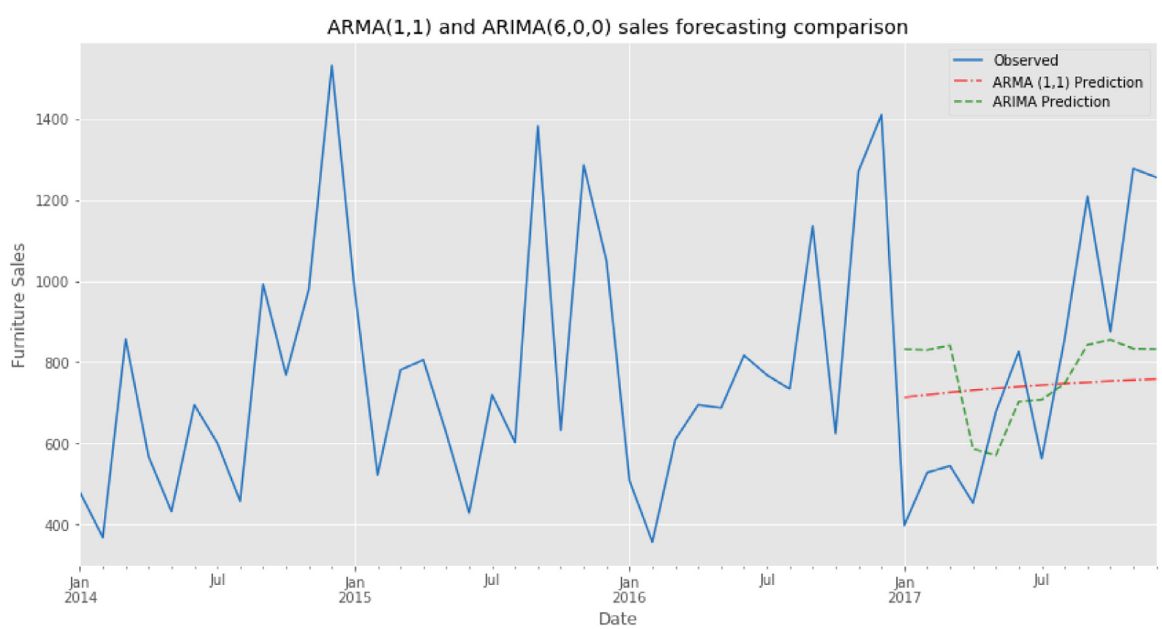

Fig. 6. ARMA and ARIMA sales forecasting.

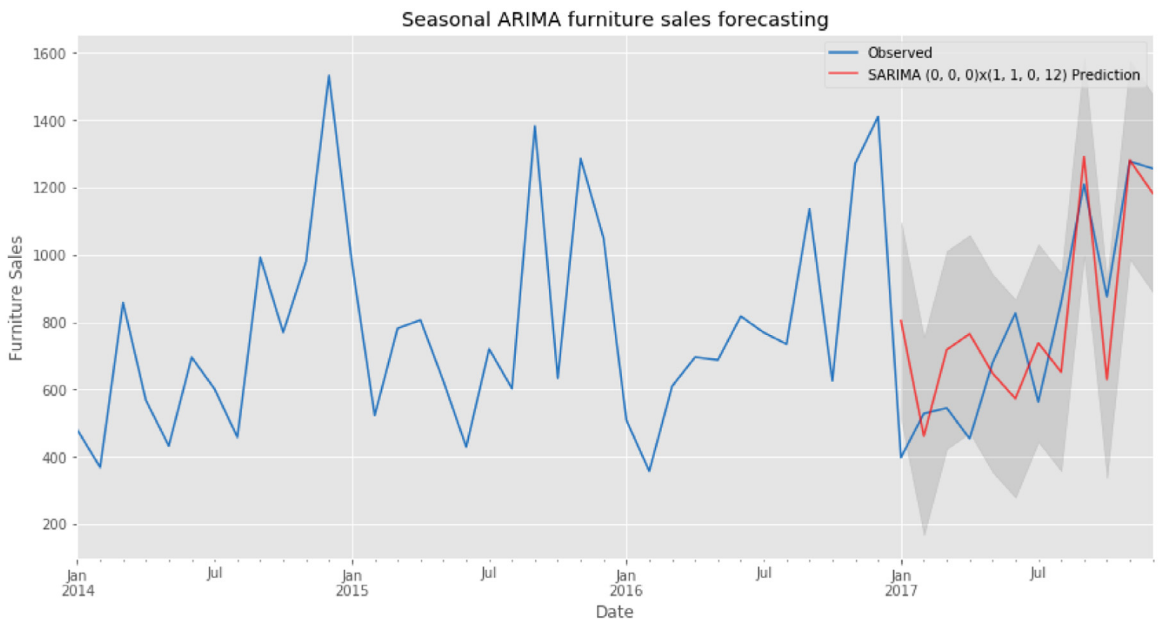

Fig. 7. SARIMA (0, 0, 0) x (1, 1, 0, 12) sales forecasting.

Fig. 8. SARIMA (0, 0, 0) x (1, 1, 0, 12) sales forecasting in far future.

the same network that has been trained on the same data. There was a major difference in the results of Vanilla LSTM and Stacked LSTM. To get stable results, the average results of the training and testing processes that has been repeated 20 times, is used as the final result. After fitting all these models, the results of prediction show the good and close performance of Vanilla LSTM and Stacked LSTM compared to the previous classical methods and Prophet models. The results indicate the superiority of Stacked LSTM. The MAPE value of Stacked LSTM is 17.34. MAPE values for Vanilla and Bidirectional LSTM are 18.39 and 31.40, respectively. The RMSE values of them are 128.51, 137.22, and 232.34. Fig. 14 illustrates the comparison of different LSTM models for furniture sales forecasting.

Another LSTM model (similar to Vanilla LSTM) is developed without using Time-series Generator (Brownlee, 2017b). The results are not as good as Vanilla LSTM results and have RMSE value of 262.05 which is near twice as the result of Vanilla LSTM. A comparison of this model and the other LSTM models is represented in Fig. 15.

5.4. CNN

The results of forecasting with CNN indicate MAPE value of 22.26 which puts this method as the third best method. However, its RMSE value is 199.85 which places this model as the fifth best model after Stacked LSTM, Vanilla LSTM, and both Prophet models. Fig. 16 shows the sales forecasting graph of this model.

5.5. Discussions

In this paper, an effort has been made to assess the ability of machine learning, deep learning, and traditional models to forecast seasonal time-series. Moreover, the performances of the models on real-world data have been evaluated by three widely used metrics including MSE, RMSE, and MAPE. Employing these metrics is a great asset in the interpretation of the results and determining the most appropriate forecasting method for the given dataset.

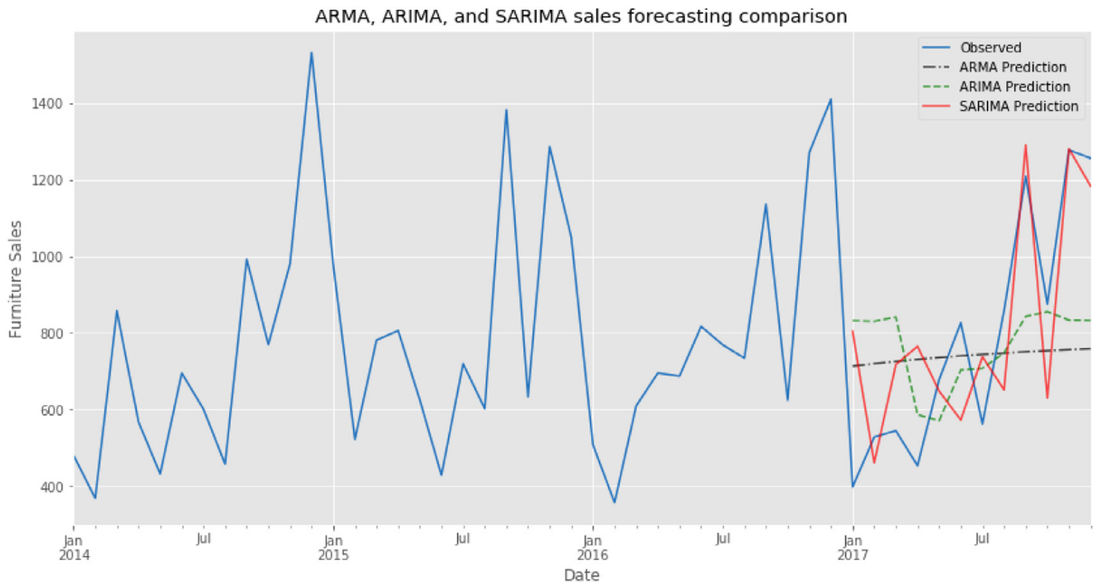

Fig. 9. ARMA, ARIMA, and SARIMA comparison sales forecasting.

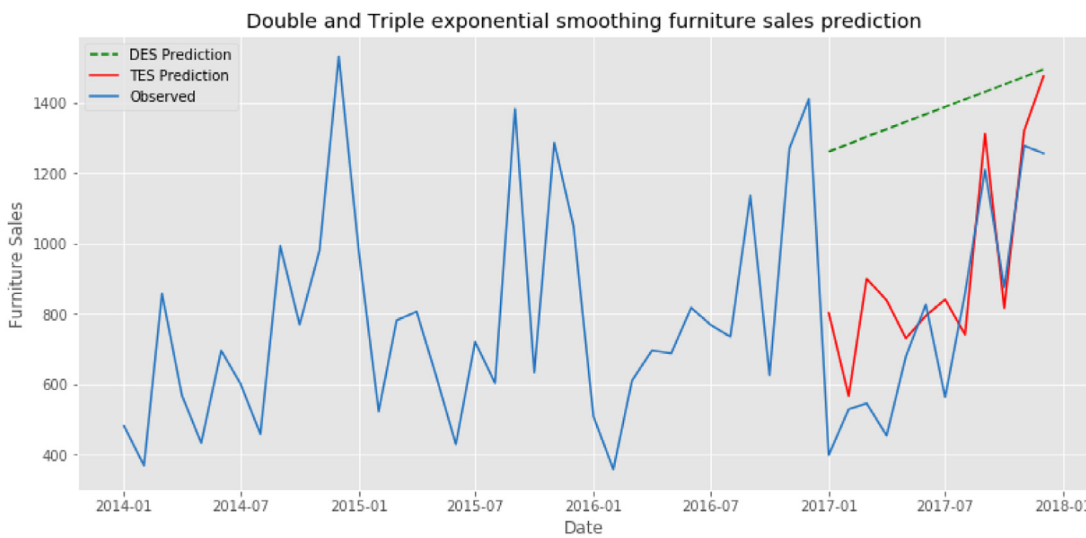

Fig. 10. Double and triple exponential smoothing furniture sales predictions.

Tables 1 and 2 include the prediction values in the test set based on the Python codes. In Table 1, the original values of the monthly sales of furniture can be seen in the TEST SET column. The next columns display the predicted values of monthly sales based on the results of classic models and Prophet. To elaborate, in January of 2017, the number of sales was $397.60. However, ARMA model has predicted this amount as $713.53. Table 2 has the similar structure, but it includes the predicted values of monthly sales based on the results of different LSTM and CNN models.

Fig. 17 illustrates the sales forecasting of the best models from each forecasting technique. They have produced mixed results in the sales prediction for the beginning of the year. However, all of them could capture the growth in the sales at the end of 2017.

In this research, the most commonly used KPIs have been utilized to measure the forecast accuracy. The first used KPI is MSE which has been discussed in Section 4.3. Another KPI in this study is RMSE which can be preferred to the MSE since it is on the same scale as the data. Finally, the third measure is MAPE which provides the advantage of being independent of scale (Hyndman & Koehler, 2006).

The performance of our models is measured with these KPIs and is summarized in Table 3. All three measures indicate the superiority of Stacked LSTM over the other models with the MAPE value of 17.34, and the RMSE value of 128.51. Among other neural networks models, the results show the good performance of Vanilla LSTM and CNN models compared to the other methods. Among the traditional models, the first applied SARIMA has outperformed the other classical methods in terms of forecasting accuracy in different performance measurements. Furthermore, the holiday factor has improved the prediction accuracy of the second Prophet model compared to the first model. In addition, it has been performed better than the SARIMA model.

5.6. Contributions to literature

The research contributions of this study are significant. First, to the best of our knowledge, this is the first study that has compared the performances of more than ten different forecasting models, such as classical and advanced techniques applied on a real-world dataset. Second, we

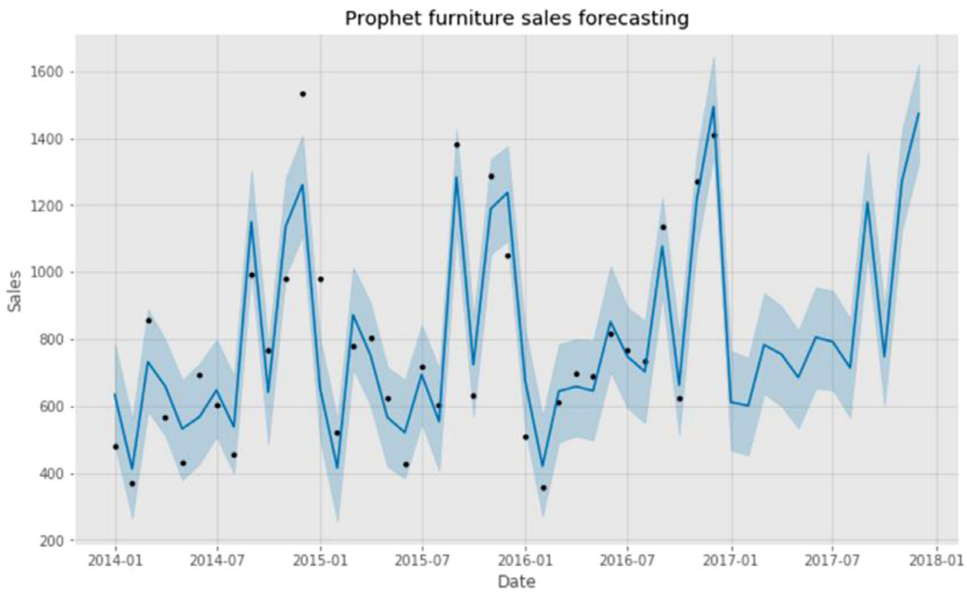

Fig. 11. Prophet classic sales prediction plot.

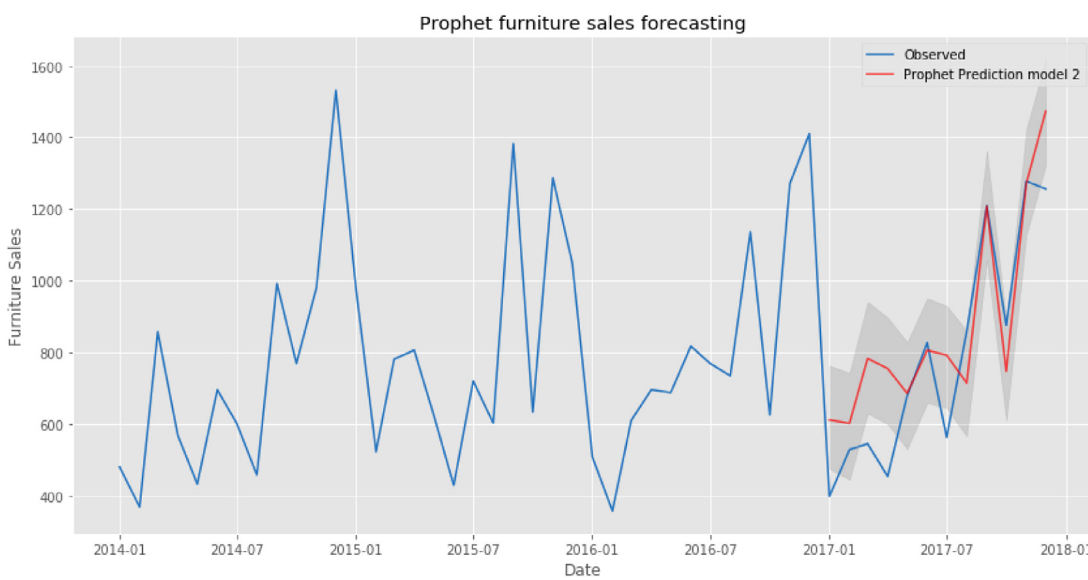

Fig. 12. Prophet sales forecasting without holidays.

Table 1
Prediction results of classical and Prophet models.

	Test Set	ARMA	ARIMA	SARIMA	DES	TES	Prophet1	Prophet2
2017–01–01	397.602	713.539	831.876	803.994	1261.313	802.530	804.654	611.315
2017–02–01	528.179	719.933	830.124	461.016	1282.552	566.047	601.206	601.273
2017–03–01	544.672	725.734	841.433	717.583	1303.791	899.554	782.905	782.874
2017–04–01	453.297	730.997	586.479	764.867	1325.030	838.253	754.585	754.559
2017–05–01	678.302	735.771	570.763	648.086	1346.269	729.941	685.352	685.312
2017–06–01	826.460	740.102	703.242	572.567	1367.508	794.136	805.507	805.412
2017–07–01	562.524	744.031	707.771	737.887	1388.747	840.477	792.293	792.174
2017–08–01	857.881	747.596	747.874	651.320	1409.986	740.652	714.448	714.293
2017–09–01	1209.508	750.829	843.043	1291.261	1431.224	1311.992	1208.615	1208.430
2017–10–01	875.362	753.763	855.413	629.973	1452.463	816.216	747.521	747.343
2017–11–01	1277.817	756.424	833.363	1281.007	1473.702	1320.823	1271.011	1270.810
2017–12–01	1256.298	758.838	832.149	1183.352	1494.941	1475.529	1473.348	1473.174

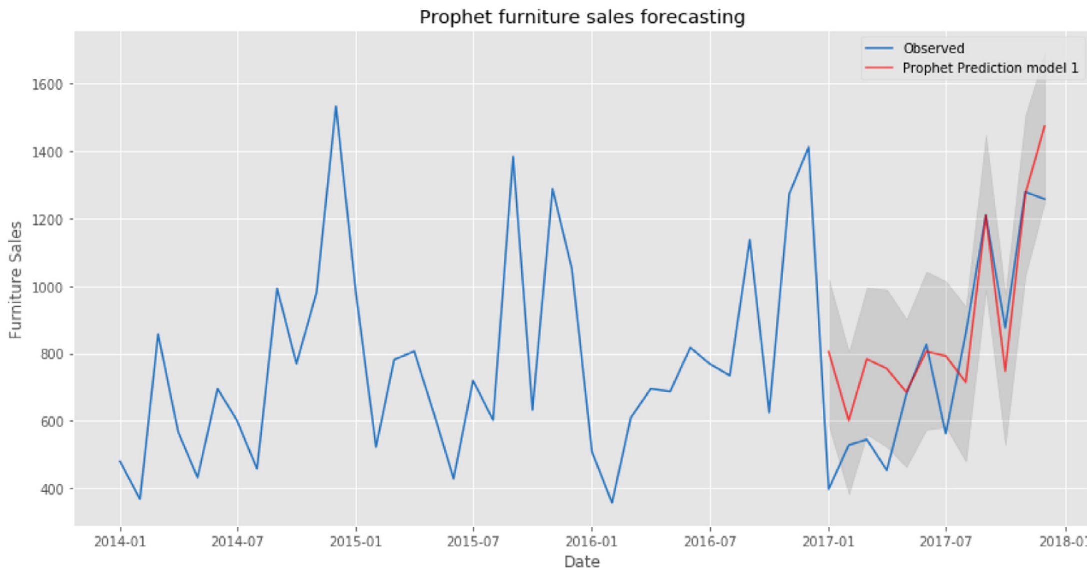

Fig. 13. Prophet sales forecasting with holidays.

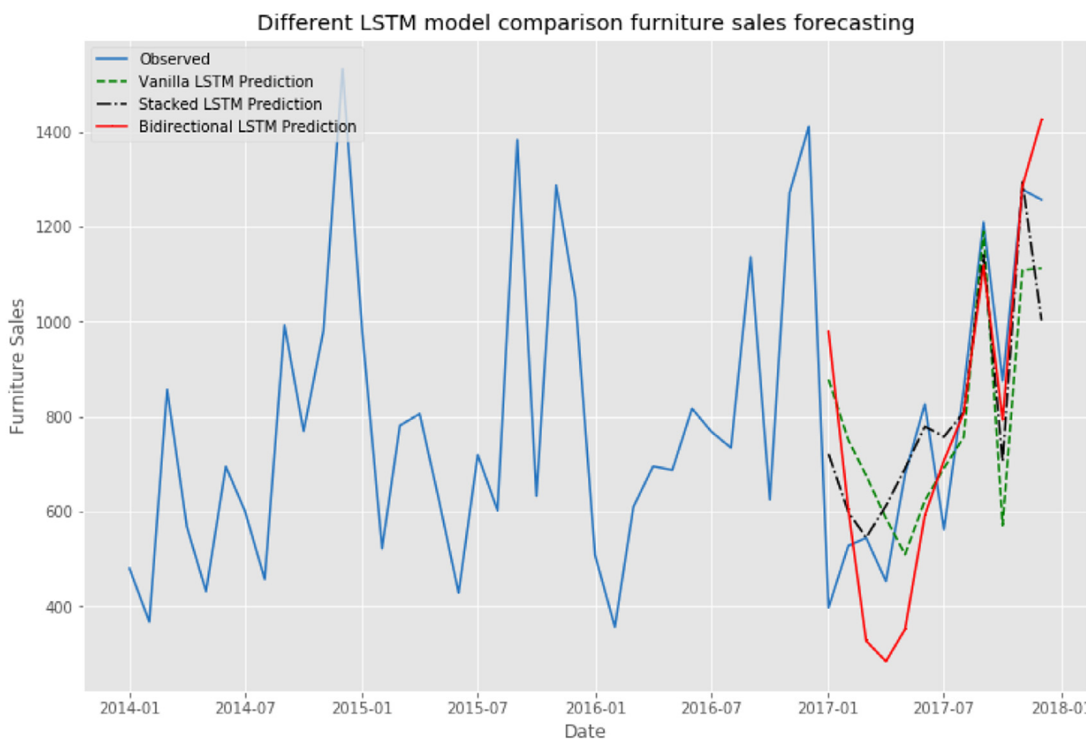

Fig. 14. The comparison of different LSTM models for furniture sales forecasting.

Table 2
Prediction results of LSTM models.

	Test Set	Vanilla LSTM	Stacked LSTM	Bidirectional LSTM	LSTM model 1	CNN
2017–01–01	397.602	878.086	721.582	978.992	1029.998	877.598
2017–02–01	528.179	751.017	598.841	604.988	589.142	556.138
2017–03–01	544.672	675.898	545.463	327.734	551.464	610.688
2017–04–01	453.297	585.882	612.004	284.586	608.583	480.613
2017–05–01	678.302	509.826	690.918	352.483	660.358	498.080
2017–06–01	826.460	623.241	779.181	591.979	697.176	461.794
2017–07–01	562.524	691.610	757.686	707.542	778.016	527.989
2017–08–01	857.881	756.711	808.759	807.887	798.324	891.835
2017–09–01	1209.508	1189.904	1140.226	1118.337	835.134	1312.435
2017–10–01	875.362	571.337	705.962	793.152	923.776	729.115
2017–11–01	1277.817	1108.735	1293.719	1285.389	907.176	1283.490
2017–12–01	1256.298	1112.389	1002.293	1425.237	1034.123	1048.552

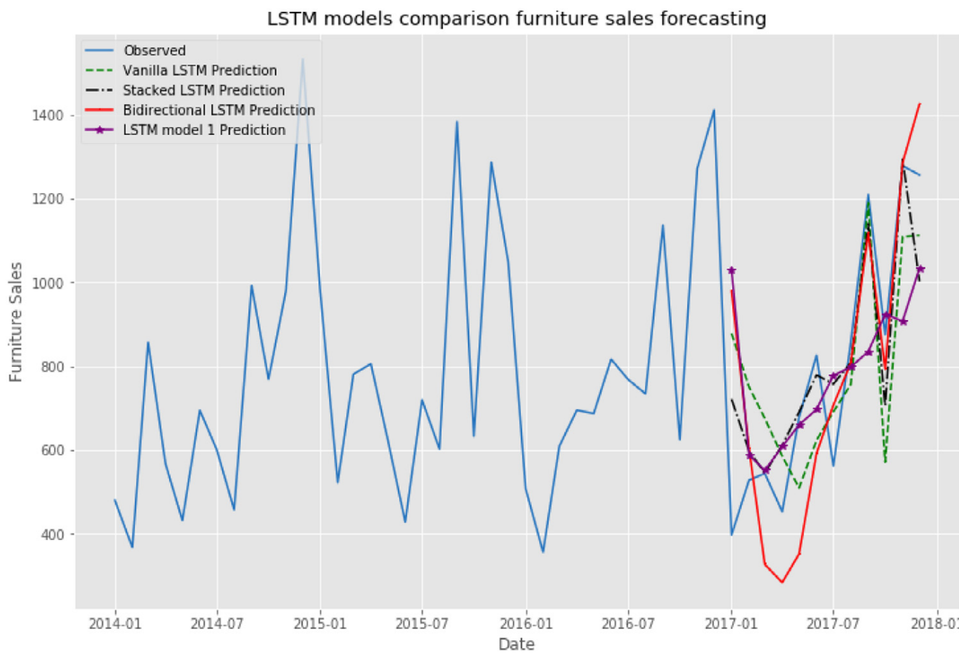

Fig. 15. Four different LSTM models comparison.

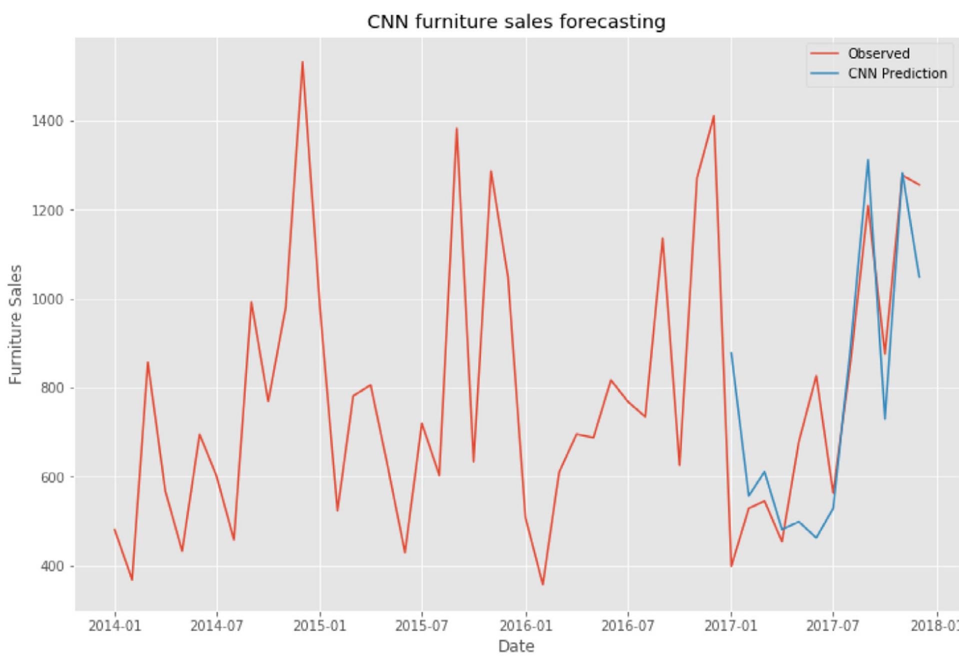

Fig. 16. CNN sales forecasting.

Table 3
Prediction results of different models.

Model	MSE	RMSE	MAPE
ARMA	87,237.01	295.36	33.88
ARIMA	79,804.31	282.50	35.07
SARIMA 1	42,305.37	205.68	28.89
SARIMA 2	55,497.86	235.58	33.50
DES	40,4596.36	636.08	98.79
TES	49,846.48	223.26	30.81
Prophet1	37,992.52	194.92	26.67
Prophet2	27,986.56	167.29	22.62
Vanilla LSTM	18,829.66	137.22	18.39
Stacked LSTM	16,515.49	128.51	17.34
Bidirectional LSTM	53,981.32	232.34	31.40
LSTM 1	68,671.50	262.05	29.40
CNN	39,938.47	199.85	22.26

have demonstrated the ability of neural networks models such as LSTM and CNN on seasonal time-series forecasting. Third, we have contributed to finding how competitive are neural networks time-series forecasting compared with traditional univariate methods. After performing a rigorous evaluation, the superior performance of LSTM was confirmed on this retail store dataset. Finally, this study has proven that CNN and the Prophet models are good candidates for predicting the future patterns of seasonal items.

5.7. Implications for practice

The derived results of this research can be applied in practice to forecast future sales of any items with a similar seasonal pattern. Moreover, by performing the necessary data pre-processing and tuning the hyper-parameters, the provided models can be applied to different datasets.

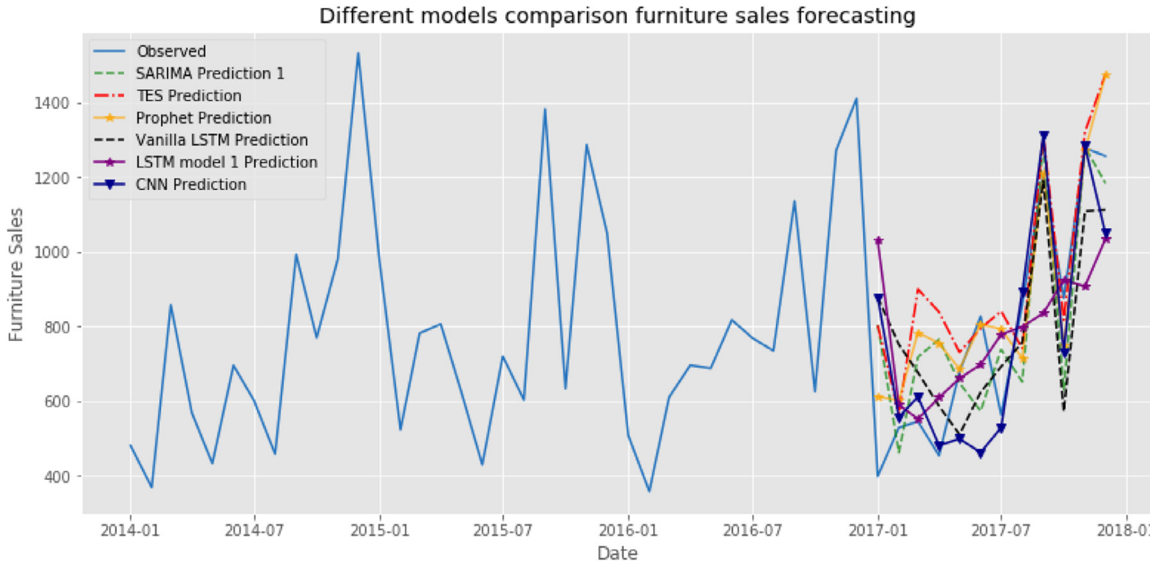

Fig. 17. Comparison of different models.

This has significant implications for advanced strategic decision-making and inventory planning.

6. Conclusions

Nowadays, sale forecasting is an inseparable part of every industry especially businesses that work with seasonal items. It is essential to know which model can produce better results because accurate estimation of the future can help businesses to boost their work. In this research, the capabilities of some neural network models for seasonal items forecasting have been investigated and compared with other forecasting methods.

After performing sales forecasting with various methods, the following conclusions can be drawn from this study. First, most of the neural network methods have performed better than the classical forecasting methods. In addition, the results have indicated the superiority of Stacked LSTM over the other models. Moreover, the CNN method which is usually applied for the task of image recognition showed a great performance in this case. In addition, the findings have highlighted the potential usefulness of Prophet model. Considering the purpose of this method, which is producing quick and accurate forecasts, this model has performed surprisingly good, and better than the other classical methods. Furthermore, the findings have indicated that using the holiday factor improves the performance of Prophet model.

In summary, five models including both of Prophet models, Vanilla LSTM, Stacked LSTM, and CNN (out of 13 models) are able to forecast the sales of furniture in the dataset within less than 200 units of the real sales. It should be mentioned that the furniture sales range from around 500 to over 1500. Therefore, a model with RMSE of 128.51 can be classified as an acceptable model.

The results of this paper depend on the selected dataset. It is possible to apply the same models to other seasonal time-series. However, it is still required to tune the parameters and find the most optimal forecasting model. As future research, it is valuable to evaluate the mentioned forecasting methods for other seasonal items and compare the performances of the applied models. Furthermore, there are other future research directions for this study. To improve the results, more complex LSTM and CNN models can be tested. Furthermore, applying multivariate time-series forecasting can be considered. In addition, developing hybrid models which are the combination of classical and modern forecasting methods can be valuable and a possible path for future research.

Acknowledgments

The authors would like to thank the editor and reviewers for the excellent comments that improved the quality of the paper significantly. This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Ackermann, N. (2018). Introduction to 1D convolutional neural networks in Keras for time sequences. [online] Medium. Available at: https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf [Accessed November 23, 2021].

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of Artificial Neural Networks and traditional methods. *Journal of Retailing and Consumer Services, 8*(3), 147–156.

Alpaydin, E. (2009). *Introduction to machine learning*. MIT press.

Arnoux, P. H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R., & Sinha, V. (2017). 25 tweets to know you: A new model to predict personality with social media. In *Proceedings of the international AAAI conference on web and social media: 11*.

Avuçlu, E., & Elen, A. (2020). Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements. *Medical & Biological Engineering & Computing, 58*(11), 2775–2788.

Box, G. E. P., & Jenkins, G. M. (1970). Control. Halden-Day, San Francisco.

Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). *Introduction to time series and forecasting*: 2. New York: Springer.

Brownlee, J. (2016). How to backtest machine learning models for time series forecasting. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/ [Accessed November 23, 2021].

Brownlee, J. (2017a). How to grid search ARIMA model hyperparameters with python. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/ [Accessed November 23, 2021].

Brownlee, J. (2017b). Time series forecasting with the long short-term memory network in python. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python/ [Accessed November 23, 2021].

Brownlee, J. (2018a). How to develop LSTM models for time series forecasting. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/ [Accessed November 23, 2021].

Brownlee, J. (2018b). A gentle introduction to exponential smoothing for time series forecasting in python. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/ [Accessed November 23, 2021].

Chakraborty, A., & Kar, A. K. (2017). Swarm intelligence: A review of algorithms. *Nature-Inspired Computing and Optimization*, 475–494.

Choi, B. (2012). *ARMA model identification*. Springer Science & Business Media.

Chopra, S., & Meindl, P. (2016). *Supply chain management: Strategy, planning, and operation*. Community.tableau.com [online] Available at https://community.tableau.com/docs/DOC-1236 [Accessed November 23, 2021].

Corrius, J. (2018). Simple stationarity tests on time series - bluekiri - Medium. [online] Medium. Available at: https://medium.com/bluekiri/simple-stationarity-tests-on-time-series-ad227e2e6d48 [Accessed November 23, 2021].

Furniture Global Market Report. (2021). COVID-19 impact and recovery to 2030. [online] Available at: https://www.researchandmarkets.com/reports/5238009/furniture-global-market-report-2021-covid-19#src-pos-1 [Accessed November 23, 2021].

Gamboa, J. C. B. (2017). Deep learning for time-series analysis. arXiv preprint arXiv:1701.01887.

Gensler, A., Henze, J., Sick, B., & Raabe, N. (2016). Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM neural networks. In *Proceeding of the IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 002858–002865). IEEE.

Ghosh, I., & Sanyal, M. K. (2021). Introspecting predictability of market fear in Indian context during COVID-19 pandemic: An integrated approach of applied predictive modelling and explainable AI. *International Journal of Information Management Data Insights, 1*(2), Article 100039.

Giering, M. (2008). Retail sales prediction and item recommendations using customer demographics at store level. *ACM SIGKDD Explorations Newsletter, 10*(2), 84–89.

Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management, 23*(5), 499–510.

Hamzaçebi, C. (2008). Improving Artificial Neural Networks' performance in seasonal time series forecasting. *Information Sciences, 178*(23), 4550–4559.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting, 20*(1), 5–10.

Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors, 2015*.

Huang, T., Fildes, R., & Soopramanien, D. (2019). Forecasting retailer product sales in the presence of structural change. *European Journal of Operational Research, 279*(2), 459–470.

Huber, J., Gossmann, A., & Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. *Expert Systems with Applications, 76*, 140–151.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer Science & Business Media.

Islam, S., & Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques. *Journal of Big Data, 7*(1), 1–22.

Islam, S., Amin, S. H., & Wardley, L. J. (2021). Machine learning and optimization models for supplier selection and order allocation planning. *International Journal of Production Economics, 242*, Article 108315.

Jain, A. (2016). Complete guide to time series forecasting (with codes in python). [online] analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/ [Accessed November 23, 2021].

Kalekar, P. S. (2004). Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi School of Information Technology, 4329008*(13).

Kaneko, Y., & Yada, K. (2016). A deep learning approach for the prediction of retail store sales. In *Proceedings of the IEEE 16th international conference on data mining workshops (ICDMW)* (pp. 531–537). IEEE.

Kang, E. (2017). Time series: Check stationarity - EugineKang - medium. [online] Medium. Available at: https://medium.com/@kangeugine/time-series-check-stationarity-1bee9085da05 [Accessed November 23, 2021].

Kar, A. K. (2016). Bio inspired computing–A review of algorithms and scope of applications. *Expert Systems with Applications, 59*, 20–32.

Kar, A. K., & Dwivedi, Y. K. (2020). Theory building with big data-driven research–Moving away from the "What" towards the "Why". *International Journal of Information Management, 54*, Article 102205.

Khashei, M., & Bijari, M. (2010). An Artificial Neural Network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications, 37*(1), 479–489.

Koprinska, I., Wu, D., & Wang, Z. (2018). Convolutional neural networks for energy time series forecasting. In *Proceedings of the international joint conference on neural networks (IJCNN)* (pp. 1–8). IEEE.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159–190.

Kotzur, L., Markewitz, P., Robinius, M., & Stolten, D. (2018). Impact of different time series aggregation methods on optimal energy system design. *Renewable Energy, 117*, 474–487.

Kusiak, A., & Zhang, Z. (2010). Short-horizon prediction of wind power: A data-driven approach. *IEEE Transactions on Energy Conversion, 25*(4), 1112–1122.

Lee, H. C., Rim, H. C., & Lee, D. G. (2019). Learning to rank products based on online product reviews using a hierarchical deep neural network. *Electronic Commerce Research and Applications, 36*, Article 100874.

Li, C., & Lim, A. (2018). A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. *European Journal of Operational Research, 269*(3), 860–869.

Loureiro, A. L., Miguéis, V. L., & da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems, 114*, 81–93.

Lv, H. R., Bai, X. X., Yin, W. J., & Dong, J. (2008). Simulation based sales forecasting on retail small stores. In *Proceedings of the winter simulation conference* (pp. 1711–1716). IEEE.

Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. arXiv preprint arXiv:1607.00148.

McCoy, T. H., Pellegrini, A. M., & Perlis, R. H. (2018). Assessment of time-series machine learning methods for forecasting hospital discharge volume. *JAMA Network Open, 1*(7), Article e184087 -e184087.

Momeny, M., Latif, A. M., Sarram, M. A., Sheikhpour, R., & Zhang, Y. D. (2021). A noise robust convolutional neural network for image classification. *Results in Engineering, 10*, Article 100225.

Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights, 1*(1), Article 100007.

Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques with the applications in supply chain management. *International Journal of Information Management, 57*, Article 102282.

Olah, C. (2015). Understanding LSTM Networks – Colah's blog. [online] Colah.github.io. Available at: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ [Accessed November 23, 2021].

Padilla, W. R., García, J., & Molina, J. M. (2021). Improving time series forecasting using information fusion in local agricultural markets. *Neurocomputing, 452*, 355–373.

Pai, P. F., Lin, K. P., Lin, C. S., & Chang, P. T. (2010). Time series forecasting by a seasonal support vector regression model. *Expert Systems with Applications, 37*(6), 4261–4265.

Papacharalampous, G., Tyralis, H., & Koutsoyiannis, D. (2018). Predictability of monthly temperature and precipitation using automatic time series forecasting methods. *Acta Geophysica, 66*(4), 807–831.

Polusmak, E. (2018). Open machine learning course. Topic 9. Part 2. Predicting the future with Facebook prophet. [online] Medium. Available at: https://medium.com/open-machine-learning-course/open-machine-learning-course-topic-9-part-3-predicting-the-future-with-facebook-prophet-3f3af145cdc [Accessed November 23, 2021].

Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing, 34*, 151–163.

Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science, 179*, 524–532.

Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. In *Proceedings of the international conference on advances in computing, communications and informatics (ICACCI)* (pp. 1643–1647). IEEE.

Sharma, S., Rana, V., & Kumar, V. (2021). Deep learning based semantic personalized recommendation system. *International Journal of Information Management Data Insights, 1*(2), Article 100028.

Singh, A. (2018). A gentle introduction to handling a non-stationary time series in python. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2018/09/non-stationary-time-series-python/ [Accessed November 23, 2021].

Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change, 163*, Article 120480.

Subashini, A., Sandhiya, K., Saranya, S., & Harsha, U. (2019). Forecasting website traffic using prophet time series model. *International Research Journal of Multidisciplinary Technovation, 1*(2), 1–8.

Tandon, C., Revankar, S., & Parihar, S. S. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics. *International Journal of Information Management Data Insights, 1*(2), Article 100035.

Taylor, J. W. (2011). Multi-item sales forecasting with total and split exponential smoothing. *Journal of the Operational Research Society, 62*(3), 555–563.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37–45.

Tkáč, M., & Verner, R. (2016). Artificial Neural Networks in business: Two decades of research. *Applied Soft Computing, 38*, 788–804.

Verstraete, G., Aghezzaf, E. H., & Desmet, B. (2019). A data-driven framework for predicting weather impact on high-volume low-margin retail products. *Journal of Retailing and Consumer Services, 48*, 169–177.

Verstraete, G., Aghezzaf, E. H., & Desmet, B. (2020). A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. *Computers & Industrial Engineering, 139*, Article 106169.

Vincent, T. (2017a). ARIMA time series data forecasting and visualization in python | DigitalOcean. [online] Digitalocean.com. Available at: https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3 [Accessed November 23, 2021].

Vincent, T. (2017b). A guide to time series forecasting with prophet in python 3 | DigitalOcean. [online] Digitalocean.com. Available at: https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-prophet-in-python-3 [Accessed November 23, 2021].

Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering, 129*(6), 664–672.

Yang, H., Li, X., Qiang, W., Zhao, Y., Zhang, W., & Tang, C. (2021). A network traffic forecasting method based on SA optimized ARIMA–BP neural network. *Computer Networks, 193*, Article 108102.

Zekić-Sušac, M., Has, A., & Knežević, M. (2021). Predicting energy cost of public buildings by Artificial Neural Networks, CART, and random forest. *Neurocomputing, 439*, 223–233.

Zhang, C., Kuppannagari, S. R., Kannan, R., & Prasanna, V. K. (2018). Generative adversarial network for synthetic time series data generation in smart grids. In *Proceedings of* *the IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm)* (pp. 1–6). IEEE.

Zhang, G. P., & Kline, D. M. (2007). Quarterly time-series forecasting with neural networks. *IEEE Transactions on Neural Networks, 18*(6), 1800–1814.

Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research, 160*(2), 501–514.

Zhao, K., & Wang, C. (2017). Sales forecast in e-commerce using convolutional neural network. arXiv preprint arXiv:1708.07946.

Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems, 11*(2), 68–75.

Zhuge, Q., Xu, L., & Zhang, G. (2017). LSTM neural network with emotional analysis for prediction of stock price. *Engineering Letters, 25*(2).