IMPROVEMENT OF SALES PREDICTION BY FITTED-TO-PRODUCT TIME-SERIES MODELS

Mohamed Sameh Belaid, Stéphane Lecoeuche, Anthony Fleury, Baptiste Herve

13[the] International Conference on Modeling, Optimization and Simulation - MOSIM'20 – November 12-14, 2020-Agadir – Morocco "New advances and challenges for sustainable and smart industries"

IMPROVEMENT OF SALES PREDICTION BY FITTED-TO-PRODUCT TIME-SERIES MODELS

Mohamed Sameh Belaid, Stéphane Lecoeuche, Anthony Fleury

IMT Lille-Douai, Douai, France
mohamed-sameh.belaid@imt-lille-douai.fr,
stephane.lecoeuche@imt-lille-douai.fr,
anthony.fleury@imt-lille-douai.fr

Baptiste Hervé

elm.leblanc, Bosch Group, Drancy, France
baptiste.herve@fr.bosch.com

ABSTRACT: *In this paper, we present an application of machine learning models with real-world data in order to enhance a company's performance. First, we present a brief literature review about the new trends in the field of supply chain predictive analytics and about the models used during this case study. We present then the industrial data-driven models to improve the overall supply chain performances in the scope of the Sales and Operations Planning (S&OP) process. We will present in our work, the considered assumptions and the methodology developed in order to implement a sales prediction engine, through several time series forecasting models. In this work, we use and mix classic time series models like ARIMA and Exponential Smoothing, and the recent Prophet algorithm. Finally, we present the results of the implementation of our methodology and the noted improvements on the predictions accuracy. Our discussions derive on how predicting customer demand for many products from different natures and on how the model parameters affected our sales forecasting process.*

KEYWORDS: *Industry 4.0, Supply Chain, Machine Learning techniques, Data Analytics, Time Series*

1 INTRODUCTION

In an economic world characterized by increased competition, companies are still seeking to strengthen their competitiveness in order to gain a foothold in the market. To achieve this goal, several areas of development are considered, among which, the improvement of the supply chain through exploiting emerging data-driven techniques.

Our paper is related to an industrial application seeking for improving the industrial performance of the company elm.leblanc (subsidiary of Bosch Group and a French market leader in the heating systems industry). We aim to implement innovative methods around data-driven modelling, which will ensure better management and efficiency of the supply chain.

In order to reduce storage costs and effectively plan the company production, we realized investigations and diagnostics over the Supply Chain stockholders. A common point reported is the poor quality of sales forecasts as it is the first step in the Sales & Operations Planning process (will be discussed later) and the conductive phase of the rest of the Supply Chain.

The sales prediction task represents multiple complexities due to the influence of internal and external environments (Sun *et al.*, 2008) such as the diversity of the products, the several variables that could affect the sales and the various non-linear patterns that can be represented in the data. Added to that, a huge data volumes are generated leading to difficulties in manually making accuracy forecasts. In order to raise these challenges, machine-learning techniques are considered as interesting approaches and suitable methods to deal with such data (Cheriyan *et al.*, 2019) through extracting useful information and insights. To respond to this industrial need, we investigate the state-of-the-art data-driven models and exploit them to implement a sales prediction engine.

Our proposition consists in a scalable methodology which aims to improve sales forecasting through selecting the accurate model for each product based on several assumptions. This is motivated by the fact that products have specific sales signatures and trends. These works illustrate an industrial experience of implementing a machine learning approach with real-world data. The purpose is to reduce costs, enhance planning processes, improve inventory management and optimize the overall supply chain to satisfy customer demand.

In this paper, we present a brief literature review about the new trends for supply chain improvements coupled with the S&OP process and its relation with these trends. We also present a literature review about different models used for time series forecasting. Then, we will introduce the framework application and the followed methodology in order to implement an accurate prediction engine for a variety of products and sales patterns. Finally, we discuss the results and experiments and our future intended works.

2 SUPPLY CHAIN PREDICTIVE ANALYTICS

2.1 Sales & Operations Planning (S&OP)

The S&OP process is an integral process that aims to balance supply and demand, and to make the link between the company's strategic and its operational plans (Thome *et al.*, 2012a). These objectives are achievable through the deployment of a vertical alignment between the planning levels (strategic, tactical, operational) (Thome *et al.*, 2012b) as well as a horizontal alignment between the different functions of the company or inter-functional alignment. This process typically involves five steps as shown in the following figure 1.

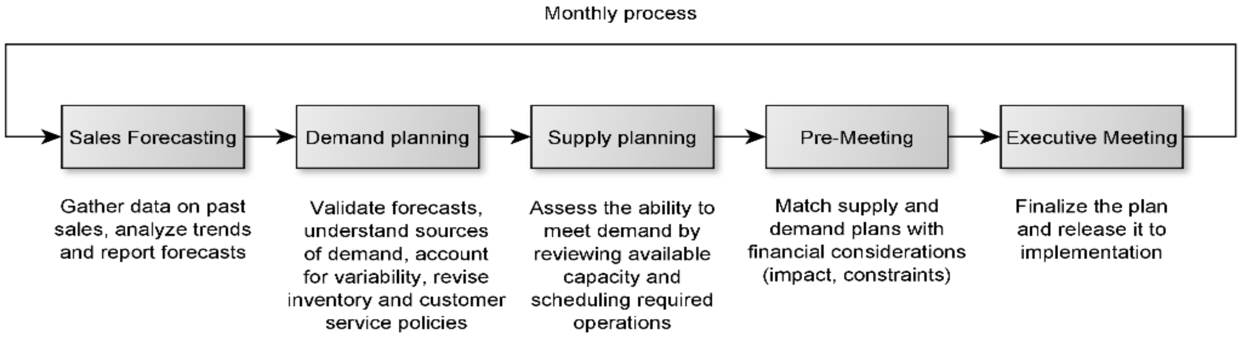

Figure 1: The S&OP process

According to (Wallace and Stahl, 2008), the demand-planning step can be the most challenging within the five steps of the S&OP for many companies. This step is preceded (and strongly influenced) by the demand forecasting, in which statistical techniques are used. (Lapide, 2011) stresses the importance of the demand forecasts for the smooth running of the S&OP process since the plans are driven by it. (Wang *et al.*, 2016) affirm that there are two widely quantitative techniques used in supply chain analytics which are "times-series analysis and forecasting" and "regression analysis". The first aims to determine existing patterns in historical data and predicts future events based on that, and the second intends to understand relationships and causality between variables. Beyond these classic methods, several models are developed by associating other machine learning techniques like Artificial Neural Networks. The rest of the paper is devoted to that objective, requesting the use of efficient time series modelling techniques.

2.2 New trends and field of research

Recent works on Supply Chain Management (SCM) field are interested in the integration of new technologies derived from Data science such as big data and predictive analytics. These trends bring out new opportunities for professionals. (Waller, Fawcett and Beane, 2013) suggest different definitions and concepts in this purpose, namely, SCM Predictive analytics and SCM data science. They are defined as the application of advanced analytic techniques including data mining, statistical analyses and predictive analytics on big datasets (Tiwari, Wee and Daryanto, 2018). They bring more possibilities to create data-driven decision-making tools in order to improve supply chain profitability.

In this context, our works focus on a new contribution of SCM predictive analytics in the objective to improve the operation planning decision process by the means of a better understanding and a better prediction of the demand. The next section gives an overview of models, used for predictive analysis and time series forecasting.

3 TIME SERIES FORECASTING MODELS

Time series, is a sequence of data points representing the evolution of a quantity over time, such as product sales. This area has attracted the attention of researchers over last few decades due to its different applications, especially in economy, finance, retail and industry. Time series forecasting fits under supervised machine learning problems and specifically regression ones. It aims to determine the way in which time series are changing by describing the relationship between its historical data points and/or other explanatory variables through a mathematical model. Several models were developed in the literature in order to model and predict time series. We present hereafter the models that were studied and used in our approach.

3.1 ARIMA and Seasonal ARIMA

ARIMA, which stands for Autoregressive Integrated Moving Average, is a linear forecasting model mainly used for stationary time series prediction (Panigrahi and Behera, 2017) .i.e. time series with time independent properties such as mean and variance. It is a combination of autoregressive (AR) and moving average (MA) models, with a term of differencing (I) used to overcome the non-stationarity of some time series. ARIMA is characterized by (p,d,q) parameters which are respectively, the number of lag observations, the size of moving average window and the number of times to differentiate the time series in order to make it stationary. ARIMA can be expressed (Hyndman, R.J. Athanasopoulos, 2019) as the following equation (1):

$$(1-\varphi_1 B-\cdots-\varphi_p B^p)(1-B)^d y_t = c+(1+\theta_1 B+\cdots+\theta_q B^q)\varepsilon_t \quad (1)$$

Where y_t is the data values at time *t*, ε_t an identically distributed random error with a null mean and a constant variance, φ_p and θ_q are respectively the Auto Regression and the Moving Average coefficients. A backshift operator *B* defined as $B^d y_t = y_{t-d}$.

The most important phase in the process of forecasting using ARIMA model is identifying the best parameters that lead to well fit the data and provide forecasts with a good accuracy. In the literature, different methods are proposed to find these parameters. The most popular way is to follow the Box-Jenkins methodology (BOX *et al.*, 2016). It relies on an iterative approach of identifying a model, estimate its parameters and diagnostic checking to determine the best model. The choice of models is generally based on ACF (autocorrelation function), PACF (partial autocorrelation function), AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) calculation. Another method is to simply try all possible values of the parameters through a grid search and then compare them by a metric like AIC. A more intelligent and automated method is the Hyndman-Khandakar algorithm described in (Robin Hyndman *et al.*, 2008). It is an automated process of best order identification through combining unit root tests, minimization of the AICc (corrected AIC) and MLE (Maximum Likelihood Estimation) (Hyndman, R.J. Athanasopoulos, 2019).

In the Seasonal ARIMA extension, additionally to the (p,d,q) ARIMA orders, (P,D,Q,m) orders represents the seasonal part of the model. "P" represents the seasonal autoregressive order, "Q" represents the seasonal moving average order, "D" is the seasonal difference order and "m" is the seasonal period of the time. SARIMA can be expressed through the following equation (2):

$$\Phi_P(B^m)\,\varphi_p(B)(1-B)^d\,(1-B^m)^D\,y_t = \Theta_Q(B^m)\,\theta_q(L)\varepsilon_t \quad (2)$$

Where Φ of order *P* and Θ of order *Q* are respectively the seasonal Auto-Regression and the seasonal Moving Average coefficients, and B the backshift operator. Comparing to ARIMA model, SARIMA(p,d,q)(P,D,Q,m) has more possible combinations, and in order to fit the best model, the parameters can be chosen through the same methods presented before for ARIMA.

3.2 Exponential Smoothing

Exponential smoothing models are classical time series forecasting models widely used in business and industry (De Gooijer and Hyndman, 2006). They include 24 different variants presented in (Hyndman *et al.*, 2002). Simple exponential smoothing consists in calculating future values based on weighted averages of past values with an exponential decreasing of the weights. It is characterized by an α parameter representing a forgetting factor (Hyndman, R.J. Athanasopoulos, 2019) i.e. the importance that will be given to the last observation. It is usually used with data representing no trend and no seasonality patterns, to predict the level i.e. the average value of the series.

This model was improved later with (Holt, 2004) works, adding the capacity of detecting the trend component in the time series through adding another component with a second parameter β. This is known as double exponential smoothing model with constant trend. The works of (Gardner, 1985), (Taylor, 2003) proposed to model a "damped" trend through introducing a damping parameter ϕ able to capture a constant long-term evolution, which forces the trend to converge to a constant value in the long-term future.

Through adding a seasonal component, the triple exponential smoothing comes with two variants: additive and multiple seasonality components (Hyndman, R.J. Athanasopoulos, 2019). The seasonal effect is determined through an exponential weighting method with a learning rate γ. (Hyndman, R.J. Athanasopoulos, 2019) presented in their book the different taxonomies and formulas of level, slope and seasonality calculation. The following equation (3) presents an example of an exponential smoothing model with additive trend and seasonality:

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \quad (3)$$

Where $\hat{y}_t$, l_t, and s_t represent respectively the predicted values, the level, the slope and the seasonal component of the series at time *t*. *h* denotes the number of predicted periods and *k* is the integer part of $\frac{h-1}{m}$. *m* is the frequency of seasonality. The choice of the best parameters (*α, β, ϕ, γ*) is done by the analyst or fitted automatically by an optimization tool. In the most of the cases, the selection is based on maximizing of the likelihood i.e. the probability of the data arising from the specified.

3.3 Prophet

Recently, (Taylor and Letham, 2018) published a new forecasting model based on the Generalized Additive Models (GAM) (Hastie, Tibshirani and Friedman, 2008). GAMs are recently used in the data science field (LARSEN, 2015). They represent an interesting model structure thanks to their interpretability, flexibility and automation with the predictors' capacity to model different behaviours. Based on this approach, Prophet model was proposed by the research team of Facebook in order to predict time series. The model is composed from three main components: trend, seasonality and holidays as presented in the equation (4).

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \qquad (4)$$

The trend $g(t)$ can represent a nonlinear behaviour by the mean of a piecewise logistic growth model, and a linear one through a piecewise linear model. An interesting feature in trend modelling is the "change points" aspect where the user can specify special date points where trend may change according to his knowledge on the time series. The seasonality aspect $s(t)$ is modelled through a standard Fourier series that can handle yearly, monthly and weekly seasonality. The Prophet model can handle additive or multiplicative seasonality in the time series. The third component $h(t)$ represents holidays or special events modelling.

3.4 Machine learning approach

Several works are presented in the literature on solving sales prediction task with machine learning techniques. In (Cheriyan *et al.*, 2019), authors evaluate the performances of three models: generalized linear model, decision tree and gradient boosted tree, and select the best algorithm to make sales prediction for an e-fashion store. In the same idea, (Gumani *et al.*, 2017) compare various machine learning models namely XGBoost, Support Vector Machines (SVM) and Regressive Neural Network (ARNN) with ARIMA, hybrid methods and STL decomposition based methods in order to figure out the most accurate model to forecast drug store company sales. They concluded that non-linear models such SVM and XGBoost performed better than ARIMA, however, hybrid models gave better performances than individual ones. Added to that, STL decomposition based methods gave the best results in terms of prediction accuracy.

An interesting approach has drawn attention in the recent years by combining multiple machine learning models to make final predictions. (Punam, Pamula and Jain, 2019) present in their works a two-level model for sales forecasting. First, predictions are made individually by the different models such as linear regression and support vector regression (SVR). Then, these predictions are combined using another model such as cubist to make more accurate final predictions. In the same idea, (Pavlyshenko, 2019) implement a multilevel model composed of several machine learning models, and the final predictions are generated through summing weighted results of the last level.

3.5 Other approaches

Other interesting approaches also exist in the state of the art and has recently emerged. Long Short-term Memory (LSTM) represents a successful application of the Artificial neural networks (ANN) for time series forecasting (Pankaj Malhotra *et al.*, 2015). However these models present some inconvenient especially their complexity and over-parameterization (De Gooijer and Hyndman, 2006). In addition, generally, they need very long time series to produce good forecasts, so that for short-time series, classic methods perform better.

Another approach emerged recently through the use of hybrid methods through combining linear and nonlinear models such as ARIMA with ANN in (Zhang, 2003) works, or combining Exponential Smoothing with ANN models (Panigrahi and Behera, 2017).

Several works in the literature have proved that classic methods such as ARIMA and Exponential Smoothing outperform ANN (Gorr, Nagin and Szczypula, 1994). However, others have demonstrated that ANN are more accurate in terms of forecasting (Aras and Kocakoç, 2016). Actually, these results depend on the application, the type, and the length of the time series in question. For our case, we are treating univariate noisy time series that are more or less short (60 data points at maximum). Therefore, ANN will not deliver accurate results as stated due to the lack of long-time series data. Added to that, our data may represent linear behaviour that can be modelled with ARIMA and simple or double exponential smoothing, and nonlinear patterns that can be modelled through Triple exponential smoothing or Prophet. Based on these facts, we decided to use classical models to implement a prediction engine for all our time series.

4 PROPOSED METHODOLOGY

4.1 Data preparation

Our first step is the data collection. It is represented by the sales information such as the quantities, products, type of customers, products family, etc. from 2015 until the end of 2019. Data are gathered from different sources namely the enterprise resource planning system, marketing department and sales department. These data are merged together and aggregated to construct time series representing monthly quantities per product. Then, the whole dataset of time series is fed as input to the algorithms as explained later in the methodology.

4.2 Assumptions

Our time series represent sales quantities of several products per month. Some products present seasonal patterns, however, others do not and these products belong to different families, which lead to different nature of time series. Another challenge is that some products have been recently launched, so with short time series and others that are older with longer time series.

As (Hyndman and Kostenko, 2007) explained in their article, the use of seasonal models requires a minimum number of observations on the data since they should be at least larger than the number of parameters by one. They highlight that the use of Holt-Winters method for example, requires at least 17 observations for monthly data, and for SARIMA(0,1,1)(0,1,1)12 at least 16 observations. Additionally, the randomness and variations also require more data to fully capture the patterns in data. Based on these statements, we decided to separate our data into two groups: short time series with fewer than 24 observations to be forecasted with non-seasonal models, and long-time series with more than 24 observations can be forecasted using seasonal models.

As mentioned in (Hyndman, R.J. Athanasopoulos, 2019), exponential smoothing methods with multiplicative trend tend to produce poor forecasts quality, however, (McKenzie and Gardner, 2010) argue that a damped multiplicative trend has performed well in numerous empirical studies. In addition, during experiments we found that for some time series a multiplicative trend produce better forecasts than other models, so we decided to keep some multiplicative trend models in the prediction engine.

The proposed methodology helps to forecast several products with different models, and as was expected, algorithms do not behave in the same way for all products. The modelling stage consists in fitting the best model for each product, model that outperforms the others. Based on these findings and assumptions, the following methodology was developed to implement the forecasting engine for the different products.

4.3 Methodology

The following diagram (figure 2) represents our global forecasting methodology:

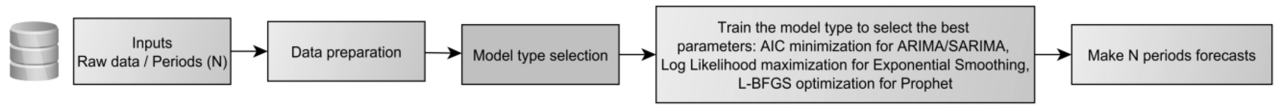

Figure 2: Global forecasting methodology

After cleaning and preparing the data, we select the most suitable model structure and fit it to the data in order to identify the best parameters. These latter are selected based on different measures for each model type such as AIC minimization for ARIMA and SARIMA models. Then, the trained model with optimized parameters will be used to forecast the N periods. The most important step is the model type selection that will be explained hereafter.

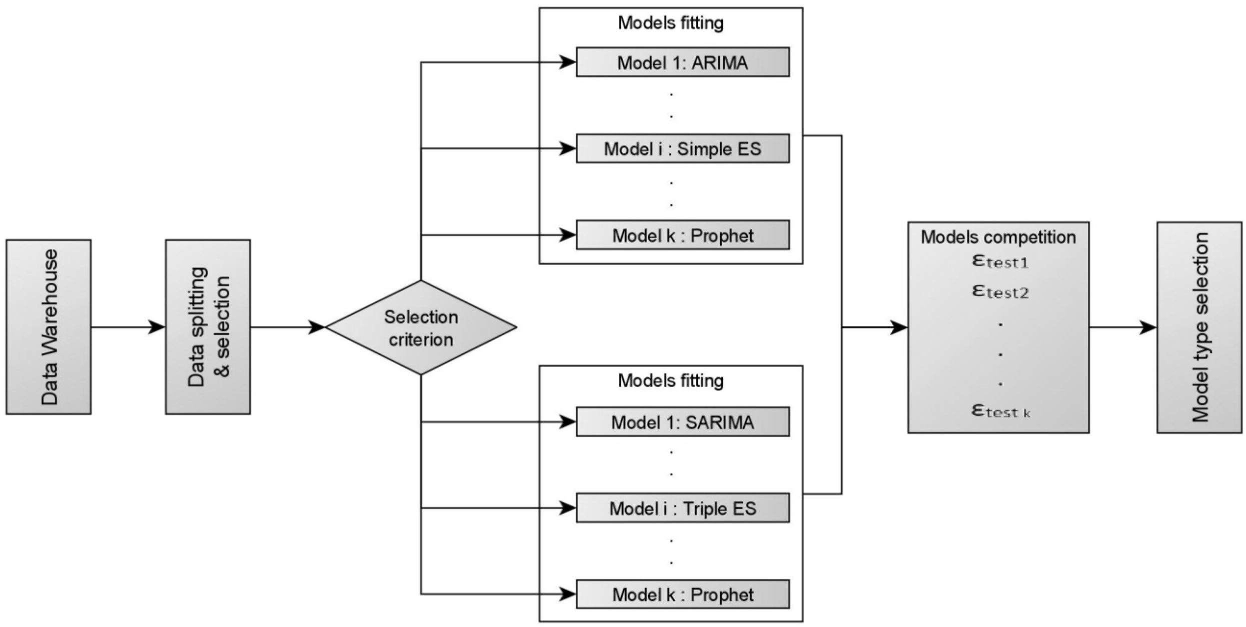

Figure 3: Model type selection methodology

Figure 3 explains our methodology for model type selection. The first phase corresponds to data splitting and preparation. As mentioned before, based on each time series observations, the appropriate group of models will be fitted on the data.

The second stage consists in the estimation of the models. For each time series, a variety of different model structures are fitted using a complete training and validation procedure. For model type, the best parameters are estimated. During the model fitting, the selection of the best ARIMA and SARIMA parameters is based on AIC measure. For exponential smoothing models, the selection of different parameters of each model is based on likelihood maximization as mentioned in the state of the art. Moreover, for Prophet, parameters tuning is based on the L-BFGS optimization.

For the next stage, the trained models are in competition during the testing phase in order to choose the best model structure with respect to the following equation:

$$Win_{model} = argmin_k\,(\varepsilon_{test\ k}) \qquad (5)$$

The error metric is used to identify the winning model. In our model selection, we used the Root Mean Squared Error (RMSE) (Chai and Draxler, 2014) in order to select the best model for a given time series. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (6)$$

Where are the observed values, the predicted values and n the number of observations.

The choice of the best model type for each product depends, actually, on the number of periods given by the user to be forecasted. In fact, if the user asks for one-month (one-step) forecasting, that means that we should choose the best model for short-term prediction. In this case, the selection procedure is done using the walk-forward (back testing) technique: we fit a model on a training data until the period t and we predict the t+1 period. Then, the predicted t+1 period will be compared to the real t+1 value from the testing set to save the error. Later, we add this real value to the training dataset and refit the model. We continue this process until we finish all the test dataset. In the case where the user asks for mid or long-term forecasts, the test is done directly overall the test set. In both cases, the model structure with the minimal RMSE will be chosen to make predictions.

The implementation of the different algorithms is done in Python using stats models library for Exponential Smoothing models, pmdarima library for ARIMA and SARIMA models and the fbprophet package for Prophet model.

5 EXPERIMENTS OF INDUSTRIAL USE CASE

5.1 Framework and experiments

These works are conducted within elm.leblanc company in their Research and Development department. To perform different experiments, we gathered sales data represented by time series of all components, products and spare parts from data warehouse. In order to simplify the context application, we decided to focus our works on a limited perimeter represented by the top sold accessories and by the wall boilers products assembled in one plant, for the French market. These boilers represents 80% of the sales, and bring the most profitability to the company. The sales are composed of three product types among them there are different product families and among every family several products considered as time series. To make forecasts, the data is split in training and testing datasets using 90% to train models and 10% to test them. In our use case study, we have monthly time series data of different natures (stationary or not) and lengths (60 months for old products to 16 months for recent ones).

The approach presented in 4 has been applied on the 316 different products using real-world data to make midterm (3 months) and short-term (1 month) predictions. Before presenting the impact of these experiments, we are interested to see how products are gathered by model type in order to more understand and categorize our time series based on forecasting models. This identification can lead to create a sort of product groups, study the similarities presented in these time series and obtain a better understanding on the product sales in order to improve forecasts.

For midterm forecasting, the table below presents the repartition of the products modelled by each model.

Models	Modelled products
Prophet	40 products
ARIMA/SARIMA	43 products
Exponential Smoothing	233 products

Table 1: Number of products for each model

We found that Prophet model was selected for 40 products which are time series representing almost a similar pattern. Figure 4 presents examples of the sales of these products.

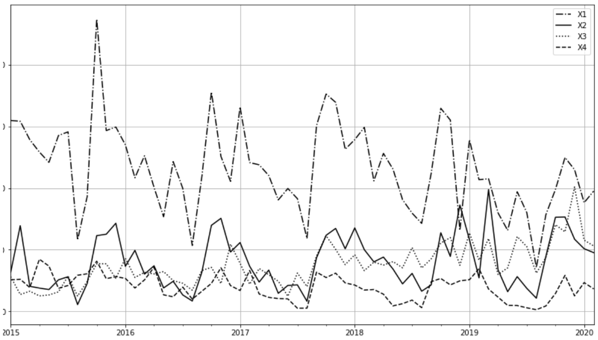

Figure 4: Examples of products modelled by Prophet.

Most of the products represent a recognizable seasonal behaviour presented by decreasing sales in August from each year and an important increase on the quantities in the end and the beginning of each year. Prophet succeeded in capturing the variability on these products better than other models. Therefore, this example supports our methodology in choosing suitable model for each time series and not to apply one model for all products.
The ARIMA model was used for 24 products representing different behaviours but we found that some products having a special behaviour is well captured by ARIMA as presented in the next figure 5.

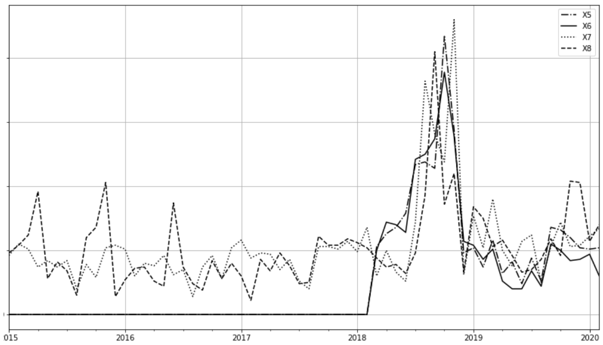

Figure 5: Examples of captured behaviour by ARIMA

We notice that this type of product presenting an important increase on sales quantities in the end of 2018 was well captured by ARIMA than by other models. Also, these time series do not represent strong seasonal patterns which justify why ARIMA performed over other seasonal models.

The most used model type were Exponential Smoothing (ES) techniques, they were selected for 233 time series, with different components. 12 Exponential Smoothing models competed to make forecasts with multitude combinations. Table 2 summarizes the repartition of the products by model.

		Seasonality		
		None	*Additive*	*Multiplicative*
Trend	*None*	-	47 pdts	-
	Additive	81 pdts	50 pdts	10 pdts
	Multiplicative	24 pdts	9 pdts	12 pdts

Table 2: Number of products for each ES model

We noticed that a model with additive trend (damped or not) was the most selected over other ES models and other models types with 141 products, which explain that most of the products have a slowly increasing trend. Some products do not present a trend, so a pure seasonal model won the competition. However, others with important trend are modelled through a multiplicative trend and a seasonal component. As we can also see, many products are modelled through an additive seasonal component, which also explains that from year to another, the seasonal pattern do not increase rapidly.

Comparing these products to others modelled by Prophet and ARIMA/SARIMA, we notice some differences between time series. Those with a clear trend but little variability were almost modelled with a double exponential smoothing. Other times series representing an important seasonality but no trend pattern was almost modelled with a simple exponential smoothing model. The figure below represents examples of products modelled with a double exponential smoothing with a multiplicative damped trend.

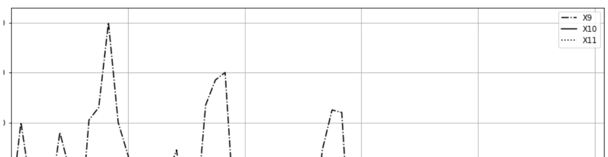

Figure 6: Examples of products with multiplicative damped trend

Through these examples, we can explain why a model can be used for a time series and not for another, which justify the importance of using several models in our sales forecasting methodology.

For short-term predictions, we applied this methodology also to predict 3 months, one month each time. The model type selection differs from month to month and we do not find the same model taxonomy as for the mid-term predictions. This fact also justifies, our interests on using different models since while we evolve on time, we get more data and the time series patterns and behaviours can change according to external factors. Therefore, this methodology permits this prediction engine to be scalable over time.

An interesting aspect also noted is that products that belong to the same product family are modelled with different methods, which prove their different natures and the necessity to be treated separately.

The next paragraph will present the impact on company sales forecasting after implementing and running the algorithms

5.2 Methodology Impact on company sales forecasting

In the following graph (figure 7), we present a sample of different products histogram of sales versus models forecasts versus sales department forecasts for the last trimester of 2019, and then we will present an example of the results on the top four sold boilers.
We note first that the actual method used in the sales department for forecasting is done manually and it is based on the intuition and the customers' demands.

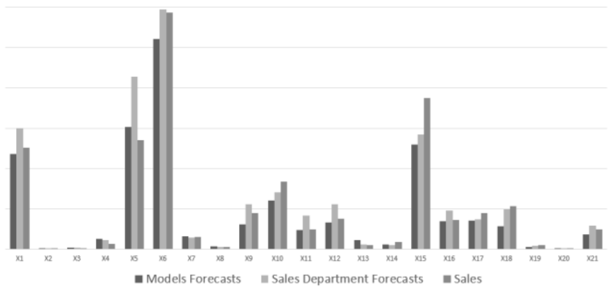

Figure 7: Sales VS Models Forecasts VS Sales Department Forecasts.

As we can notice, for several products models outperform the actual forecasts especially for products with important quantities. However, for some products models were less powerful. These products are generally recently launched, so with short time series.

In the following table 3, we present a sample of results on the top four sold boilers and we present the forecasting errors of the models and the sales department. In the last row, we present the total of absolute errors over all boilers for both horizons also.

	Last trimester 2019		**A month each time for last 3 months of 2019**	
Products	**Models Error**	**Sales Dpt Error**	**Models Error**	**Sales Dpt Error**
X5	842	1572	1076	1681
X6	871	991	1038	1449
X15	1148	910	1083	641
X1	154	489	123	600
All products	5588	6258	5852	6759

Table 3: Forecasts errors comparison for short and midterm.

We also studied the 4 boilers that were the most erroneously forecasted by the sales department and, surprisingly, they were the same products that are the most sold. As we can understand from the tables, our methodology outperformed sales department in 3 products in both horizons and sometimes with much less error. Models were less good on one product which presented huge variability over years.

As we can see, for many products, we could have an important improvement especially that among them there are "high runner" products, which means important quantities and then important cost saving specially in storage. However, the models did not well performed for some other products and this especially due to either lack of data for the recently launched ones, or for unpredictable events such as unexpected shutdowns or other external factors that may influence the sales.

The table mentioned above is an example of the results that we had and it does not represent the whole products studied. However, the overall noted improvements will be discussed in the next paragraph.

The overall improvement rate for all products in the midterm is 12 % compared to actual forecasts made by sales department. Moreover, for the short-term forecasting, we estimate an improvement rate by 15 % for all products. One can say that this represents a small improvement but these are significant improvements compared to the volume and the value of the products and the storage costs. More, these results are based only on historical data, with some lacks of data and without taking in consideration any other external or internal data. In this work, we considered to avoid the use of external data. As an example, commercials are in a direct contact with customers and can bring explanatory factors. In its industrial application, we plan to integrate this additional information in our proposed methodology in order to improve our models and our forecasts. In addition, the presented work will be improved through multiple ways that will be discussed in the perspective section.

From a managerial point of view, the sales forecasting tool has improved the efficiency of the S&OP process through enhancing the demand-planning step by bringing an additional information captured from the data. It provides a solid base to conduct discussions in this most challenging step of the process. An accurate demand plan will facilitate the development of production and replenishment planning, and the control of inventory costs.

6 CONCLUSION & PERSPECTIVES

In this paper, we made a literature review on supply chain analytics field and the potential applications of emerging technologies such as machine learning and data mining to improve supply chain processes. A specific process on which we have focused is the Sales & Operations Planning. We then proposed a way to improve the implementation of such process through the enhancing the sales prediction step. In this purpose, we studied several models used for time series forecasting and we proposed a new methodology to select the best structure model according to the type of product sales.

This is justified by the fact that even the products could be seen as similar, in a wide product catalogue, each product has specificity in terms of sales. We designed a prediction engine following the developed methodology in order to make forecasts for several products with different natures. The results seem to be very encouraging since we note a good progress in terms of global prediction accuracy compared to the existing methods. In order to reduce the scope, we performed the different experiments on specific products. These preliminary results represent a gain of performance and a success for the company but the methodology has to be extended to other products and components. This represents a challenge since there have dozens of thousands of time series to forecast.

These preliminary results can be improved through different ways. In these works, the main limitation is related to the selection of a unique "best model". As improvement, a decision procedure could be added in order to compare scores for several models. We can use a cooperation/competition of model results in order to define clusters of products that have similar sales model structures, and to provide forecasts with more accuracy. In addition, the data granularity can be increased through using weekly data instead of monthly data then aggregate the results. This may enlarge the time series length, which let us use models that are more sophisticated. Also, other external factors that may influence the sales of the products have to be studied and then multivariate time series models could be requested.

REFERENCES

Aras, S. and Kocakoç, T. D. (2016) 'A new model selection strategy in time series forecasting with artificial neural networks: IHTS', *Neurocomputing*. Elsevier, 174, pp. 974–987.

BOX, G. E. P. *et al.* (2016) *TIME SERIES ANALYSIS : Forecasting and Control*.

Chai, T. and Draxler, R. R. (2014) 'Root mean square error (RMSE) or mean absolute error (MAE)?', *Geoscientific Model Development Discussions*. Copernicus GmbH, 7(1), pp. 1525–1534.

Cheriyan, S. *et al.* (2019) 'Intelligent Sales Prediction Using Machine Learning Techniques', in *Proceedings - 2018 International Conference on Computing, Electronics and Communications Engineering, iCCECE 2018*. Institute of Electrical and Electronics Engineers Inc.

Gardner, E. S. (1985) 'Exponential smoothing: The state of the art', *Journal of Forecasting*, 4(1), pp. 1–28.

De Gooijer, J. G. and Hyndman, R. J. (2006) '25 years of time series forecasting', *International Journal of Forecasting*. Elsevier, 22(3), pp. 443–473.

Gorr, W. L., Nagin, D. and Szczypula, J. (1994) 'Comparative study of artificial neural network and statistical models for predicting student grade point averages', *International Journal of Forecasting*. Elsevier, 10(1), pp. 17–34.

Gumani, M. *et al.* (2017) 'Forecasting of sales by using fusion of machine learning techniques', in *2017 International Conference on Data Management, Analytics and Innovation, ICDMAI 2017*. Institute of Electrical and Electronics Engineers Inc., pp. 93–101.

Hastie, T., Tibshirani, R. and Friedman, J. (2008) *The Elements of Statistical Learning Data*.

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

Hyndman, R.J. Athanasopoulos, G. (2019) *Forecasting: principles and practice, 3rd edition*, *OTexts: Melbourne, Australia.*

Hyndman, R. J. *et al.* (2002) 'A state space framework for automatic forecasting using exponential smoothing methods', *International Journal of Forecasting*. Elsevier, 18(3), pp. 439–454.

Hyndman, R. J. and Kostenko, A. V. (2007) 'Minimum sample size requirements for seasonal forecasting models'.

Lapide, L. (2011) 'S&OP: The Linchpin Planning Process', *Supply Chain Management Review*.

LARSEN, K. (2015) *GAM: The Predictive Modeling Silver Bullet*.

McKenzie, E. and Gardner, E. S. (2010) 'Damped trend exponential smoothing: A modelling viewpoint', *International Journal of Forecasting*. Elsevier, 26(4), pp. 661–665.

Panigrahi, S. and Behera, H. S. (2017) 'A hybrid ETS–ANN model for time series forecasting', *Engineering Applications of Artificial Intelligence*. Elsevier Ltd, 66, pp. 49–59.

Pankaj Malhotra *et al.* (2015) 'Long Short Term Memory Networks for Anomaly Detection in Time Series', in *European Symposium on Artificial Neural Networks*.

Pavlyshenko, B. (2019) 'Machine-Learning Models for Sales Time Series Forecasting', *Data*. MDPI AG, 4(1), p. 15.

Punam, K., Pamula, R. and Jain, P. K. (2019) 'A two-level statistical model for big mart sales prediction', in *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018*. Institute of Electrical and Electronics Engineers Inc., pp. 617–620.

Robin Hyndman *et al.* (2008) *Forecasting with Exponential Smoothing - The State Space Approach*. 1st edn. Springer-Verlag Berlin Heidelberg.

Sun, Z. L. *et al.* (2008) 'Sales forecasting using extreme learning machine with applications in fashion retailing', *Decision Support Systems*. North-Holland, 46(1), pp. 411–419.

Taylor, J. W. (2003) 'Exponential smoothing with a damped multiplicative trend', *International Journal of Forecasting*. Elsevier, 19(4), pp. 715–725.

Taylor, S. J. and Letham, B. (2018) 'Forecasting at Scale', *American Statistician*. American Statistical Association, 72(1), pp. 37–45.

Thome, A. M. T. *et al.* (2012a) 'Sales and operations planning : A research synthesis', *Int . J . Production Economics*, 138, pp. 1–13.

Thome, A. M. T. *et al.* (2012b) 'Sales and operations planning and the firm performance', *International Journal of Productivity and Performance Management*.

Tiwari, S., Wee, H. M. and Daryanto, Y. (2018) 'Big data analytics in supply chain management between 2010 and 2016 : Insights to industries', *Computers & Industrial Engineering*. Elsevier, 115(October 2017), pp. 319–330.

Wallace, B. T. and Stahl, B. (2008) 'The Demand Planning Process in Executive S&OP', *The Journal of Business Forecasting*, pp. 19–24.

Waller, M. A., Fawcett, S. E. and Beane, B. (2013) 'Click Here for a Data Scientist : Big Data , Predictive Analytics , and Theory Development in the Era of a Maker Movement Supply', *Journal of Business Logistics*, 34(4), pp. 249–252.

Wang, G. *et al.* (2016) 'Big data analytics in logistics and supply chain management : Certain investigations for research and applications', *Intern. Journal of Production Economics*. Elsevier, 176, pp. 98–110.

Zhang, G. P. (2003) 'Time series forecasting using a hybrid ARIMA and neural network model', *Neurocomputing*. Elsevier, 50, pp. 159–175.