International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Machine Learning in Demand Forecasting

Trang T. M. Ho[1], Ly V. Tran[1], Hieu M. Tran[2], Son V.T. Dao[1*]
[1]School of Industrial Engineering and Management, International University, Vietnam National University, Ho Chi Minh City 700000, Vietnam
[2]School of Electrical Engineering, International University, Vietnam National University, Ho Chi Minh City 700000, Vietnam
Corresponding author: Son V.T. Dao, email: dvtson@hcmiu.edu.vn

Abstract— *In the context of intensely competitive retail industry, retail companies taking advantage of data, the role of demand forecasting is increasingly important. This paper uses machine learning to predict retail demand at 1C Company with thousands of products. The model used in this study includes traditional statistical techniques and machine learning techniques, specifically hybrid Support Vector Machine. This paper also compares and analyzes the results of the models, considering the improvement of the machine learning method compared to the traditional time series forecasting methods.*

Keywords— *Demand forecasting, multi-time series, machine learning, Support Vector Machine (SVM), supply chain.*

I. Introduction

A supply chain is defined as the entire network of organizations that flows resources, capital, and information from the manufacturer to the customer in the production process. Supply chain managers are facing with optimization problem in several terms as inventory, operation planning, procurement, sourcing, and transportation to increase their competitiveness with maximum productivity and minimum cost. In the economical context, Mentzer et al., 2001 mention that the control of products, services, information, financial resources, and demand forecasting are potential value to achieve the main goals such as customer satisfaction, profitability, value, competitive advantages [7].

Many recent studies in recent years have focused on demand forecasting because of its direct effects on supply chain productivity, especially the downstream part of supply chain. In particularly, the uncertainties of demand have the greats effects on production scheduling inventory planning, and transportation that influences supply chain performance in the real life [12]. Anticipation of customer demand helps retailers to control stock levels of inventory, production schedule, product distribution so that company may avoid negative impacts such as loss of sale revenue, decreasing service level or, excess holding cost. For this purpose, the idea of demand forecasting was motivated. In a survey paper, Seyedan and Mafakheri (2020) shows that the number of publications in demand forecasting from 2005 to 2019 is steadily increasing in Figure 1 [8].

In the context of technological innovation, the production and trading activities are easily created by various technological approaches. Therefore, the large scale of dataset in supply chain has increasing every day including the information of customers, sales, stores, orders, and product. Many supply chain managers are dealing with the problem of using the huge dataset that call big data in demand forecasting that cannot be captured by conventional techniques. Big data is described as a massive amount of unstructured and structured data that grows at an exponential rate [10]. Structured data refers to databases that are organized in rows or columns so that all entities are easy to collect, whereas unstructured data refers to data from various sources that are difficult to categorize, such as videos, photos, webpages, blogs, etc. Conventional forecasting techniques are simple and easy to approach but they highly rely on domain knowledge and skill of planner to choose the appropriate techniques and parameters in a particular context. In the data-driven context, there is a significant advantage of advanced techniques in supply chain demand forecasting which are big data analytics/machine learning techniques in comparison to conventional approaches [4]. One way to take advantage of big data and incorporate non-linear models is to use machine learning techniques to predict customer demand.

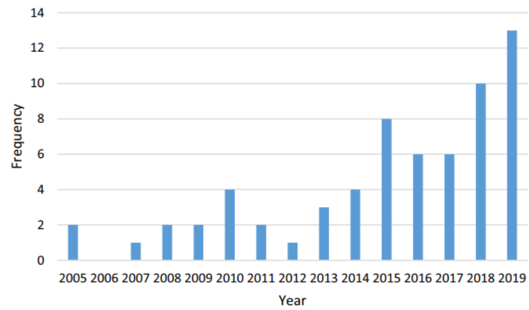

Fig. 1. Distribution of literature in supply chain demand forecasting from 2005 to 2019.

A. *Case study: 1C company*

1C Company was founded in 1991 that specialize in development, distribution, licenses, computer software, video games and support related services with the headquarters in Moscow, Russia. The company is considered as one of the largest companies in business software in their country. At present, there are more than 1200 staffs and 10millions business partners at 25 countries in the world. Consequently, the company is handling a huge dataset of customers, stores, and products. In this thesis project, customer demand of the company will be focused to predict the total sales for every product so that the company can control the production and stock schedule and satisfy their value customers.

B. *Objectives of study*

According to the case study, the aim of this paper is to solve the problem of demand forecasting with multi-time series dataset. In particular, the method is using traditional techniques and machine learning technique which is hybrid Support Vector

Trang T. M. Ho, Ly V. Tran, Hieu M. Tran, and Son V.T. Dao, "Machine Learning in Demand Forecasting," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 3, pp. 225-233, 2022.

Machine to anticipate the total sales for every item and store in the next month based on previous devised model and suitable approaches. The result of hybrid SVM model is compared to the traditional models.

C. *Scope and Limitations*

The scope of this work focus on predicting total sales of every items of whole based on traditional and machine learning approach. The methods will be discussed in detail in the literature review section. The data fields consists the lists of items, shops, and item categories combined with their ID, names, prices, date and the number of product sold per day which is also the measure expected to be predicted monthly. The result of hybrid SVM techniques application is analyzed and compared to traditional time series forecasting techniques.

There are some limitations of the dataset and forecasting technique. Total amount of sales product will be anticipated based on the historical data from January 2013 to October 2015. The prediction is executed by employing an approach machine learning technique as hybrid Support Vector Machine (or one more advance technique will be applied if I have enough time).

II. Literature Review

In the literature of demand forecasting, the method is classified into two main assemblages: qualitative forecasts based on forecasts of grassroots, market research, and expert estimation; while quantitative forecast based on casual demand and time-series forecast [1]. Time series methods are based on how demand changes over time and look for trends in the data. Regarding to quantitative forecast focused on this study, there are several traditional forecasting methods which are simple and easy to understand can be approached, such as naïve forecast, moving average, average, trend, and multiple linear regression.

One of the most basic forecasting methods is Naïve Forecast which is commonly used to compare to other approaches as a base classifier [2]. Naïve forecast used to predict the future value by the most recent value of the interest data. The next technique is moving average in which planners take the average of previous numbers in periods to forecast the future demand with long term or short term [2]. Trend forecasting use time as an independent variable to forecast demand based on a simple regression model [2]. Linear regression models use a number of demand changes in the past as predictor variables [2].

Furthermore, Chusyairi et al. stated that the Simple Exponential Smoothing method can capture the data with no trend or seasonal patterns and Holt's linear method can capture the seasonal data [13]. However, these two models are not able to properly handle the seasonal data with their own. Holt-Winter approach was early extended to directly support the seasonal data [6]. Additionally, there was a number of attractive features such as: it was simple and quick to compute by programming; the storage data needed is lesser and the weight of old data is decreasing. Holt-Winter was considered as the better method with respect to Exponential Smoothing and Holt's linear model [3]. To compare with Exponential Smoothing (ES), there is no covariates employed in the model assuming that such data are self-explanatory. For the significantly nonlinear problem having uncertain functional connection, ES models are not able to handle the issue [3]. Therefore, choosing suitable techniques is challenges to fit the data.

ARIMA models was early developed in 1970's by Box and Jenkins [14]. Autoregressive (AR) models and Moving Average models might be effectively linked together to produce the Autoregressive Moving Average (ARMA) models. This model is a general and useful class of time series models and they can be extended to non-stationary series which called Autoregressive Integrated Moving Average (ARIMA) [3]. However, using an ARIMA model for the real out-of-sample forecasting might not be accurate as the forecasting intervals because of few reasons. According to Makridakis et al. (1998) [6] accounting the uncertain parameter estimations has been ignored, then the mathematical formulas is too complicated to allow the additional uncertainty. Additionally, there are several assumptions in an ARIMA model that may not be validated like samples of historical data might be stationary during the forecasting period [6].

In term of machine learning methods, there are five most frequently techniques used in supply chain demand forecasting as "Neural Network," "Regression", "Time-series forecasting (ARIMA)", "Support Vector Machine", and "Decision Tree" methods [8]. The forecasting error of conventional models is not easy to be lower than 20%, while advanced machine learning techniques have proven better accuracy within their specific context. The literature review of this study focuses on Support Vector Machine technique. The advantages of this method will be discussed in the next section.

TABLE I. The five most commonly machine learning techniques used in supply chain demand forecasting from 2005 to 2019 (Seyedan and Mafakheri, 2020).

Rank	Technique	Frequency
1	Neural networks	30
2	Regression	27
3	Time-series forcasting (ARIMA)	13
4	Support vector machine	8
5	Decision tree	8

A. *Support vector machines*

In the literature of machine learning, Seyedan (2020) [8] depicted support vector machine (SVM) as an algorithm that a training dataset can be converted into data classes using nonlinear mapping. SVM is early applied for supply chain demand forecasting by Carbonneau, 2008 who mentioned that SVM is important in producing a higher dimension in order to maximize the margins among classes or minimize the error [2]. While Artificial Neural Network (ANN) models comply to the principle of empirical risk minimization, SVM models focus on reducing the upper bound of generalization error instead of the training error [9]. SVM models are useful for forecasting problem by converting complex nonlinear to linear regression problems in high dimensional feature space [5]. SVM models have been expanded to handle regression issues and have become another significant tool for predicting difficulties with the development of the ε-insensitive loss function [11].

Trang T. M. Ho, Ly V. Tran, Hieu M. Tran, and Son V.T. Dao, "Machine Learning in Demand Forecasting," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 3, pp. 225-233, 2022.

Lu et al. employed SVM to predict the sales of computer products based on weekly sales data and the error were 4.09%-8.62% [5]. In addition, many studies have found that SVMs models performed better than ANN models and other models based on different kernel functions [10]. SVM models was also discovered that their potential capability can be used to develop hybrid forecasting models with statistical-based models to outperform other models [14]. Zhu et al., (2019) proposed a hybrid Holt-Winter and SVM model to forecast onion seed demand that performed better than statistical-based method and other machine learning methods [9].

III. Methodology

A. Approaches Comparison and Selection:

Demand forecasting

Generally, the demand forecasting method are grouped into two main categories which are qualitative and quantitative. The qualitative predictions focus on qualitative data that the predictions base on expert judgments or awareness of specific occurrences. In contrast, the quantitative forecasts base on the changes of date that occur over time which called time-series.

In term of these two main categories, qualitative method includes three forecasting techniques such as participant observation (grassroots), market research, and expert judgments, while quantitative forecasting methods are time series forecasts and Causal demand forecasts. In this research, the time series forecast in quantitative method is selected to predict the future sales based on problem identification and historical sale data of the 1C Company.

Holt-Winter model

The Holt-Winter (HW) model are classified by two different methods based on the way to model the seasonality that are additive and multiplicative [9]. They developed the model based on three smoothing formulas for trend, level, and seasonality as follow:

Level l_t: $$l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (1)$$

Trend b_t: $$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \quad (2)$$

Seasonality s_t: $$s_t = \frac{\gamma y_t}{l_{t-1}+b_{t-1}} + (1-\gamma)s_{t-m} \quad (3)$$

Forecast: $$\hat{y}_{t+h/t} = (l_t + b_t h)s_{t-m+h_m^+} \quad (4)$$

where:

m= length of seasonality (i.e, number or month or quarter in year)

l_t= the level of the series

b_t= the trend

s_t= seasonal component

$\hat{y}_{t+h/t}$= the forecast for h period ahead

$h_m^+ = [(h-1) \bmod m]$+1

$\alpha, \beta^*, \gamma =$ Parameters are usually restricted between [0,1]

Autoregressive Integrated Moving Average (ARIMA)

ARIMA (p, d, q) is known as the general nonlinear seasonal model, where: p is the order of the autoregressive (AR) part, d presents the degree of differencing involved, q is order of the moving average (MA) part [6]. The simplest equation of the models is ARIMA (1, 1, 1) as follows:

$$(1-\phi_1 B)(1-B)Y_t = c + (1-\phi_1 B)e_t \quad (5)$$

where:

$(1-\phi_1 B)$ is the AR(1) part,

$(1-B)Y_t$ is the first difference

$c + (1-\phi_1 B)e_t$ is the MA(1) part.

Support Vector Machine

SVM models, as discussed in the previous section, are based on the structural risk minimization concept and seek to reduce the upper bound of the generalization error instead of the training error by converting complex nonlinear to linear regression problems in high dimensional feature space [5].

The regression function: $y = \omega\varphi(X) + b$ (6)

where: $\varphi(X)$ the feature of data that are non-linearly mapped into space X

ω and b are the coefficients estimated by minimizing $R(C)$ function (the risk function):

$$Minimize\ R(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2}\|\omega\|^2 \quad (7)$$

$$s.t\ L_\varepsilon(d,y) = \{|d-y| - \varepsilon,\ \ |d-y|\ \varepsilon\ 0,\ \ otherwise \quad (8)$$

where:

d_i is actual demand of period i

N is entire data length

C is the regularized constant that defines the trade-off between the empirical error and the regularization term

$L_\varepsilon(d,y)$ is called the ε-intensive loss function

The empirical error is represented by the objective in Eq (7), while the function flatness is represented by the constraint Eq (8).

The problem can be expressed by introducing Lagrange multipliers, $\alpha_i, \alpha_i^*(\alpha_i\alpha_i^* = 0, \alpha_i, \alpha_i^* > 0)$, and letting the partial derivatives of $\omega, b, \zeta_i^* = 0$, the problem might be presented as follow:

$$Maximize\ \ R(\alpha_i - \alpha_i^*) = \sum_{i=1}^{N} d_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*).K(X_iX_j)$$

$$s.t.\ \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0; \quad (10)$$

$$\alpha_i, \alpha_i^* \in [0, C], \alpha_i, \alpha_i^* = 0;\quad i = 1,2,\dots,N.$$

The regression function from Eq (6) above can be written as [9]:

$$y = f(X, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*).K(X_iX_j) + b \quad (11)$$

Where, $K(X_iX_j) = \varphi(X_i)\varphi(X_j)$ is the kernel function. Lu et al., 2014 noted that any function that meets Mercer's condition might be employ as the kernel function [5].

Hybrid Support Vector Machine

Zhu et al. (2019) recently proposed a potential forecasting technique using hybrid statistical-based model and machine learning model, namely Holt-Winter model and SVM model [9]. In this research, this hybrid model is applied to forecast the future sales of 1C Company. Figure 2 illustrate the process of employing the hybrid model. The training set of data is inputted into HW model to achieve the statistical forecasting result that return the linearity and seasonal trends of historical data. Then, the SVM model will be prepared to input the result of HW model as variables. The key point of applying hybrid SVM model is the result of HW model [9]. Seasonal trends in sales data are learnt in the SVM model using the RBF kernel, which would be identified as outliers in statistical-based models. As a result, the RBF kernel is utilized to get predicting values during training [9]

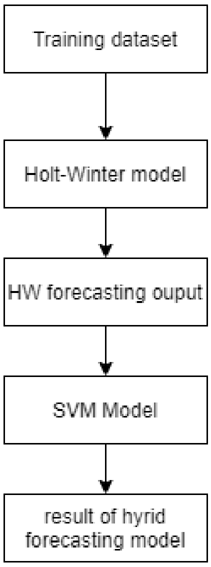

Fig. 2. The flow chart of hybrid SVM methodology.

B. Conceptual design

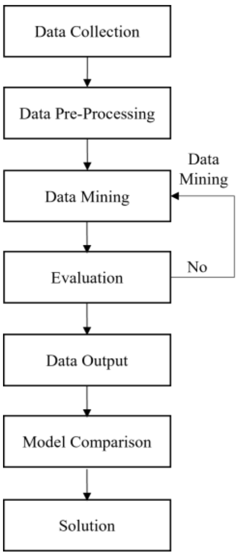

Fig. 3. The research process of the thesis.

Evaluation metrics

The performance of hybrid model is evaluated by comparing with statistical based-line models as Moving Average, HW, and ARIMA with respect to their results. In this study, there are three error measurements used to evaluate forecasting performance. That are mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE) and mean absolute percentage error (MAPE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - A_i|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2$$

$$RMSE = \sqrt{MSE}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{F_i - A_i}{A_i}\right| \times 100\%$$

Where: F_(i)and A_(i) are the forecasting values and actual demand values.

To compare the future and the actual value, the root mean square error (RMSE) is better to mean square error (MSE), since it measured in the same units as the data. MAE is the same units as the data too, is usually smaller than RMSE. MAPE is genetic percentage term which is guaranteed to be strictly positive. MAE and MSE metrics can be used to compare different forecasting methods with the same data set, while MAPE is used to compare different forecasting method with different data sets [9]

IV. Modelling

A. Defining problem

In the case of 1C Company, the problem is forecasting the total sales of next month based on a multi-time series historical dataset of more than twenty-one thousand items in 60 different shops. The list of items and shops slightly changed every month. The data explosion is a big challenge for the operation and planning department of the company. In such situation, a suitable model forecasting technique need to be applied to handle this problem and improve forecasting accuracy.

B. Data collection

The dataset was collected on website Kaggle.com. The online community was established for people who interest in data scientists and machine learning. They allows users to collect and publish datasets. The dataset of 1C Company has been chosen as the suitable data for the research topic of demand forecasting in supply chain problem of the thesis.

The dataset include historical sale data of more than 21807 items in 60 different shops from 2013 [16].

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day
0	02.01.2013	0	59	22154	999.00	1.0
1	03.01.2013	0	25	2552	899.00	1.0
2	05.01.2013	0	25	2552	899.00	-1.0
3	06.01.2013	0	25	2554	1709.05	1.0
4	15.01.2013	0	25	2555	1099.00	1.0
...	...	...	...	...	...	...
2935844	10.10.2015	33	25	7409	299.00	1.0
2935845	09.10.2015	33	25	7460	299.00	1.0
2935846	14.10.2015	33	25	7459	349.00	1.0
2935847	22.10.2015	33	25	7440	299.00	1.0
2935848	03.10.2015	33	25	7460	299.00	1.0

2935849 rows × 6 columns

Fig. 4. The list of index and description in the dataset.

Trang T. M. Ho, Ly V. Tran, Hieu M. Tran, and Son V.T. Dao, "Machine Learning in Demand Forecasting," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 3, pp. 225-233, 2022.

TABLE II. List of features in the dataset

Index	Descriptions
ID	An Id that represents a (Shop, Item) tuple within the test set
shop_id	Unique identifier of a shop
item_id	Unique identifier of a product
item_category_id	Unique identifier of item category
item_cnt_day	Number of products sold. A monthly amount of this measure will be predicted.
item_price	Current price of an item
date	Date in format dd/mm/yyyy
date_block_num	A consecutive month number, used for convenience. January 2013 is 0, February 2013 is 1,..., October 2015 is 33
item_name	Name of item
shop_name	Name of shop
item_category_name	Name of item category

C. *Data pre-processing*

Read and analyze data

This initial step is an important step to have an overview of data and analyzes main features of the dataset before developing any forecasting model. The reading data step is to see the data characteristics to analyze and visualize. There are some time series will be chosen to input to model for developing before input the whole dataset. There are some characteristics as:

- Size of data frame: 2935849 rows and 6 columns
- The head of data:
- Number of shop and item
- The top shop (with the greatest number of sold items): shop_id 31
- The top item (the item was sold the most in all shops): item_id 20949

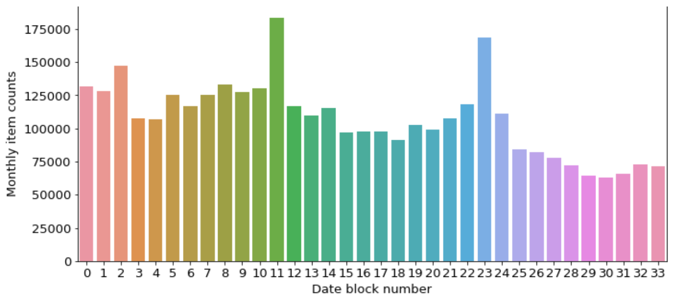

Fig. 5. Historical data of items sold by month.

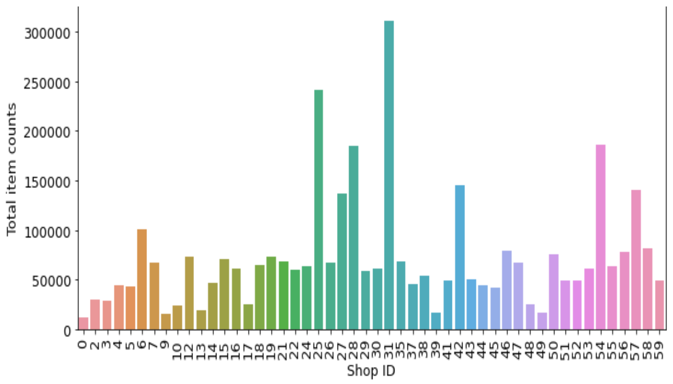

Fig. 6. Historical of total items sold by shops ID

Pre-process data

This step is to solve some problem of data as missing value and data framework. The missing value should be fill by zero to avoid the problem when implementing the model by coding. The data should be rearranged to see the number of items sold per month of every item in a shop

shop_id	item_id \ date_block_num	0	1	2	3	4	5	6	7	8	9	...	24	25	26	27	28
0	1830	14.0	8.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
	2445	NaN	25.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
	2753	NaN	24.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
	2808	9.0	10.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
	5822	9.0	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
59	16790	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2.0	2.0	NaN	NaN	NaN
	17717	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	27.0	22.0	26.0	19.0	13.0
	20949	NaN	NaN	NaN	25.0	116.0	123.0	109.0	94.0	122.0	106.0	...	49.0	76.0	68.0	48.0	32.0
	21376	NaN	NaN	NaN	26.0	12.0	3.0	1.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
	22088	5.0	NaN	5.0	3.0	6.0	8.0	7.0	4.0	4.0	NaN	...	6.0	3.0	4.0	3.0	3.0

Fig. 7. The data before filling the missing value

shop_id	item_id \ date_block_num	0	1	2	3	4	5	6	7	8	9	...	24	25	26	27	28
31	26	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
	27	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
	28	3.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0
	29	2.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
	30	0.0	112.0	65.0	13.0	10.0	3.0	4.0	1.0	4.0	2.0	...	1.0	1.0	1.0	0.0	0.0
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22160	1.0	2.0	2.0	0.0	1.0	2.0	0.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0
	22162	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	32.0	25.0	7.0	2.0
	22163	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
	22164	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	24.0	8.0	2.0	3.0
	22167	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	11.0	...	1.0	1.0	2.0	2.0	2.0

Fig. 8. The data after filling the missing value

The figure below illustrates the distribution of {shop_id, item_id} (or time series) which have no item was sold in a month. The horizontal axis describes number of months that a time series has zero sale, the vertical axis describes the percentage of these types of time series that occupies the entire sale data. The graph shows that most of products in shops were not sold in many months, and there was a small part of {shop_id, item_id} sold with high number of months of sales. Each time series has 34 periods, if the time series have less than 15 months with zero sale are chosen, the percentage of time series satisfy this condition is only 1.3% in quantity of (shop_id, item_id} (namely 5811 time series). However, these 5811 time series occupies 21% of entire products sold of the company.

This study chooses these 5811 {shop_id,item_id} as the sample set to execute the predictions.

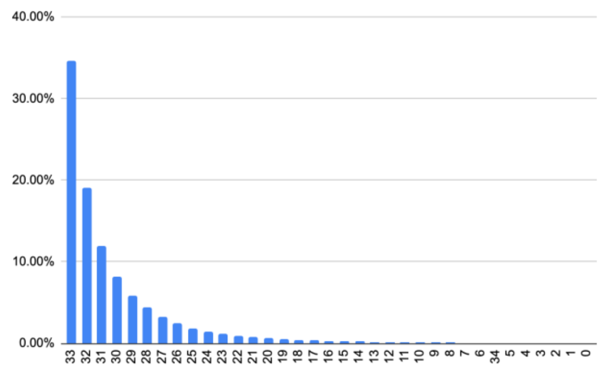

Fig. 9. The distribution plot of {shop_id, item_id} that no item sold in a month

Trang T. M. Ho, Ly V. Tran, Hieu M. Tran, and Son V.T. Dao, "Machine Learning in Demand Forecasting," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 3, pp. 225-233, 2022.

D. Forecasting method

Holt Winter Forecasting Model

The HW model are classified by two different methods based on the way to model the seasonality that are additive and multiplicative (Winter, 1960). Winter (1960) developed the model based on three smoothing formulas for trend, level, and seasonality as follow:

Level l_t: $$l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (12)$$

Trend b_t: $$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \quad (13)$$

Seasonality s_t: $$s_t = \frac{\gamma y_t}{l_{t-1}+b_{t-1}} + (1-\gamma)s_{t-m} \quad (14)$$

Forecast: $$\hat{y}_{t+h/t} = (l_t + b_t h)s_{t-m+h_m^+} \quad (15)$$

where:

m= length of seasonality (i.e, number or month or quarter in year)

l_t= the level of the series

b_t= the trend

s_t= seasonal component

$\hat{y}_{t+h/t}$= the forecast for h period ahead

$h_m^+ = [(h-1) \bmod m]+1$

α, β^*, γ = Parameters are usually restricted between [0,1]

ARIMA Forecasting Model

ARIMA (p, d, q) is known as the general nonlinear seasonal model, where: p is the order of the autoregressive (AR) part, d presents the degree of differencing involved, q is order of the moving average (MA) part [6]. The simplest equation of the models is ARIMA (1, 1, 1) as follows:

$$(1-\phi_1 B)(1-B)Y_t = c + (1-\phi_1 B)e_t \quad (16)$$

where:

$(1-\phi_1 B)$ is the AR(1) part,

$(1-B)Y_t$ is the first difference,

and $c + (1-\phi_1 B)e_t$ is the MA(1) part.

Support Vector Machine Model

The regression function: $y = \omega\varphi(X) + b \quad (17)$

where:

$\varphi(X)$ the feature of data that are non-linearly mapped into space X

ω and b are the coefficients estimated by minimizing $R(C)$ function (the risk function):

$$Minimize\ R(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2}\|\omega\|^2 \quad (18)$$

$$s.t\ L_\varepsilon(d,y) = \{|d-y| - \varepsilon,\ |d-y| \geq \varepsilon\ 0,\ otherwise \quad (19)$$

where:

d_i is actual demand of period i

N is entire data length

C is the regularized constant that defines the trade-off between the empirical error and the regularization term

$L_\varepsilon(d,y)$ is called the ε-intensive loss function

The empirical error is represented by the objective in Equation (18), while the function flatness is represented by the constraint Equation (19).

The problem can be expressed by introducing Lagrange multipliers, $\alpha_i, \alpha_i^*(\alpha_i\alpha_i^* = 0, \alpha_i, \alpha_i^* > 0)$, and letting the partial derivatives of $\omega, b, \zeta_i^* = 0$, the problem might be presented as follow:

$$Maximize:$$

$$R(\alpha_i - \alpha_i^*) = \sum_{i=1}^{N} d_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) \ldots$$
$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*).K(X_iX_j) \quad (20)$$

$$s.t.\ \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0; \quad (21)$$
$$\alpha_i, \alpha_i^* \in [0, C], \alpha_i, \alpha_i^* = 0;$$
$$i = 1,2,\ldots,N.$$

The regression function from Equation (17) above can be written as [9]:

$$y = f(X, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*).K(X_iX_j) + b \quad (22)$$

where: $K(X_iX_j) = \varphi(X_i)\varphi(X_j)$ is the kernel function. Lu et al., 2014 noted that any function that meets Mercer's condition might be employ as the kernel function [9].

Evaluation metrics

The performance of hybrid model is evaluated by comparing with statistical based-line models as Moving Average, HW, and ARIMA with respect to their results. In this study, there are three error measurements used to evaluate forecasting performance. That are mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE) and mean absolute percentage error (MAPE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - A_i| \quad (23)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2 \quad (24)$$

$$RMSE = \sqrt{MSE} \quad (25)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{F_i - A_i}{A_i}\right| \times 100\% \quad (26)$$

V. Results

A. Results of traditional forecasting method

Holt-Winters

For the Holt-Winters model, both multiplicative and additive methods were applied to forecast the future sales. However, only forecasting result of additive method give small accuracy errors with four evaluation metrics which are showed in the table below. The additive method is suitable for trend and seasonality which does not change over time, while the multiplicative method works best with data that has a trend and seasonality that grows over time. In the current research, the additive model works best with more than five thousand historical time series of 1C Company which are not increase over time. The figures below show the sample forecasting result of some time series in the data set.

TABLE III. Performance metrics of Holt-Winters Model.

Holt-Winters Model	Performance metrics			
	MSE	RMSE	MAE	MAPE
Additive model	291.05	17.06	3.55	223.856
Multiplicative model	1.13736E+22	1.06647E+11	1352662333	1.22222E+11

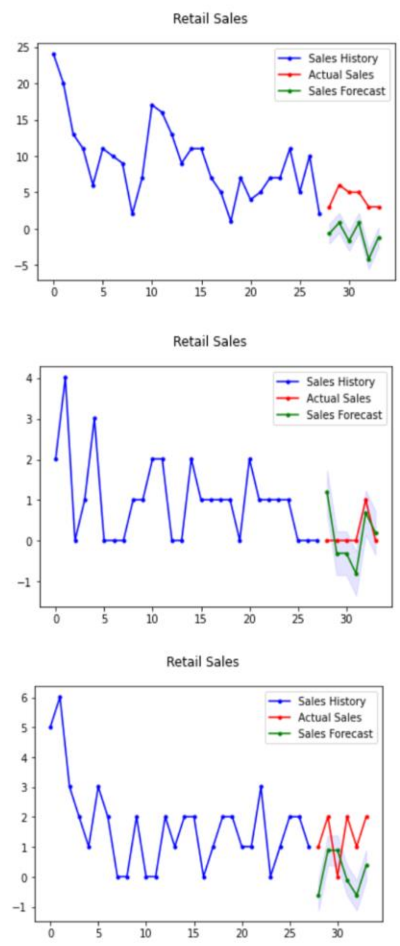

Fig. 10. Some forecasting result sample of Holt-Winters Model.

Auto Regressive Integrated Forecasting Average

In this session, ARIMA model was executed using a python library (statsmodels.tsa.arima.model.ARIMA) to approach the forecasting output of test set, the sample result of a time series achieved by ARIMA(1,0,1) as bellow.

After the training test was inputted to four ARIMA (1, 1, 1), ARIMA (1, 0, 1), ARIMA (2, 0, 2), ARIMA (2, 1, 2) which are identified in the modelling chapter, the forecasting was executed to obtain four error measurements.

TABLE IV. Performance metrics of four ARIMA models.

Forecasting model	Performance metrics			
	MSE	**RMSE**	**MAE**	**MAPE**
ARIMA(1,0,1)	153.836	12.403	1.647	99.732
ARIMA(1,1,1)	164.957	12.844	1.987	132.449
ARIMA(2,1,2)	777.275	27.880	1.946	121.682
ARIMA(2,0,2)	81462.388	285.416	4.383	251.596

The forecasting result shows that the result of ARIMA(1,0,1) model is over the three model such as ARIMA(1,1,1), ARIMA(2,1,2), ARIMA(2,0,2) for all four performance standards. The second ranking is result of ARIMA(1,1,1) whose all good result, while the two remaining model with number of lag are 2 that ARIMA(2,1,2), ARIMA(2,02) show the clearly not good results.

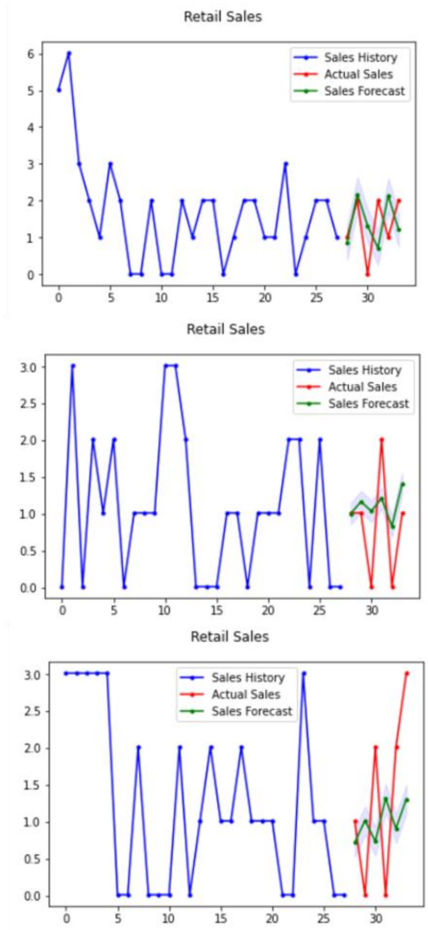

Fig. 11. Some forecasting result sample of ARIMA.

B. Results of machine learning forecasting method

Testing sample forecasting result of hybrid SVM – Holt Winters model

A machine learning method is defined as a mathematical which has some parameters that must be learned from the dataset. However, there are some hyper-parameters that cannot be learnt immediately. These hyper-parameters are usually selected based on some judgment or evaluation of researchers for testing before employing the actual training. These parameters demonstrate their significance by enhancing the model's performance, such as its complexity or learning rate. The model might include a large number of hyper-parameters (parameters), and determining the optimal combination of parameters can be regarded a search issue.

SVM also includes several hyper-parameters to employ (such as C or gamma values), and determining the best hyper-parameter is a complex undertaking. However, it may be discovered by attempting all combinations and determining which one works best. The basic concept is to generate a grid of hyper-parameters and just attempt all their combinations. GridSearchCV accepts a dictionary detailing the parameters that can be used to train a model. The parameter grid is defined as a dictionary, with keys representing parameters and values representing settings to be verified. The GridsearchCV method of the Scikitlearn library was used in the study to automatically discover the best kernel with C and gamma values (Specifically,

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

the options including: C values 1e0, 1e1, 1e2, 1e3; Gamma values: 1.e-02, 1.e-01, 1.e+00, 1.e+01, 1.e+02; Kernel: RBF kernel and Polynomial kernel)

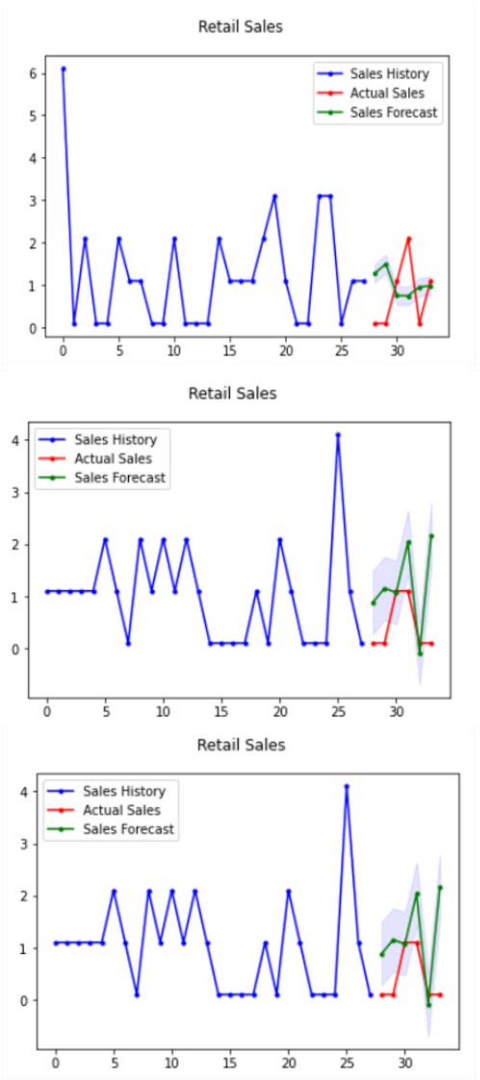

Fig. 12. Some forecasting result sample of hybrid SVM-Holt Winters model.

TABLE V. Performance metrics of hybrid SVM-Holt Winters model.

Forecasting model	Performance metrics			
	MSE	RMSE	MAE	MAPE
Hybrid SVM-HW model	111.225	10.546	2.403	145.457

The hybrid SVM – Holt Winters gives satisfactory result with respect to independent Holt Winters model and ARIMA models presented in previous sections. With RMSE evaluation metric, the root mean square error of Hybrid SVM – Holt Winters is better than the error of ARIMA(1,0,1) model (RMSE of ARIMA(1,0,1) is 12,403). With respect to MAE and MAPE metric, the result of ARIMA(1,0,1) is better.

However, with comparing to the MAE metric, the RMSE metric penalizes the predicted values with large errors that exceed the actual values. Therefore, the hybrid SVM – Holt Winters models has the better result for these cases.

Testing sample forecasting result of hybrid SVM –ARIMA model.

TABLE VI. Performance metrics of hybrid SVM-ARIMA model

Forecasting model	Performance metrics			
	MSE	RMSE	MAE	MAPE
SVM - ARIMA(1,0,1)	109.246	10.452	2.395	145.036
SVM - ARIMA(1,1,1)	109.349	10.457	2.398	145.268

Similar to independent ARIMA model, the hybrid SVM – ARIMA (1,0,1) model shows better results that the hybrid SVM – ARIMA(1,1,1) model. Hybrid SVM – ARIMA(1,0,1) model result is also better than the hybrid SVM – Holt Winter model result for all four evaluation metrics as MSE, RMSE, MAE, and MAPE. For forecast demand sample size at 1C enterprise, the study finds evidence that using machine learning will perform better on RMSE, MSE metrics than traditional ARIMA model.

TABLE VII. Forecasting accuracy of all testing model

Forecasting model	Performance metrics			
	MSE	RMSE	MAE	MAPE
Hybrid SVM - ARIMA(1,0,1)	**109.246**	**10.452**	2.395	145.036
Hybrid SVM - ARIMA(1,1,1)	109.349	10.457	2.398	145.268
Hybrid SVM - HW	111.225	10.546	2.403	145.457
ARIMA(1,0,1)	153.836	12.403	**1.647**	**99.732**
ARIMA(1,1,1)	164.957	12.844	1.987	132.449
Holt-Winters Model	291.05	17.06	3.55	223.856
ARIMA(2,1,2)	777.275	27.880	1.946	121.682
ARIMA(2,0,2)	81462.388	285.416	4.383	251.596

VI. CONCLUSIONS

The time series forecast in quantitative method is selected to predict the future sales based on problem identification and historical sale data of the 1C Company. With regard to the independent Holt Winters model and ARIMA models, the hybrid SVM-Holt Winters provides good results. With The RMSE measure, in contrast to the MAE metric, penalizes the projected values with significant errors that are greater than the actual values. As a result, in these circumstances, the hybrid SVM-Holt Winters models produce better results. The hybrid SVM-ARIMA (1,0,1) model performs better than the hybrid SVM-ARIMA(1,1,1) model, similar to the independent ARIMA model. For all four assessment metrics—MSE, RMSE, MAE, and MAPE—the hybrid SVM-ARIMA(1,0,1) model result is also superior than the hybrid SVM-Holt Winter model result. The study discovers evidence that machine learning would perform better on RMSE, MSE metrics for predict demand sample size at 1C enterprise than traditional ARIMA model.

ACKNOWLEDGEMENT

This research is funded by International University, VNU-HCM under grant number T2021-02-IEM.

REFERENCES

[1] Chambers, J.C., Mullick, S.K., & Smith, D.D. (1971). How to Choose the Right ForecastingTechnique. *Harvard Business Review*, 45(1), 45-71.

[2] Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research, 184(3), 1140–1154.*
[3] Da Veiga, C. P., da Veiga, C. R. P., Puchalski, W., dos Santos Coelho, L., & Tortato, U. (2016). Demand forecasting based on natural computing approaches applied to the foodstuff retail segment. *Journal of Retailing and Consumer Services*, *31*, 174-181.
[4] Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. *The International Journal of Logistics Management*, *29(*2), 739–766.
[5] Lu, C. J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing*, 128, 491-499.
[6] Makridakis, S., Wheelwright, S.C., & Hyndman, R.J. (2008). *Forecasting Methods and Applications, Third Edition.* John Wiley & Sons.
[7] Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, 22(2), 1–25.
[8] Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Journal of Big Data, 7(1).*
[9] Zhu, Y., Zhao, Y., Zhang, J., Geng, N., & Huang, D. (2019). Spring onion seed demand forecasting using a hybrid Holt-Winters and support vector machine model. *PloS one*, *14*(7), e0219889.
[10] Zillner, S., Lasierra, N., Faix, W. and Neururer, S. (2014). User needs and requirements analysis for big data healthcare applications. *Proceeding of the 25th European Medical Informatics Conference (MIE 2014).* Istanbul, Turkey.
[11] Müller, K. R., Smola, A. J., Rätsch, G., Schökopf, B., Kohlmorgen, J., & Vapnik, V. (1999). Using support vector machines for time series prediction. *In Advances in kernel methods: support vector learning* (pp. 243-253).
[12] Tosarkani BM, Amin SH. A possibilistic solution to configure a battery closed-loop supply chain: multi-objective approach. *Expert Syst Appl. 2018;92*:12–26.
[14] A. Chusyairi, N. S. P. Ramadar and Bagio, "The use of exponential smoothing method to predict missing service e-report," *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2017, pp. 39-44, doi: 10.1109/ICITISEE.2017.8285535.
[15] Harvey, A.C. (1990). ARIMA Models. In: Eatwell, J., Milgate, M., Newman, P. (eds) Time Series and Statistics. The New Palgrave. Palgrave Macmillan, London. https://doi.org/10.1007/978-1-349-20865-4_2
[16] https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data

Trang T. M. Ho, Ly V. Tran, Hieu M. Tran, and Son V.T. Dao, "Machine Learning in Demand Forecasting," *International Research Journal of Advanced Engineering and Science*, Volume 7, Issue 3, pp. 225-233, 2022.