Content list available at JournalPark

Turkish Journal of Forecasting

Journal Homepage: tjforecast.xyz

A Poisson-Regression, Support Vector Machine and Grey Prediction Based Combined Forecasting Model Proposal: A Case Study in Distribution Business

F. Yiğit[1,*], Ş. Esnaf[2], B. Yalcin Kavus[3]

[1]*İstanbul Arel University, Faculty of Engineering and Architecture, Department of Industrial Engineering, İstanbul, Turkey*
[2]*İstanbul University-Cerrahpaşa, Faculty of Engineering, Department of Industrial Engineering, İstanbul, Turkey*
[3]*Nişantaşı University, Faculty of Engineering and Architecture, Department of Industrial Engineering, İstanbul, Turkey*

A R T I C L E I N F O

Article history:

Received	25	June	2021
Revision	09	August	2021
Accepted	06	September	2021
Available online	26	December	2021

Keywords:

Grey prediction
Poisson-Regression
Support vector machine
Combined forecasting
Sales demand forecasting

R E S E A R C H A R T I C L E

A B S T R A C T

Demand forecasting is a complicated task due to incomplete data and unpredictability. Accurate demand forecasting has a direct impact on the performance of a company. The goal of the study is to present a new two-stage combination model named Hybrid-2-Best, for accurate demand forecasting. The model combines three forecasting models in a single combined forecast. The Hybrid-2-Best model uses a two-stage algorithm to achieve better-performing forecasts. Case study showed that the proposed Hybrid-2-Best model performs the best forecast performance among other combination techniques and individual methods. Furthermore, GP integration in the first and second stages gives flexibility. Experimental results indicate that the proposed Hybrid-2-Best model is a promising alternative for sales demand forecasting. MAPE of the proposed model is 0,13. This is a good result and better than compared other models. Proposed model performed better than other compared models in MASE and MSE as well.

Turkish Journal of Forecasting by Giresun University, Forecast Research Laboratory is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

1. Introduction

Forecasting plays a critical role in Business-to-Business (B2B) and Business-to-Customer (B2C) environments. B2B refers to the marketing or sales practices that companies carry out among themselves. Companies operating in B2B, basically work with the principle of improving each other's services. B2C refers to the commercial relationship that companies establish directly with consumers. All purchases we make individually are within the scope of this business model. The long-term strategic decisions such as investments, short-term decisions such as inventory management or production planning are affected by the accuracy of forecasts.

Investments will be idle if the forecasts do not meet the actual sales volumes. Similarly, production, inventory, or backorder costs may increase due to inaccurate forecasts. Increased production costs may occur due to overtime or wrong hire-fire decisions, or production levelling decisions. Inventory costs occur due to increased stock levels, perished products, wastes, and additional financial costs. Backorder costs occur due to lost orders and customers or expedited shipping costs to cover the gap between forecasted and actual demand. Hidden costs may also occur due

[*] Corresponding author.
E-mail addresses: fatihyigit@arel.edu.tr (Fatih Yiğit)

https://doi.org/10.34110/forecasting.957494
2618-6594/© 2021 Turkish Journal of Forecasting. All rights reserved.

to lost customers and the negative effect on the company's overall status among other potential customers. As a result, sales demand forecasting is an area that gains increased focus. Combining forecasts produced by combining different algorithms increases the accuracy of forecasting [1].

This study aims at establishing a sales forecast model that incorporates multiple models to achieve better results. The proposed model uses forecasts calculated by Poisson-regression (PR), support vector machines (SVM), and grey prediction (GP). The outputs of the best two models based on Mean Absolute Percentage Error (MAPE) are used as inputs for the second stage. The weights are calculated using optimization with the goal of minimization of MAPE. The second stage uses these inputs for the calculation of sales forecasts.

The results are compared according to MAPE, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Scaled Error (MASE) between three separate models (PR, SVM, GP) and the proposed model. The results proved that the proposed Hybrid-2-Best outperforms individual forecasts. The sales figures of a distribution company are used to analyse the accuracy. Results show the better performance of the proposed model compared to three well-known alternatives.

A combination of forecasts generally improves the performance [1]. Hybrid-2-Best also outperforms other combination methods. The comparisons prove the higher performance of the Hybrid-2-Best method. Hybrid-2-Best outperforms widely used combination methods in the given case study. The higher performance compared to other individual and combined models are achieved. This is the main contribution of this study. Also, the forecasting techniques used for the study (PR, SVM, and GP) are easy to implement. Hybrid-2-Best performs better than individual forecasts. The Hybrid-2-Best model chooses the best of two of three forecast techniques for the second stage. The first stage allows the best performing methods are chosen for the second stage. The low-performing method will be eliminated after the first stage. Additional methods can easily be implemented into the first stage. This way, based on the needs of the specific case, the proposed model can easily be adjusted accordingly. To the best of our knowledge, the two-stage model is new in the area of combined forecasting. Also, to the best of our knowledge, optimization is used for the first time for combining forecasts. The idea proved to be efficient, according to the results of our study. The GP method is well-known for achieving good performance with small data. As a result, the proposed model can be used when not enough data is available or in the presence of grey systems, as suggested by [2]. Additional or alternative forecasting techniques can be incorporated. This feature would increase flexibility and improve the performance of the proposed Hybrid-2-Best model.

This study is organized as follows. In Section 2, we review the literature and outline the contributions. The details of the proposed model and compared alternative models are given in Section 3. In Section 4, the details of the case study and experimental results are given. In Section 5, the conclusion is given.

2. Literature Review

Because sales forecasting is a vital part of the business, many forecasting studies are in the literature. This section will both cover the literature review on sales demand forecasting and the methodologies used in this study.

Forecasting is an integral part of any supply chain. Many methodologies have implemented the use of well-established methods. Moving or weighted average, smoothing applied on time-series with trend and seasonal patterns is a widely used method [3]. These methods have been applied for decades producing satisfactory results. However, the inherent linearity of the approaches comes in contradiction with the type of relationship between model variables in real life. This issue caused an increased number of scientific publications aiming to address this issue. The most common approach to the problem is utilizing the power of artificial neural networks (ANNs). ANN models are usually non-linear, and as such, they give more efficient results in classification, sample recognition, and forecasting problems than linear models [4]. This realization created an increased number of publications comparing ANNs and traditional linear methods for sales forecasting. In that direction, to benefit from both linear and non-linear models simultaneously, research efforts turned towards combined models for demand forecasting [4]. Such studies lead to improved forecasting accuracy for a supermarket retail chain. Also, enhancing ANNs with more search and learning capabilities integrating genetic algorithms (GA) and back-propagation generates more accurate forecasts at the expense of execution speed [5]. Combined models are used for different demand forecasting problems other than sales demand forecasting. Zhang et al. [6] proposed a combined method to solve emergency patient arrivals. The study uses ARIMA and SVM methods simultaneously. The conclusion shows that the proposed model outperforms the other competing methods. It affects directly and significantly the manufacturers and wholesalers around the world [7]. Forecasting is also used for forecasting financial bubbles. Kabran and Unlu [8] combined SVM with machine learning to forecast financial bubbles in S&P 500. The results show the high performance of the proposed model.

PR is a method used when the independent variables are count numbers. It is widely used for the forecasts associated with different topics. Wahiduzzaman and Yeasmin [9] used PR for rainfall forecast over the North Indian Ocean rim countries. The outcome is an %31 improvement above climatology. Graff et al. [10] used PR to forecast daily wildfire activities. The results show that regression models significantly outperform traditional persistence-based models used in operational smoke forecasting applications at both the cell and regional levels.

Similarly, in the hot topic of the Covid-19 Pandemic, PR is used to forecast deaths as well [11]. Due to these broad application areas and high performance, PR is chosen as one method in the first stage of the proposed model. SVM is a machine learning technique used for the classification of data. Besides a method used for classification, SVM is a promising tool for forecasting in an area. As an emerging tool for forecasting, it is commonly used for different the area of forecasting. Pan et al. [12] used SVM with improved ant colony optimization (ACO) to forecast photovoltaic power generation. The study shows that the combined usage of SVM and ACO achieved highly accurate forecasts. As a result, SVM is integrated into the first stage of the proposed combined model. Both SVM and PR are widely used in combined methods as well. Bilişik et al. [13] applied this SVM and PR for the revenue management area. The study used SVM and PR for comparison. Based on the outcomes, price-based revenue functions are generated. SVM is also used for different type of forecasts. Sahoo et al. [14] used SVM to detect DDOS attacks in software defined networks.

GP is based on the grey systems proposed by Deng Ju-long [15]. A system that lacks information such as structure message, operation mechanism, and behaviour document is called a grey system. Grey system theory is widely in literature for modelling many economic activity [16]. Other applications of grey theory are also used for different purposes. Kose et al. [17] used grey relational analysis to rank Turkey's cities and find the most liveable city. GP, sometimes referred to as grey forecasting, is a methodology derived from grey systems. The most crucial advantage of GP is the ability to make forecasts with little data. GP gained popularity in the past decade because of its simplicity and ability to characterize an unknown system using a few data points [18]. Due to this, flexibility and advantage GP is used in different studies. Gao [19] used an improved GP model for the medium-term load forecasting of a distribution network. The simulation executed showed that the forecast model has higher forecast accuracy for the mid-term forecast of load and obtains the forecasted value of electricity consumption and maximum load in a region from 2018 to 2021. Chai et al. [20] used the GP model to estimate the market value of the tutoring industry in Taiwan. The study proposed that the results can be used for strategic decisions that should be taken for the upcoming years in Taiwan. One of the main advantages inherent in GP is to perform forecast with a very low amount of data.

Combined models are widely used in forecasting. The goal is to use the high-performing attributes of the particular method to achieve a higher-performing combined model. The number of combined models and structures of combination can be different based on needs and applications. The applications can vary from sales demand forecasting, load forecasting, workforce requirements, patients’ arrival, or electricity demand forecasting. There are still problems with forecasting algorithms using separate models. No single approach can gain widespread industry acceptance since every model has its advantages and disadvantages [21]. However, some improvements for forecasting precision have been achieved by including specific algorithms [22]. Wang et al. [23] proposed a model to combine different methods. The application is performed on oil price forecasting. The results showed that the combined method performs better than individual methods. Qin and Cheng [24] combined Least-Square-Error (LSE) method with Auto-Regressive Integrated Moving Average (ARIMA) method. According to the results of the study, the proposed combined method improved the performance of a single model. Kartrika et al. [25] used a combined model to forecast short-term load in India. The study concluded that the model has supplemented its advantages effectively and showed that the combined model provided better accuracy than two separate models. In a recent study, combined forecasting approach is used to forecast short-term wind forecasting. [26] The proposed model shows that the performance is superior to other compared models in terms of stability and accuracy.

The study by Thomson et al. [27] proposed an analytical framework. The framework can be used to enhance the assessment of forecast performances. The forecasters are pooled to this framework in order to obtain an effectively combined forecast. Diebold and Shin [28] proposed an alternative to the "partially egalitarian" (peLASSO) procedure. Contrary to peLASSO that discards some forecasts, the proposed methodology uses "best average" combinations. The study showed that they seem highly competitive for out-of-sample forecasting. Kourentzes et al. [29] showed that the performance of forecast combination is a theoretically superior but cumbersome estimation of a full mixture model. Mamun et al. [30] studied the methodologies of combination and methods used. The combination of two or more forecasting models showed higher accuracy compared to individual forecasts. SVM, ANN, and their relevant models have been proven fruitful.

The proposed model incorporates two stages. In the first stage, the best two are chosen to be used in the second stage. To the best of our research, this two-stage methodology is proposed for the first time in literature. GP also allows accurate forecasting in case of small data is present. The flexibility of the proposed system selects the better performing alternative forecasting technique among alternatives. This system allows better performance compared to single alternatives. Also, additional forecasting techniques can be used additionally as a replacement or addition to the existing methods.

3. Methods and The Proposed Model

In this section, we give information about SVM, PR, and GP. These three methods are used in the first stage of the proposed combined method. Hybrid-2-Best model is a two-stage model to combine SVM, PR, and GP. Details of SVM are given 3.1, PR in 3.2, GP is given in 3.3, and the combined model is given in Section 3.4.

3.1. Support Vector Machines

The SVM model was originally proposed by Vapnik [31]. The objective is to minimize the upper bound of the generalization error by using a subset of data points called support vectors. The training set has pairs of (x_1, y_1), (x_2, y_2) …, and (x_n, y_n), where x_i ∈Rn and y_i ∈Rn are the input vectors and associated output values, respectively. Let xi be mapped into a high dimensional feature by a non-linear function φ(x). The SVR function can be expressed as Eq. (1),

$$f(x) = w^T\varphi(x) + b \tag{1}$$

where f(x) denotes the forecast output, φ (x) is the feature function, w and b are the corresponding coefficients. In SVM, the overall goal is to find a function f(x), whose deviation from each yi is at most ε for all the training data. Thus, the SVR is an optimization problem. The problem will be expressed as Eq. (2) and Eq. (3)

$$min\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi + \xi^*) \tag{2}$$

$$subject\ to \begin{cases} y_i - w^T\varphi(x) - b \le \varepsilon + \xi \\ w^T\varphi(x_i) + b - y_i \le \varepsilon + \xi^* \\ \xi \ge 0,\ \xi^* \ge 0 \end{cases} \tag{3}$$

where n represents the sample size, ξ and ξ* represent the upper and lower training errors subject to ϵ insensitive $|y_i\text{-}w^T\varphi(x_i)\text{-}b| < \epsilon$,and C>0 is the regularization factor.

3.2. Poisson Regression

Poisson Regression is a method to forecast count data. This method assumes a conditional Poisson distribution for y, modelling the logarithm of the expected counts as a linear function of the input variables [29]. The model is characterized as given in Eq. (4).

$$\log\Big(E\big(Y\,\big|\,\mathrm{x}\big)\Big) = a + bx \tag{4}$$

In Eq. (4), a ϵ R and b ϵ Rn. From Eq. (4), we can deduct Eq. (5) the forecasted mean of the associated Poisson distribution.

$$\Big(E\big(Y\,\big|\,\mathrm{x}\big)\Big) = e^{a+bx} \tag{5}$$

3.3. Grey Prediction

GP is a specialized application of grey Systems. Ju-long [15] first proposed grey systems. Grey system refers to a system in which information is partially known. Grey system theory fits into the internal functioning mechanism of the environmental system, picking out ambiguous grey quantum difficult to be quantified in the form of a mathematical model, and determines the parameter of differential equation by application of time series data [32].

GP is a model known for its simplicity and ability to characterize unknown systems by using a few data points [16]. GM (1,1) is a basic model that is used for forecast. GM (1, 1) represents first-order one-variable grey model.

Grey forecasting uses the accumulated generation operation (AGO) accumulated to reduce variation among the original data series and build differential equations by linearly transforming data series. GM(1, 1) and Verhulst models are mainly used for forecast [33]. GM (1,1) is considered a pure model commonly applied to academic research and industrial applications. [16]

The GM(1, 1) model constructing process is described as follows. [34]

Let's assume that the original data sequence by

$$x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\right) \tag{6}$$

The number of the data observed in Eq. (6) defined as "n". The AGO is a part of GP. The aim is to reduce the randomness of data. The AGO formation of $x^{(0)}$ is given in Eq. (7).

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)\right) \tag{7}$$

where

$$x^{(1)}(k) = \sum_{m=1}^{k} x^{(0)}(m), \; k = 1,2, \dots, n \tag{8}$$

The GM (1, 1) model can be constructed by establishing a first order differential equation for $x^{(1)}(k)$ as;

$$\frac{dx^{(1)}(k)}{dk} + ax^{(1)}(k) = b \tag{9}$$

Therefore, the solution of Eq. (9) can be obtained by using the least square method. That is,

$$\hat{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right) e^{-\hat{a}(k-1)} + \frac{\hat{b}}{\hat{a}} \tag{10}$$

where

$$\left[\hat{a}, \hat{b}\right]^T = (B^T B)^{-1} B^T X_n \tag{11}$$

and

$$B = \begin{bmatrix} -0.5\left(x^{(1)}(1) + x^{(1)}(2)\right) & 1 \\ -0.5\left(x^{(1)}(2) + x^{(1)}(3)\right) & 1 \\ \vdots & \vdots \\ -0.5\left(x^{(1)}(n-1) + x^{(1)}(n)\right) & 1 \end{bmatrix} \tag{12}$$

$$X_n = \left[x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), \dots x^{(0)}(n)\right],^T \tag{13}$$

From the Eq. (10), $\hat{x}^{(1)}$ is obtained. Let $\hat{x}^{(0)}$ be fitted and forecasted series.

$$\hat{x}^{(0)} = \left(\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \hat{x}^{(0)}(3), \dots, \hat{x}^{(0)}(n)\right) \tag{14}$$

where $\hat{x}^{(0)}(1) = x^{(0)}(1)$.

Applying the inverse AGO, we then have Eq. (15)

$$\hat{x}^{(0)}(k) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right)\left(1 - e^{\hat{a}}\right)e^{-a\widehat{(k-1)}}, k = 2, 3, \dots \tag{15}$$

where $\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \hat{x}^{(0)}(3), \dots, \hat{x}^{(0)}(n)$ are called the GM(1, 1) fitted sequence, while $\hat{x}^{(0)}(n+1)$, $\hat{x}^{(0)}(n+2)$ … are called the GM(1, 1) forecast values.

3.4. Two-Stage Forecast Combination Model

The proposed method (Hybrid-2-Best) used to forecast sales demand is a combined model that incorporates PR, SVM, and GP. In the first stage, the training data is used to forecast the first test data. MAPE values are used for the performance measurement of the methods. The best two methods are chosen for the second stage.

The two-stage model allows the user to integrate different methods in the first stage. Based on the experience of the practitioner and the specific needs of the case, different and additional methods can be integrated. The forecasts received are used to minimize the MAPE of the first stage, according to the Eq. (16).

$$minMAPE(c_x(B1) + d_x(B2) + e_x) \tag{16}$$

where c_x, $d_x \geq 0$. In Eq. (16), B1 represents the best method, and B2 represents the second-best method according to stage 1. cx, dx, and ex represent the variables calculated for each item x.

In the second stage, the first stage results are used to forecast the second test stage of the combined method via GP. The workflow is given in Figure 1.

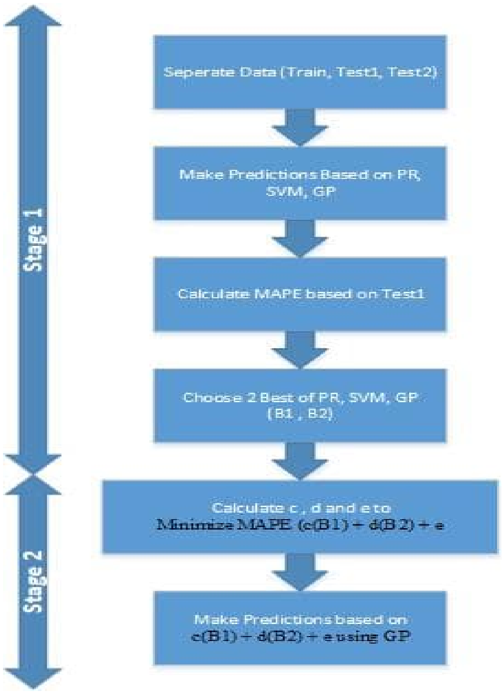

Figure 1. Workflow of the two-stage combined method (Hybrid-2-Best)

4. Case Study in Distribution Business

The numerical study is applied for the data that is received from a distribution company. The company deals with the sales of distribution products. Due to long lead times, accurate forecasting is vital for the success of a high-performing inventory management system. Sales demand forecasting is vital for the company as inventory directly affects net working capital (NWC). Besides, due to high competition, possible stock-outs may undermine the general performance due to lost sales or customers. The Hybrid-2-Best model is applied for 30 items. According to ABC analysis, the sample of 30 items covers the most important or, in other words, class A items. All sales data is received from the ERP system covering 60 months of data.

4.1. Application of the Hybrid-2-Best Model on a Real-Life Example

The Hybrid-2-Best model is applied to real data to assess its performance. Data is received from the distribution company that serves B2B customers. Data for each item covers 60 months. The first 48 period that covers between 01.2015 and 12.2018 is chosen for the training of the first stage. The next six months' data between 01.2019 and 06.2019 is used as the validity of the Stage 2. The next six months' data covering between 07.2019 and 12.2019 is used to test the proposed model.

The first training data is used to calculate the forecasts of Stage 1 by PR, SVM, and GP. This information is used for the second stage. All calculations are executed on MATLAB R2017b working on the following configuration: Intel Core i5 processor at 2.6 GHz with 8GB RAM. We used the MATLAB code given in the study by Zhang and Li [32] for calculating the GP model. SVM calculations are made using WEKA software with the default options. For the calculation of PR, MATLAB embedded function is used.

The results show that SVM showed the best performance among the three alternatives. Figure 2 shows the number of best and second-best models in Stage 1 according to MAPE calculations.

The best two out of three forecasting methods for each item are chosen, namely B1 and B2. The coefficients cx, dx, and ex are calculated according to Eq. (16). The evolutionary algorithm of MS Excel Solver is used. The default configuration is used for the calculation

Table 1. c_x, d_x and, e_x calculation results

Item	c_x	d_x	e_x
1	1.013	0.000	0.000
2	0.000	0.830	1.178
3	0.359	0.007	8.849
…	…	…	…
28	0.000	1.640	4.568
29	0.498	0.077	-30.385
30	1.690	0.096	-403.241

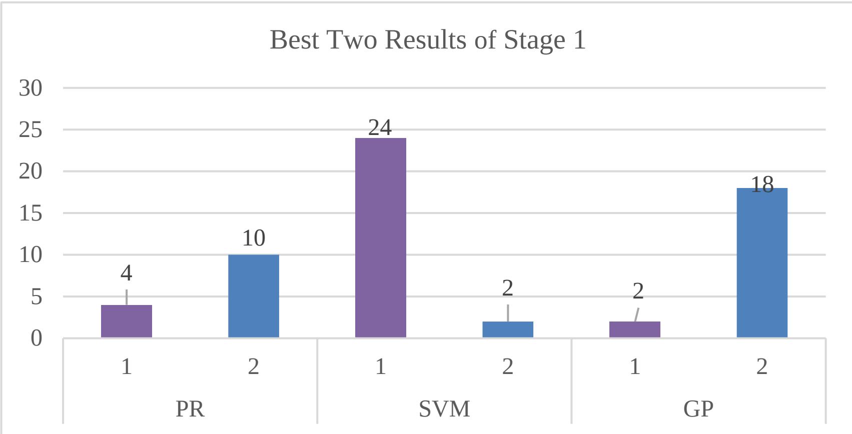

Figure 2. Best-two results of stage 1

In this second stage, the forecasts are calculated using GP. The outputs of the first stage 1 are used to generate forecasts. This 6-period forecast covers the period between 55-60. The comparison between individual methods and Hybrid-2-Best is made based on MAPE, MAE, RMSE, and MASE. Eq. 17- 20 show the calculation of this metrics.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{e_t}{y_t} \right| \tag{17}$$

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} e_t^2} \tag{19}$$

$$MASE = \frac{\sum_{t=1}^{n} |e_t|}{\frac{n}{n-1} \sum_{t=2}^{n} |Y_t - Y_{t-1}|} \tag{20}$$

The results are given in Figures 3, 4, 5 and 6 according to performance criteria MAPE, MAE, RMSE and MASE. As seen from the results, the proposed combined model performs better compared to individual alternatives. In Table 2, the results of the second stage are given. The comparison of individual methods and Hybrid-2-Best are given in Table 3.

Table 2. Results of the second stage

	Item No								
	1		2		3		...	30	
Month	**Actual**	**Forecast**	**Actual**	**Forecast**	**Actual**	**Forecast**		**Actual**	**Forecast**
55	103,094	**98,121**	29,891	**31,417**	12,350	**11,877**	...	3,777	**3,670**
56	94,661	**100,202**	29,144	**31,281**	12,409	**11,785**	...	3,870	**3,616**
57	95,087	**102,326**	30,520	**31,145**	13,232	**11,694**	...	3,700	**3,562**
58	107,198	**104,496**	34,689	**31,010**	12,835	**11,604**	...	3,885	**3,509**
59	120,047	**106,711**	37,364	**30,876**	14,665	**11,514**	...	4,760	**3,457**
60	125,101	**108,974**	40,884	**30,742**	15,439	**11,425**	...	4,865	**3,405**

Table 3. Average performance results of the second stage

	SVM	Poisson-Regression	SAGF	Hybrid-2-Best
MAPE	0,22	0,38	0,38	0,13
MASE	2,27	4,61	4,45	1,00
MAE	1.992	3.150	3.043	1.533
RMSE	9.771	10.529	10.456	9.663

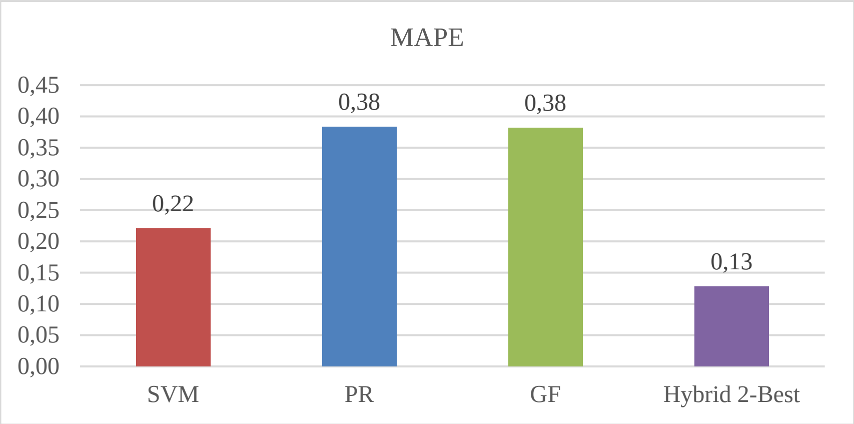

Figure 3. Average MAPE results of individual methods vs. Hybrid-2-Best

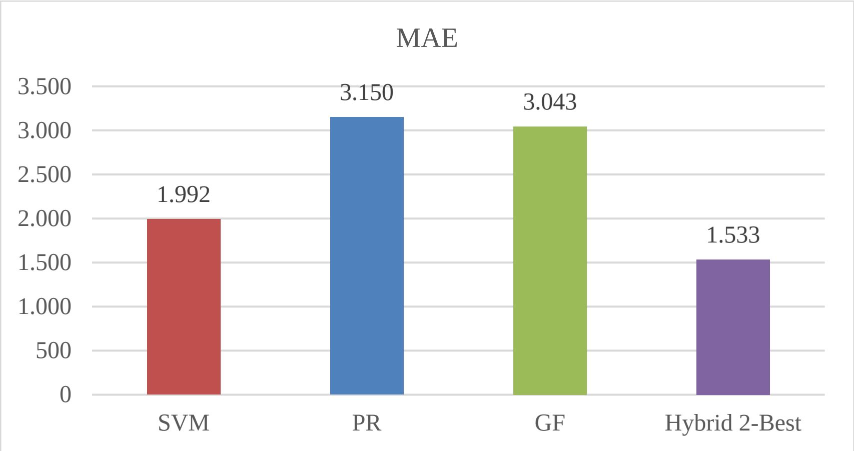

Figure 4. Average MAE results of individual methods vs. Hybrid-2-Best

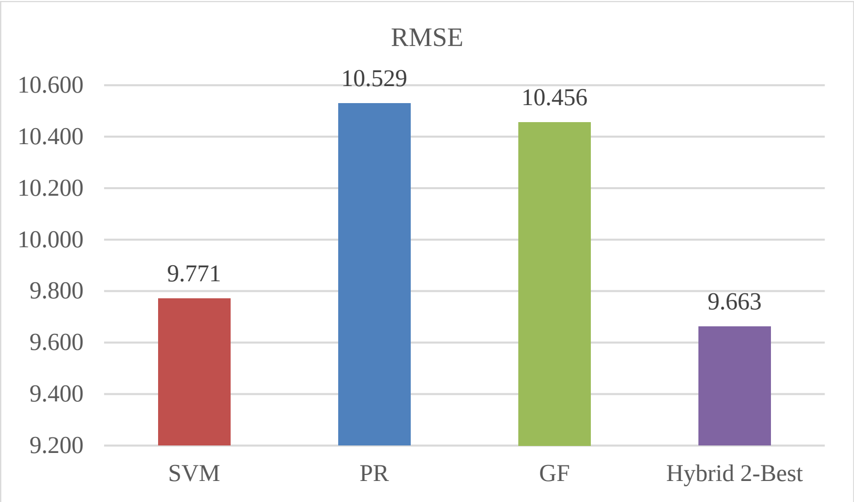

Figure 5. Average RMSE results of individual methods vs. Hybrid-2-Best

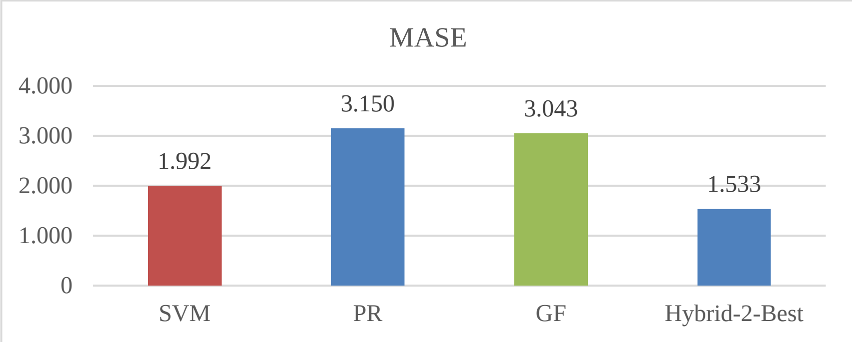

Figure 6. Average MASE results of individual methods vs. Hybrid-2-Best

4.2. Comparison of Hybrid-2-Best vs. Other Combination Methods

Generally, combined methods improve forecasting. It is essential to compare a proposed method with other combination methods. In this section, a comparison between well-known combination methods and Hybrid-2-Best is made.

The combination methods used are as follows. [35]

In Equal Weights (EW) approach, equal weights are assigned to each forecast. We chose w_1= 1/3, w_2 = 1/3, and w_3 = 1/3. It is a straightforward approach.

Inverse of Forecast Errors (INV) assign weights proportional to their individual inverse of the forecast errors, i.e., mean square error (MSE). Eq. (21). Shows the calculation of INV.

$$MSE_k = \frac{1}{n}\sum_{j=1}^{n}(e_j)^2 \tag{21}$$

$$w_x = \frac{1/MSE_x}{\sum_{y=1}^{n} 1/MSE_y} \tag{22}$$

In Eq. (22), w_x represents the weights of each individual forecast used for combined forecasting. N represents the total number of individual forecast types.

In Variance Based (VAR) approach assigns higher weight to forecasts with lower variability. It is widely used approach. Eq. (23) and Eq. (24) shows the calculation of VAR.

$$w_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \tag{23}$$

$$w_2 = 1 - w_1 \tag{24}$$

In Rank Based (RANK) approach is used to combine forecasts based on their ranks. The weights are defined based on the times model 1 outperforms model2 and vice versa. Eq. (25) shows the calculations of weights. $RANK_x$ represents the number of times that the model outperforms other methods.

$$w_x = \frac{RANK_x}{\sum_{y=1}^{n} RANK_y} \tag{25}$$

Weights of the Discounted Mean Square Forecast Error (MSFE) are assigned according to recent forecast errors. The weights of discounted MSFE combination are given in Eq. (26) and Eq. (27).

$$w_{it} = \frac{m_{it}^{-1}}{\sum_{j=1}^{n} m_{it}^{-1}} \tag{26}$$

$$m_{it} = \sum_{s=T_0}^{t-h} \partial^{t-h-s} \left(Y_{s+h}^{h} - \hat{Y}_{i,s+h|s}^{h}\right)^2 \tag{27}$$

where δ is the discount factor. T_0 is the start time out-of-sample forecasts, Y and $\hat{Y}$ are the actual value and forecasted value.

The comparison showed that the proposed method outperforms all other combined methods. The results are given in Fig. 7-10.

The comparisons showed that the proposed method outperforms all other combined methods. The proposed model has an average MAPE of 0.13. This ratio is considered a good result [36]. The proposed model is the only forecast in this category.

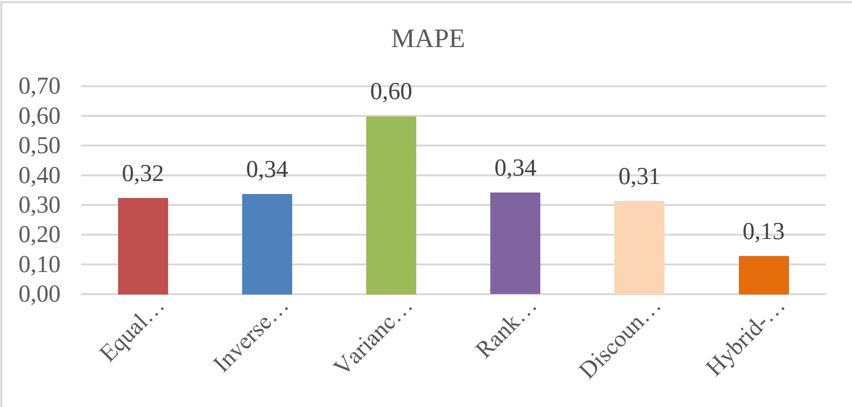

Figure 7. Average MAPE results of combined methods vs. Hybrid-2-Best

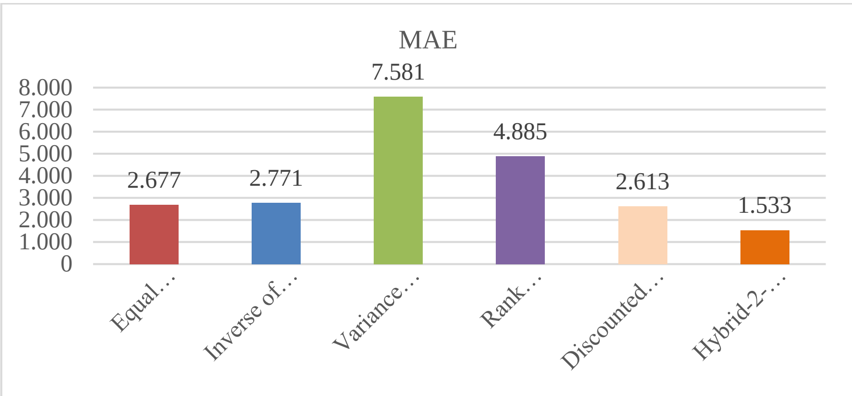

Figure 8. Average MAE results of combined methods vs. Hybrid-2-Best

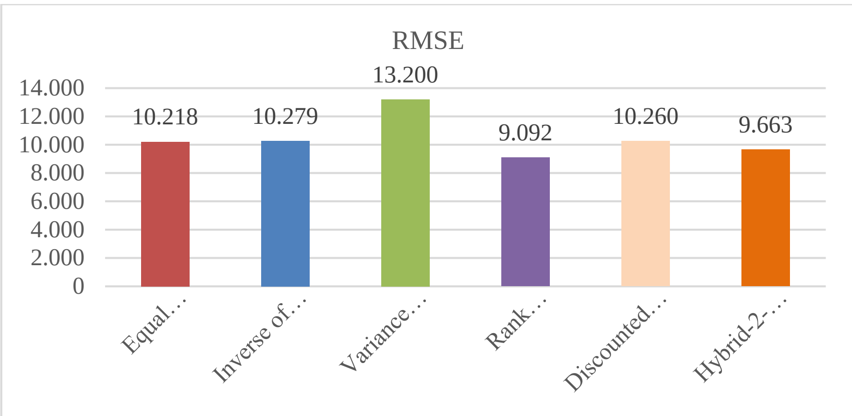

Figure 9. Average RMSE results of combined methods vs. Hybrid-2-Best

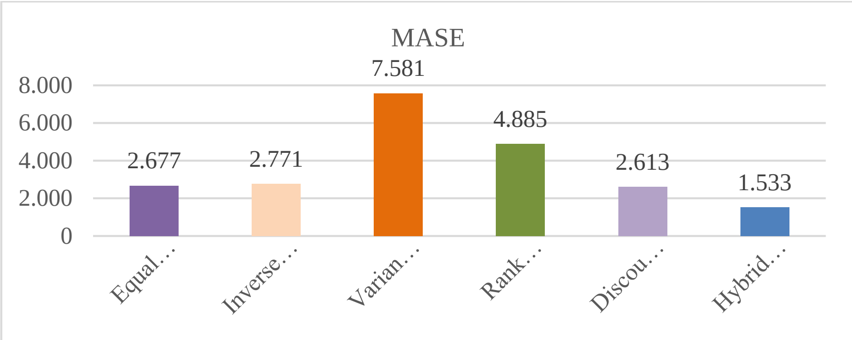

Figure 10. Average MASE results of combined methods vs. Hybrid-2-Best

5. Conclusion

Sales demand forecasting plays a vital role in the success of a company. Accurate forecasting contributes to improved customer service, improved inventory management, meeting return of investment expectations. Hybrid-2-Best model for sales demand forecasting is proposed in this study. The proposed Hybrid-2-Best aims to improve the forecasting procedure. It incorporates two stages. The first stage executes three separate forecasting methods, namely PR, SVM, and GP. The outputs of the first stage are used to generate input to the second stage to calculating variables. GP is also used in the second stage to generate final forecasts. The numerical study showed that the proposed method performed better based on MAPE, MAE, RMSE, and MASE. Hybrid-2-Best showed better performance in all these four criteria. Proposed Hybrid-2-Best is a combined method. The comparison is made between Hybrid-2-Best and other combination methods given in the study. The results showed the higher performance of the proposed combination method. The result of the mean MAPE results of 30 items is 0.13. According to our study, it is considered a good forecast. The proposed method was the only individual or combined model in this category.

There are other advantages besides the improved performance. The second advantage is incorporating the GP model as a part of the first stage. GP has an advantage of applicability when data has a "small sample size or poor quality" [14]. This feature is an important advantage as the business is characterized as volatile, uncertain, complex, and ambiguous (VUCA). In such environments, the application of GP will help to deal with VUCA. The third advantage is utilizing three different methods and selecting the best two for the second stage. Therefore, the best performing models are chosen for the second stage. This elimination allows the proposed model to use the highest performing methods in Stage 1. The results also show that the proposed model Hybrid-2-Best performs better than individual models. Although the Hybrid-2-Best model is promising, it has limitations. The performance is dependent on the success of the three forecast methods. (PR, SYM, GP). Also, when data size is not enough, the performance depends only on the GP method's performance.

Future research will focus on these areas for improvement. Additional models can be integrated to improve the performance, such as improved GM (1, 1), ARIMA, or other combined models. To extend the application areas and improve performance, seasonality factors will be integrated into the model for further studies.

The proposed model is applied in a real-life case. Although the model performed superior compared to separate models, application with a bigger dataset may be valuable to see the comparison with a different dataset. So, we aim to use the proposed model with a competition dataset such as M-Competition.

Optimization in the model is achieved with the goal of minimizing MAPE. MAPE method has an important drawback. It may cause excessive penalty for overshooting of the forecasts. So, we plan to accommodate other performance metrics to overcome this drawback in our future studies.

References

[1] J.S. Armstrong, Combining forecasts: The end of the beginning or the beginning of the end?, Int. J. Forecast. 5 (1989) 585–588. doi:10.1016/0169-2070(89)90013-7.

[2] L. Si-Feng, Emergence and Development of Grey System Theory and Its Forward Trends [J], J. Nanjing Univ. Aeronaut. \& Astronaut. 2 (2004) 266–272.

[3] E.S. Gardner, E. Mckenzie, Forecasting Trends in Time Series, Manage. Sci. 31 (1985) 1237–1246. doi:10.1287/mnsc.31.10.1237.

[4] A. Aslanargun, M. Mammadov, B. Yazici, S. Yolacan, Comparison of ARIMA, neural networks and hybrid models in time series:

Tourist arrival forecasting, J. Stat. Comput. Simul. 77 (2007) 29–53. doi:10.1080/10629360600564874.

[5] W.K. Wong, Z.X. Guo, A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm, Int. J. Prod. Econ. 128 (2010) 614–624. doi:10.1016/j.ijpe.2010.07.008.

[6] Y. Zhang, L. Luo, J. Yang, D. Liu, R. Kong, Y. Feng, A hybrid ARIMA-SVR approach for forecasting emergency patient flow, J. Ambient Intell. Humaniz. Comput. 10 (2019) 3315–3323. doi:10.1007/s12652-018-1059-x.

[7] P. Lalou, S.T. Ponis, O.K. Efthymiou, Demand Forecasting of Retail Sales Using Data Analytics and Statistical Programming, Manag. Mark. 15 (2020) 186–202. doi:10.2478/mmcks-2020-0012.

[8] F. Başoğlu Kabran, K. Demirberk Ünlü, A two-step machine learning approach to predict S&P 500 bubbles, J. Appl. Stat. (2020). doi:10.1080/02664763.2020.1823947.

[9] M. Wahiduzzaman, A. Yeasmin, Statistical forecasting of tropical cyclone landfall activities over the North Indian Ocean rim countries, Atmos. Res. 227 (2019) 89–100. doi:10.1016/j.atmosres.2019.04.034.

[10] C.A. Graff, S.R. Coffield, Y. Chen, E. Foufoula-Georgiou, E. Foufoula-Georgiou, J.T. Randerson, P. Smyth, Forecasting daily wildfire activity using poisson regression, IEEE Trans. Geosci. Remote Sens. 58 (2020) 4837–4851. doi:10.1109/TGRS.2020.2968029.

[11] T. KIm, B. Lieberman, G. Luta, E. Pena, Prediction Regions for Poisson and Over-Dispersed Poisson Regression Models with Applications to Forecasting Number of Deaths during the COVID-19 Pandemic, (2020). http://arxiv.org/abs/2007.02105.

[12] M. Pan, C. Li, R. Gao, Y. Huang, H. You, T. Gu, F. Qin, Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization, J. Clean. Prod. 277 (2020). doi:10.1016/j.jclepro.2020.123948.

[13] M.T. Bilişik, F.H. Sezgin, Ş. Esnaf, Dynamic Pricing for Revenue Management in Retailing Using Support Vector Machine , Poisson Regression and Nonlinear Programming, Eurasian Bus. Econ. J. 8 (2017) 11–34.

[14] K.S. Sahoo, B.K. Tripathy, K. Naik, S. Ramasubbareddy, B. Balusamy, M. Khari, D. Burgos, An Evolutionary SVM Model for DDOS Attack Detection in Software Defined Networks, IEEE Access. 8 (2020) 132502–132513. doi:10.1109/ACCESS.2020.3009733.

[15] D. Ju-Long, Control problems of grey systems, Syst. Control Lett. 1 (1982) 288–294.

[16] D. Camelia, Grey systems theory in economics--bibliometric analysis and applications' overview, Grey Syst. Theory Appl. (2015).

[17] E. Kose, D. Vural, G. Canbulut, The most livable city selection in Turkey with the grey relational analysis, Grey Syst. Theory Appl. (2020).

[18] L. Suganthi, A.A. Samuel, Energy models for demand forecasting - A review, Renew. Sustain. Energy Rev. 16 (2012) 1223–1240. doi:10.1016/j.rser.2011.08.014.

[19] F. Gao, Application of improved grey theory prediction model in medium-term load forecasting of distribution network, in: Proc. - 2019 7th Int. Conf. Adv. Cloud Big Data, CBD 2019, Institute of Electrical and Electronics Engineers Inc., 2019: pp. 151–155. doi:10.1109/CBD.2019.00036.

[20] Y.-C. Chai, Y.-F. Huang, H.-S. Dang, An application of grey prediction to estimate the market size of tutoring industry in Taiwan, in: 2017 Int. Conf. Appl. Syst. Innov., IEEE, 2017: pp. 1740–1743.

[21] M. Negnevitsky, P.L. Johnson, S. Santoso, Short term wind power forecasting using hybrid intelligent systems, (2007).

[22] T. Hui, N. Dongxiao, Combining simulate anneal algorithm with support vector regression to forecast wind speed, in: 2010 Second IITA Int. Conf. Geosci. Remote Sens., IEEE, 2010: pp. 92–94.

[23] S.P. Wang, A.M. Hu, Z.X. Wu, Y.Q. Liu, X.W. Bai, Multiscale combined model based on run-length-judgment method and its application in oil price forecasting, Math. Probl. Eng. 2014 (2014). doi:10.1155/2014/513201.

[24] P. Qin, C. Cheng, Prediction of Seawall Settlement Based on a Combined LS-ARIMA Model, Math. Probl. Eng. 2017 (2017). doi:10.1155/2017/7840569.

[25] S. Karthika, V. Margaret, K. Balaraman, Hybrid short term load forecasting using ARIMA-SVM, in: 2017 Innov. Power Adv. Comput. Technol., IEEE, 2017: pp. 1–7.

[26] Z. Liu, P. Jiang, L. Zhang, X. Niu, A combined forecasting model for time series: Application to short-term wind speed forecasting, Appl. Energy. 259 (2020) 114137. doi:10.1016/j.apenergy.2019.114137.

[27] M.E. Thomson, A.C. Pollock, D. Önkal, M.S. Gönül, Combining forecasts: Performance and coherence, Int. J. Forecast. 35 (2019) 474–484. doi:10.1016/j.ijforecast.2018.10.006.

[28] F.X. Diebold, M. Shin, Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives, Int. J. Forecast. 35 (2019) 1679–1691. doi:10.1016/j.ijforecast.2018.09.006.

[29] N. Kourentzes, D. Barrow, F. Petropoulos, Another look at forecast selection and combination: Evidence from forecast pooling, Int. J. Prod. Econ. 209 (2019) 226–235. doi:10.1016/j.ijpe.2018.05.019.

[30] A. Al Mamun, M. Sohel, N. Mohammad, M.S. Haque Sunny, D.R. Dipta, E. Hossain, A Comprehensive Review of the Load Forecasting Techniques Using Single and Hybrid Predictive Models, IEEE Access. 8 (2020) 134911–134939.

doi:10.1109/access.2020.3010702.

[31] Vapnik, Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media., (1995).

[32] Y. Zhang, M. Li, Grey system forecasting based on MATLAB and its example application, in: 2010 2nd IEEE Int. Conf. Inf. Manag. Eng., IEEE, 2010: pp. 48–52.

[33] D. Julong, Introduction to grey system theory, J. Grey Syst. 1 (1989) 1–24.

[34] H.M. Hsu, C.T. Chen, Aggregation of fuzzy opinions under group decision making, Fuzzy Sets Syst. 79 (1996) 279–285. doi:10.1016/0165-0114(95)00185-9.

[35] L. Ling, D. Zhang, A.W. Mugera, S. Chen, Q. Xia, A Forecast Combination Framework with Multi-Time Scale for Livestock Products' Price Forecasting, Math. Probl. Eng. 2019 (2019). doi:10.1155/2019/8096206.

[36] C.D. Lewis, Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting, Butterworth-Heinemann, 1982.