International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Hybrid Machine Learning Model for Demand Forecasting: Combination of K-Means, Elastic-Net, and Gaussian Process Regression

Doohee Chung*[1], Chan Gyu Lee[2], Sungmin Yang[3]

Submitted: 10/02/2023 **Revised:** 14/04/2023 **Accepted:** 08/05/2023

Abstract: Accurate prediction of demand assists companies in responding flexibly to uncertain market conditions. However, companies face difficulties due to the wide variety of product types and the uncertainty of their scale. To solve this problem, this study proposes a hybrid model that combines K-means, ElasticNet, and Gaussian process regression (GPR). GPR effectively addresses the issue of nonlinear prediction, and its performance is excellent when the model is trained with clusters of similar data instead of training the entire dataset at once. Moreover, the model enhances its training process by extracting crucial and contributing variables for each cluster. To implement these techniques, this study utilizes K-means and ElasticNet in combination with GPR. The model is applied to a case study of a U.S. manufacturing company and is compared with other benchmarking models, including single GPR, K-means+GPR, and ElasticNet+GPR, to evaluate its performance. The hybrid model achieved the best prediction accuracy, recording a mean absolute error (MAE) of 5.57, demonstrating its potential as foundational research for constructing an efficient demand forecasting model under uncertain conditions.

Keywords: *Demand Forecast, Gaussian Process Regression, K-means, ElasticNet, Hybrid Machine Learning*

1. Introduction

Demand forecasting is one of the most essential techniques for enhancing corporate competitiveness. Accurate demand forecasting forms the foundation for optimizing scheduling, managing capacity, inventory management [2], assortment planning [12], implementing advanced order policies, and optimizing sales management [24].

In particular, accurate demand forecasting enables companies to proactively prepare necessary products based on market demand, thereby preventing inventory shortages or excess inventory and maximizing sales volume throughout the product life cycle [11]. This prevents companies from incurring costs and inefficiencies across the supply chain [22]. Furthermore, demand forecasting offers additional benefits such as customer retention, acquisition, and improved sales [17].

However, accurate demand forecasting poses significant challenges. First, sales patterns are complex, irregular, and uncertain. Product demand is influenced by environmental and hidden factors, in addition to historical trends. Incorporating these factors into the data or capturing them accurately poses difficulties in demand forecasting [35]. Moreover, the shrinking product life cycles resulting from technological advancements further intensify the challenge of accurately predicting demand [22].

Demand forecasting research also has data-related limitations, such as disorganized historical data and sparse sales data [4]. In many cases, the lack of demand data leads to biased prediction results. Therefore, advanced research in demand forecasting focuses on tasks like data supplementation and model development using small-sized data [6, 8].

Meanwhile, numerous studies have applied data science models, such as time series models, machine learning algorithms, and deep learning techniques, to predict future demand based on historical data [29]. Time series models mainly consist of Moving Average (MA) models, Auto-Regressive (AR) models, and Auto-Regressive Integrated Moving Averages (ARIMA) models [37]. These models require data stationarity as a precondition. Machine learning models, including Random Forest, XG Boost, and Artificial Neural Network (ANN), are also widely applied to demand forecasting problems [16, 30, 34]. Deep learning, which increases the accuracy of demand forecasting, has gained prominence [10, 20]. A previous study proposed the use of the LSTM (Long-Short Term Memory) model, a type of deep learning model, for market demand forecasting [19]. Compared to linear regression and MLP regressor, the LSTM model effectively addresses unstructured demand forecasting problems.

Despite the great performance of these models in specific

[1] Corresponding Author, School of Global Entrepreneurship and Information Communication Technology (ICT), Handong Global University - 05319, Republic of Korea
Email: profchung@handong.edu
ORCID ID : 0000-0002-7413-0778
[2] Graduate School of Data Science, Seoul National University - 05319, Republic of Korea
[3] School of Global Entrepreneurship and Information Communication Technology (ICT), Handong Global University - 05319, Republic of Korea

situations, there is a limitation to their general applicability due to distinct strengths and weaknesses. In other words, while these models may be suitable for specific scenarios, they cannot be applied to general cases.

In recent years, the use of hybrid prediction models, which combine multiple models in a cooperative manner, has demonstrated outstanding performance. The objective is to address the limitations of individual models by supplementing them with other models or maximizing performance through the integration of their respective strengths.

In a previous study, the performance of a hybrid model combining LSTM and clustering techniques was tested for predicting product sales in the e-commerce industry [4]. Although LSTM can capture the nonlinear demand relationship of products, it has been noted that a single LSTM model may not be ideal when dealing with different sales patterns among product portfolios. Therefore, to enhance the training process of LSTM, clustering techniques such as K-means were utilized to group several products, and each group was trained individually. In this study, using online sales data from Walmart, the combined clustering and LSTM model demonstrated superiority over the single model.

Another study proposed a hybrid PROPHET-SVR model for time-series demand forecasting in the manufacturing industry [14]. This study represented the temporary and rapid changes in time series data as parameters using the PROPHET model. These parameters were then utilized in the SVR model to enhance prediction performance. The research suggests that by considering forecast residuals, the hybrid model can be customized to account for the effects of holidays and seasons, resulting in higher performance compared to other demand forecasting methods (GA-LSTM, ARIMA-LSTM, HPF-GM, SGM, DSGM, SARIMA-SVR3, EMD-LSSVR, CD-LSSVR, X12-A-LSSVR, X12-M-LSSVR, X12-A-SAL).

One study proposed the 'Demand Forest,' a hybrid model for demand forecasting of new products [36]. This model employs K-means clustering to group data based on demand patterns and utilizes a random forest model to train each data cluster. Additionally, the research incorporates QRF (Quantile Regression Forest), a quantile regression system, to predict the total demand amount for the four months following product launch and estimate weekly demand by incorporating this information into the demand pattern. Through the integration of these models, successful demand forecasting for new products in the absence of historical data was achieved.

Another study introduced the STL-DADLM model for tourism demand forecasting [42]. This model combines the STL decomposition and the DLM model to address the issue of overfitting and enhance the accuracy of demand forecasting in both the short and long term. The research highlights that the STL-DADLM model effectively mitigates the problem of overfitting and demonstrates superior performance compared to general machine learning models.

This study acknowledges the effectiveness of hybrid forecasting models and aims to explore an even more superior combination model for demand forecasting. Therefore, it is crucial to construct a suitable model set for the demand forecasting problem and determine the appropriate role for each model.

To begin, it is important to note that product demand is influenced by market trends and exhibits different characteristics depending on the season. Additionally, the sales patterns vary across different product categories. Thus, a more realistic approach is to train and predict each category separately, taking into account their distinct patterns. In this study, the first core process involves applying clustering techniques to historical product data based on demand patterns. For this purpose, the widely used K-means clustering technique is employed.

Furthermore, Gaussian process regression (GPR) is utilized in this study, where the model is applied individually to each cluster. The GPR model leverages the mean and covariance to determine the distribution that corresponds to the confidence interval of the predicted value. It calculates the variance of the distribution to make predictions. This approach simplifies the analysis of unstructured and non-parametric data.

The advantage of this method is its ability to capture the non-linear relationship inherent in irregular markets, surpassing general machine learning models. As a result, it has gained popularity among researchers in the field of predictive model research [38, 42]. Hence, the second core process of this research is to construct a model with advanced predictive performance using GPR.

On the other hand, GPR also has some drawbacks. It utilizes a kernel function to create a posterior prediction distribution and derive a suitable predicted value. However, this process involves a significant increase in calculation and consumes a large amount of time and computing memory. Additionally, as the predictive model is constructed, the dimension of independent variables increases due to feature generation, resulting in a substantial increase in computational power. Therefore, identifying relevant features for efficient model training becomes a key task. Consequently, the third core process of this study focuses on identifying optimized feature sets for each cluster, aiming to build a more efficient demand forecasting model. To accomplish this, the study utilizes the ElasticNet algorithm for feature selection and model optimization.

In summary, this study proposes a new hybrid model that combines K-means, ElasticNet, and GPR to effectively address atypical and complex demand forecasting problems. To evaluate the model's performance, comparisons are made with the single GPR model and benchmarking models such as K-means+GPR and ElasticNet+GPR.

The model presented in this study yields several valuable insights. Through K-means clustering, data with similar time-series properties are grouped together, confirming previous studies' findings regarding data characteristics and their utilization. Additionally, this study focuses on improving prediction accuracy and model efficiency by utilizing ElasticNet, which selects the most contributing features from the full dataset. Lastly, the study aims to enhance the model's performance by incorporating the GPR model and considering the confidence intervals of the predicted results.

Above all, the core contribution of this study lies in presenting a methodology for a hybrid demand forecasting model that surpasses the predictive performance of existing research that relies on single machine learning models.

2. Research Methodology

The advanced demand forecasting model proposed in this study consists of K-means, ElasticNet, and GPR. The concepts and mechanisms of each methodology are as follows.

2.1. K-means

Clustering is a method used to identify data characteristics by grouping the entire dataset based on similar properties or categories [9, 33]. The primary objective of clustering is to divide the collected sample data into multiple clusters, where data within the same cluster exhibit higher similarities, while data in different clusters show heterogeneity.

One of the most well-known clustering techniques is K-Means, which was proposed by Ref [21]. K-Means has the advantage of being conceptually simple and easy to apply and operate. Additionally, it is widely used in clustering due to its ability to improve the prediction accuracy of models [3, 9, 27]. The K-Means algorithm follows a specific process to determine the optimal clusters for the data. First, users specify the value of 'k,' which represents the initial number of clusters. Each data sample is then assigned to the nearest cluster center point. By calculating the average distance of data samples within each cluster, the model updates the cluster center points iteratively to minimize the average distance. This iterative process may result in some data points moving from one cluster to another. The model repeats this process until either the data's cluster arrangement is finalized or the change in the cluster center points becomes negligible.

K-means uses the Euclidean distance formula to calculate the distance between each data sample and the cluster center. The formula for Euclidean distance can be expressed as follows.

K-means uses the Euclidean distance formula to calculate the distance between each data sample and the cluster center. The formula for Euclidean distance can be expressed as follows.

By assuming the following vector x and y as follows, $X = (x_1, x_2, \dots, x_n)^T, Y = (y_1, y_2, \dots, y_n)^T$ the Euclidean distance between X and Y are expressed as below.

$$Euclidean\ Distance = \|X - Y\| = \sum_{i=1}^{n} (x_i - y_i)^2 \quad (1)$$

In the above process, each data sample is assigned based on the calculated distance, allowing users to identify the common characteristics of each cluster by observing the data samples located near the cluster center. Reference [15] demonstrated the utility of applying K-Means clustering techniques in time series analysis, using an example of detecting structural changes in the weighted stock index of the Taiwan Stock Exchange. K-means offers several advantages, including its ease of use, ability to handle duration, capability to distinguish temporary changes, and its prediction function. Leveraging these advantages, this study aims to enhance prediction performance by utilizing K-means for demand forecasting problems.

2.2. ElasticNet

Dealing with all the explanatory variables that affect the response variables involves several difficulties and constraints. These include reducing the model's performance, slowing down the model training process, increasing the complexity of the model, and raising the cost of training. Discovering the relationship between explanatory variables and response variables is an essential task in overcoming the challenges and constraints associated with high-dimensional data today [7].

ElasticNet offers an effective solution for complex high-dimensional data problems [45]. It is a hybrid regression model that combines the features of Ridge and Lasso models, incorporating both normalization from Ridge and variable selection from Lasso. Additionally, ElasticNet provides grouping effects that Lasso cannot achieve. While Lasso performs variable selection by shrinking the coefficients of less important variables to zero [32], ElasticNet identifies variables with high correlations and assigns similar weights to them. This process allows ElasticNet to decide whether to select or discard variables based on the importance of high correlation. The variable selection capability of ElasticNet is particularly useful in the presence of correlated explanatory variables. Therefore, in today's high-dimensional data with significant correlation between variables, ElasticNet can deliver outstanding performance.

While $Y = (y_1, y_2, \dots, y_n)^T$ is a response variable, $\mathrm{X} = (x_1, x_2, \dots, x_n)^T$ is an explanatory variable, and $X = (x_1|x_2| \dots |x_n)$ is the design matrix, the ElasticNet could be expressed in the formula below.

$$\hat{\beta} = argmin_{\beta}(Y - X\beta)(Y - X\beta)^T + (1 - \alpha)\sum_{i=1}^{n}|\beta_i| + \alpha\sum_{i=1}^{n}\beta^2{}_i \qquad (2)$$

$(1-\alpha)\sum_{i=1}^{n}|\beta_i| + \alpha\sum_{i=1}^{n}\beta^2{}_i$ in the above formula represents the loss function. At this point, the range of α is 0, then the ElasticNet becomes Lasso Model. By selecting the optimal value of α, the model can derive the greatest result.

2.3. Gaussian Process Regression (GPR)

Most forecasting problems in industries are complicated because they involve the unknown relationship between explanatory variables and response variables. Gaussian process regression (GPR) effectively addresses these complex problems by uncovering the hidden relationships [28]. Unlike parametric approaches, GPR tackles the regression problem in a non-parametric manner, utilizing a large number of parameters to capture the intricate relationships between variables. Moreover, GPR employs Bayesian Theory to determine the level of complexity [13, 39]. By providing an estimate of prediction uncertainty through probability distribution, GPR offers the advantage of achieving high performance by training parameters with the kernel [26].

The prediction process of GPR is as follows. First, assign the training latent values X and testing latent values $\acute{X}$ into the joining GP prior distribution. Second, a likelihood function, $p(Y|X)$ is combined to the Bayes' theorem to calculate the Gaussian process' joint prior distribution.

$$p(X, \acute{X}|Y) = \frac{p(X,\acute{X})p(Y|X)}{p(Y)} \qquad (3)$$

By removing unnecessary training values in this process, the desired distribution of Gaussian process is obtained.

$$p(X, \acute{X}) = N\left(0, \begin{bmatrix} K_{x,x} & K_{x,x} \\ K_{x,x} & K_{x,x} \end{bmatrix}\right) p(X, \acute{X}) = N(X, \sigma^2 I) \qquad (4)$$

In the above formula, K, I, σ^2 represent covariance function, identity matrix, and noise, respectively.

GPR uses a covariance matrix as a kernel's parameter, and it identifies similarities for all training values. The similarity calculated here is used in the process of granting the weight of the predictive model. Finally, the distribution above derives predicted values [39]. While GPR calculates similarities between data points and construct the model, its amount of computation increases as $O(n^3)$. n in this formula refers to the size of the training data, therefore, the amount of computation increases rapidly as the data size gets larger [26].

3. Model Construction

This study proposes an advanced hybrid model that accurately predicts product demand by combining K-means, ElasticNet, and GPR. The process of model construction is explained in the framework depicted in Figure 1. The prediction model is constructed following the procedure outlined below.

3.1. Model Formation

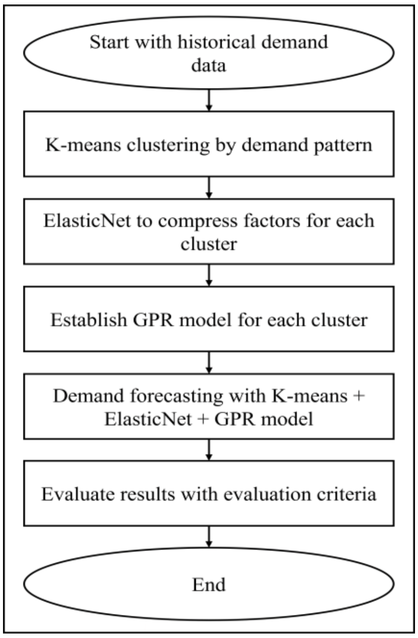

Fig. 1. Model Framework

[Step 1] First, the model construction starts by preparing the data. Once the historical data on product demand is prepared, the feature engineering technique is applied to create necessary variables and select appropriate ones among them.

[Step 2] The next step is to use K-means to divide the dataset into time-series clusters by considering several demand patterns. During this process, the model considers some specific properties of the demand pattern, such as strength of trend, strength of spikiness, strength of linearity, strength of curvature, ACF (Auto-correlation Function), and spectral entropy. As K-means need a direction of a specific number of clusters, this study uses Davies-Bouldin index, Silhouette coefficient, and Calinski-Harabasz criterion to calculate the optimal number of clusters [36].

[Step 3] After the data clustering according to time series characteristics, appropriate variables are selected for demand forecasting in each cluster using the ElasticNet. In this process, unimportant variables are removed, and independent variable sets are uniquely constructed for each cluster.

[Step 4] After the variable selection process through the ElasticNet, GPR prediction model trains each clustered data and construct a predictive model. This is the demand prediction model optimized for each cluster.

[Step 5] By using the hybrid model combining K-means, the ElasticNet, and GPR, the predicted value of future demand is calculated.

[Step 6] In order to analyze the performance of the predictive model, the results are evaluated according to the set criteria.

3.2. Model Evaluation

To confirm the superiority of the model proposed in this study, the prediction performance is evaluated as follows. The main evaluation criteria used in predictive models are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Root Mean Square Log Error (RMSLE). Each of these error metrics are calculated by the following equation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i^{pre} - Y_i| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^{pre} - Y_i)^2} \quad (6)$$

$$RMSLE = \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^{pre} - Y_i)^2}\right) \quad (7)$$

4. Case Study

4.1. Data

T This study designed a case analysis to demonstrate the effectiveness and feasibility of the proposed model. For the case experiment, this study prepared product demand data from US manufacturing companies. The company has sales channels all over the world and has production facilities in four different locations.

For the data set, one production facility was selected, and sales data was collected by items of 16 categories of products. The data is collected from January 5, 2014 to January 5, 2017 and it includes a total number of 244 products and 38,552 observations.

Input variables to predict weekly sales of products are additionally created through feature engineering techniques based on each item’s attributes and its time-series sales volume information. The variables are shown in the following table.

The training and validation data used the data from January 2014 to July 2016, and the test data used the data from August 2016 to January 2017 for a 20-week period.

Table 1. List of Independent Variables

Product Category	***Name of Variable***	***Type***	***Explanation***
Target Variable	Demand	int	Sales quantity
	Date	date	Sales date
Sales Property	Year	int	Sales year
	Month	int	Sales month
	Product_Category	str	Product Category
Product Property	Product_Code	str	Specific code
	Warehouse	str	4 types of warehouses
	lag1	int	weekly demand on time difference (1 week)
	lag2	int	weekly demand on time difference (2 weeks)
	lag3	int	weekly demand on time difference (3 weeks)
	lag4	int	weekly demand on time difference (4 weeks)
Feature Engineering	lag_mean	num	weekly demand average
	lag1_count	int	sales count on time difference (1 week)
	lag2_count	int	sales count on time difference (2 weeks)
	lag3_count	int	sales count on time difference (3 weeks)
	lag4_count	int	sales count on time difference (4 weeks)

lag_mean_count	num	weekly sales count average
change_lag1	int	sales difference by time(1 week)
change_lag2	int	sales difference by time (2 weeks)
change_lag3	int	sales difference by time (3 weeks)
sum_q	int	quarterly total sales
max_q	int	quarterly maximum sales
min_q	int	quarterly minimum sales
lag1year	int	weekly demand on time difference (1 year)
lag2year	int	weekly demand on time difference (2 years)
MovingMean_12	num	Moving average (12 weeks)
MovingMean_4	num	Moving average (4 weeks)
derived1	num	squared value of lag1
derived2	num	squared value of change_lag1
newProduct	num	new product or not (0.1)
recall	num	product recalled or not (0,1)
product_compete	num	number of competing products
upgrade	num	product upgraded or not (0,1)
derived3	num	squared value of lag2
derived4	num	squared value of change_lag2

4.2. Model Construction

This study conducted an experiment to predict product demand using the proposed hybrid model of K-means + ElasticNet + GPR. The parameters applied to this model are shown in Table 2. To compare the performance of the K-means, ElasticNet, and GPR models, several benchmarking models were set up as follows. First, previous research mainly focused on constructing a single GPR model for demand prediction. Using this single GPR model as a benchmarking model allows us to validate the effectiveness of clustering and feature selection in our research. The second benchmarking model is K-means + GPR, excluding ElasticNet. This benchmarking model validates the feature selection function of the model. The third benchmarking model is ElasticNet + GPR, excluding K-means. This model confirms the validity of the clustering technique.

Table 2. List of Parameters by Model

Model	***Parameters***
K-means	center = 3, nstart = 100

ElasticNet	Cluster1: alpha = 0.005 lambda = 2029.042 Cluster2: alpha = 0.010 lambda = 754.9816 Cluster3: alpha = 0.014 lambda = 2032.402
GPR	kernel = Vanilladot (linear) scaled = True

4.3. Experimental Result and Analysis

4.3.1. K-means Result

As a result of identifying the optimal number of clusters using the K-means algorithm, the entire dataset could be divided into three similar demand patterns.

Fig. 2 is the result of visualizing the demand pattern of each cluster using the sales data for 20 weeks.

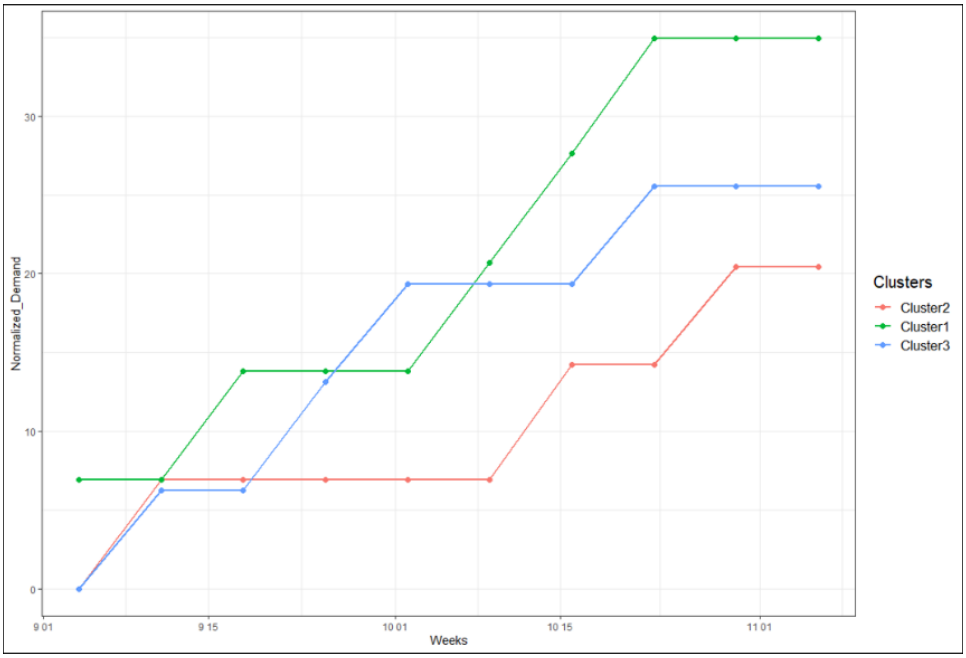

Fig. 2. Cumulated Demand Pattern for Each Cluster

4.3.2. ElasticNet Result

After clustering, the ElasticNet algorithm begins optimal variable selection for each cluster. Firstly, a grid search technique is used to find the optimal ratio, α, between the two penalty functions in the ElasticNet. The evaluation criterion is the RMSE value of each ElasticNet model built with a specific α value and the number of selected variables.

The value of α is sequentially applied from 0 to 1 at intervals of 0.001, and the penalty degree is set to the value that minimizes the Mean Square Error when cross-validating each ElasticNet model. The results of the RMSE and the number of variable selections for each cluster are shown in Fig. 3.

This process of searching for the optimal α is repeated for all three clusters, selecting an appropriate α value that minimizes data loss and RMSE for each cluster. The selected variables for each cluster are presented in Table 3 below.

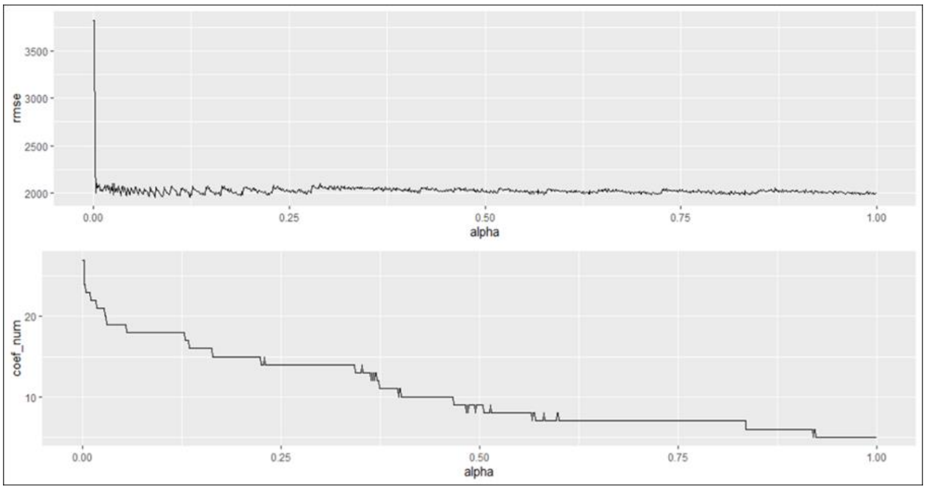

Fig. 3. ElasticNet Variable Selection Process

Table 3. ElasticNet List of Selected Variables

Cluster	***Num_variable***	***Selected Variables***
Cluster1	24	Product_Code, Date, Demand, lag1, lag2, lag3, lag4, lag_mean, lag1_count, lag2_count, lag3_count, lag4_count, change_lag1, change_lag2, change_lag3, max_q, min_q, lag1year, lag2year, sum_q, MovingMean_12, MovingMean_4, derived1, derived2
Cluster2	23	Product_Code, Date, Demand, lag1, lag2, lag3, lag4, lag_mean, lag1_count, lag3_count, lag4_count, change_lag1, change_lag2, change_lag3, max_q, min_q, lag1year, lag2year, sum_q, MovingMean_12, MovingMean_4, derived1, derived2
Cluster3	24	Product_Code, Date, Demand, lag1, lag2, lag3, lag4, lag_mean, lag1_count, lag2_count, lag3_count, change_lag1, change_lag2, change_lag3, max_q, min_q, lag_mean_count, lag1year, lag2year, sum_q, MovingMean_12, MovingMean_4, derived1, derived2

4.3.3. GPR Result

To train and test demand prediction, a GPR model is constructed for each data cluster. During this process, the proposed hybrid model is compared with benchmarking models, including the single GPR, K-means + GPR, and ElasticNet + GPR models, to evaluate its prediction performance. The performance of each model using RMSE, MAE, and RMSLE is presented in the table below. Since these are error indicators, lower values indicate better prediction performance.

Upon analysis, the K-means + ElasticNet + GPR hybrid model proposed in this study demonstrated the highest prediction accuracy. The effectiveness of each algorithm used in the hybrid model was validated by comparing it with other models. The results confirm that training product data by clusters with similar demand patterns through K-means is more effective than training with a single GPR model using the entire dataset. Additionally, it is evident that reducing dimensions by identifying key variables through ElasticNet contributes to improved model performance compared to using all variables. In conclusion, this experiment confirms that the K-means + ElasticNet + GPR model, which clusters the dataset and selects key variables for the training process, achieves the best prediction performance.

Table 4. Model Evaluation

Model	*RMSE*	*MAE*	*RMSLE*
GPR	18.049	6.548	2.256
K-means + GPR	16.912	6.057	1.228
ElasticNet + GPR	17.855	5.766	1.251
K-means + ElasticNet + GPR	**16.905**	**5.569**	**1.228**

These hybrid models not only improve prediction accuracy but also enhance efficiency by reducing computational resources and time consumption required for the training process. In other words, they can be considered as excellent models that overcome the limitations of Gaussian process regression, which involves exponential computations despite its high prediction performance.

5. Conclusions

As customer demand is uncertain and the market is complex, effectively responding to irregular market conditions can be challenging. Therefore, accurate demand forecasting methods have been in high demand. Despite recent attempts to forecast demand using data science [29, 30], existing prediction models still have limitations in accurately predicting nonlinear demand patterns. To address these limitations, this study proposed a hybrid model, K-means + ElasticNet + GPR, which combines excellent models to enable accurate demand forecasting.

GPR is a highly effective model for handling time series forecasting problems, particularly nonlinear ones. It can accurately predict complex situations compared to general machine learning models. However, GPR has the drawback of significantly increasing computation requirements as the data size grows. Consequently, relying solely on a single GPR model can result in substantial time and cost during the training process, depending on the data size. Moreover, a single GPR model has limitations in accurately predicting diverse demand patterns and item compositions.

To overcome these challenges, this study adopted a method that maximizes model performance by constructing an efficient and effective data structure using K-means and ElasticNet. K-means divides the time-series dataset into clusters with similar characteristics, and ElasticNet is employed to select important variables for each cluster. By utilizing these techniques, GPR can effectively train the model and establish a hybrid machine learning model.

Through an analysis conducted on a specific case of a U.S. manufacturing company, it was confirmed that the K-means + ElasticNet + GPR model presented in this study outperformed other benchmarking models, such as single GPR, K-means + GPR, and ElasticNet + GPR.

This study holds several significant contributions. Firstly, it is the first instance where the combination of K-means + ElasticNet + GPR has resulted in an improved performance in demand forecasting. This achievement is noteworthy as it introduces a new approach to advance the field of demand forecasting research and presents an innovative model that can enhance business performance in the industry.

Additionally, the utilization of K-means for clustering time series data reaffirms the findings of previous studies regarding the effectiveness of this method in identifying and leveraging data characteristics. The study further validates the efficacy of this model by demonstrating that the accuracy and efficiency of the demand forecasting process can be enhanced through the selection and normalization of variables using ElasticNet.

Lastly, this study highlights the strengths of a hybrid approach that judiciously combines models with their respective advantages. While individual machine learning models have their limitations, a hybrid model exhibits a crucial advantage by overcoming the constraints of a specific model and creating synergies through the integration of different models. The research reinforces the notion that a hybrid model can generate substantial synergistic effects, particularly in the context of uncertain demand forecasting.

On the other hand, this study has some limitations as well. First, the predictive model was constructed using a specific case of sales data from U.S. manufacturers. To ensure the generalizability of the model, it is essential to incorporate data from other situations or industries for testing. Furthermore, it is crucial to provide guidance on appropriate clustering techniques and dimension reduction techniques for different industries, considering that data characteristics vary across sectors. Depending on the nature of variables and the complexity of items, the significance of clustering or dimension reduction techniques may vary in the model. Exploring these aspects will be a future direction for further development and refinement of the ideas presented in this study.

Acknowledgements

This research was supported by Impactive-AI. We thank Impactive-AI for providing research motivation, insights

and expertise that greatly assisted the research.

Author contributions

Doohee Chung: Conceptualization, Project administration, resources, software,supervision, validation, Writing & Editing

ChanGyu Lee: Software, Methodology, Investigation, Visualization, Writing & Editing

Sungmin Yang: Software, Methodology, Investigation, Writing, Investigation, Writing

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] F. Amini and G. Hu, "A two-layer feature selection method using genetic algorithm and elastic net," *Expert Systems with Applications*, vol. 166, 2021, 114072.

[2] Y. Aviv, "On the benefits of collaborative forecasting partnerships between retailers and manufacturers," *Management Science*, vol. 53, no. 5, 2007, pp. 777-794.

[3] P. K. Bala, "Improving inventory performance with clustering based demand forecasts," *Journal of Modelling in Management*, vol. 7, no. 1, 2012, pp. 23-37.

[4] K. Bandara, P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran and B. Seaman, "Sales demand forecast in e-commerce using a long short-term memory neural network methodology," in *International conference on neural information processing*, vol. 11955, 2019, pp. 462-474.

[5] Y. Bengio, O. Delalleau and N. Roux, "The curse of highly variable functions for local kernel machines," *Advances in neural information processing systems*, vol. 18, 2005.

[6] J. Bibaud-Alves, P. Thomas and H. Bril, "Demand forecasting using artificial neuronal networks and time series: application to a French furniture manufacturer case study," in *11th International Joint Conference on Computational Intelligence, IJCCI'19 (11th International Conference on Neural Computation Theory and Applications, NCTA)*, vol. 1, 2019.

[7] V. Bolón-Canedo, N. Sánchez-Maroño and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and information systems*, vol. 34, no. 3, 2013, pp. 483-519.

[8] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *Journal of Artificial Intelligence Research*, vol. 70, 2021, pp. 245-317.

[9] I. F. Chen and C. J. Lu, "Demand Forecasting for Multichannel Fashion Retailers by Integrating Clustering and Machine Learning Algorithms," *Processes*, vol. 9, no. 9, 2021, 1578.

[10] E. Choi, S. Cho, and D. K. Kim, "Power demand forecasting using long short-term memory (LSTM) deep-learning model for monitoring energy sustainability," *Sustainability*, 12(3) (2020) 1109.

[11] M. Christopher, and H. Lee, "Mitigating supply chain risk through improved confidence," *International journal of physical distribution & logistics management*, vol. 34, no. 5, 2004.

[12] D. Dzyabura, and S. Jagabathula, "Offline assortment optimization in the presence of an online channel," *Management Science*, vol. 64, no. 6, 2018, pp. 2767-2786.

[13] S. J. Gershman, and D. M. Blei, "A tutorial on Bayesian nonparametric models," *Journal of Mathematical Psychology*, 56(1) (2012) 1-12.

[14] L. Guo, W. Fang, Q. Zhao, and X. Wang, "The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality," *Computers & Industrial Engineering*, vol. 161, 2021, 107598.

[15] K. H. Huarng, T. H. K. Yu, and T. T. Kao, "Analyzing structural changes using clustering techniques," *International Journal of Innovative Computing, Information and Control*, vol. 4, no. 5, 2008, pp. 1195-1202.

[16] S. Ji, X. Wang, W. Zhao, and D. Guo, "An application of a three-stage XGBoost-based model to sales forecasting of a cross-border e-commerce enterprise," *Mathematical Problems in Engineering*, vol. 2019, 2019.

[17] M. A. Khan, S. Saqib, T. Alyas, A. U. Rehman, Y. Saeed, A. Zeb, M. Zareei and E. M. Mohamed, "Effective demand forecasting model using business intelligence empowered with machine learning," *IEEE Access*, vol. 8, 2020, pp. 116013-116023.

[18] K. Krauth, E. V. Bonilla, K. Cutajar and M. Filippone, "AutoGP: Exploring the capabilities and limitations of Gaussian process models," *arXiv preprint arXiv:1610.05392*, vol. 1, 2016.

[19] B. Lakshmanan, P. S. N. Vivek Raja and V. Kalathiappan, "Sales demand forecasting using LSTM network," in *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, vol. 1056, 2020, pp. 125-132.

[20] R. Law, G. Li, D. K. C. Fong, and X. Han, "Tourism demand forecasting: A deep learning

approach," *Annals of tourism research*, vol. 75, 2019, pp. 410-423.

[21] J. MacQueen, "Classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Statist. Probability*, vol. 1, 1967, pp. 281-297.

[22] M. Nagashima, F. T. Wehrle, L. Kerbache and M. Lassagne, "Impacts of adaptive collaboration on demand forecasting accuracy of different product categories throughout the product life cycle," *Supply chain management: an international journal*, vol. 20, no. 4, 2015.

[23] J. Noh, H. J. Park, J. S. Kim and S. J. Hwang, "Gated recurrent unit with genetic algorithm for product demand forecasting in supply chain management," *Mathematics*, vol. 8, no. 4, 2020, 565.

[24] S. Punia, K. Nikolopoulos, S. P. Singh, J. K. Madaan and K. Litsiou, "Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail," *International journal of production research*, vol. 58, no. 16, 2020, pp. 4964-4979.

[25] J. Quinonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *The Journal of Machine Learning Research*, vol. 6, 2005, pp. 1939-1959.

[26] J. Quinonero-Candela, C. E. Rasmussen and C. K. Williams, "Approximation methods for Gaussian process regression," in *Large-scale kernel machines*, MIT Press, Cambridge, MA, USA, 2007, pp. 203-223.

[27] T. Salloom, O. Kaynak and W. He, "A novel deep neural network architecture for real-time water demand forecasting," *Journal of Hydrology*, vol. 599, 2021, 126353.

[28] E. Schulz, M. Speekenbrink and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, vol. 85, 2018, pp. 1-16.

[29] M. Seyedan, and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," *Journal of Big Data*, vol. 7, no. 1, 2020, pp. 1-22.

[30] T. Tanizaki, T. Hoshino, T. Shimmura and T. Takenaka, "Demand forecasting in restaurants using machine learning and statistical analysis," *Procedia CIRP*, vol. 79, 2019, pp. 679-683.

[31] S. Thomassey, "Sales forecasts in clothing industry: The key success factor of the supply chain management," *International Journal of Production Economics*, vol. 128, no. 2, 2010, pp. 470-483.

[32] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, 1996, pp. 267-288.

[33] C. E. Rasmussen and C. K. Williams, "Gaussian processes for machine learning," MIT Press, Cambridge, MA, USA, 2006, pp. 4.

[34] L. Rokach, and O. Maimon, "Clustering methods," in *Data mining and knowledge discovery handbook*, Springer, Berlin, 2005, pp. 321-352.

[35] M. Rosienkiewicz, E. Chlebus and J. Detyna, "A hybrid spares demand forecasting method dedicated to mining industry," *Applied Mathematical Modelling*, vol. 49, 2017, pp. 87-107.

[36] J. Van Belle, T. Guns and W. Verbeke, "Using shared sell-through data to forecast wholesaler demand in multi-echelon supply chains," *European Journal of Operational Research*, vol. 288, no. 2, 2021, pp. 466-479.

[37] R. M. Van Steenbergen and M. R. Mes, "Forecasting demand profiles of new products," *Decision support systems*, 139 (2020) 113401.

[38] F. Wang, M. Li, Y. Mei and W. Li, "Time series data mining: A case study with big data analytics approach," *IEEE Access*, vol. 8, 2020, pp. 14322-14328.

[39] Y. Wang, B. Feng, Q. S. Hua and L. Sun, "Short-term solar power forecasting: A combined long short-term memory and gaussian process regression method," *Sustainability*, vol. 13, no. 7, 2021, 3665.

[40] C. K. Williams, "Prediction with Gaussian processes: From linear regression to linear prediction and beyond," in *Learning in graphical models*, vol. 89, 1998, pp. 599-621.

[41] M. Yedla, S. R. Pathakota and T. M. Srinivasa, "Enhancing K-means clustering algorithm with improved initial center," *International Journal of computer science and information technologies*, vol. 1, no. 2, 2010, pp. 121-125.

[42] Y. Zhang, G. Li, B. Muskat and R. Law, "Tourism demand forecasting: A decomposed deep learning approach," *Journal of Travel Research*, vol. 60, no. 5, 2021, pp. 981-997.

[43] Z. Zhang, H. Qin, Y. Liu, Y. Wang, L. Yao, Q. Li, J. Li and S. Pei, "Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction," *Energy Conversion and Management*, vol. 192, 2019, pp. 37-51.

[44] Z. Zhang, C. Wang, X. Peng, H. Qin, H. Lv, J. Fu and H. Wang, "Solar Radiation Intensity Probabilistic Forecasting Based on K-Means Time Series Clustering and Gaussian Process Regression," *IEEE Access*, vol. 9, 2021, pp. 89079-89092

[45] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, 2005, pp. 301-320.