Methodology for Improving the Performance of Demand Forecasting Through Machine Learning

Daeyoung Yoon
AI Convergence & Entrepreneurship, Handong Global University
Seonyeong Park
AI Convergence & Entrepreneurship, Handong Global University
Yechan Song
AI Convergence & Entrepreneurship, Handong Global University
Jihun Chae
AI Convergence & Entrepreneurship, Handong Global University
Doohee Chung (profchung@handong.edu)
AI Convergence & Entrepreneurship, Handong Global University

Research Article

Keywords: Demand forecasting, K-means, LASSO, LSTM, Hybrid Model

Posted Date: March 3rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2637740/v1

Additional Declarations: No competing interests reported.

Methodology for Improving the Performance of Demand Forecasting Through Machine Learning

Daeyoung Yoon*, Seonyeong Park*, Yechan Song*, Jihun Chae*, Doohee Chung**

**AI Convergence & Entrepreneurship, Handong Global University, Pohang, South Korea*

***Corresponding author, AI Convergence & Entrepreneurship, Handong Global University, Pohang, South Korea*

Methodology for Improving the Performance of Demand Forecasting Through Machine Learning

ABSTRACT

Accurate demand forecasting is crucial for industries to make strategic decisions and maintain their competitive edge. However, existing demand forecasting methods have prodigious problems, especially when it comes to handling the uncertainty, complexity, and nonlinearity of demand forecasting. In addition, the lack of historical data and data biases can create unreliable sources, which discourages the utilization of demand forecasting at a higher level of implementation in businesses. In addition, lack of historical data and data biases can create unreliable sources, which discourages utilization of demand forecasting at higher level of implementation in businesses. The proposed hybrid model aims to improve demand forecasting performance by combining the strengths of existing methods such as K-means clustering, LASSO regression, and LSTM deep learning. By leveraging these techniques, the model can overcome the limitations of each method and improve the accuracy of demand forecasting in various industries. K-means clustering helps to group similar data points, LASSO regression helps to select the most relevant features, and LSTM deep learning helps to capture the temporal dependencies in the data. The combination of these techniques can result in a more accurate and robust demand forecasting model. The model was tested on 2,548 retail products, and outperformed three benchmarking models using the mMAPE, RMSE, and MAE indicators. The proposed model can be used in the retail industry to improve management performance and decision-making, and its ability to optimize variables for each cluster can improve resource allocation.

Subject Areas: Demand forecasting, K-means, LASSO, LSTM, Hybrid Model

INTRODUCTION

Especially in business operations, accurate demand forecasting is essential for enhancing competitiveness, which creates the foundation for key strategic decisions for companies (Bandara, Shi, Bergmeir, Hewamalage, Tran, & Seaman, 2019). The process of estimating the future demand for a product or service, based on historical data, trends, and other relevant factors helps companies to grasped hold their plannings on inventory management, production planning, and promotions (Bohanec, Borštnar, & Robnik-Šikonja, 2017). The task of discovering a more precise method for demand forecasting has been of significant importance to numerous companies and scholars.

Traditional demand forecasting algorithms, such as regression, exponential smoothing (ETS), and auto-regressive integrated moving average (ARIMA), rely on linear movements to make predictions. However, the demand data in many industries often has complex, nonlinear characteristics, making it difficult to accurately forecast using these methods. To address this challenge, machine learning techniques, such as artificial neural networks (ANN), recurrent neural networks (RNN), and support vector machines (SVM), have been introduced as a more flexible and accurate way to forecast nonlinear and complex data. For example, Chalapathy and Chawla (2019) used ANN to develop a demand forecasting model for Walmart, a multinational retailer.

Carbonneau, Laframboise, and Vahidov (2008) found that RNN and SVM models had the best predictive accuracy when applied to forecasting distorted demand in manufacturing supply chain management. Faber and Finkenrath (2021) proposed a model for predicting heat demand for heating using existing decision tree-based regression algorithms such as AdaBoost and Random Forest (RF), as well as ANN, feed forward neural networks (FNN), and Long Short-Term Memory (LSTM). They also proposed an optimized power plant layout method based on this model. Ermawanto and Kurniati (2021) implemented Pareto analysis to identify and prioritize problems and tasks for an efficient inventory management method for spare parts required for maintenance in the process industry. They reviewed demand patterns using four methods and ADI-CV according to failure probability and restocking period. The reliability-based method had the most influence on the demand pattern and, unlike the other methods, it considers the standard deviation for each item (Wang & Wang, 2021).

There are limitations in existing machine learning-based demand forecasting methods. These methods often fail to capture the unique characteristics of individual items or categories, leading to less accurate predictions. Additionally, these methods require a lot of time and memory to create and verify features through feature engineering, making them difficult to implement in practice. Furthermore, these methods only perform well on data that does not show specific patterns or have large deviations. As a result, research has shifted towards developing hybrid models that strategically combine multiple models to achieve better performance than single models (Ren, Chan, & Siqin, 2020). These hybrid models aim to combine the strengths of different models to create synergies and overcome the disadvantages of individual models. A hybrid model is a combination of different models that are used together to improve predictive performance. These models can include preprocessing, feature selection, parameter optimization, clustering, error correction, and postprocessing. For example, Zhang, Li, Muskat, and Law (2021) proposed a hybrid model combining Seasonal and Trend decomposition using Loess (STL) and Dynamic Aggregate Demand Learning Model (DADLM) to predict tourism demand. The STL-DADLM model was able to overcome the problem of overfitting and outperformed other models such as Dynamic Linear Model (DLM), Artificial Neural Network (ANN), Autoregressive Integrated Moving

Average (ARIMA), Support Vector Regression (SVR), Extreme Gradient Boosting Tree Regressor (XGBTR), Error, Trend, Seasonality (ETS), and Naïve models.

Sun, Huang, Wong, and Jang (2016) proposed a model that improves the accuracy of a single Multi-Layer Perceptron (MLP) by combining it with a Least Absolute Shrinkage and Selection Operator (LASSO). The study showed that the combination of these two models allowed for more precise reduction of the MLP's input weights, and that it could be applied more easily to nonlinear industrial processes. Van Steenbergen and Mes (2020) attempted to predict demand for new products by combining K-means clustering, RF, and Quantile Regression Forest (QRF). This model clustered products based on their demand patterns and was able to predict demand for new products without using past data by applying a quantile regression-type RF model.

METHODOLOGIES

The hybrid model aims to enhance the performance of demand forecasting by leveraging the strengths of each method. K-means clustering is used to group similar products based on demand patterns, allowing the model to identify similarities and differences in demand behaviors across different product groups. LASSO regression is used to reduce the input weights of the LSTM model, which helps to improve the accuracy of the model by reducing the impact of irrelevant features in the input data. LSTM deep learning is then used to capture the long-term dependencies in demand data and make predictions based on the identified patterns. LSTM is particularly effective at modeling sequential data, which makes it well-suited for demand forecasting, where the data is often temporal in nature. By combining these three techniques, the proposed hybrid model can better capture the complex relationships between demand data and other relevant factors, resulting in more accurate and robust demand forecasting performance in various industries.

K-Means

K-means is a widely used clustering algorithm for its simplicity and effectiveness, since it does not need labeled data and automatically structures particular patterns in the data. Thus, it can be utilized to identify key demand patterns and unique characteristics into clusters based on this pattern. The ultimate goal of K-means is to convene similarities into different chunks of clusters and dissemble the data by their groups.

Niennattrakul and Ratanamahatana (2007) used the K-means-ANN algorithm to predict PM10 and PM2.5 concentrations at coastal locations in New Zealand. By using K-means to identify key factors influencing concentration patterns, they were able to develop a model that accurately predicted concentrations based on meteorological parameters. This demonstrates the effectiveness of using K-means to improve the performance of machine learning models.

Vadyala, Betgeri, Sherer, and Amritphale (2021) proposed a time series model that combined K-means clustering with LSTM to predict the number of COVID-19 patients. The proposed model used K-means to classify data based on regional characteristics, and then used LSTM to make predictions based on the classified data. The model showed better performance than the Susceptible-Exposed-Infectious-Recovered model (SEIR), which is commonly used to predict the spread of infectious diseases. This demonstrates the usefulness of using K-means to improve the performance of machine learning models for time series prediction.

The K-means algorithm is a method for dividing a dataset of N dimensions into K similar clusters. To use the algorithm, K cluster centers are randomly chosen from the dataset. Then, for each data value in the dataset, the distances to each of the *K* cluster centers are calculated. The data value is assigned to the cluster with the closest center. The distance between two points is typically calculated using the Euclidean distance, which is the straight-line distance between two points in *N*-dimensional space. The formula for Euclidean distance is shown in Equation 1 below:

Equation 1 (1.1)

$$\|p - q\| = \sqrt{(p - q) \cdot (p - q)} = \sqrt{\|p\|^2 + \|q\|^2 - 2p \cdot q}$$

$$p = (p_1, p_2, \dots, p_n) \;\&\; q = (q_1, q_2, \dots, q_n)$$

where (*p, q*) is the Euclidean distance between points *p* and *q, q1, q2*, ..., q_n are the coordinates of point *q*, and *p1*, *p2*, ..., p_n are the coordinates of point p.

The center point is recalculated based on the assigned data. The process above is repeated until the cluster to which each data value belongs does not change. The process is to finally find *Si* that minimizes the value of Equation 2 below. μ_i is the center of the *ith* cluster, and *Si* is the set of points belonging to the cluster.

Equation 2 (1.2)

$$V = \sum_{i=1}^{k} \sum_{x_i \in S_i} \left| x_j - \mu_i \right|^2$$

LASSO

The utilization of numerous explanatory variables to predict response variables, such as sales volume, can lead to a reduction in the model's learning rate and an increase in its complexity, thus impeding performance. To combat this issue, a plethora of methods have been devised for identifying critical variables within high-dimensional data sets and selecting a subset of these variables for model use. Among these techniques, LASSO, originally proposed by Tibshirani (1996), holds a prominent position. By integrating the least squares method that minimizes the predictive model's error with a formula that encompasses the sum of absolute values of each coefficient in the model, LASSO efficiently distinguishes and elects the most important variables in high-dimensional data.

LASSO is distinguished by its capacity to reduce the coefficients of less significant variables to zero, a capability that enables the selection of a subset of critical variables to be utilized in the model. The application of this process yields notable improvements to the model's performance, particularly by streamlining its complexity and optimizing its learning rate.

Equation 3 (2.1)

$$min_{\beta_1,\beta_2,\ldots,\beta_N}\left\{\sum_{i=1}^{N}\left(Y_i - \beta_0 - \sum_{j=1}^{N}\beta_j X_{ij}\right)^2 + \lambda\sum_{j=1}^{N}\left|\beta_j\right|\right\}$$

At the core of the LASSO model is the tuning parameter lambda (λ), which regulates the extent of the penalty imposed upon the coefficients of the explanatory variables. As the value of λ diminishes, the model selects an increasing number of variables, thereby augmenting the model's complexity. Conversely, as λ rises, the regression coefficients of less consequential variables tend towards zero, leading to a decrease in model complexity. By adeptly manipulating the value of λ, LASSO selectively identifies and utilizes only the most pertinent variables in the final model, improving the performance of the model by streamlining its complexity and enhancing its learning rate.

The LASSO regression analysis seeks to estimate the regression coefficients beta (β) by determining the β values that minimize the expression in Equation 3. Since β_0 relies on the estimated value of the response variable and therefore remains fixed, the remaining coefficients β_1 , β_2 , ..., β_N are judiciously selected to minimize the expression for β_N . This mechanism of coefficient selection empowers LASSO to effectively identify and distinguish the most pivotal variables in the model, while also enabling the coefficients of less consequential variables to be diminished to zero. In accomplishing this, LASSO serves to enhance the performance of the model by simplifying its complexity and augmenting its learning rate.

LSTM

Deep learning represents a subtype of machine learning that has demonstrated proficiency in achieving favorable results on an array of prediction tasks. Among the deep learning models that have been utilized for demand forecasting, Recurrent Neural Network (RNN) has been custom-built to analyze sequential data. However, RNN suffers from a deficiency in its capacity to accurately capture long-term dependencies in the data, which is crucial for achieving precise predictions. In an effort to overcome this challenge, Hochreiter and Schmidhuber (1997) proposed the LSTM model. The fundamental feature of LSTM is the inclusion of gates, which permits the model to record and retain information from previous steps while also regulating the flow of information through the network. This property enables LSTM to extend its memory and more effectively capture long-term dependencies in the data. Since its inception, researchers have advanced LSTM through various variations, such as LSTM without a forget gate, LSTM with a forget gate, and LSTM with a peek connection. These variations of LSTM have been demonstrated to enhance the memory capacity and performance of the model across various tasks.

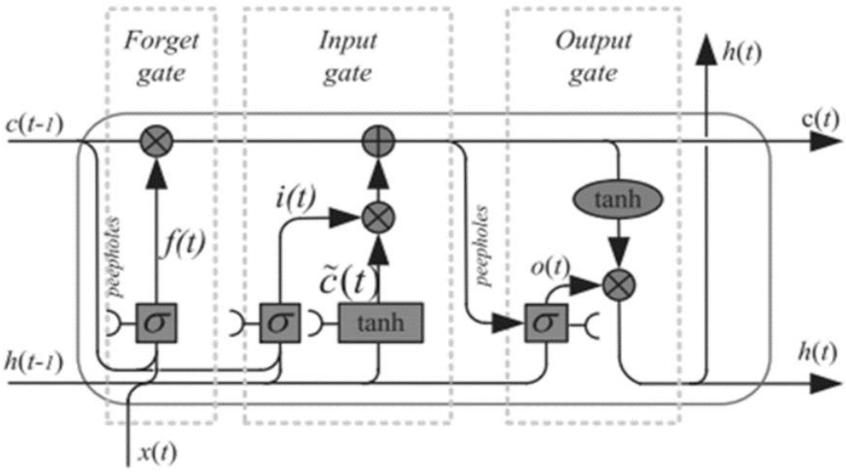

Figure 1: LSTM Framework

LSTM networks leverage a forgetting gate to retain relevant past information for a specific period. The forgetting gate decides the amount of past information to retain by computing the product and sum of weights $\boldsymbol{W_{ih}}$ and $\boldsymbol{W_{ix}}$, respectively, with the present information $\boldsymbol{x_t}$ and hidden layer value $\boldsymbol{h_{t-1}}$. The output value is then passed through a sigmoid function to determine the information that is to be retained, which is then multiplied by $\boldsymbol{c_{t-1}}$. The closer the forgetting gate value $\boldsymbol{f(t)}$ is to 1, the more information is retained, and the closer it is to 0, the more information is discarded.

The input gate in LSTM networks plays a crucial role in deciding which new information to store in the cell state. It does this by applying a sigmoid function and a hyperbolic tangent function to determine the significance of the current input candidate $\boldsymbol{x_t}$. These results are then employed to update the cell state, accounting for the forgetting gate decisions taken in the previous step, which aids the LSTM network in storing and accessing vital information efficiently over time.

Finally, the LSTM process computes the output value by applying a sigmoid function to the input data and multiplying the result by the cell state's output, which is passed through a hyperbolic tangent function. This generates the output value $\boldsymbol{h_t}$, which serves as the value of the past hidden layer for the subsequent cell. This mathematical operation is represented as Equation 4.

Equation 4 (3.1)

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i)$$

$$\tilde{c}_t = \tanh\,(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}})$$

$$c_t = c_{t-1} + i_t \cdot \tilde{c}_t$$

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o)$$

$$h_t = o_t \cdot \tanh\,(c_t)$$

The LSTM model has demonstrated efficacy as a versatile instrument across various prediction tasks, exhibiting superiority over competing approaches. For instance, the utilization of an LSTM network for short-term traffic prediction, as showcased in a study by Hochreiter and Schmidhuber (1997), resulted in superior performance compared to alternative algorithms. Furthermore, Mujeeb, Javaid, Ilahi, Wadud, Ishmanov, and Afzal (2019) demonstrated that the implementation of a deep LSTM neural network for predicting electricity prices and loads achieved enhanced performance relative to conventional neural networks. Despite the successes of the LSTM model, it has its limitations, such as its substantial memory bandwidth demands and challenging parallel processing due to its recursive nature. An approach that has emerged as a promising solution to circumvent these limitations is to merge the LSTM model with dimensionality reduction techniques, such as LASSO, to improve both efficiency and performance while retaining its ability to address the vanishing gradient problem.

PROPOSED MODEL

Model Framework

The proposed hybrid model combines K-means clustering, LASSO regression, and LSTM for demand forecasting. The framework of the model is shown in the Figure 2. This model leverages the strengths of each of these methods to improve the accuracy of demand forecasting. K-means is used to divide the data into clusters, LASSO is applied to reduce the dimensionality of the data and select relevant features, and LSTM is used to capture temporal patterns in the data and make predictions. This combination of methods allows the model to effectively handle large and complex datasets and make accurate demand forecasts. To prepare for modeling, demand data containing sales history information for each item is collected. This data includes basic variables such as item attributes and sales volume, as well as derived variables created through feature engineering to improve the accuracy of demand forecasting. Next, the K-means algorithm is used to perform clustering on the data. The K-means algorithm adjusts the centers of clusters to minimize the distance between them, and then performs clustering based on similarity. This allows data with high levels of similarity to be grouped into the same cluster. To determine the optimal number of clusters, techniques such as the Davies-Bouldin index, silhouette coefficient, Calinski Harabasz index, and majority vote can be used (Van Steenbergen & Mes, 2020). The model is then trained on the clustered data.

When the number of variables is large, multicollinearity among variables can reduce the accuracy of predictions. In addition, excessive calculations can increase time complexity and memory consumption. To address these issues, this study applies LASSO to select the key variables that play a significant role in prediction for each cluster. Once the relevant variables have been identified for each cluster using LASSO, an LSTM-based demand forecasting model is built for each cluster. This model is then used to make demand forecasts. The predictive performance of the model is evaluated according to a set of criteria to assess its accuracy and reliability.

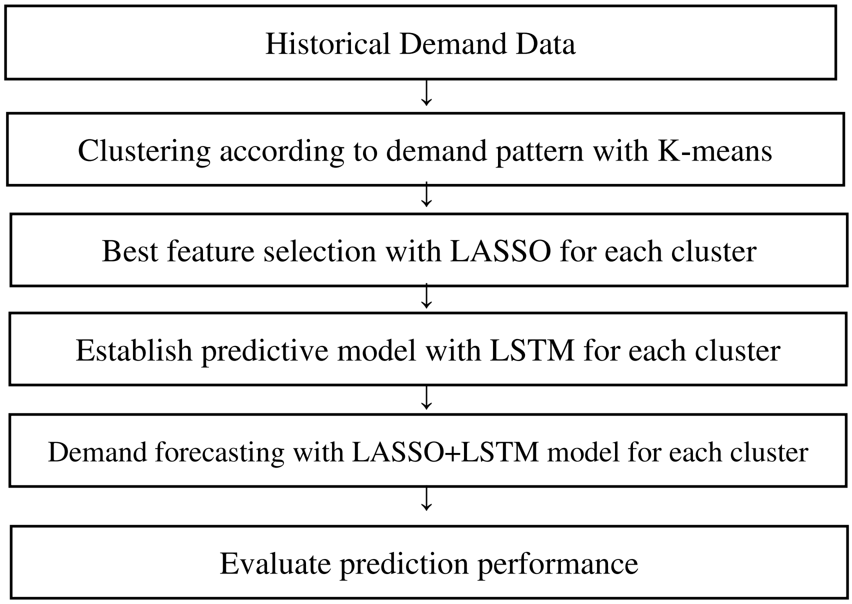

Figure 2: Proposed model framework

Data

The present study employs product demand data from US retail companies, spanning from January 1, 2014 to March 26, 2016, and containing weekly unit demand data, culminating in a total of 116 weekly demand data points. The final dataset, with a size of 2,548 products and 312,388 observations, is utilized to predict the demand for each product over the course of 12 weeks. The training dataset comprises 104 weeks of demand data over a 2-year period, while the verification dataset consists of 12 weeks of demand data for a 3-month period. These datasets are utilized for both training the model and evaluating its performance.

Evaluation Index

In order to assess the efficacy of the proposed predictive model, three metrics of error are employed: modified Mean Absolute Percentage Error (mMAPE), Root Mean Square Error

(RMSE), and Mean Absolute Error (MAE). Among these, mMAPE is a commonly used evaluation index in the retail and supply chain industries due to its ability to address the limitations of the Mean Absolute Percentage Error (MAPE) index, which struggles to provide accurate assessment when the actual value is 0. The formula for mMAPE is as follows, where $\boldsymbol{i}$ represents the time at which the actual demand occurs, $\boldsymbol{y_i}$ refers to the predicted demand value for the corresponding time $\boldsymbol{i}$, $\boldsymbol{y_i}$ represents the actual demand value, and n denotes the length of the entire demand period.

Equation 5 (4.1)

$$mMAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|\hat{y}_i - y_i|}{1 + |y_i|}\right)$$

Root Mean Square Error (RMSE) serves as a frequently used assessment metric for discerning the deviation between predicted and actual values in a model. RMSE is mathematically derived by computing the square root of the mean squared error (MSE) between predicted and actual values. The utilization of MSE in this context provides greater emphasis to substantial discrepancies, thereby offering a less vulnerable assessment of anomalous data points. The salient advantage of RMSE is its capacity to evaluate the performance of a model in handling anomalous data points, a capacity that is especially pertinent when outliers are present in the data.

Equation 6 (5.1)

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

Mean Absolute Error (MAE) is an evaluation metric frequently employed to measure the average absolute discrepancy between the predicted and actual values in a model. MAE's utility stems from its use of the absolute value of the error, which confers greater robustness to outliers than other measures such as Root Mean Square Error (RMSE). By employing MAE, one can more intuitively appraise the accuracy of the prediction model, making it a valuable tool for assessing overall performance.

Equation 7 (6.1)

$$MAE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}|\hat{y}_i - y_i|}$$

Benchmarking Settings

A comparative benchmarking model was set up to evaluate the performance of the K-means+LASSO+LSTM demand forecasting methodology proposed in this paper. First, we evaluate the predictive performance of a single LSTM model. Through this, the validity of the hybrid method can be verified. Next is K-means+LSTM. The effect of clustering can be confirmed when comparing this with the basic model, and the validity of LASSO can be verified when compared with the proposed model.

The LASSO+LSTM model integrates feature selection to the LSTM model, and its performance can be compared with the solitary LSTM model to assess the advantages of feature selection. Furthermore, this sub-model can be compared with the proposed hybrid model to ascertain the advantages of clustering in the hybrid model. By contrasting the performance of these benchmarking models with the proposed hybrid model, the overall effectiveness of the proposed methodology can be demonstrated.

Hyperparameter Settings

The K-means+LASSO+LSTM forecast model was established by selecting optimal parameters, which were determined by tuning using the training and validation datasets. Hyperparameters for the K-means model were derived using the Silhouette Score, while Grid Search and Early Stopping were employed for the LASSO and LSTM models. Table 1 presents the hyperparameters used in this study.

Table 1: Parameter setting

Model	Parameter	
K-means	centers = 3 nstart = 100	
LASSO	ALL	alpha = 0.0001 max_iter = 10000
	Cluster1	lambda = 0.0001 max_iter = 7775
	Cluster2	lambda = 0.00015 max_iter = 7740
	Cluster3	lambda = 0.00012 max_iter 2032.402
LSTM	scaled = True BATCH_SIZE = 128 BUFFER_SIZE = 10000 EVALUATION_INTERVAL = 200 validation_steps = 50 optimizer = RMSprop val_loss = mae early_stopping(min_delta = 0, patience = 2) EPOCHS = early_stopping	

RESULTS

Clustering Through K-means

The analysis using the K-means+LASSO+LSTM model yielded several noteworthy findings. The initial step in the analysis involved employing the K-means algorithm to cluster the 2,548 products in the dataset based on their time series characteristics and the Silhouette score. The resulting clustering produced three distinct clusters, as illustrated in Figure 3. Specifically, Cluster 1 contained 1,560 products that displayed an initial increase in sales volume followed by a relatively stable demand pattern. In contrast, Cluster 2 included 645 products that exhibited a pattern of relatively irregular changes in demand. Lastly, Cluster 3 contained 434 products with low levels of demand that displayed large fluctuations at the beginning, but a pattern that gradually became smoother.

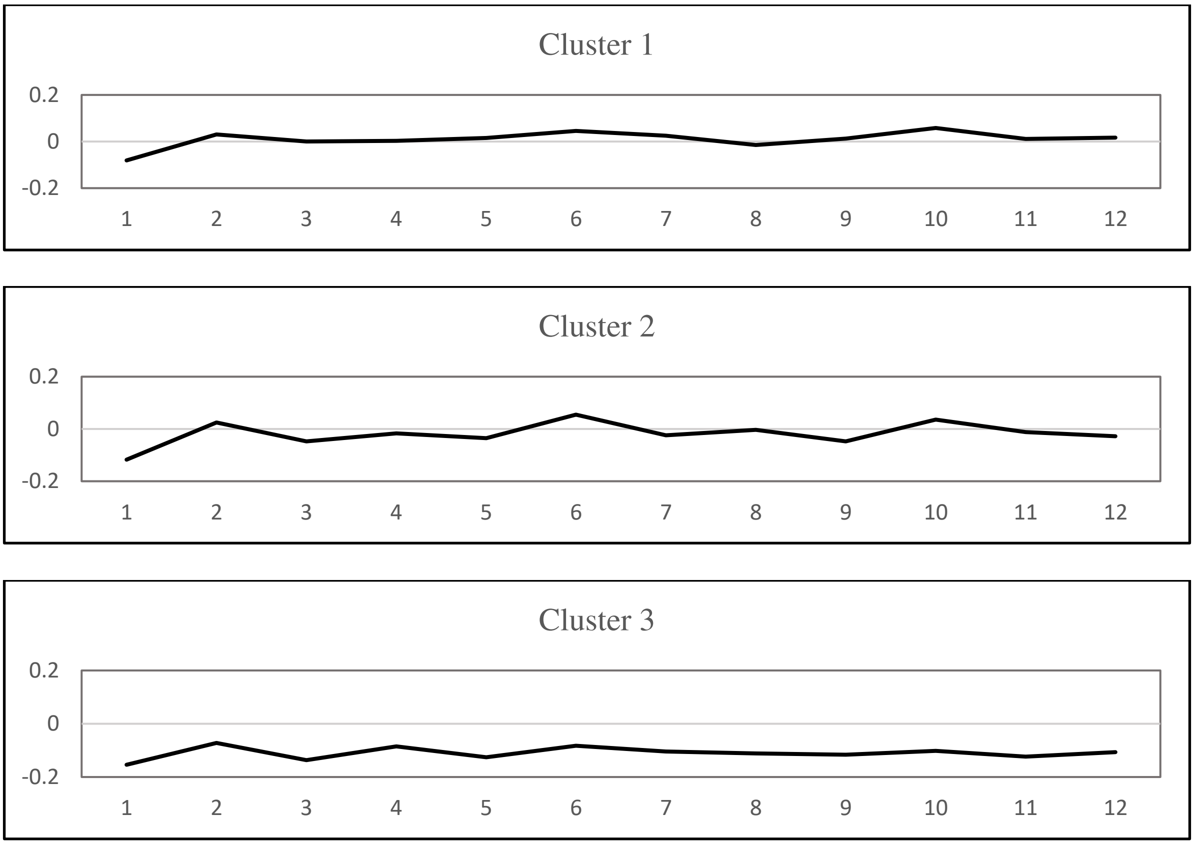

Figure 3: Classification through K-means

Result of Variable Selection Through LASSO

LASSO was applied to identify the key variables that affect demand in each cluster. This involved measuring the coefficient value, which is the effect of each variable on demand, to extract the most effective variables. The results of this analysis are shown in Table 2 below.

Overall, derived variables such as lag variables and moving averages, which were created through feature engineering, tended to be selected as effective variables for each cluster. A total of 20 variables were extracted for cluster 1, 22 variables for cluster 2, and 21 variables for cluster 3, with a detailed breakdown of the variables in each cluster.

Table 2: Results of LASSO variable selection

Cluster	Variable Number	Variables
Cluster 1	20	‘Year’, ‘Month’, 'lag1', 'lag4', 'lag_mean', 'change_lag1', 'change_lag2', 'change_lag3', ‘sum_q’, 'max_q', 'min_q', 'lag1year', 'MovingMean_12', 'WeightMovingMean_12', 'perceptionsOfCorruption', 'lifeExpectancy', 'Warehouse_Whse_S', 'Warehouse_Whse_C', 'Warehouse_Whse_A', 'category’
Cluster 2	22	‘Year’, 'Month', 'lag2', ‘lag3’, ‘lag4’, 'lag_mean', 'change_lag1', 'change_lag2', 'change_lag3', 'max_q', 'min_q', 'lag1year', 'MovingMean_12', ‘MovingMean_4’, 'WeightMovingMean_12', ‘wheat’, 'perceptionsOfCorruption', 'lifeExpectancy', 'Warehouse_Whse_S', 'Warehouse_Whse_C', 'Warehouse_Whse_A', 'category’
Cluster 3	21	‘Year’, 'Month', 'lag1', ‘lag2’, ‘lag3’, ‘change_lag1', 'change_lag2', 'change_lag3', 'sum_q', 'max_q', ‘min_q’, 'lag1year', 'MovingMean_12', ‘MovingMean_4’, ‘wheat’, 'perceptionsOfCorruption', 'lifeExpectancy', 'Warehouse_Whse_S', 'Warehouse_Whse_C', 'Warehouse_Whse_A', 'category’

LASSO LASSO+LSTM Demand Forecast Result

Looking at the prediction results, the proposed K-means+LASSO+LSTM combination model was found to be the best in all three evaluation indicators compared to other benchmarking models. The single LSTM model showed the lowest performance with mMAPE 0.627, RMSE 1.282, and MAE 0.654. Next, it can be seen that the performance of the model improves in the order of the K-means+LSTM combination model and the LASSO-LSTM combination model. This shows that the performance of the model that has undergone the clustering or variable selection process is better than that of a single deep learning model. However, in the case of demand forecasting, the hybrid model that performs LSTM prediction after applying both K-means clustering and LASSO variable selection shows the best performance with mMAPE 0.356, RMSE 0.958, and MAE 0.387. These results prove that the direction proposed in this model is sufficient for demand forecasting.

Table 3: Performance evaluation of predictive model

Model	mMAPE	RMSE	MAE
LSTM	0.627	1.282	0.654
K-means + LSTM	0.528	1.165	0.560
LASSO + LSTM	0.440	1.080	0.467
K-means + LASSO + LSTM	**0.356**	**0.958**	**0.387**

CONCLUSION

Demand for more accurate demand forecasting continued amid complex and uncertain market changes. In this study, a new demand forecasting methodology using a hybrid model combining K-Means, LASSO, and LSTM was proposed. The proposed model performs clustering based on the time series characteristics of the variable demand data through K-means, extracts key variables from each cluster through LASSO, and then uses LSTM to analyze specific data for one quarter (12 weeks). It is a model that predicts demand. In this study, a case study was conducted to measure the performance of the proposed model using retail company data. As a result of quantifying the performance of the predictive model through three evaluation indicators, the proposed hybrid model showed the best predictive performance compared to the three benchmarking models: single LSTM model, LASSO and

LSTM combined model, and K-means and LSTM combined model. In the proposed hybrid model, K-Means and LASSO are applied step by step to a single LSTM model to gradually improve the accuracy.

In particular, since K-means clustering clusters items with similar patterns in consideration of time series characteristics, it was confirmed that the learning effect is superior to modeling the entire data at once. In addition, it was confirmed that accuracy can be improved by removing noise and interference between variables when constructing a prediction model by differentially selecting key variables that have the most impact on demand forecasting for each cluster. This study proves that the role of K-means and LASSO is a preliminary device that can maximize the learning effect of LSTM. The hybrid model presented in this study is expected to be applied to various companies in the retail field to produce excellent demand forecasting results, and based on this, it is expected to increase value by being applied to efficient inventory and production management of companies.

However, although the proposed model demonstrated excellent performance, it also has some limitations. In this study, performance was evaluated using data from a company in the US retail industry. In order to have validity in terms of generalization, whether this model can be used universally in the retail field, it is necessary to conduct analysis by applying it to various business cases. It will be possible to improve the feasibility of application to the field by applying the proposed model to various situations such as detailed fields, sales scale, types of products handled, and diversity of data attributes held.

In addition, in this study, in order to improve the performance of demand forecasting, significant variables that affect demand volume were additionally created in the feature engineering stage. Since products are greatly influenced by the market environment, this study also included environmental variables such as price and purchase freedom. However, in order to make learning about environmental factors more effective, it is necessary to include more diverse environmental variables such as macroeconomic indicators, consumer sentiment, and sales conditions of competitors. In addition, the proposed model of this study attempted prediction using only a single LSTM model after selecting key variables for each cluster through LASSO. However, since the characteristics of the variables selected for each cluster are different, an attempt can be made to find an optimal model other than LSTM. Therefore, it is expected that it will be a meaningful study to explore the performance of various models that can be combined with K-means and LASSO in the future.

REFERENCES

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. and Seaman, B., 2019. Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26* (pp. 462-474). Springer International Publishing.

Bohanec, M., Borštnar, M.K. and Robnik-Šikonja, M., 2017. Explaining machine learning models in sales predictions. *Expert Systems with Applications*, *71*, pp.416-428.

Chalapathy, R. and Chawla, S., 2019. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*.

Carbonneau, R., Laframboise, K. and Vahidov, R., 2008. Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), pp.1140-1154.

Ermawanto, D. and Kurniati, N., 2021. Examining Inventory Management for Spare-Part Warehouse. *International Conference on Industrial Engineering and Operations Management*. IEOM Society International.

Faber, T. and Finkenrath, M., 2021, February. Load forecasting in district heating systems using stacked ensembles of machine learning algorithms. In *14th International Renewable Energy Storage Conference 2020 (IRES 2020)* (pp. 1-4). Atlantis Press.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.

Mujeeb, S., Javaid, N., Ilahi, M., Wadud, Z., Ishmanov, F. and Afzal, M.K., 2019. Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities. Sustainability, 11(4), p.987.

Niennattrakul, V. and Ratanamahatana, C.A., 2007, April. On clustering multimedia time series data using k-means and dynamic time warping. In *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)* (pp. 733-738). IEEE.

Ren, S., Chan, H.L. and Siqin, T., 2020. Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. *Annals of Operations Research*, *291*, pp.761-777.

Sun, K., Huang, S.H., Wong, D.S.H. and Jang, S.S., 2016. Design and application of a variable selection method for multilayer perceptron neural network with LASSO. *IEEE transactions on neural networks and learning systems*, *28*(6), pp.1386-1396.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), pp.267-288.

Vadyala, S.R., Betgeri, S.N., Sherer, E.A. and Amritphale, A., 2021. Prediction of the number of COVID-19 confirmed cases based on K-means-LSTM. *Array*, *11*, p.100085.

Van Steenbergen, R.M. and Mes, M.R., 2020. Forecasting demand profiles of new products. *Decision support systems*, *139*, p.113401.

Wang, J. and Wang, J., 2021. A new hybrid forecasting model based on SW-LSTM and wavelet packet decomposition: a case study of oil futures prices. *Computational Intelligence and Neuroscience*, *2021*.

Zhang, Y., Li, G., Muskat, B. and Law, R., 2021. Tourism demand forecasting: A decomposed deep learning approach. *Journal of Travel Research*, *60*(5), pp.981-997.