14th Annual International Conference on Industrial Engineering and Operations Management
Dubai, United Arab Emirates (UAE), February 12-14, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/AN14.20240177
Published: February 12, 2024

A Hybrid Approach to Prescribe a Dynamic Operational Location for Stationing a Delivery Truck of LPG Cylinder for Efficient Last Mile Delivery

Prithvirajan D and Mathirajan M
Department of Management Studies
Indian Institute of Science
Bangalore 560012, India

Abstract

Last Mile Delivery (LMD) relates to all product delivery activities to the end customer in the supply chain. Timely delivery of products to customers is one of the primary objectives of businesses rendering LMD services. LPG cylinder distribution in India is one such business that provides LMD to its subscribed customers through delivery agents employed by distributors. As per the existing system, the cylinders are brought to the office premises from the warehouse using Delivery Trucks (DT) and are delivered to customers using small delivery vehicles (SDV). The orders delivered each day by the SDV exceed their stock holding capacity. So, SDV must reach the office location from the last delivered customer's location to restock. This current practice of restocking increases the overall distance travelled by SDV and increases. Further, this increases the delivery lead time, which directly affects customer satisfaction. To overcome these logistics issues in the existing system, this study proposes a dynamic operational location (DOL) for stationing each DT around the area, where the delivery is going to happen each day, considering the customers ordering pattern, instead of all the DT stationed at the office premise. To prescribe DOL for stationing each of the DT this study proposes a hybrid approach and does efficient LMD of LPG cylinders. The proposed hybrid approach involves a multi-step process. In the first step, a forecasting model is developed to analyze the customer's order interval pattern and predict their next ordering date range (i.e., the next order interval). Customers whose next ordering date range is within the next ten days from the decision-making epoch/date are segregated and clustered in the second step. The total number of clusters equals the total number of DT operated. In the final step, the centroid location of the generated cluster is prescribed as the DOL to station a DT. Based on the proposed DOL for each DT, the SDV will stock the cylinders from the DT and proceed for LMD. The proposed hybrid approach is demonstrated using real-life data from a large-scale LPG cylinder distributor in a metro city in Tamilnadu. Accordingly, the dataset used for this study has the customer's unique consumer numbers, geocoded addresses, and order placement dates for the past five years. Using data pre-processing and feature engineering techniques, the ordering intervals for each customer are obtained, and a forecasting model is developed to predict the next order interval for each customer. Customers' geolocations with predicted ordering date range within the next ten days are clustered using the K-Means clustering algorithm, and their centroid location is prescribed as DOL for each DT. Finally, both existing practice and the proposed hybrid approach for LMD are compared in terms of the distance travelled by an SDV in the current system with the proposed system.

Keywords
LPG Cylinder, Last Mile Delivery, Dynamic Operational Location, Forecasting, Clustering Algorithm, Centroid Method

1. Introduction

Last Mile Delivery (LMD) relates to all product delivery activities to the end customer in the supply chain. LPG cylinder distribution in India is one such business that provides LMD to its subscribed customers through delivery agents employed by the distributor. Petroleum companies market and distribute the LPG cylinders through distributors. Customers are provided with a unique consumer number by the petroleum company upon subscription for LPG cylinders and will be assigned a distributor nearby for LMD. The government of India has limited the maximum number of bookings to two cylinders per month per customer. Timely delivery of products is one of the primary objectives of businesses rendering LMD services.
In the current practice of LPG cylinder delivery, two types of vehicles are used. Delivery Truck (DT) will transport the cylinders from the warehouse to the office location, and Small Delivery Vehicles (SDV) will stock the LPG cylinder from the office location and deliver the cylinders to the customer's doorstep. The total number of cylinders delivered by each SDV per day is greater than its stock holding capacity. So, after delivering the last filled LPG cylinder to the customer, the SDV will approach the office location to restock. Due to this, each SDV makes multiple trips between the office and the last customer's location during delivery day. Further, the sequence in which the cylinders are delivered to customers is an arbitrary decision made by the SDV deliverymen.

This current delivery practice has certain logistics issues that hinder the distributors from attaining their delivery goal of delivering cylinders to customers within 24 hours from the time of booking. The overall distance travelled by each SDV increases due to multiple trips from the last customer's location and office location to restock. This causes fatigue to SDV deliverymen, reducing their delivery capacity and increasing the fuel consumption of SDV. In addition, the delivery lead time increases, which negatively impacts customer satisfaction. To overcome the aforementioned logistic issues, this study proposes a hybrid approach to prescribe a dynamic operational location (DOL) for stationing each DT around the area, where the delivery is going to happen each day, considering the customers ordering pattern, instead of all the DT stationed at the office premise.

The paper is organized as follows. Section 2 presents the review on closely related literature. Section 3 presents the proposed hybrid approach for determining DOL for each DT. Section 4 presents the computational results and analysis. Finally, Section 6 concludes with research in progress.

2. Closely Related Literature Review

To the best of our knowledge, no specific study has contributed to determining operational location (OL) or dynamic OL (DOL) for efficient LMD of LPG cylinders. However, in the following sections, we will review the literature on operational location determination considering clustering algorithms and forecasting models to predict the customer's next purchase date in general.

2.1 A Brief Review on OL Determination Considering Clustering Algorithms

Clustering is an unsupervised learning tool used in data mining, which groups data based on similarity in their characteristics. Klose et al. (2005) and Ghoseiri and Ghannadpour (2009) approached the OL determination problem as a p-median problem where p facilities must be located and recommended the centroid of clusters generated as the optimal facility location. Guersola et al. (2017) clustered their customers into groups for each truck they operate to reduce the distance travelled by truck in LMD. Suntanto et al. (2017) stated that the K-Means-based approach performed better in clustering customer locations and determining the OL for facilities. Rautela et al. (2019) focused on minimizing the distribution cost of dairy products, concluded that the k-means clustering algorithm produced better results than agglomerative clustering.

Based on the performance evaluation of the k-means algorithm, Ahmed et al. (2020) listed mobile storage positioning as one of the recent applications of k-means and concluded that k-means and its variants perform efficiently in determining optimal facility location. Borba et al. (2022) used clustering analysis evaluated by a Management Information System to determine if the centroid is the optimal location of the police station to cover maximum geographic areas. Wang and Biljecki (2022) and Lin et al. (2022) located logistic distribution centers using a hybrid k-means clustering algorithm and described it to be practical, scientific, and effective in determining the location of logistic distribution centers. Wang et al. (2022) stated that K-Means, Self-Organising Maps (SOM), and Density Based Spatial Clustering of Application with Noise (DBSCAN) as the top 3 frequently used clustering techniques for facility location applications.

2.2 A Brief Review on Forecasting Customer's Next Purchase Date

Forecasting is a scientific method of predicting the future using historical data as input. Forecasting as an application has made significant operational improvements in the supply chain by predicting demand uncertainties to assist in inventory planning (Fildes et al. (2008)). Demand forecasting as a crucial tool for efficient planning in operations management. Further, the traditional forecasting method, like exponential smoothening, is best fit for smooth, high-volume demands and not for erratic demands (Nenni et al. (2013)). In a study to identify the gaps between theoretical and practical gaps in forecasting, Syntetos et al. (2016) state that the aggregation of demand from a specific customer or a specific region is highly relevant in marketing and sales initiatives. Chauhan and Singh (2018) stated that in forecasting, the choice between machine learning and deep learning is application-oriented, and machine learning algorithms can train effectively on smaller datasets than deep learning.

Tanizaki et al. (2019) forecasted the number of customer demand in restaurants using machine learning algorithms and concluded that ensemble learning models are efficient and practically applicable. Wu et al. (2020) and Anitha and Patil (2022) used the K-Means clustering algorithm to cluster and segment customers based on their recent purchase date, frequency of purchase, and the monetary returns to the company for identifying potential customers in an enterprise. Huber and Stuckenschmidt (2020) employed machine learning algorithms to predict retail demand forecasting on special days and stated that machine learning algorithms provide accurate predictions and are also suitable for application in large-scale forecasting. Hussain et al. (2021) applied Random Forest Regressor for time series forecasting and concluded that it outperformed other algorithms, including Decision Tree, Gradient Boosted Tree, and ARIMA, in predicting future gold prices. Spiliotis et al. (2022) compared statistical and machine learning methods in daily demand forecasting at warehouses and concluded that machine learning is superior and provides better forecasts in terms of accuracy and bias.

Based on the closely related literature review carried out, to the best of our knowledge, no specific study has been addressing the determination of DOL for DT in LMD of LPG cylinders.

3. Proposed Hybrid Approach for Determining DOL for DT in LMD

The proposed hybrid approach has a machine learning-based forecasting model to analyze and predict the next order interval for each customer. Each customer's next ordering date range is computed from the predicted order interval. A clustering model using the K-Means clustering algorithm is used to cluster customers whose predicted order date range is within the next ten days (arbitrary decision) from the decision-making date. The centroid of each cluster is prescribed as OL for each DT. The proposed hybrid approach is demonstrated using real-life data from a large-scale LPG cylinder distributor who is operating in a metro city in Tamilnadu. The ordering dates for the past five years and the geocoded location database for all the customers are obtained from the distributor for this study. The proposed hybrid approach is graphically represented in Figure 1.

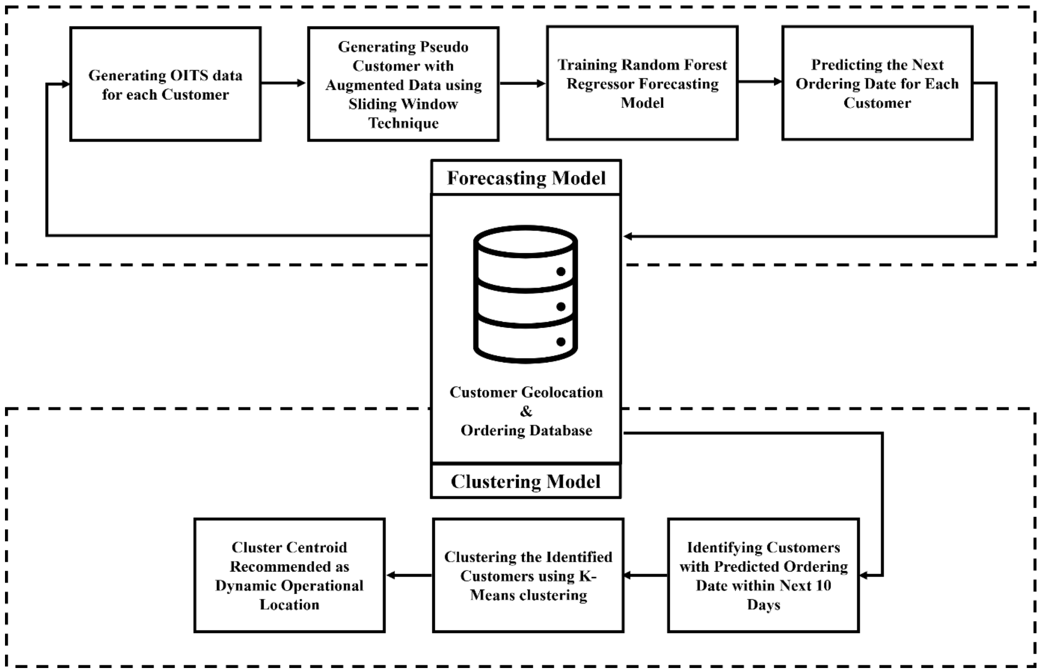

Figure 1. The Proposed Hybrid Approach to Prescribe DOL of Delivery Trucks

3.1 Data Preprocessing and Feature Engineering

Customer's ordering pattern / behaviour are independent and erratic in nature. The total number of customers served by the distributor is 10140, which is the population size in this study. If a customer has not placed any order for the past year, they are categorized as dormant and are not included in this study. Nearly 3.7% of the customers are categorized as dormant, reducing the population size to 9761 customers.

The ordering dates for each customer are extracted and sorted in chronological order from the database. From the chronologically sorted ordering dates, the order interval between two consecutive orders is computed and recorded as time series data for each customer and is referred to as Order Interval Time Series (OITS) data in this study. The number of order intervals in the OITS data ranges between 0 to 66. The next order interval for each customer is predicted based on their OITS data. Accordingly, this study considers only customers with OITS data with order intervals greater than or equal to six. Approximately 5% of the current population had order intervals of less than six, so this further reduced the population size to 9283, which is approximately 91.5% of the total customers served by the distributor.

In certain cases, customers have made an order after an abnormal interval, which is considered an outlier in this study. The outlier in each OITS is identified using the mean and standard deviation of the OITS. Let e be an order interval in the OITS and μ and σ be the mean and standard deviation of the OITS, respectively. The following condition is used to identify if e is an outlier or not,

$$\mu - 2\sigma \leq e \leq \mu + 2\sigma$$

If an order interval in an OITS data is identified to be an outlier, it is replaced with the mean value of order intervals that are not outliers in the respective OITS data.

3.2 Machine Learning-based Forecasting Model

The minimum number of order intervals in OITS data in our study is 6, where each order interval is a feature to the forecasting model. A forecasting model must analyze each customer's ordering interval pattern and forecast the next ordering interval from which their next ordering date will be computed. The characteristics of the study data with less features, have ruled out deep learning based forecasting models in this study. An ensemble learning algorithm called Random Forest Regressor (RFR) is used as a forecasting model in this study because, when compared with other machine learning algorithms, RFR involves less difficulty in training (Rodriguez-Galiano et al. (2015).

3.2.1 Random Forest Regressor (RFR)

In Ensemble learning, accuracy, and resilience enhancement are achieved by merging multiple models. RFR is one such ensemble learning algorithm in which the prediction is the average of decisions made by multiple decision tree models built into it, enabling it to perform effectively without stationarity transformation and data scaling (Houssainy et al. (2021)). In addition, the over-fitting problem faced by the Decision Tree algorithms is addressed in RFR by combining multiple decision trees and through random data splitting, providing relatively better predictions. The total number of decision trees in an RFR is determined by the parameter n-estimator, whose default value is 100. RFR parameters are fine-tuned for efficient performance, thereby eradicating the need for hyperparameter tuning.

3.2.2 Input Data for the Forecasting Model

The OITS data has a wide range of order intervals, with a minimum of six and a maximum of 66 order intervals in each. This might make the forecasting model overfit customers with more order intervals and underfit customers with fewer order intervals. A good forecasting model must generalize well by gaining knowledge of the order interval patterns of all the customers. To address this limitation, a data augmentation technique called the sliding window is used, in which the OITS data of an original customer is split to generate multiple pseudo customers with relatively smaller OITS data. A window of constant length, i.e., the number of order intervals to be in the OITS of the pseudo customer, will slide over the original customer's OITS data with a unit step and continue till the last order interval. This OITS data of pseudo customers is provided as training data in the forecasting model, where the first five-order intervals are treated as features and the sixth-order interval as a target.

The least number of order intervals in OITS data in this study is six; accordingly, the window length is set to be 6. All the pseudo customers generated by the sliding window technique have a fixed dimension of 6 order intervals. The OITS data of the first pseudo customer will have the first order interval of the OITS data of the original customer as its first order interval, and the next pseudo customer will have the second order interval of the OITS data of the original customer as its first order interval, and so on until the last order interval of the OITS data of the original customer becomes the last order interval of the OITS data of the last pseudo customer. If N is the number of order intervals in the OITS data of the original customer and n is the number of order intervals in the OITS data of the pseudo customer, then the total number of pseudo customers M is derived from an original customer as follows:

$$M = N - n + 1$$

After training the forecasting model on the training data, it will be used to predict the next order interval of each customer. For this purpose, the testing dataset is generated by extracting each customer's last six order intervals from their OITS data. Similar to the training dataset, the first five order intervals are treated as features and the last order interval as target. The forecasting model, based on the knowledge gained earlier with the training dataset, predicts the next order interval for each customer. The model's Mean Absolute Error (MAE) is adjusted with the predicted order interval to obtain the order interval range for each customer. The next order date range is computed by adding the predicted order interval range to the recent order date for each customer.

3.3 Clustering Algorithm

Clustering algorithms are used to group and segment data based on their similarity in characteristics. In this study, we have used the K-Means clustering algorithm to cluster customers based on their geographic coordinates. All the customers whose predicted next ordering date range is within the next ten days from the decision-making date, are segregated and provided as input to the clustering model. The number of clusters will be equal to the number of DT operated by the distributor. Once the number of clusters is decided, the clustering algorithm will randomly generate points called cluster points, equal to the number of clusters.

Each customer is assigned to a cluster point based on their proximity. The K-Means clustering uses the Euclidean distance algorithm to determine the distance between each customer and the cluster points. In each iteration, the cluster points will adjust to be closer to the maximum number of customers clustered. Once the customers are clustered, the centroid location of each cluster is determined to decide the DOL to the DT. The SDV will approach the nearby DT positioned at the prescribed DOL for restocking, thereby reducing the distance travelled by an SDV for restocking in current practice. Once in every ten days, the DOL is determined only for delivery awaiting customers.

4. Computational Results and Analysis

This study aims to reduce the overall distance travelled by SDV by stationing the DT at the DOL determined by the proposed hybrid model. The proposed hybrid approach has an RFR-based forecasting model and a K-Means clustering model to prescribe DOL for the DT. Each DT will be stationed at a DOL instead of stationing all the DT at the office premises. The model was developed using Python programming on a system with an i7 8th-generation processor and 16GB RAM.

The training dataset generated using the sliding window technique had 274133 samples with five features and a target each. The RFR forecasting model consumed 1.45 minutes to train on training data, and the model's R^2 score [the statistical measure for goodness of fit for a regression model] of 91.00% was obtained with the model's MAE of 6 days. The statistical description of the error in model prediction states that 25% of customer's next order interval was predicted with an error of 2 days, 50% of customers with an error of 4 days, and 75% of customers with an error of 8 days. With these results from the model, the predicted order interval of each customer is adjusted with the model's MAE and is used to compute each customer's next order date range using their recent order placement date. If a customer's predicted next order date range falls within the next ten days from the decision-making date, they are segregated and provided as input to the clustering algorithm. In this study, the decision-making date, was set as the 1st of July 2023.

The total number of clusters equals the number of DT operated by the distributor. In this study, the number of clusters was set to be 3. Once the customers are assigned to a cluster, using the K-Means clustering algorithm, the centroid location for each cluster is determined and prescribed as DOL for each DT. The graphical representation of forecasted potential customers with the DOL for DT is shown in Figure 2.

In real-time, the routes are asymmetric in nature, i.e., distance from node A-B ≠ B-A. This study estimates the real-time distance between two nodes using the Google API. Considering the current practice, the real-time distance between the farthest customer to the office location and vice versa was found to be 43.9 km and 42.9 km, respectively. Considering the proposed system, the same customer location from the nearest cluster and vice versa was found to be 31.7 km and 38.6 km, respectively. Relative to the office location, the distance travelled by an SDV to the farthest customer from the nearest DOL and vice versa has been reduced by 11.2 km and 5.3 km, respectively.

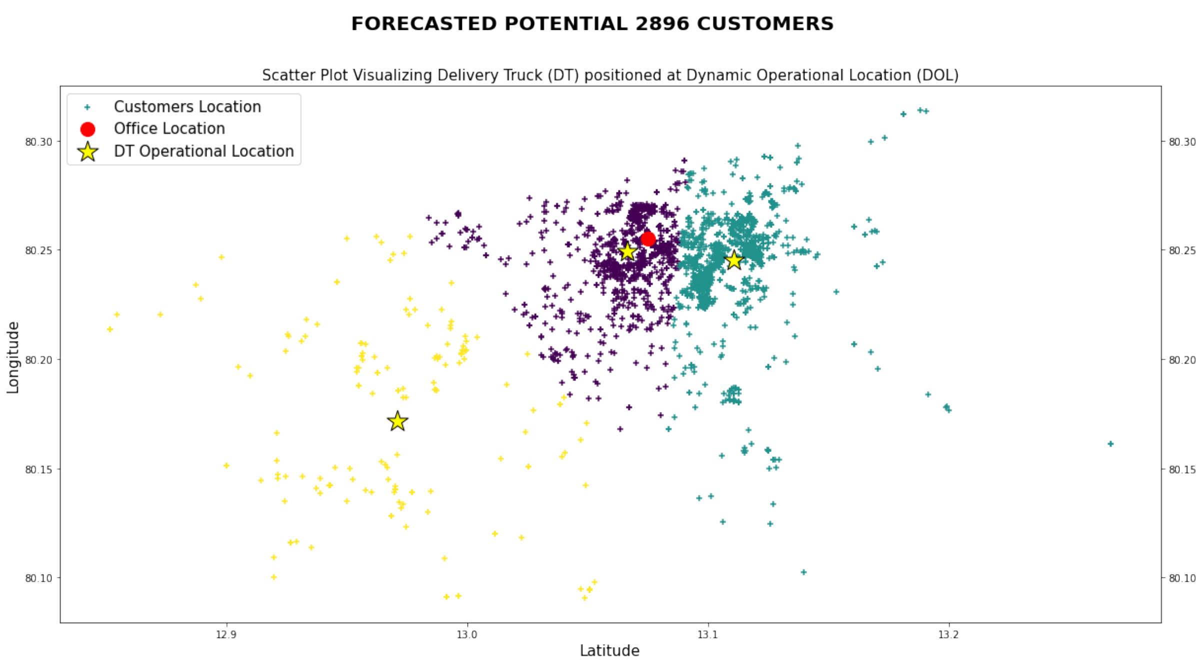

Figure 2. Scatter Plot Visualizing DT Positioned at DOL

The forecasting model has missed approximately 5.7% of customers who have placed orders within ten days from the study date. Accordingly, considering the current practice, among the missed customers, the distance between the farthest customer and the office location and vice versa was estimated to be 25.6 km and 35.4 km, respectively. Whereas, considering the proposed system, the distance between the same customer and the nearest cluster and vice versa was found to be 8.9 km and 8.3 km, respectively. Relative to the office location the distance travelled by an SDV to and from the farthest customer's location missed by the forecasting model to the nearest DOL has been reduced by 16.7 km and 27.1 km, respectively. The relative reduction in distance travelled by an SDV from the nearest DOL is given by,

$$\text{Distance Reduced}(\%) = \frac{(\text{Distance to or from Office Location} - \text{ Distance to or from Nearest DOL})}{\text{Distance to or from Office Location}}$$

The comparison of distance travelled by an SDV to the farthest customer in the current practice and the proposed system are computed and tabulated in Table 1.

Important note: The DT will be stationed at the prescribed DOL as per the proposed system or near the prescribed DOL if there is any public disturbances due to parking the DT at the prescribed OL and this decision will be made dynamically by the deliveryman operating the DT.

Table 1. Comparison of Distance Travelled by SDV in Current Practice and the Proposed System

	Office Location (km)		DOL1 (km)		DOL2 (km)		DOL3 (km)		Distance Reduced (%)	
	To	From	To	From	To	From	To	From	To	From
Forecasted Farthest Customer Serviced	43.9	42.9	40.7	42.3	66.3	66.6	**31.7**	**38.6**	**27.80**	**10.00**
Missed Farthest Customer	25.6	35.4	24.4	24.0	43.8	45.3	**8.9**	**8.3**	**65.23**	**76.55**

5. Conclusion

This study has proposed a hybrid approach to prescribe the DOL for temporarily stationing the DT every day in the LMD of LPG cylinders. The hybrid approach has a machine learning-based forecasting model developed using the RFR algorithm, and a clustering model developed using the K-Means clustering algorithm. A data augmentation process called the sliding window technique was used to train the RFR forecasting model with the available small dataset. The forecasted order interval was adjusted with the model's MAE to determine the ordering date range for each customer. The customers whose ordering date range falls within the next ten days from the decision-making date were segregated and clustered. The centroid location of each cluster was prescribed as DOL for each DT. The distance travelled by an SDV from the DOL prescribed by the proposed system was identified to be less than the distance travelled in current practice. The comparative study on various forecasting models like Exponential Smoothening, ARIMA model, and other regression-based machine learning models like the Regularised Linear Regression model, Support Vector Machine Regressor, and Multi Layer Perceptron are yet to be analyzed and is under progress of the research agenda towards the thesis/dissertation work of the first author of this study.

References

Ahmed Mohiuddin, Raihan Seraj and Syed Mohammed Shamsul Islam, The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics, Vol.9(8)*, 2020.

Anitha P and Patil Malini M, RFM model for customer purchase behavior using K-Means algorithm, *Journal of King Saud University – Computer and Information Sciences, Vol. 34, pp. 1785–1792,* 2022.

Borba Bruno Ferreira da Costa Borba, Ana Paula Henriques de Gusmao, Tharcylla Rebecca Negreiros Clemente and Thyago Celso Cavalcante Nepomuceno, Optimizing Police Facility Locations Based on Cluster Analysis and the Maximal Covering Location Problem, *Applied System Innovation, Vol.5(4),* 2022.

Chaudhuri Neha, Gupta Gaurav, Vamsi Vallurupalli, and Bose Indranil, On the platform but will they buy? Predicting customers' purchase behavior using deep learning, *Decision Support Systems, Vol. 149,* 2021.

Chauhan, Nitin Kumar, and Krishna Singh, A review on conventional machine learning vs deep learning, *International conference on computing, power and communication technologies (GUCON). IEEE,* 2018.

Ciric Milica, Predic Bratislav, Stojanovic Dragan, and Ciric Ivan, Single and Multiple Separate LSTM Neural Networks for Multiple Output Feature Purchase Prediction, *Electronics, Vol. 12,* 2023.

Fildes R, Nikolopoulos K, Crone S.F., and Syntetos A.A., Forecasting and operational research: a review, *Journal of Operational Research Society, Vol. 59, pp. 1150–1172,* 2008.

Ghoseiri Keivan and Ghannadpour Seyed Farid, An efficient heuristic method for capacitated P-Median problem, *International Journal of Management Science and Engineering Management, Vol.4(1), pp. 72-80,* 2009.

Guersola Mariana, Maria Teresinha Arns Steiner and Cassius Tadeu Scarpin, A methodology for minimizing LPG transportation impact, *Management of Environment Quality: An International Journal Vol.28(6), pp. 807-820,* 2016.

Huber Jakob and Stuckenschmidt Heiner, Daily retail demand forecasting using machine learning with emphasis on calendric special days, *International Journal of Forecasting, Vol. 36, pp. 1420-1438,* 2020.

Houssainy EL A, Rady A, Fawzy Haitham and Fattah Amal Mohamed Abdel, Time series forecasting using tree based methods, *Journal of Statistics Applications and Probability, Vol. 10 (1), pp. 229-244,* 2021.

Klose Andreas, and Andreas Drexl, Facility location models for distribution system design. European *Journal of Operational Research, Vol.162(1), pp. 4-29,* 2005.

Maria Elena Nenni, Luca Giustiniano and Luca Pirolo, Demand Forecasting in the Fashion Industry: A Review, *International Journal of Engineering Business Management,* 2013.

Rautela Anubha, Sharma S.K, and Bhardwaj P, Distribution planning using capacitated clustering and vehicle routing problem – A case of Indian cooperative dairy, *Journal of Advances in Management Research, Vol. 16 (5), pp. 781-795,* 2019.

Rodriguez-Galiano. V, Sanchez-Castillo. M, Chica-Olmo. M and Chica-Rivas. M, Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines, *Ore Geology Reviews, Vol. 71, pp. 804-818,* 2015.

Spiliotis Evangelos, Makridakis Spyros, Semenoglou Artemios-Anargyros, and Assimakopoulos Vassilios, Comparison of statistical and machine learning methods for daily SKU demand forecasting, *Operational Research, Vol. 22 (3), pp. 3037-3061,* 2022.

Suntanto G. R, S. Kim, and H. Sutato, A heuristic approach to handle capacitated facility location problem evaluated using clustering internal evaluation, *IOP Conference Series: Materials and Engineering, Vol.332,* 2017.

Syntetos Aris A, Babai Zied, Boylan John E, Kolassa Stephan, and Nikolopoulos Konstantinos, Supply chain forecasting: Theory, practice, their gap and the future, *European Journal of Operational Research, Vol. 252, pp. 1-26,* 2016.

Tanizaki Takashi, Hoshino Tomohiro, Shimmura Takeshi, and Takenaka Takeshi, Demand forecasting in restaurants using machine learning and statistical analysis, *Procedia CIRP, Vol. 79, pg. 679-683,* 2019.

Wang Jing and Biljecki Filip, Unsupervised machine learning in urban studies: A systematic review of applications. *Cities, Vol.129,* 2022.

Wang Ping, Xianjun Chen, and Xuebin Zhang, Research on Location of Logistics Distribution Center Based on K-Means Clustering Algorithm, *Security and Communication Networks, Vol.2022,* 2022.

Wu Jun, Li Shi, Lin Wen-Pin, Tsai Sang-Bing, Li Yuanyuan, Yang Liping, and Xu Guangshu, An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm, *Mathematical Problems in Engineering, Vol. 2020,* 2020.