[1]Anjanadevi B
[2]B S Nagakishore
[3]T Sujith
[4]V Nagesh

Smart Supply Chain Management Using Machine Learning Algorithms

Journal of Electrical Systems

Abstract: - Currently, supply chains are the most widely used in marketing around the world to increase productivity and to add more features to make supply chain management easy to handle. For smart supply chain management, the main focus is on analyzing a product demand and supply chain with fraud suspects of customers and delays in product delivery. RFID can be further extended to authenticate items used to track and manage products.

Demand forecasting is a crucial part of any company or supply chain. It aims at predicting and estimating the future demand of products to help in better decision-making and fraud detection is useful because it allows businesses to identify and prevent unauthorized activities. With the integration of machine learning and the IoT, supply chain management can effectively increase the demand for productivity.

Keywords: Supply chain management (SCM), Radio frequency identification (RFID – IoT), Machine Learning, Delay – In – Delivery

I.INTRODUCTION

Supply chain management (SCM) is the centralized management of the flow of goods and services and includes all processes that transform raw materials into final products. By managing the supply chain, companies can reduce excess costs and deliver products to the consumer faster and more efficiently. Good supply chain management keeps companies out of the headlines and away from expensive recalls and lawsuits. The five most critical elements of SCM are developing a strategy, sourcing raw materials, production, distribution, and returns. A supply chain manager is tasked with controlling and reducing costs and avoiding supply shortages.

Modern day marketing has evolved so much along with strategies to sell a particular product. Every day and every second a count for market sellers, as their main motto is to attract increasingly more of the customers and to sustain the inside market at their best. The present Indian market (like the supermarket) is striving to sell products with the help of best strategies. The IoT (Internet of Things) and ML (machine learning) are apt to help this retail marketing, where the IoT acts as an initiator to record and maintain the details via sensors, and ML provides the full advantages of real-time data (by the IoT) and prediction analysis (by ML). In addition to the abstract hidden for business purposes, every transaction performed by a customer is recorded, and the aggregate of those is maintained regularly to ensure that the sales market cannot lose that valuable customer. By calculating the aggregate of those transactions along with ML, the sales market with the best offers can attract customers. In [1], the review explored analytical and conceptual approaches for addressing various risks in supply chain management. The current work focuses on specific studies that are accepted in industry.

RFID (radio frequency identification) is a form of wireless communication that incorporates the use of electromagnetic or electrostatic coupling in the radio frequency portion of the electromagnetic spectrum to uniquely identify an object, animal or person. The RFID system consists of three components: a scanning antenna, a transceiver and a transponder. When the scanning antenna and transceiver are combined, they are referred to as an RFID reader. It uses radio waves to transmit signals that activate the tag. Once activated, the tag sends a wave back to the antenna, where it is translated into data.

The Transponder is in the RFID tag itself, type of reader, RFID frequency and interference in the surrounding environment or from other RFID tags and readers. Tags that have the strongest power source also have a longer read range.

[1] Review of RFID and IoT integration in supply chain management (2022) Describes the current state-of-the-art literature and potential trends in the application of RFID-IoT in SCM. Few research subjects used inventory

[1] *Associate Professor, MVGR College of Engineering, AP, India.
[2] Registrar, Mangalayatan University, [3] Associate Professor Siddartha Educational Academy Group of Institutions
[4] Professor MVGR College of Engineering

management or dynamic inventory tests. There is a lack of studies that specify the changes required in SCM to adopt RFID-IoT in the future.

[2] A Case Study on Supply Chain Management System (2016) This work mainly focused on the growth of SCM as a tool in marketing and how technology plays an important role in supply chains. Suppliers, intermediaries, and third-party service providers were used. XYZ Automobile Ltd. and ABC Industries Ltd. were used as the datasets. The data collected are collected verbally; hence, the accuracy of the data may vary. This survey touches the intent of the present human generation much less.

[3] Supply Chain Management 4.0: A literature review and research framework (2022) This work mainly focuses on supply chain management 4.0 and the interaction between digital technologies and supply chain management. Few databases were utilized to search for relevant papers to determine the products. There is no strategic or explicit implementation for SCM 4.0. There is a pause in the study and implementation of SCM 4.0 challenges.

[4] Forecasting new product demand using machine learning. This work mainly focuses on how machine learning will optimize production, management and logistics to maximize profits and minimize costs. Here, they used gradient tree boosting for regression. The dataset consisted of more than 4.5 billion samples of sales made from 2012 to 2020. It is planned to improve accuracy by using features extracted from other types of data, such as images and videos.

[5] Demand forecasting using statistical and machine learning algorithms This study aims to compare the performance of traditional statistical and machine learning algorithms in forecasting the demand for 50 products by using ARIMA (autoregressive integrated moving average), the Theta model, multilayer perceptron (MLP) neural networks, and H2O driverless AI. The monthly productwise units sold for the year 2018 were used as the dataset.

[6] Supply chain management practices: A classification based on the literature review (2017). This study presents practices that can support companies achieving higher levels of maturity in their SCM. CMMI capability maturity model integration literally with no dataset used. This study is the first step toward understanding practices related to supply chain management and management maturity.

Supply chain management (SCM) using the Internet of Things (IoT) is a rapidly growing field that leverages interconnected devices and sensors to enhance visibility, traceability, and efficiency in supply chain operations. IoT technology allows various physical objects and devices to be connected to the internet, enabling them to collect, exchange, and analyze data in real time. The following are some key applications and benefits of using the IoT in supply chain management:

- Real-time Tracking and Visibility: IoT-enabled sensors can be attached to products, containers, and vehicles to track their location and movement in real time. This level of visibility helps in monitoring shipments, identifying potential delays, and ensuring timely delivery.

- Inventory Management: IoT devices can monitor inventory levels, providing accurate and up-to-date information about stock levels and automatically triggering reordering points when the inventory reaches predefined thresholds. This reduces stockouts and overstock situations.

- Condition Monitoring: IoT sensors can monitor the condition of goods, such as temperature-sensitive products or perishable items, during transportation. Any deviations from optimal conditions can be immediately detected, allowing timely corrective actions to prevent spoilage or damage.

- Predictive Maintenance: In the case of transportation vehicles or equipment, IoT sensors can monitor health and performance. Predictive maintenance based on IoT data can help prevent breakdowns and optimize maintenance schedules, reducing downtime and improving overall efficiency.

- Supply Chain Analytics: IoT-generated data can be combined with other data sources to perform advanced analytics and gain valuable insights into supply chain operations. This can lead to better decision-making, process improvements, and cost reductions.

- Demand Sensing: IoT data can be used to capture real-time demand signals from consumers and retail outlets, helping companies respond quickly to changes in demand and adjust production and distribution accordingly.

- Security and Risk Management: IoT devices can enhance supply chain security by providing continuous monitoring and alerting for potential security breaches, theft, or tampering of goods during transportation.

- Sustainable Supply Chain: IoT technology can facilitate sustainability initiatives by optimizing routes, reducing fuel consumption, and minimizing waste and emissions.

- Collaboration and Integration: IoT-enabled devices and platforms can foster better collaboration among supply chain partners by sharing data seamlessly and enabling efficient information exchange.
- Customer Experience: Improved supply chain visibility and efficiency through the IoT can lead to better customer service and satisfaction due to accurate delivery estimations and faster response times.

The technology used for proposed system is machine learning. Machine Learning is a branch of artificial intelligence and computer science that focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.
Using machine learning in supply chain management, businesses can build applications that can do the following:

- Forecast demand
- Recognize packages
- Detection of misplaced items
- Improving logistics operations
- Prevent fraud
- Manage automated guided vehicles
- Detection of errors

The aim of using machine learning in our project, "smart supply chain management", is fraud detection and delay in delivery.

A. Fraud detection:

Machine learning is a collection of artificial intelligence algorithms trained with our historical data to suggest risk rules. We can implement the rules to block or allow certain user actions, such as suspicious logins, identity theft or fraudulent transactions. Here, we learn and determine how fraud differs based on the attributes and features and the different kinds of fraud involved. These include products with fraud detection, the highest fraud detection and regions with the highest fraud.

B. Delays in delivery:

When a product/service is delayed or not performed at the right time in supply chain management, a delay in delivery occurs. We can determine the reasons or predict when this would happen using machine learning classification and regression algorithms. By applying different data mining and machine learning techniques, prediction techniques were developed to predict delays in delivering a service or a product. In our project, we detected delays in delivery based on a few factors, such as the location of the production, customer, shipment and shipping modes.

C. Dataset used:

The dataset consists of approximately 180k transactions from supply chains used by the company DataCo Global for 3 years. The dataset can be downloaded from:
https://data.mendeley.com/datasets/8gx2fvg2k6/5
The considered dataset consists of different attributes with different values, which are useful for determining the desired outcome for both fraud detection and delay in delivery. Along with the above attributes, we have other attributes or features that are useful for other applications and classification or regression models. Now, let us look at the process we need to follow to obtain the desired outcomes.

II.PROCESS FLOW STEPS

The following section describes the process flow steps of proposed system.

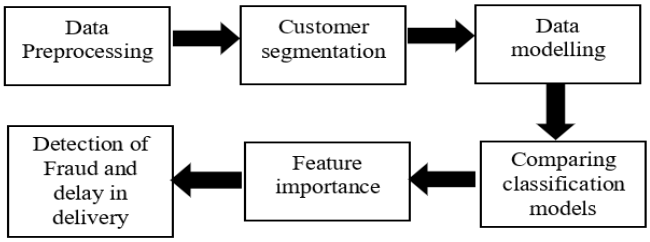

Fig 1.1: Proposed system process flow diagram

- Data Preprocessing:

This step includes data collection, data cleaning and data visualization. However, first begin the complete process by importing all the required libraries

- Data Cleaning

The data consist of some missing values from certain columns that should be removed or replaced before proceeding with the analysis. Additionally, since there is a chance that different customers might have the same first name or same last name, a new column with a 'customer full name' is created to avoid any ambiguities. To make the analysis easier, some unimportant columns were removed.

Let us consider a certain column and see how it works. There are 3 missing values in the customer zipcode column. Since the missing values are only zip codes, which are not very important, they are replaced with zero before proceeding with the data analysis.

- Data Visualization

To identify important parameters, data correlation is performed. Based on the data, we can observe that the product price is strongly correlated with sales and the total order item.

As the data used for analysis are related to the supply chain, let us determine which region has the most sales. Using the groupby method, similar market regions are segregated together, and all sales for that particular region are added using the sum function.

Using a similar method, we determined which category of products had the highest sales. Since the correlation between Price and Sales was high, let us see how price impacts sales for all products.

It can be observed that prices have a linear relation with sales. Let us determine which quarter recorded the most sales. The order of time was divided into years, months, weekdays, and hours to better observe the trend.

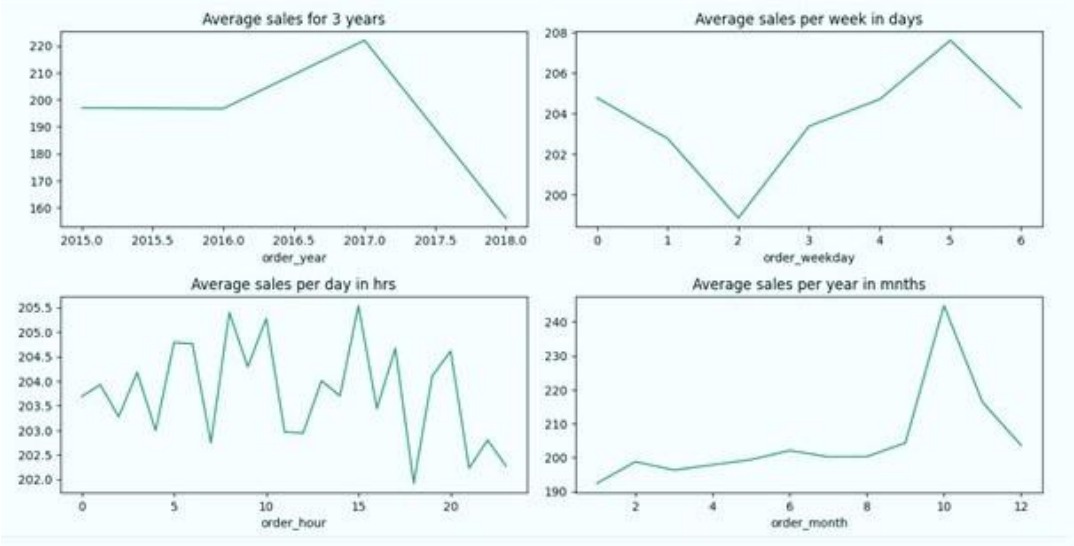

Fig 1.2: Average sales in days, months and years basis

Here, we can observe how price impacts sales and when and which products have more sales. The greatest number of orders came in October, followed by November, and the orders for all other months were consistent.

The highest numbers of orders were placed by customers in 2017. On Saturday, the highest number of average sales was recorded, and on Wednesday, the lowest number of sales was recorded. The average sales are consistent throughout the day irrespective of time, with a std of 3.

It is also important to know what type of payment method is preferred by people to buy all these products in all regions. The .unique() method is used to determine different payment methods. There are our types of payment methods are used. Let us determine which payment method is preferred the most by people in different regions.

The use of the Debit type is the most preferred payment method by people in all regions, and the use of the Cash type is the least preferred method. Some products have negative benefits per order, which indicates that the orders are generating loss of revenue to the company. Let us find what products are those

The total loss sales are approximately 3.9 million, which is a large amount. Cleats is the category with the most loss sales, followed by men's footwear. Most lost sales are occurring in Central America and Western Europe. This loss of sales may have occurred due to suspected fraud or late deliveries.

- Fraud detection:

Finding which payment method is used to conduct fraud can be useful for preventing fraud from occurring in the future.

It can be clearly seen that there is no fraud conducted with the DEBIT, CASH, or PAYMENT methods, so all the suspected fraud orders are made using wire transfer, probably from abroad. Let us determine which region and what product is most suspected to be fraudulent.

The highest number of suspected fraud orders is from Western Europe, which accounts for approximately 17.4% of the total orders, followed by Central America, with 15.5%. Let us determine which product is suspected to be fraud the most.

The total amount was almost 102 K, which is a very large amount. Since Mary was using different addresses every time when placing orders, a new customer ID was issued each time, which made it difficult to identify the customer and ban them. All these parameters should be taken into consideration to improve the fraud detection algorithm so that fraud can be identified more accurately.

- Delays in delivery:

Delivering products to customers on time without late delivery is another important aspect for a supply chain company because customers will not be satisfied if products are not delivered on time. Let us see what category of products are being delivered late the most:

Thus, it can be concluded that for all products with late delivery risk irrespective of region, the product is actually being delivered late; to avoid late delivery, the company can ship products faster using a better shipping method or schedule more days for shipment, so customers will know in advance when the products will reach them. It will be interesting to see the number of late delivered orders for different types of shipment methods in all regions.

III.CUSTOMER SEGMENTATION

Understanding customer needs and targeting specific clusters of customers based on their needs is one way for a supply chain company to increase the number of customers and to gain more profits. Since the purchase history of customers is already available in the dataset, RFM analysis can be used for customer segmentation. Although there are many different methods for customer segmentation, RFM analysis is being used because it utilizes numerical values to show customer recency, frequency and monetary value, and the output results are easy to interpret.

R_Value(Recency): Recency indicates how much time elapsed since a customer's last order.

F_Value(Frequency): Frequency indicates how many times a customer is ordered.

M_Value(Monetary value):Monetary value indicates how much a customer has spent purchasing items.

3. Data modeling:

To measure the performance of different models, the machine learning models are trained to detect fraud and late delivery is used for classification. In addition, sales and order quantity are predicted for regression-type models.A new dataset is created with a copy of the original data for training and validation.Two new columns are created for orders with suspected fraud and late delivery, making them binary classified, which in turn helps to better measure the performance of different models.

To measure machine models accurately, all the columns with repeated values are dropped, such as the late_delivery_risk column, because all the products with late delivery risk are known to be delivered late. The order status column shows that because a new column for fraud detection is created, there is a chance that the machine learning model might take values directly from these columns to predict the output.

It is important to check the type of variables in the data because machine learning models can only be trained with numerical values. There are some columns with object type data that cannot be trained in machine learning models, so all the object type data are converted to the int type using a preprocessing label encoder library.

IV.COMPARING CLASSIFICATION MODELS

We compared all the classification models and determine the best model via the following steps:

The process is started by considering all the columns that are present in our dataset. Then, the data were divided into 2 parts, training data (80%) and testing data (20%), for the model. The data are now ready to be used in machine learning models since many different models are compared, and training every model from the beginning is complicated; therefore, a function is defined to simplify the process. The output is in binary classification format, so all the models are measured with accuracy, recall and F1 score metrics.

Accuracy score: One of the widely used metrics for computing the performance of classification models is accuracy. The percentage of labels that our model successfully predicted is represented by accuracy. For instance, if our model accurately classified 80 of 100 labels, its accuracy would be 0.80.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

TN- True Negative, TP- True Positive, FP- False Positive and FN- False Negative.

Recall score:

The model recall score represents the model's ability to correctly predict the positives out of the actual positives. This is unlike precision, which measures how many predictions made by models are actually positive out of all positive predictions made.

$$Recall = \frac{TP}{TP + FN}$$

F1 score:

The F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Confusion Matrix:

A confusion matrix presents a table layout of the different outcomes of the prediction and results of a classification problem and helps visualize its outcomes. It plots a table of all the predicted and actual values of a classifier.

To measure the performance of different models, the F1 score is used as the main metric because it is the harmonic mean of the precision score and recall score. All the scores are multiplied by 100 for clarity.

I. Support vector machine:

The SVM algorithm chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence, the algorithm is termed the support vector machine.

Algorithm:

1.The SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called a hyperplane.

2. Finds the closest point of the lines from both the classes. These points are called support vectors.

3.The distance between the vectors and the hyperplane is called the margin. The goal of SVM is to maximize this margin.

4. The hyperplane with the maximum margin is called the optimal hyperplane.

II. K Nearest Neighbors:

The K-nearest neighbor algorithm is one of the simplest machine learning algorithms and is based on the supervised learning technique.The K-NN algorithm assumes the similarity between the new case/data and available cases and puts the new case into the category that is most similar to the available categories.

Algorithm:

1.Select the number K of neighbors.

2.The Euclidean distance of K neighbors is calculated.
3.Take the K nearest neighbors as per the calculated Euclidean distance.
4.Among these k neighbors, the number of data points in each category is counted.
5.The new data points are assigned to the category for which the number of neighbors is maximal.

III. Random forest classification:
The random forest regression learning algorithm uses an ensemble learning method for regression. The ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.
Algorithm:
1.Picks at random k data points from the training set
2.A decision tree associated with these k data points is constructed
3.Choose the number N of trees you want to build and repeat the approximately 2 steps
4.For a new data point, make each one of your N-trees predict the value of y for the data point in question and assign the new data point to the average across all the predicted y values.
IV.eXtreme Gradient Boosting Classification
XGBoost, which stands for extreme gradient boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.
Algorithm:
1.The data are split into training and testing sets.
2.Set the model parameters and train the model.
3.Predicting class labels on training and test data using our model.
4.Generate model summary statistics.
5.The model is run, and the results are displayed.

V.Comparison Table for Classification Scores:
To obtain the best classification model, we compare every model and construct a comparison table as follows:
Scores obtained for Fraud Detection:

S No	Model Used	Accuracy	Recall	F1_Score
1	Support Vector Machines	97.75	56.89	28.42
2	K-nearest neighbors	97.36	41.9	35.67
3	Random Forest	98.48	93.18	54.67
4	eExtreme gradient boosting	98.12	89.89	73.22

The following scores were obtained for delay in delivery:

S No	Model Used	Accuracy	Recall	F1_Score
1	Support Vector Machines	98.84	97.94	98.96
2	K-nearest neighbors	80.82	83.45	82.26
3	Random Forest	98.6	97.52	98.74
4	eExtreme gradient boosting	99.24	98.44	99.31

Here, we can conclude that eExtreme gradient boosting is the best fit for our algorithm.

- Feature importance:

The feature importance is the importance given to each and every feature present in our dataset, based on which we can predict either fraud detection or delay in delivery.
Given that Fraud detection is not related to Days for Shipping (real), it is surprising to see that it was given an importance of 0.12. All other important parameters, such as customer full name, shipping mode, and type of payment used, are given an importance of 0.7, which helps the company to detect fraud accurately when the same customer is conducting fraud. (The values are determined from the lower graph).
Similarly, which variables were important for the prediction of late delivery was found.

- Detection of Fraud and Delays in Delivery

From the above, we can conclude that after analyzing the DataCo Company dataset, it has been discovered that both Western Europe and Central America are the regions with the highest number of sales, but the company lost the most revenue from these regions only. Both of these regions are suspected to have the highest number of fraudulent transactions and orders with more late deliveries.
The total sales for the company were consistent until the 2017 Quarter 3 and 10% increase in total sales by quarter and then suddenly decreased by almost 65% in 2018 quarter 1. October and November are the months with the most sales in the total year. Most people prefer to make payments through debit cards, and all fraudulent transactions occur with wire transfers; thus, companies should be careful when using wire transfers because they are scammed by more than 100,000 single customers.
All the orders at risk of late delivery are delivered late every time. Most of the orders with cleats, men's footwear, and women's apparel products cause late delivery, and these products are suspected to cause fraud the most. However, the neural network classifier model trained for fraud detection outperformed all machine learning classifier models, with an F1 score of 0.96. The random forest and eXtreme Gradient Boosting models outperform the Neural Network model.

VI.CONCLUSION

The proposed system effectively predicted the details of sold items, such as delays in delivery and fraud detection, along with authentication using RFID and reasons for delays in delivery using the XGBoost algorithm. In addition, RFID technology is used for tracking products. Experimentation is performed in the IoT by using a single product tag, and authentication details are obtained. By further extending the use of demand-based product tags for this process, we can store all the item details for the decision-making process.In the future, large collections of product data generated through RFID will be stored for analysis with neural network models, which will help in predicting and analyzing delays in delivery and fraud detection.

REFERENCES

[1] Supply Chain Management: Strategy, Planning, and Operation, Sunil Chopra and Peter Meindl.

[2] The effect of supply chain integration on operational and financial performance: Evidence from the retail industry, A. Gunasekaran et al, International Journal of Production Economics, 2015.

[3] The impact of supply chain integration on responsiveness: The moderating role of using an agile information system, K. M. Dooley and C. R. Lacity, Information Systems Frontiers, 2012.

[4] Supply Chain Risk Management: Review, Classification, and Future Research Directions, M. S. Sodhi and C. S. Tang, Production and Operations Management, 2012.

[5] A conceptual framework for understanding the outsourcing risks in the supply chain, M. C. B. Rekik and S. M. A. Belhassen, International Journal of Production Economics, 2015.

[6] Sustainable supply chain management: A literature review, S. Seuring and M. Müller, Corporate Social Responsibility and Environmental Management, 2008.

[7] Sustainable supply chain management and the transition toward a circular economy: Evidence and some applications, C. Estampe et al, International Journal of Production Economics, 2013.

[8] Big data analytics in supply chain management: A. K. M. Najmul Islam et al, Computers & Industrial Engineering, 2017.

[9] Supply chain analytics: A survey, P. K. Kannan et al, International Journal of Production Research, 2018.

[10] Supply chain network design: A critical review and research agenda, G. Barbarosoğlu and E. Özceylan, European Journal of Operational Research, 2016.

[11] A review of optimization models and algorithms for emergency logistics planning and operations, S. A. Torabi et al. (Transportation Research Part E: Logistics and Transportation Review, 2012.

[12] The impact of technology and digitalization on supply chain management: A literature review, M. B. Hashem et al. (International Journal of Production Economics, 2018.

[13] The role of Industry 4.0 technologies in enabling supply chain agility: A systematic literature review, M. F. Hassini et al. International Journal of Production Research, 2019.

[14] Machine Learning in Demand Forecasting: A Comprehensive Survey, V. Syntetos et al, International Journal of Forecasting, 2016.

[15] Machine Learning Applications in Supply Chain Management: A Review, Y. Q. Gao et al, Expert Systems with Applications, 2018.

[16] Machine Learning in Inventory Management: A Review, Y. Shen et al, European Journal of Operational Research, 2019.

[17] Machine Learning Techniques Applied to Inventory Optimization in Supply Chain Management, R. A. Marques et al, Operations Research Perspectives, 2019.

[18] Machine Learning Techniques in Supplier Selection: A Comprehensive Review, H. Jahanshahi et al, Expert Systems with Applications, 2019.

[19] Machine Learning in Supplier Selection and Order Allocation, H. Yang et al. Expert Systems with Applications, 2018.

[20] Machine Learning in Transportation and Logistics: A Survey, A. N. Brown et al, Transportation Research Part C: Emerging Technologies, 2017.

[21] Machine Learning Applications in City Logistics: A Comprehensive Reviewm M. Wang et al, Transportation Research Part C: Emerging Technologies, 2019.

[22] Machine Learning in Warehouse Operations: A Survey, L. V. R. D. N. Senarathne et al, Computers & Industrial Engineering, 2018.

[23] Machine Learning in Warehouse Order Picking: A Literature Review, B. Ma et al, Computers & Industrial Engineering, 2020.

[24] Machine Learning in Supply Chain Risk Management: A Review of Research, T. S. A. Shafei et al, International Journal of Production Research, 2019.

[25] Machine Learning in Supply Chain Risk Management: A Systematic Literature Review, B. K. Ramesh et al, International Journal of Information Management, 2020.

[26] Machine Learning in Supply Chain Network Design: A Comprehensive Review and Future Directions, K. Zhang et al, Computers & Industrial Engineering, 2019.

[27] Machine Learning in Multi-Echelon Inventory Optimization: A Comprehensive Review" by A. V. Araújo et al., Computers & Operations Research, 2021.