Supply Chain Orchestration Tool: using Data Driven Decision Modelling

Vishal Makode
200875978
Dr. Flynn Castles
Msc IOT (Data)

Abstract — **There are several publications on the application of predictive analytics in Supply Chain Management (SCM). However, most of them are limited to specialised supply chain management tasks, such as Logistics, Transportation, Demand prediction, Demand management, Inventory management or other pure research elements. The primary focus of the report is on improving visibility to manage complexity and assist decision-making for handling risk and disruptions along the supply chain. Demand management and Procurement are the two main areas of SCM in which predictive analytics is often applied. This report has two objectives: first, to provide an overview of supply chain management functions (SCMF), apply data-driven predictive analytics, highlighting practical approaches, algorithms, or models in SCM through a comparative review of several machine learning approaches and secondly present a comprehensive supply chain orchestration tool encompassing all those finding into production-ready functional software. For these reasons, pertinent literature information was obtained and evaluated. Accordingly, this report will present the data pipeline, data insights, cloud architectures, and machine learning models for predictive analytics and their performances. These will form the basis for the SCM orchestration tool, which includes all the necessary components to implement SCMF with an intuitive dashboard for supply chain managers and data analysts.**

Keywords — **Supply chain management, Forecasting, Neural networks, Big Data Analytics (BDA), Predictive Analytics, Machine learning (ML), Recurrent Neural Network (RNN), Random Forest, Kubernetes, cloud, Dashboard, Docker, Mongo DB**

I. Introduction

Supply chains contribute significantly to user satisfaction, budget control, and a corporation's responses to market advantages and risks. While examining their cost considerations, timeframes, and stock control, firms pursue efficiency, dependability, and reproducibility (Zhang *et al.*, 2016) to evaluate and avoid occurrences and circumstances affecting logistics management, from the most frequent (production delays, manufacturing faults) to the most significant (social turmoil, environmental disasters, manufacturers' financial distress). Several characteristics could complicate the nature of production processes in contexts where uncertainty is already present. As a result of the swift advancement of technology, the lifespan of products is decreasing at an unprecedented rate. Organisations throughout the globe are utilising reverse supply chain (RSC) tactics to circumvent laws and create profit-generating options. Generally, the production planning procedure can identify the ideal implementation plan for a supply chain's production and logistical activities. A computer program that facilitates the detection of concealed data within collections. While combining data mining and optimization, after the confidential information is recovered from the collection, the feature selection of an optimization problem could be decreased (Aria and Cuccurullo, 2017). Thus, a practical or high-quality response can be identified in a small amount of time and computational performance. To exemplify the efficacy of this strategy, an optimal machine learning technique has been applied to a limited optimal control problem, and the outcomes are clarified.

As the economy has been growing, the quality of human existence has continually improved. In addition, as the market for various items rises, individuals are placing an increased significance on product durability and safety (Han and Zhang, 2021). The supply chain management fundamentally is a technique which combines several conceptual frameworks (Turkulainen and Swink, 2017), as several scholars have long predicted that integration will be an essential factor of supply chain research (Luo and Yu, 2019). The commercial supply chain has shifted from a single regional vertical clustering to an internally strategic alliance and will evolve into a stage characterized by several supply chain functional relationships (Gholamian and Taghanzadeh, 2017).

Fig. 1. Understanding supply chain.

As per (Mentzer *et al.*, 2001), the supply chain is the network of businesses having many actions and procedures that produce wealth in the shape of goods and services given to the end user. These links, operations, and activities depend on ecosystem modelling, planning, and management to function effectively to produce more flexible and sustainable supply chain. Artificial intelligence (AI) capabilities have evolved in numerous industries in recent decades, especially in supply chains (Borges *et al.*, 2021). AI enables computers to make intelligent decisions and complete operations without personal interaction. Industries use AI and machine learning to understand various parameters, including logistics, supply chain features, and storage. Supply chain automation definitions change depending on the standpoint from which they are formulated (Zhang, Pee and Cui, 2021).

TABLE I. DataCo Dataset.

1	Type	Days for shipping	Days for shipment	Benefit per order	Sales per customer	Delivery Status	Late_delivery	Category Id	Category Name	Customer City	Customer Country	Customer Segment	Customer Street	Department	Latitude	Longitude
2	DEBIT	3	4	91.25	314.6400146	Advance shipping	0	73	Sporting Goods	Caguas	Puerto Rico	Consumer	5365 Noble Nectar	Fitness	18.2514534	-66.037056
3	TRANSFER	5	4	-249.0899963	311.3599854	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	Consumer	2679 Rustic Loop	Fitness	18.2794514	-66.037064
4	CASH	4	4	-247.7799988	309.7200012	Shipping on time	0	73	Sporting Goods	San Jose	EE. UU.	Consumer	8510 Round Bear (	Fitness	37.2922325	-121.88128
5	DEBIT	3	4	22.86000061	304.8099976	Advance shipping	0	73	Sporting Goods	Los Angeles	EE. UU.	Home Office	3200 Amber Bend	Fitness	34.1259461	-118.29102

1	Market	Order City	Order Country	order date (DateOrders)	Order Item Disco	Order Item Discount Rate	Order Item Product Price	Order Item Profit Ratio	Order Item Quantity	Order Item Total	Order Profit Per Order	Order Region	Order State	Order Status
2	Pacific Asia	Bekasi	Indonesia	1/31/2018 22:56	13.10999966	0.039999999	327.75	0.289999992	1	314.6400146	91.25	Southeast Asia	Java Occiden	COMPLETE
3	Pacific Asia	Bikaner	India	1/13/2018 12:27	16.38999939	0.050000001	327.75	-0.800000012	1	311.3599854	-249.0899963	South Asia	Rajast·n	PENDING
4	Pacific Asia	Bikaner	India	1/13/2018 12:06	18.03000069	0.059999999	327.75	-0.800000012	1	309.7200012	-247.7799988	South Asia	Rajast·n	CLOSED
5	Pacific Asia	Townsville	Australia	1/13/2018 11:45	22.94000053	0.07	327.75	0.079999998	1	304.8099976	22.86000061	Oceania	Queensland	COMPLETE

II. Related Work

The data-driven supply chain is an integral part of today's business to move products from one site to another. This type of supply chain optimizes the exposure of the product from unprocessed materials to its usage. Such visibility enables sophisticated supply chains to achieve improved service performance and real-time supply chain knowledge. On the other hand, complete integration has not yet been attained; It is a challenging problem due to the significance of integration of operations and requirement of vast amount of data and customisation for product category. This demonstrates the capability of data driven SCM orchestration to facilitate intelligent business scenarios in manufacturing and management.

A study by (Hauser et al., 2017) intended to develop a cloud platform that enables the creation of services for the administration of collaborative planning systems among supply chain individuals. Additionally, this study proposed a concept structured via the platform's five principal services: simulation, detection, evaluation, transformation, and workflow orchestration. First, it introduced the first 4 essential services via the establishment of rules for data analysis, automatic simulation of a supply chain scenario evaluation of deviations, and modelling of an adaption method. Finally, it discussed the principle of orchestration techniques. Another study done by Dalmolen et al, 2015 provided an unique method for supply chain choreography to assist supply chain businesses in generating chain integration that is smooth in practice. Initially, the author developed a body of knowledge by merging supply chain collaboration and constraints literature with scientific observation acquired from applied research and commercial projects. Next, they presented a semantic model that enables fair transition and the creation of an ecosystem in which customizable logistics are the future.

III. Background Research

Despite having a substantial number of scientific articles in the domain of machine learning (ML) and supply chain management individually, not enough attention has been made to the uses of ML algorithms in SCM (Bertolini et al, 2021). This section analyses the utilization of the most well-known machine learning classifiers to the SCM related challenges, such as supplier evaluation and segmentation, supply risk detection, market and sale projection, manufacturing, stock control, and transportation. The examined literature displayed several contributions regarding analysing datasets and capturing demand changes in real-time for sensing and forecasting demand. The dataset explored for this project – Table 1. This section will also explore cloud architecture strategies and tools to integrate the data predictive analysis logic into a working software with some visualisations.

A. Overview of Dataset

A Dataset of Supply Chains by the company DataCo Global was used for the analysis (Table 1). Dataset of Supply Chain, which allows the use of Machine Learning Algorithms and R Software. Areas of important registered activities: Provisioning, Production, Sales, Commercial Distribution. It also allows the correlation of Structured Data with Unstructured Data for knowledge generation (Constante, Silva and Pereira, 2019) This dataset contains a compilation of their financial facts (profit, loss, total sales, etc.), sold items, shipping details, and customer details, including sales, demographics, and transaction details. The data covers 91 MB and contains the details of 200,520 clients across 40

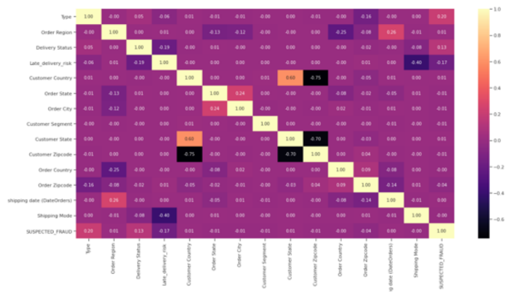

Fig. 2. Correlation matrix of DataCo dataset.

columns, majorly about product categories like Clothing, Sports, and Electronic Supplies. Fig 2. shows the feature correlation heat map from the dataset.

B. Demand Sensing

The phrase "demand sensing" refers to the translation of demand information with minimal delay to minimise prediction inaccuracy, improve operational efficiency, and increase inventory accuracy (Pham et al., 2020). For example, Liu (Liu, Shin and Burns, 2019) developed a strategy employing BDA and natural language processing to assess consumer involvement and the influence of social media marketing for the top 15 luxury companies. The author utilised a semantic analysis method and a FE model for the dataset, including 3.78 million tweets retrieved from Twitter in a 60-month period. The result indicated that interaction, entertainment (Pham et al., 2020),
and trendiness dimensions have a significant impact on customer engagement, while the customization dimension does not. Assessing the impact of social media marketing

elements on customer engagement are essential for improving the effectiveness of marketing campaigns. The authors used content analysis to transform and combine unstructured textual input into a keyword index. Then, three textual data qualities (highest, moderate, and minimal impact) were classified using decision and classification algorithms for grey circumstances and fuzzy (GFuzzy) (Liu, 2018) and lastly, successful marketing strategies were established. Decision-making by supply chain manager or Project Management Office (PMO) is crucial because it enables more efficient customer service and market expansion.

In conclusion, a dataset was obtained from the interaction between enterprises and customers, proving helpful in detecting demand via Comprehensive Efficiency Value (CEV).

C. Demand forecasting

The definition of *demand forecasting* is an accurate assessment of product demand based on the relationship between a product and a collection of independent input factors (Pham et al., 2020). Unsurprisingly, as demand forecasting is now receiving the most research interest concerning the function of demand management. Jiang (Jiang, Liu, and Gao, 2019) presented the proposed approach using various statistical score criteria to estimate power consumption accurately. The researchers proved that the proposed model is more error-free than the reference models. Accurate energy demand forecasting is vital for the operation and management of power systems (Pham et al., 2020).

Similarly, three publications propose several approaches for improving the performance and precision of power demand forecasts. First, Runge presented multi-step-ahead short-term forecasting models designed to improve supply fans forecasting ability by employing black-box and hybrid grey-box methodologies to predict the future supply airflow rate and power demand (Runge, Zmeureanu and Le Cam, 2019). Time series analysis is conduction on demand values to observe its behaviour over time.

D. ML Algorithms in SCM

Supplier selection is the most important aspect of the purchasing function (Pang, 2017). Due to the significance of suppliers in terms of cost, time, reliability, supply chain executives have invested significant effort in the supplier selection process. The selection procedure is driven by MCDM (Multi-Criteria Decision-Making Methods) strategies that incorporate competing considerations. Consequently, achieving a balance between these aspects is a crucial challenge for buying managers. MCDM approaches aid decision makers in assessing a group of choices (Guo, 2009). MCDM approaches assist decision-makers in analysing and selecting among a collection of possibilities. In some instances, the number of prospective suppliers and criteria is much greater than what MCDM approaches can handle effectively. On the other hand, MCDM methods are classified as descriptive and static methods, like most other conventional methods. However, in today's competitive marketplace, data analysis approaches are unquestionably more valuable than qualitative research methods. In this era, ML algorithms outperform the before mentioned methods significantly.

TABLE II. State of the art ML algorithms for SCM

Types of Machine Learning	**Classifiers**	**Detail**
Supervised learning	Decision Trees	Decision trees used to classify features into distinct nodes for classification objectives. Additionally, to analyse uncertain decisions (Song, 2015).
	Random Forest	It implements decision trees on various examples and utilizes the maximum vote for classification (Biau, 2016).
	SVM	Support vector machine is most effective for classification purposes (Shmilovici, 2009) when it is used to calculate margins
	K-nearest neighbour	Generally, the classifier uses the training data. When the classifier is presented with the test data, both are compared. Here, the K most correlated training data are extracted (Kramer, 2013).
	PCA	Principal component analysis (PCA) can make calculations faster and simpler by reducing the dimension of the data (Abdi, 2010).
Unsupervised learning	K-means clustering	K-means clustering is used to discover data object groupings within a dataset (Likas, 2003)
	Boosting	Boosting utilizes two types of variables including weak and strong learners. By combining weak learners and transforming them into strong classifiers, the method attempts to reduce bias and variations (Zhou, 2009).
Ensemble learning	Boosting	Boosting utilizes two types of variables including weak and strong learners. By combining weak learners and transforming them into strong classifiers, the method attempts to reduce bias and variations (Zhou, 2009), In ensemble technique to correct the errors in existing models, new models are added.
	Bagging	Bagging is another technique that is be used to reduce variations and improve the precision and consistency of ML (Lemmens, 2006).

E. Software and Cloud Strategies

Docker and Kubernetes have changed DevOps and cloud deployment as both are today's most used container orchestration tools. Managing an increase in the need for scalability and self-healing of the network and virtual instances is always difficult. Because microservices operating on containers cannot interact with one another, container management is always a challenge for any organisation. Kubernetes is a container management platform where containers may be Docker services or other container types. Kubernetes orchestrates, administers, and establishes a communication channel amongst these containers and includes container deployment, scaling and descaling of services, and container load balancing. It is a framework for managing containerized services, as well as for automating the deployment, scaling, and orchestration of applications. It is all about managing many containers in pods. If the container can execute the programme, then Kubernetes can run the application as well.

IV. Methodology

Several studies have compared the performance of neural networks with traditional linear forecasting methodologies. Carbonneau (Carbonneau, Laframboise and Vahidov, 2007) compared forecasting time series such as moving average and linear regression to RNN and SVM and determined that recurrent neural networks performed the best. Reports like (Ghamsari, Rezaei and Vakili, 2020) evaluated the performance of popular machine learning and deep learning algorithms for the classification on IoT-related datasets. They concluded that Random Forests performed better than other machine learning models among deep learning models, while Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) produced more intriguing results. Other researchers, such as (Ahmed, Atiya, and Gayar, 2010), compared regression models and determined that the Multilayer perceptron (MLP) and Gaussian process models are the two best models for regression-type data. However, not much research was discovered that compared Classification type ML models and Regression type ML models with Neural Network models using the same dataset.

A. Model comparision

To measure the performance of different models F1 score is utilised as the primary metric as it is the harmonic mean of the precision score and recall score.

TABLE III. Classification scores.

Classification Model		Accuracy Score for Fraud Detection	Recall Score for Fraud Detection	F1 Score for Fraud Detection	Accuracy Score for Late Delivery	Recall Score for Late Delivery	F1 Score for Late Delivery
0	Logistic	97.80	59.40	31.22	98.84	97.94	
1	Gaussian Naive bayes	87.84	16.23	27.92	57.27	56.20	
2	Support Vector Machines	97.75	56.89	28.42	98.84	97.94	
3	K nearest Neighbour	97.36	41.90	35.67	80.82	83.45	
4	Linear Discriminant Analysis	97.88	56.57	49.20	98.37	97.68	
5	Random Forest	98.48	93.18	54.57	98.60	97.52	
7	gradient boosting	98.93	89.89	73.22	99.24	98.65	

B. Cloud Architecture

As seen in the above table, Decision Tree classifiers are marginally better for classification than the rest. This classifier model will be deployed into a prediction and analysis service or "Business logic", which would function at the back end of the SCM tool. Raw data would be received at

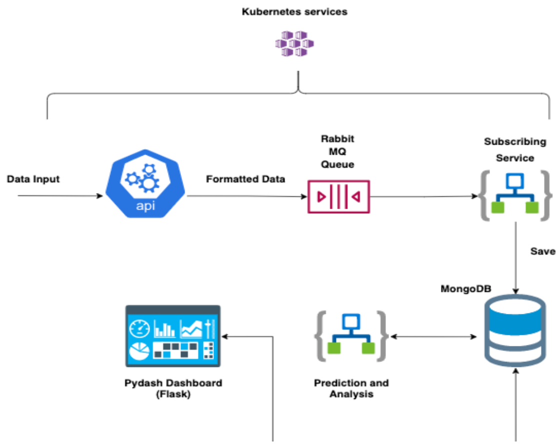

Fig. 3. Holistic cloud architecture for SCM tool.

the data input API service. Data engineering and cleaning will happen in subscribing service, pushing the data into a MongoDB database. The HCI would be a dashboard with all the analytics. Kubernetes will orchestrate all the services as shown in Fig 3.

C. Wider Supply Chain

Fig 4. displays a prototype of the wider supply chain with either a participation bottleneck or an increased supply imbalance. The latter might be described as the component of the supply chain upon which stakeholders do not provide specific forecast information (Würtz *et al.*, 1996). Therefore, we aim to anticipate future consumption using only historical orders from the producer. Data analyst shall study the

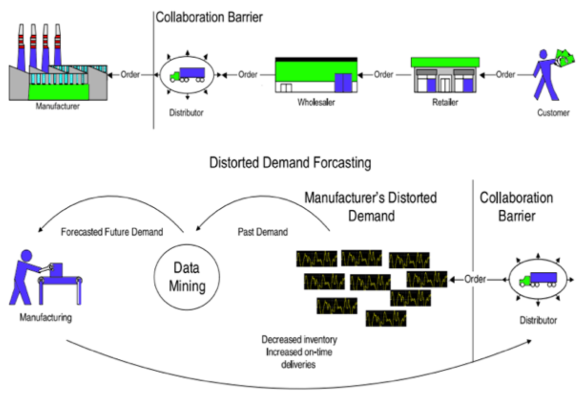

Fig. 4. Collaboration in supply chain.

relevance of sophisticated machine learning algorithms for this purpose (Plus, 2002). As a result, researchers hypothesize that if forecasting accuracy can be improved, expenses will be decreased due to a reduction in supply, and customer loyalty will enhance due to an increase in on-time shipments. This study employs the introductory study of time series (Box, Jenkins and Reinsel, 1970) as a conventional "standard" methodology using which the effectiveness of all other sophisticated methods will be measured. Neural Networks, Recurrent Neural Networks, and Machine Learning are examples of such machines.

V. IMPLEMENTATION

A. Architecture

Architecture has been proposed for predictive analysis in the tool with a rationale of individual layers or stages. A comprehensive diagram showing SCM's methodologies, components, and objectives for each stage and function is constructed.

Optimized model generation: After the pre-processing phases, the data is bifurcated into two sets: the testing dataset and the training dataset.
The training dataset is used to train and establish the optimized model by determining the ideal parameters, whilst the testing dataset is used to validate the correctness of system output. This dataset decomposition applies to all SCMF's research questions except partner selection in procurement functions. Regarding partner selection, the normalized dataset will be immediately used for a specific approach or model to provide results.
Implementation of Prediction/Classification and Evaluation: The testing dataset will be used in the optimized model to detect and anticipate outcomes such as a risk prediction for SCRM (Supply Chain Risk Management), customer engagement (demand sensing), future demand (demand forecasting), etc. In addition, for partner selection, the normalized dataset will be immediately used in the approach or model to estimate the supplier's potential risk.
Output and dashboarding: Output from all the functions i.e., the target values are visualised and assessed interactively by the supply chain manager or a data analyst.

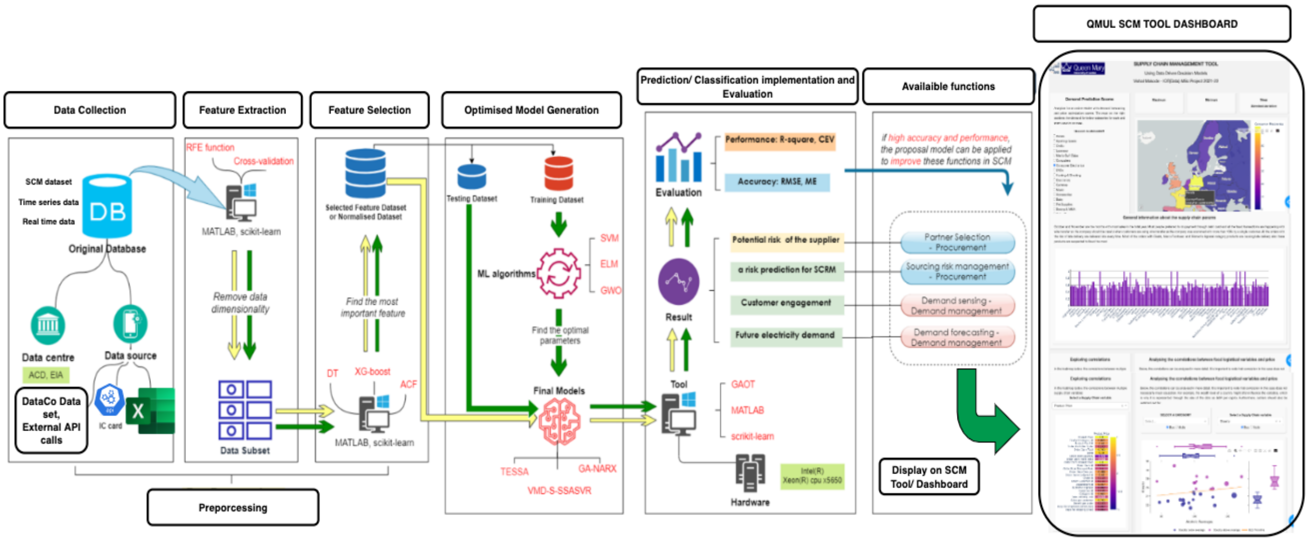

Fig. 5. The application of Predictive analysis using machine learning in the SCM diagram.

This architecture's structure splits into five primary stages:
Data gathering: Data collection is comprehending and obtaining information regarding factors associated with SCM operations, such as sourcing risk and demand forecasting. However, excessive data will take longer to normalize, and incorrect data will skew the system's overall findings. Identifying the data collection source is therefore crucial to the system's effectiveness, making this step the most critical part of the entire system. As mentioned above dataset by DataCo global was used for this case which deals with product categories like Clothing, Sports, and Electronic Supplies.
Feature extraction: This is a process which includes the use of several techniques, such as Reinforcement learning (REF), cross-validation, plotting correlation matrix etc., to extract specific essential characteristics from a dataset. A subset derived from the original data will produce a more precise conclusion than the original data. Accordingly, all the gathered data must undergo this feature extraction procedure.
Feature Selection: After the data has been deconstructed into subsets with features, the data with the most significant feature (using correlation matrix) must be picked using the Feature selection procedure. DT, XG-boost, ACF, and other standard approaches will be utilized in this step.

B. Data (dataset) injection

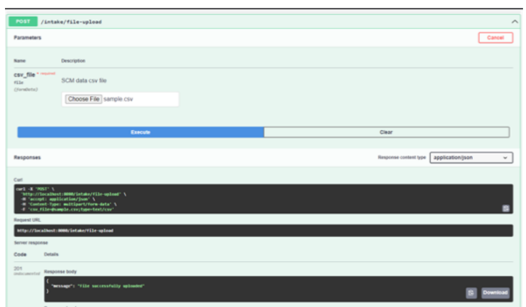

Fig. 6. API endpoint for data injection.

This custom API endpoint was made to inject multiple data types into the system such as CSV, API calls, etc. This information will be retrieved through the business logic and will be presented to ML models.

C. SCM tool Dashboard

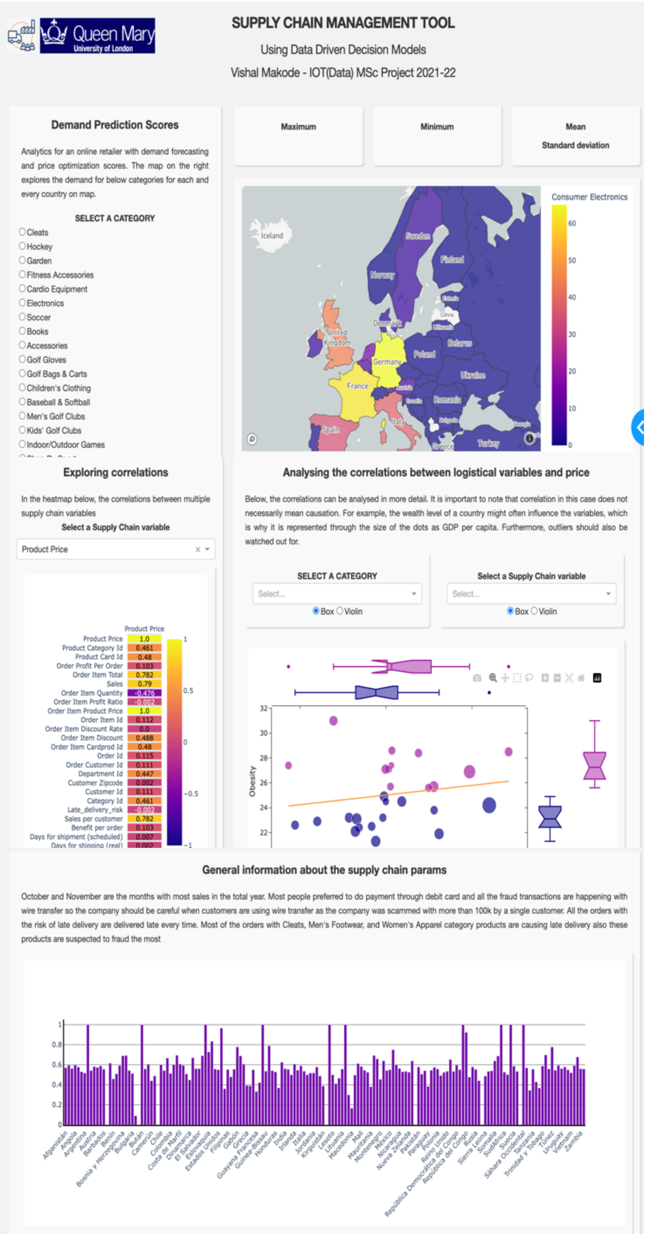

Fig. 7. Supply Chain Management Tool Dashboard.

After the models are trained to detect fraud and late delivery (classification type), the target values are scaled and plotted onto an intuitive user interface called SCM-dashboard. "Good HCI (Human-computer interaction) practices" were followed to make the dashboard simple to assess and act upon. An immersive world map is placed for information associated with demographics. Ex. future product demand for geography. Some preliminary requirements with the dashboard were: Ensuring that the user always looks at the latest information by having dashboards interact directly in real-time with the source data. The below graphs show risk in delivery, which corresponds highly to Fastest Shipment and average shipment duration in days and a correlation matrix table. The dashboard has customisable options for business and the specific driving needs of the organisation.

D. Model building and analysis for late delivery prediction

The methodology involved in the proposed forecasting model can be described in the following steps.

Step-1: List of demographic details of the customers and other details depicting purchase behaviour is prepared. These details are used as attributes to describe customers. It is to be noted that all these attributes are not equally important in describing the intended behaviour of customers (Bala, 2010). Target features selected in this case for model building – Days of shipping (real) and Days of shipment (scheduled).

Step-2: Construction of classes of customers is done for the item/SKU considered for demand forecasting. For demand forecasting, classes are to be performed based on the units of purchase for the SKU. For example, two classes of customers may be those who purchase one unit and those who purchase two or more units (Bala, 2010). Here the problem type in multi-output regression.

Step-3: Based on the target classes, feature selection is performed on the database to select the top few dominant attributes for classification.

Step-4: Based on the dominant attributes, a decision tree is inducted for classification (Bala, 2010).

Step-5: The original database is segregated based on the classification described by the decision tree induced. It is important to note that the description of a customer may fall in the class of purchasing two or more units, while he/she may purchase one unit in some of the transactions. Similarly, description of another customer may fall in the class of purchasing one unit, while he/she may purchase two or more units in some of the transactions(Bala, 2010).

Step-6: For each segregated class, ARIMA (with predictors) is used for forecasting (Bala, 2010).

Step-7: To forecast the overall demand of the item/SKU, forecasts for various classed are summed up (Bala, 2010).

```
new_model_l=tree.DecisionTreeClassifier()
New_classifiermodel(new_model_l,new_xl_train, new_xl_test,new_yl_train,new_yl_tes
t)
```

```
Model paramters used are : DecisionTreeClassifier(ccp_alpha=0.0, class_weight
=None, criterion='gini',
                       max_depth=None, max_features=None, max_leaf_nodes=Non
e,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort='deprecated',
                       random_state=None, splitter='best')
Accuracy        : 83.65832040771106 %
Recall score        : 83.65832040771106 %
Conf Matrix         :
 [[13259  3048]
 [ 2852 16945]]
F1 score       : 83.64916773492568 %
```

Fig. 8. Decision classifier function implemented with F1 scores.

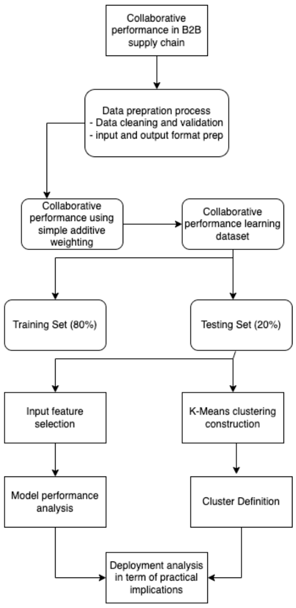

Fig. 10. The methodology of B2B-SC predictive collaborative performance evaluation model (Derrouiche, Holimchayachotikul and Leksakil, 2011).

VI. Testing And Insights

A. RFM analysis

In SCM, an effective customer-oriented strategy is very important, because it helps to increase the relationship between the customers and business. To identify the loyal customers, mostly use to methods. One is demographic variables (gender, age, etc.). Another one is interactive customer behaviours that are communicated with RFM (Recency, frequency, monetary) (Sheshasaayee and Logeshwari, 2018).
The characteristics of three variables are:

- Recency-Last purchase it refers between customer's consumption interval (R value).
- Frequency- It refers, In particularly to periodic number of transactions (F value).
- Monetary- How much do they spend in the period (M value).

RFM analysis correctly determines customer loyalty and donations. Techniques based on customer segmentation or clustering are crucial for calculating RFM analysis. Several customer groups are segmented on segmentation techniques to determine which consumers are most likely to respond to a campaign. In practical marketing, clustering algorithms may be used to analyse client behaviour (Sheshasaayee and Logeshwari, 2018).
In DataCo Dataset, understanding customer demands and targeting specific client clusters is one method for a supply chain organisation to improve its customer base, efficiency, and profitability. Since client purchase history is present in the dataset, RFM analysis may be used to segment consumers. Even though there are other approaches for

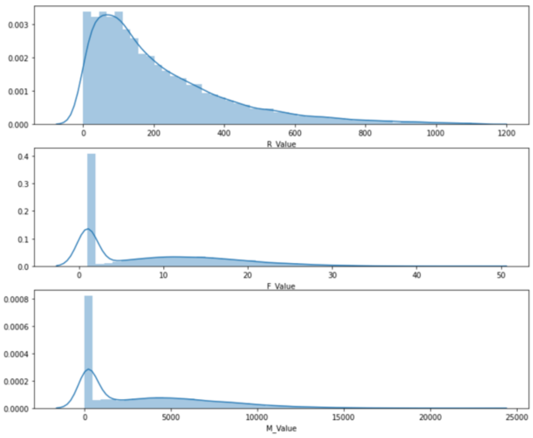

Fig. 9. Distribution plot of R, F, M values from dataset.

customer segmentation, RFM analysis is employed because it employs numerical values to demonstrate Customer recency, frequency, and monetary values, and the output findings are straightforward to comprehend. F value, M Value should be high since they indicate frequency and total value of purchase. Function is defined to indicate quantiles as numerical values and The R Value is low because recent customer activity indication.

B. Cross-Validation Scores

```
Linear Regression :  0.7852647150644007
Ridge :  0.785264849136577
Lasso :  0.7805693158852282
RandomForest :  0.9995414205129233
XGBoost :  0.999651640861491
Decision Tree :  0.9333894519285579
```

K-fold cross validation scores are used in machine learning to gauge the effectiveness of the learning model on unseen data and estimate how the model might perform when a fresh (not used in training) dataset is used to make prediction. This score is popular in scikit. Evidently enough Decision Tree have the highest cross validation score.

VII. Result And Conclusion

The purpose of this tool was to investigate the efficacy of sophisticated quasi-machine learning approaches in anticipating the deformed demand indications in the broader supply chain. The conclusions are significant for supply chain circumstances in which parties might communicate for the reasons explained at the commencement of the study. In such situations, the possibility of increasing forecasting performance would result in reduced expenses and enhanced satisfaction among customers due to more delivery products. And although advanced technologies produced superior results worldwide, they did not significantly outperform more "conventional procedures" like decision trees and other classifiers.

Consequently, researchers can infer that the application of algorithms for machine learning and MLR for estimating distorted consumption indications in the broader supply chain results in more weather predictions than traditional estimation methods (such as naive, trend, and moving average). Researchers did not observe. However, machine methodologies consistently outperform the regression model.

For example, the Decision Tree model did a fantastic job at recognising orders with delayed delivery and detecting fraudulent transactions with a high F1 score. For regression-type data, while the Linear Regression model performed better at forecasting sales revenue, Random forest and eXtreme Gradient Boosting Regression projected demand more precisely. However, the difference between the MAE and RMSE scores of the Neural Network regressor model and those of these ML models is marginal. It is baffling that the Random Forest and eXtreme Gradient Boosting models surpass the Neural Network model. In reality, the slight reliability improvement of RNN models must be evaluated against the intellectual and computation simplicity of the model for linear regression.

The study seeks to investigate the effects of intelligence exchange on prediction accuracies, such as through the Internet and other e-business technologies, as a way for businesses to collaborate choices with their many counterparties), integration of strategic planning will continue to be constrained so long as impediments persist. Considering our concept, such limitations may be reduced.

VIII. Future Work

There are numerous applications of machine learning in the management of supply chains. Therefore, designers highlight the ones that supply chain executives find most valuable. Managing suppliers, facilities, and global logistics partnerships can challenge supply chain management. However, machine learning and artificial intelligence technology can aid in all phases of managing a supply chain. Machine learning mechanisms will accurately predict demand, enhance logistics processes, decrease documentation, and automate human procedures. Consequently, SCM managers will get end-to-end information into your production process, ensuring that it operates more smoothly, incurs fewer expenses, and is less susceptible to disturbances. For fraud detection and prevention, machine learning systems can evaluate vast volumes of data and identify trends for every corporation. Henceforward, all the above-mentioned learning models can be compared with different datasets to verify the same machine learning models are performing comparatively better. These models can be improved with hyperparameter tuning as well.

To the best of my knowledge, this is the first article to provide a holistic insight into predictive analytics and a recipe for a production-ready and cloud deployable SCM tool. However, a few limitations of this approach are that the prediction depends significantly on the dataset. Architecture and the dashboard need much work to become deployable in the real world. Hence, the recommendation for the future is to extend the research to make the prediction more general and data agnostic. In addition, with more exhaustive research in Machine Learning and Artificial Intelligence in future, many prominent models or methods with high accuracy and performance will be detected and analyzed.

IX. References

Abdi, H. and Williams, L.J., 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), pp.433-459.

Ackerson, J. M., Dave, R. and Seliya, N. (2021) 'Applications of recurrent neural network for biometric authentication \& anomaly detection', *Information*, 12(7), p. 272.

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. Econometric Reviews, 29(5-6), 594–621.

Aria, M. and Cuccurullo, C. (2017) 'bibliometrix: An R-tool for comprehensive science mapping analysis', *Journal of informetrics*, 11(4), pp. 959–975.

Bandte, H. (2007) *Komplexität in Organisationen: organisationstheoretische Betrachtungen und agentenbasierte Simulation*. Springer-Verlag.

Bertolini, M., Mezzogori, D., Neroni, M. and Zammori, F., 2021. Machine Learning for industrial applications: A comprehensive literature review. *Expert Systems with Applications*, *175*, p.114820.

Biau, G. and Scornet, E., 2016. A random forest guided tour. *Test*, *25*(2), pp.197-227.

Borges, A.F.S. *et al.* (2021) 'The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions', *International Journal of Information Management*, 57, p. 102225.

Box, G. E., Jenkins, G. M. and Reinsel, G. (1970) 'Time series analysis: forecasting and control Holden-day San Francisco 1970'.

Burges, C. J. C. (1998) 'A tutorial on support vector machines for pattern recognition', *Data mining and knowledge discovery*, 2(2), pp. 121–167.

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

Constante, F., Silva, F. and Pereira, A. (2019). DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS. *data.mendeley.com*, [online] 5. doi:10.17632/8gx2fvg2k6.5.

Dalmolen, S., Moonen, H.M., van Hillegersberg, J., Stoter, A.J.R. and Cornelisse, E., 2015, May. Supply chain orchestration and choreography: Programmable logistics using semantics. *In 2015 4th International Conference on Advanced Logistics and Transport (ICALT)* (pp. 76-81). IEEE.

Gholamian, M.R. and Taghanzadeh, A.H. (2017) 'Integrated network design of wheat supply chain: A real case of Iran', *Computers and Electronics in Agriculture*, 140, pp. 139–147.

Grossman, I. *et al.* (2022) 'Can machine learning improve small area population forecasts? A forecast combination approach', *Computers, Environment and Urban Systems*, 95, p. 101806

Han, C. and Zhang, Q. (2021) 'Optimization of supply chain efficiency management based on machine learning and neural network', *Neural Computing and Applications*, 33(5), pp. 1419–1433.

Hauser, F., Pomponne, V., Jiang, Z., Lamothe, J. and Benaben, F., 2017, June. Processes orchestration for preventing and managing shortages in a supply chain a dermo-cosmetics use case. *In 2017 International Conference on Engineering, Technology and Innovation* (ICE/ITMC) (pp. 1227-1234). IEEE.

J. W. Liu "Using big data database to construct new GFuzzy text mining and decision algorithm for targeting and classifying customers" Computers & Industrial Engineering vol. 128 pp. 1088-1095 2019 [online] Available: https://doi.org/10.1016/j.cie.2018.04.003.

Kramer, O., 2013. K-nearest neighbors. In *Dimensionality reduction with unsupervised nearest neighbors* (pp. 13-23). Springer, Berlin, Heidelberg.

Likas, A., Vlassis, N. and Verbeek, J.J., 2003. The global k-means clustering algorithm. *Pa ptics express*, 27(5), pp. 7405–7425.

Liu, X., Shin, H. and Burns, A.C. (2019). Examining the Impact of Luxury brand's Social Media Marketing on Customer engagement: Using Big Data Analytics and Natural Language Processing. *Journal of Business Research*, 125, pp.815–826. doi:10.1016/j.jbusres.2019.04.042.

Luo, T. and Yu, J. (2019) 'Friends along supply chain and relationship-specific investments', *Review of Quantitative Finance and Accounting*, 53(3), pp. 895–931.

Makridakis, S. and Hibon, M. (2000) 'The M3-Competition: results, conclusions and implications', *International journal of forecasting*, 16(4), pp. 451–476.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2020) 'The M4 Competition: 100,000 time series and 61 forecasting methods', *International Journal of Forecasting*, 36(1), pp. 54–74.

Mentzer, J. T. *et al.* (2001) 'Defining supply chain management', *Journal of Business logistics*, 22(2), pp. 1–25.

Mukherjee, S., Osuna, E. and Girosi, F. (1997) 'Nonlinear prediction of chaotic time series using support vector machines', in *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pp. 511–520.

Müller, K.-R. *et al.* (2018) 'An introduction to kernel-based learning algorithms', in *Handbook of Neural Network Signal Processing*. CRC Press, pp. 1–4.

MathWorks, I. 2000. Using MATLAB, Version 6, MathWorks, Inc

Pang, Q., Yang, T., Li, M. and Shen, Y., 2017. A fuzzy-grey multicriteria decision making approach for green supplier selection in low-carbon supply chain. *Mathematical Problems in Engineering*, *201*

Pang, Z., Niu, F. and O'Neill, Z. (2020) 'Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons', *Renewable Energy*, 156, pp. 279–289.

Papacharalampous, G. *et al.* (2022) 'Massive feature extraction for explaining and foretelling hydroclimatic time series forecastability at the global scale', *Geoscience Frontiers*, 13(3), p. 101349.

Pham, X.V., Maag, A., Senthilananthan, S. and Bhuiyan, M. (2020). *Predictive analysis of the supply chain management using Machine learning approaches: Review and Taxonomy*. [online] IEEE Xplore. doi:10.1109/CITISIA50690.2020.9371842.

Plus, S. (2002) 'Markets, Basic Statistics, Date and Time Management'.

Runge R. Zmeureanu and M. Le Cam "Hybrid short-term forecasting of the electric demand of supply fans using machine learning" Journal of Building Engineering vol. 29 pp. 101-144 2020 [online] Available: https://doi.org/10.1016/j.jobe.2019.101144

Scholkopf, B. *et al.* (1997) 'Comparing support vector machines with Gaussian kernels to radial basis function classifiers', *IEEE transactions on Signal Processing*, 45(11), pp. 2758–2765.

Shmilovici, A., 2009. Support vector machines. In *Data mining and knowledge discovery handbook* (pp. 231-247). Springer, Boston, MA.

Song, Y.Y. and Ying, L.U., 2015. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), p.130.

Turkulainen, V. and Swink, M.L. (2017) 'Supply chain personnel as knowledge resources for innovation—a contingency view', *Journal of Supply Chain Management*, 53(3), pp. 41–59..

Vakili, M., Ghamsari, M., & Rezaei, M. (2020). Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification. arXiv preprint arXiv:2001.09636.

Vapnik, V., Golowich, S. and Smola, A. (1996) 'Support vector method for function approximation, regression estimation and signal processing', *Advances in neural information processing systems*, 9.

Wang, Y. *et al.* (2022) 'Predrnn: A recurrent neural network for spatiotemporal predictive learning', *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Würtz, D. *et al.* (1996) 'Analyse und Vorhersage von Finanzmarktdaten', in *Finanzmarktanalyse und-prognose mit innovativen quantitativen Verfahren*. Springer, pp. 253–298.

Yang, F. *et al.* (2019) 'State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network', *Energy*, 175, pp. 66–75.

Zhang, S. *et al.* (2016) 'ha', *Expert Systems with Applications*, 65, pp. 87–99.

Zhang, D., Pee, L. G. and Cui, L. (2021) 'Artificial intelligence in E-commerce fulfillment: A case study of resource orchestration at Alibaba's Smart Warehouse', *International Journal of Information Management*, 57, p. 102304.

Zhou, Z.H., 2009, June. When semi-supervised learning meets ensemble learning. *In International Workshop on Multiple Classifier Systems* (pp. 529-538). Springer, Berlin, Heidelberg

kaggle.com. (n.d.). *Comparison of ML models with RNN*. [online] Available at: https://www.kaggle.com/code/jaswanthhbadvelu/comparison-of-ml-models-with-rnn [Accessed 21 Aug. 2022].

kaggle.com. (n.d.). *Supply Chain Data Analysis-99% Accuracy*. [online] Available at: https://www.kaggle.com/code/nilufarhosseini/supply-chain-data-analysis-99-accuracy#Customer-Segment [Accessed 21 Aug. 2022].

kaggle.com. (n.d.). *E-Commerce Multi- Output Models {Project_CSE07}*. [online] Available at: https://www.kaggle.com/code/sukanthen/e-commerce-multi-output-models-project-cse07/notebook [Accessed 21 Aug. 2022].

P. Jiang R. Li N. Liu and Y. Gao "A novel composite electricity demand forecasting framework by data processing and optimized support vector machine" Applied Energy vol. 260 pp. 306-321 2020 [online] Available: https://doi.org/10.1016/j.apenergy.2019.114243.

Bala, P.K. (2010). Decision tree based demand forecasts for improving inventory performance. *2010 IEEE International Conference on Industrial Engineering and Engineering Management*. doi:10.1109/ieem.2010.5674628.

Derrouiche, R., Holimchayachotikul, P. and Leksakul, K. (2011). *Predictive performance model in collaborative supply chain using decision tree and clustering technique*. [online] IEEE Xplore. doi:10.1109/LOGISTIQUA.2011.5939435.