Machine learning-based technique for predicting vendor incoterm (contract) in global omnichannel pharmaceutical supply chain

KUMAR DETWAL, Pankaj, SONI, Gunjan, KUMAR JAKHAR, Suresh, KUMAR SHRIVASTAVA, Deepak, MADAAN, Jitendar and KAYIKCI, Yasanur <http://orcid.org/0000-0003-2406-3164>

Machine Learning-Based Technique for Predicting Vendor Incoterm (contract) in Global Omnichannel Pharmaceutical Supply Chain

Pankaj Kumar Detwal
Malaviya National Institute of Technology Jaipur, India
pankajrayan7@gmail.com

Gunjan Soni*
Malaviya National Institute of Technology Jaipur, India
gsoni.mech@mnit.ac.in

Suresh Kumar Jakkar
Indian Institute of Management, Lucknow, India
skj@iiml.ac.in

Deepak Kumar Shrivastava
Indian Institute of Management, Tiruchirappalli, India
srideepak1978@gmail.com

Jitendar Madaan
Indian Institute of Technology, Delhi, India
jmadaan@dms.iitd.ac.in

Yasanur Kayikci
Sheffield Business School, Sheffield Hallam University, Sheffield, UK
Science Policy Research Unit, University of Sussex Business School, Brighton, UK
y.kayikci@shu.ac.uk, yk327@sussex.ac.uk, yasanur.kayikci@gmail.com

* Corresponding author

Machine Learning-Based Technique for Predicting Vendor Incoterm (contract) in Global Omnichannel Pharmaceutical Supply Chain

Abstract: The importance of supply chain management to business operations and social growth cannot be overstated. Modern supply chains are considerably dissimilar from those of only a few years ago and are still evolving in a vastly competitive environment. Technology dealing with the rising complexity of dynamic supply chain processes is required. Robotics, machine learning, and rapid information dispensation can be supply chain transformation enablers. Quite a few functional supply chain applications based on Machine Learning (ML) have appeared in recent years; however, there has been minimal research on applications of data-driven techniques in pharmaceutical supply chains. This paper proposes a machine learning-based vendor incoterm (contract) selection model for direct drop-shipping in a global omnichannel pharmaceutical supply chain. The study also highlights the critical factors influencing the decision to select a vendor incoterm during the shipment of pharmaceutical goods. The findings of this study show that the proposed model can accurately predict a vendor incoterm (contract) for given values of input parameters. This comprehensive model will enable researchers and business administrators to undertake innovation initiatives better and redirect the resources regarding the direct drop shipping of pharmaceutical products.

Keywords: data-driven; omnichannel; pharmaceutical supply chain; vendor incoterm machine learning; direct drop-shipping

1. Introduction

The structure of the supply chain (SC) significantly impacts price, quality, customer perception, and a company's adaptability to market opportunities. A supply chain continuously changes, involving a continuous flow of products, data, and money between various phases. According to (Chopra & Meindl, 2007), "A supply chain comprises all parties involved, directly or indirectly, in satisfying a customer need. The supply chain contains suppliers, manufacturers, carriers, warehouses, merchants, and consumers. Within everybody, such as a manufacturer, the supply chain includes all functions involved in receiving and filling a customer request. These functions comprise, but are not restricted to, new product development, marketing, operations, distribution, finance, and customer service."

The Global Health Supply Chain aims to provide commodities in the correct quantity, at the right quality, at the right place, at a reasonable cost, and on time. Global healthcare supply chains differ from traditional supply chains due to the diversity, high complexity, customization of services provided, the dynamic environment, high-value medical supplies, uncertainty underlying fundamental processes, and, most pertinently, the fact that they deal with human lives. Unanticipated endemics, epidemics, and pandemics throw system actors off balance (Meijboom et al., 2011) and drive the global healthcare system into bottlenecks, resulting in system-related adverse consequences and social and economic disruptions. Being proactive rather than reacting to the situation is the best way to overcome such obstacles, especially in epidemic or pandemic scenarios. The development of functional health supply chains is essential in such an environment. Technology-driven solutions have the potential to significantly aid these healthcare systems in overcoming inefficiencies and being more responsive. Many healthcare entities have begun to use omnichannel commerce to connect with their customers (Dahl et al., 2021). The word "Omni" means "everywhere," and "omnichannel" implies "through all accessible or feasible channels."

Manufacturers (primary and secondary), distributors, wholesalers, and retailers make up the health supply chain (Haszlinna Mustaffa & Potter, 2009). The direct manufacturer's job is to obtain essential medical constituents into the health value chain. Secondary manufacturers then turn these materials into medical items, including medication, drugs, vaccines, and medical or surgical equipment. Hospitals or other healthcare units, and pharmacies purchase these medical goods from distributors or wholesalers if a considerable quantity is required. Hospitals and other health service providers, pharmaceuticals and chemists, blood centers, medical and clinical equipment benefactors, and so on are all vital components of health supply chains. The health supply chain often starts with vendors of these medical apparatus and ends with patients who consume these products and health services. Since patients are the final customer in the healthcare system, errors can jeopardize lives and have irrevocable consequences.

Data analytics has been used to improve and optimize a variety of supply chain operations, including demand prediction and shaping (Roßmann et al., 2018), risk management (Huang & Handfield, 2015), supplier selection, logistics planning (Suma et al., 2017), quality management (Li et al., 2018), evaluating coordination complexity (Dolgui et al., 2018), and evaluating concealed costs. The application & invention of these cutting-edge technologies are still inadequate (Clauson et al., 2018). Data-driven technologies are still nascent when applied in supply chains, with many promises we don't fully comprehend. Global healthcare technologists should consider the increased adoption of data-driven techniques to advance beyond the current mode of operation.

When the mutual agreement between vendor and customer countries occurs whenever medical goods are traded across countries through direct drop shipping (Schaefer, 2017), the arrangement, often known as 'incoterm,' incorporates the responsibilities of both the buyer and supplier during the entire process of shipping. It's vital that all stakeholders involved in trade clearly understand these terms and how they apply to global health supply chains. This research paper proposes a data-driven model to predict Vendor Incoterm (Contract) for a worldwide omnichannel pharmaceutical supply chain that involves direct drop-type shipping of therapeutic goods.

Research Question 1- How to predict an appropriate Vendor Incoterm for input parameters such as Product Group, Sub classification, Vendor, Freight cost (per kg.), and lead time?

Research Question 2- What factors influence the selection of Vendor Incoterm in Pharmaceutical SC?

The remaining part of this research paper is laid out in the following manner: Section 2 discusses the literature Review and its highlights. Section 3 covers the fundamentals of used ML algorithms. The detailed methodology is presented in Section 4, followed by the application case in Section 5. In Section 6, we will go through the results and analysis. The managerial implications of this research are emphasized in section 7, and the conclusion is propounded in Section 8.

2. Literature Review

2.1 Data-driven techniques

Data-driven techniques look for appropriate feature sets and decision criteria from observed data (Falamarzi et al., 2019). These methodologies include- machine learning (ML) models and statistical models.

Table 1 Comparison of Data-Driven Techniques

Technique	Advantage	Disadvantages/Limitations	Reference(s)
Regression Models	Simple and easy to interpret	Prior information about the data needed to achieve the best fit	Zhou et al., 2021; Zhu et al., 2021
Time series models (e.g., moving average method, etc.)	Easy to interpret	Time series required to be stationary	Pacella & Papadia, 2021
Probability Distribution Models	Easy to implement and high interpretability	Based on a specific hypothesis	Islam & Amin, 2020
Bayesian methods	Stable and gives better results with a small size dataset	Used for independent predictors and distribution assumptions is a prerequisite (based on conditional probability)	Sakib et al., 2021
Neural network-based models	Capability, no expertise is required	Time-consuming and poor interpretability	Han & Zhang, 2021; Pereira & Frazzon, 2021
Tree-based model (e.g., random forest, decision trees)	Easy to interpret	Overfitting of meaningless data	Islam & Amin, 2020; Shahbazi & Byun, 2020
Support vector machine (SVM)	High efficiency with small datasets and can tackle nonlinear characteristics.	Interpretability is poor	Micol Policarpo et al., 2021
K-nearest neighbor	Simple and easy to interpret	Determination of parameter k is not easy sensitivity towards data distribution.	Konovalenko & Ludwig, 2021

Statistical models infer associations between variables, but machine-learning models generate the most precise predictions (Falamarzi et al., 2019). Techniques such as Regression Modelling, Probability Distribution Model, Time Series Model, and Bayesian Models can be classified under Statistical models. In contrast, ML models include Artificial Neural Networks (ANN), Support Vector Machines (SVM), tree-based models (e.g., random forest, decision tree), K-nearest neighbor (KNN), etc. Statistical models are preferred when we are supposed to interpret the relationships between parameters. Conversely, machine learning makes predictions as accurate as possible using regression or classification.

The choice of data with diverse characteristics and the problem under consideration hugely impacts the performance of data-driven methods. Further, the effectiveness of data-driven methodologies is determined by selecting appropriate data preprocessing and analysis models. No absolute algorithm will outperform all other algorithms over all datasets and for all circumstances (Mohamed-Iliasse et al., 2020). To better understand these techniques, a comparison of various data-driven methods is presented in Table 1. We have compared these algorithms based on advantages, disadvantages, and references to SC situations where these algorithms have been utilized.

In recent times, data-driven methods have influenced decision-making in many business verticals, particularly in manufacturing, transportation, services, financial institutions, and healthcare. Chang et al., 2023 proposed an integrated framework that facilitates a data-driven model to handle the omnichannel healthcare supply chain. Konovalenko & Ludwig, 2021 analyzed the performance of ML methods concerning temperature monitoring in the pharmaceutical industry. Zhou et al., 2021 designed an XG boost method-based model for supply chain fraud prediction. Sakib et al., 2021 created a model leveraging Bayesian Network (BN) to predict and evaluate disasters in the oil and gas supply chain (OGSC). Researchers have developed data-driven solutions to address several supply chain uncertainties. Pereira & Frazzon, 2021 adopted a data-driven approach to synchronize demand and supply in omnichannel retail chains. Zhu et al., 2021 developed a novel demand forecasting framework and tested it using a machine learning-based model to achieve superior performance in the pharmaceutical supply chain.

Islam & Amin, 2020 adapted distributed random forest and gradient-boosting machine learning techniques for predicting potential back-order scenarios in the supply chain. Shahbazi & Byun, 2020 proposed a blockchain machine learning-based system to track and monitor perishable food along the supply chain. Han & Zhang, 2021 developed a model supported by machine learning and neural network technology for risk management in the supply chain. So, it is safe to say that data-driven methods can contribute to logistic and supply chain decision-making in a further uninterrupted and stout manner.

2.2 Vendor Incoterms

It is a matter of deciding the conditions of sale when selling a product to a foreign consumer as a part of an export pricing plan. Numerous businesses utilize Incoterms, which are internationally accepted commercial terms. The International Chamber of Commerce produces these standards fundamental to International Commercial Law. Governments and legal bodies all around the world acknowledge these standards. Incoterms-2020 is the latest edition of Incoterms issued by the International Chamber of Commerce.

Incoterms are the conditions under which the buyer and seller of products agree to sell and provide items internationally. Incoterm specifies which responsibilities, expenses, and risks are the seller's responsibility and which are transferred to the buyer (Vogt & Davis, 2020). Incoterms also determine when the seller's costs and risks are passed to the buyer.

Each incoterm is represented by a three-letter abbreviation that includes the delivery location. Incoterm is an instrument for determining who is accountable for paying for and handling various aspects of the shipping process (Stojanović & Ivetić, 2020), such as shipment transportation, terminal charges, loading, unloading, export clearance (including security documentation and export license if needed), cargo insurance, export regulations (such as pre-shipment inspections and packing), customs clearance, import documentation (including import license if required), cargo unloading, import taxes, duties and delivery to destination. Incoterms also highlight when the risk of loss or damage to items passes from the seller to the buyer. There are other risks to contemplate in an export deal, such as the cost of a likely customs delay and the obligation for export compliance responsibilities.

Whether the shipment employs any mode of transport or solely maritime and inland waterway transit determines which of the 11 Incoterms is used (Bergami, 2013). Seven out of these eleven Incoterms: Ex Works (EXW), Carriage Paid To (CPT), Carriage and Insurance Paid To (CIP), Free Carrier (FCA), Delivered at Place Unloaded (DPU), Delivered at Place (DAP), and Delivered Duty Paid (DDP) are for any mode of transport. On the other hand, four Incoterms are related only to sea and inland waterway transports: FAS, FOB, CIF, and CFR (*INCOTERMS® 2020 INTERNATIONAL CHAMBER OF COMMERCE (ICC)*, n.d.).

	Freight Collect Teams				Freight Prepaid Teams						
Groups	Any mode or modes of transport		Sea and Inland Transport				Any mode or modes of transport				
Incoterm	**EXW**	**FCA**	**FAS**	**FOB**	**CFR**	**CIF**	**CPT**	**CIP**	**DAP**	**DPU**	**DDP**
Transfer of risk	At buyer's disposal	On buyer's transport	Alongside ship	On board vessel	On board vessel	On board vessel	At carrier	At carrier	At named place	At named place unloaded	At named place
Obligations and charges											
Export Packaging	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller
Loading charges	Buyer	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller
Deliver to Port/place	Buyer	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller
Export duty, taxes & custom clearance	Buyer	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller
Origin Terminal Charges	Buyer	Buyer	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller
Loading on carriage	Buyer	Buyer	Buyer	Seller	Seller	Seller	Seller	Seller	Seller	Seller	Seller
Carriage charges	Buyer	Buyer	Buyer	Buyer	Seller	Seller	Seller	Seller	Seller	Seller	Seller
Insurance	Negotiable	Negotiable	Negotiable	Negotiable	Negotiable	Seller	Negotiable	Seller	Negotiable	Negotiable	Negotiable
Destination Terminal Charges	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Seller	Seller	Seller	Seller	Seller
Delivery to Destination	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Seller	Seller	Seller
Unloading at destination	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Seller	Buyer
Import duty, Taxes & custom clearance	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Buyer	Seller

Figure 1 Vendor Incoterm Chart

The chart shown in figure 1 displays the Incoterms-2020 in a format that is easy to comprehend. This chart is made to simplify understanding the terms and conditions associated with each incoterm. It is advisable to refer to the official website of the International Chamber of Commerce for a full version of Incoterms. In figure 1, the top row shows the categories of shipment, Sea and Inland waterway transport, and Any Mode/modes of transportation. The subsequent row displays the different Incoterms, beginning from left to right. The next row indicates the freight collection terms and the prepaid freight terms. The exporter or seller of the items starts on the left side, at the seller's site or warehouse. As we go to the right, the product leaves the warehouse, is loaded onto a vessel or plane at the loading port, is sent to the destination port by customs in the receiving country and is finally delivered to the customer's location. Along the route, incoterms determine which risks and expenses the seller agrees to bear, and which are passed on to the buyer. The chart's left side shows the various obligations and costs paid by the seller and carried by the customer. The first incoterm on the left is EXW, which stands for Ex-works (Rosal, 2016) or Ex-Warehouse. If the customer and vendor agree to sell products on an ex-works basis, the seller's responsibilities are straightforward. The seller will meet only the cost of the items and export packing. As a result, the seller will produce the items, package them, and have them available for pickup from their warehouse.

After then, the buyer is responsible for all extra expenses and risks associated with transportation from the facility. Under FCA, or Free-To-Carrier, the supplier will handle export duties, taxes, and customs clearance, to make the product ready for export. The seller will bear the origin terminal port handling expenses under FAS or Free-Alongside-Ship. Then there's FOB or Free-On-Board. FOB is the most often used incoterm in containerized trading. When FOB conditions are agreed upon, the seller's responsibility is to supply the products and pay for all additional costs associated with loading the commodities onto the vessel for export. This means the supplier will bear all the above mentioned expenses and the costs of loading the items onto the ship for export. All additional costs and risks are transferred to the buyer as soon as the products are on board. The buyer is accountable for the international freight and any subsequent expenses.

If the vendor agrees to cover the freight and carriage costs, the items can be sold under CFR, CPT, CIF, or CIP incoterms (Stojanović & Ivetić, 2020). CFR and CIF are incoterms that encompass both sea and inland waterway transportation. The CIF incoterm can be used if the seller agrees to cover the insurance cost during international maritime freight. CIP and CPT are comparable incoterms that apply to any method or mode of transportation when the seller agrees to pay for terminal/port handling costs at the desired location. The vendor can sell items under the CIP incoterm if they decide to bear the expense of insurance during transportation. DAP, DDP, and DPU are incoterms that define how things are delivered, unloaded, and customs approved in the destination country. The vendor will bear the delivery price to the destination under DAP or Delivered-At-Place. If the seller agrees to DPU (Delivered-at-Place-Unloaded), the seller will also cover the unloading costs at the destination. Finally, if the seller agrees to DDP (Delivery-Duty-Paid), the seller will also pay taxes, import charges, and customs clearance in the destination country. Before products are ready to be shipped internationally, the buyer and seller must agree on the incoterms that the goods are sold under. To avoid misunderstandings, the incoterms must be clearly stated in sales contracts and countersigned by both parties. In case of any dispute between the parties will refer to the details included in the documentation.

So far, through the review, we have learned about various data-driven techniques and their applications in the context of supply chain management. We have also understood the meaning of multiple incoterms and their importance in global trade. After analyzing the existing literature, we conclude this section by highlighting the research gaps. The problem of supplier performance has been pointed out in the current literature (Islam & Amin, 2020). Still, very few studies have come up with the idea of vendor incoterm selection in the global omnichannel pharmaceutical supply chain by using ML/data-driven techniques. The availability of pharmaceutical supplies at the right time in the correct quantity from the right source was a huge healthcare challenge during the COVID-19 pandemic (O'Brien et al., 2022). So there lies the scope of "developing a model for vendor incoterm selection using data-driven/ML techniques in Global Pharmaceutical SC (Zhu et al., 2021).

3. Machine Learning algorithms

One of the ways to define Machine Learning can be "It is associated with the design and development of algorithms and practices that allow computers to learn. ML research's major focus is automatically extracting data from data by statistical and computational methods. Hence, it is closely related to data statistics and data mining"(Vildósola & Pearson, n.d.). While various machine learning methods are available, we chose three for our research based on our needs: the K-Nearest Neighbor algorithm (KNN), the Decision tree algorithm, and the Random Forest algorithm.

3.1 K- Nearest Neighbors (KNN) is the most basic ML algorithm available. Constructing the model encompasses only storing the training dataset. The method searches the closest data points in the training dataset, referred to as "nearest neighbors," to create an estimate for a new data point (Müller & Guido, n.d.). The pseudo-code for the KNN algorithm is given in figure 2 below.

Classify (**P, Q,** r) // **P**: training data, **Q**: class labels of **P**, r: unknown sample **for**
i=1 **to** n **do**
Compute distance d ($\mathbf{P_i}$, r)
end for
Determine set I having indices for the k smallest distances d ($\mathbf{P_i}$, r)
return majority label for {$\mathbf{Q_i}$ where i ε I}

Figure 2 Pseudo Code of KNN Algorithm

The KNN algorithm, in its most basic form, only analyses one nearest neighbor, which is the nearest training data point concerning the point we need to estimate. The prediction is then merely the known output for this training point. A stepwise execution of the KNN algorithm is described below for a better understanding-

i. Load the data

ii. Initialize the value of k

iii. Iterate from 1 to the total number of training data points to determine the projected class.

  a. Calculate the distance between each row of training data and the test data. We'll use Euclidean distance as our distance metric since it's the most typical way. Cosine, Chebyshev, and other metrics can be used as well.

  b. Sort the computed distances in ascending order

  c. Get the first k rows of a sorted array

d. Get the most frequently occurring class of these rows

e. Return the predicted class

3.2 Decision tree- Decision trees are the most favored ML algorithms for categorization and regression applications. They learn a series of if/else questions that leads to a choice. Learning a decision tree is learning the sequence of if/else questions that leads to the actual answer in the quickest time possible (Müller & Guido, n.d.). In machine learning, these questions are called tests (not to be confused with the test set, the data we utilize to see how generalizable our model is). Data has frequently been described as continuous rather than a binary yes/no features. Tests of the kind "Is feature I greater than value a?" are used on continuous data. To generate a decision tree, the procedure iterates all possible tests to choose the one that provides the maximum information about the target variable.

A pseudo-code for the Decision tree algorithm is represented below in figure 3

Figure 3 Pseudo-code of Decision Tree Algorithm

```
GenDecTree (Sample C, Features V)
Steps:
    Ifstopping_conditions (C, V) = true then
        Leaf = createNode()
        leafLablel = classify(c)
        return leaf
    root = createNode()
    root.test_condition = findBestSplit(C, V)
    O = {o\o a possible outcomecfroot.test_condition}
    For each value o Є O:
        Cₒ = {c\ root.test_conditions(c) = o and c Є C};
        Child = TreeGrowth (Cₒ,V);
        Group child as descent of root and mark and label the edge {root → child}
        as o
    return root
```

We begin at the tree's root using the decision tree approach to predict a class label for a record. The values of the record's attribute and the root attribute are compared. We watch the branch corresponding to that value and move on to the next node based on the comparison. We correspond our data's attribute values to the tree's internal nodes until we locate a leaf node with the anticipated class value. We've already understood how to utilize the modeled decision tree to predict the desired class or value.

3.3 Random Forest- A random forest is a set of decision trees that are somewhat different. Random forests are based on the idea that while each tree may do a good job predicting, it will almost inevitably overfit some data. We can reduce overfitting by averaging the results of several trees performing well and overfitting in distinct ways. Using rigorous mathematics, this reduction in overfitting while keeping the predictive potential of the trees may be demonstrated (Müller & Guido, n.d.). Random forests receive their name because they use randomization in tree construction to ensure that each tree is unique. Figure 4 exhibits the pseudo-code for the random forest algorithm.

The trees in a random forest may be randomized in two ways: picking up the data points utilized to construct the tree and choosing the features in each split test. The algorithm initially predicts using

```
To generate n classifiers:
for i = 1 to n do

        randomly sample the training dataset D with replacement to produce Di
        Create a root node, Ni containing Di
        Call BuildTree(Ni)
end for
BuildTree(N):
if N comprises occurrences of only one class then
        return
else
        Randomly select y% of the probable splitting features in N
        Choose the feature F with the maximum information gain to split on
        Make f child nodes of N, N1,….Nf, where F has f conceivable values
        (F1,…Ff)
        for i = 1 to f do
                Set the contents of Ni to Di, where Di is all occurrences in N that
                match
                Fi
                Call BuildTree(Ni)
        end for
end if
```

every tree in the forest to make a prediction using the random forest. We can average these results to achieve our final prediction in regression.

The basic steps executed in the random forest algorithm (Garg et al., n.d.) are described below-

Figure 4 Pseudo-code of Random Forest Algorithm

The following is applicable for a = 1 to A:

(a) Create a bootstrap sample Z of size N from the training data,

(b) Fit the bootstrapped data to a random-forest tree by iteratively repeating the succeeding steps for each terminating node of the tree until the minimal node size n is attained.

i. Pick m variables at random from the predictor variables.

ii. Select the best variable/split-point among them.

iii. Creating two child nodes from a single node

1) Create the tree ensemble $\{\boldsymbol{T_a}\}_{\boldsymbol{1}}^{\boldsymbol{a}}$

To predict a new point x, the following syntax may be used:

Regression:

$$\boldsymbol{f}_{\boldsymbol{rf}}^{\boldsymbol{A}}(\boldsymbol{x}) = \frac{\boldsymbol{1}}{\boldsymbol{A}} \sum_{\boldsymbol{a}=\boldsymbol{1}}^{\boldsymbol{A}} \boldsymbol{T_a}(\boldsymbol{x})$$

Classification: Let $\hat{C}_a(x)$ be the prediction class of the a_{th} random-forest tree.

Then $\mathbf{C}_{\boldsymbol{rf}}^{\boldsymbol{A}}(\boldsymbol{x}) = majorityvote\ \{\hat{\mathbf{C}}_{\boldsymbol{a}}(\boldsymbol{x})\}_{\boldsymbol{1}}^{\boldsymbol{A}}$

4. Methodology

So, based on the research gap identified at the end of the literature review and our understanding of the fundamentals of data-driven approaches, we present a data-driven model for predicting a vendor incoterm for direct drop shipping of items in a Global Omnichannel Pharmaceutical SC. The dataset was initially gathered from an open source. The data collection included logistic information for the shipment of pharmaceuticals to client countries. The following phase was feature engineering, in which we determined which characteristics would serve as input variables for the model, with vendor incoterm being chosen as the model's target feature. Figure 5 depicts a comprehensive methodology flow chart that illustrates the steps followed throughout this model's creation.

The next step was to clean or organize the data points so they could be input into machine learning algorithms (data preprocessing). We did data preparation on direct drop-type shipment data (4920 data points). The difference between the PO sent date, and the recorded delivery date was used to calculate the shipment's lead time. Before moving further with the ML algorithm deployment, we eliminated the following items: entries with negative or zero lead time (685 data points), data points with missing freight cost (1317 data points), entries with no or zero weight (10 points), and entries with the missing mode of shipping (1 data point). Since air is the most extensively utilized method of transportation among the available data points, we created a model for air as a shipment mode. Finally, we were left with 2612 data points.

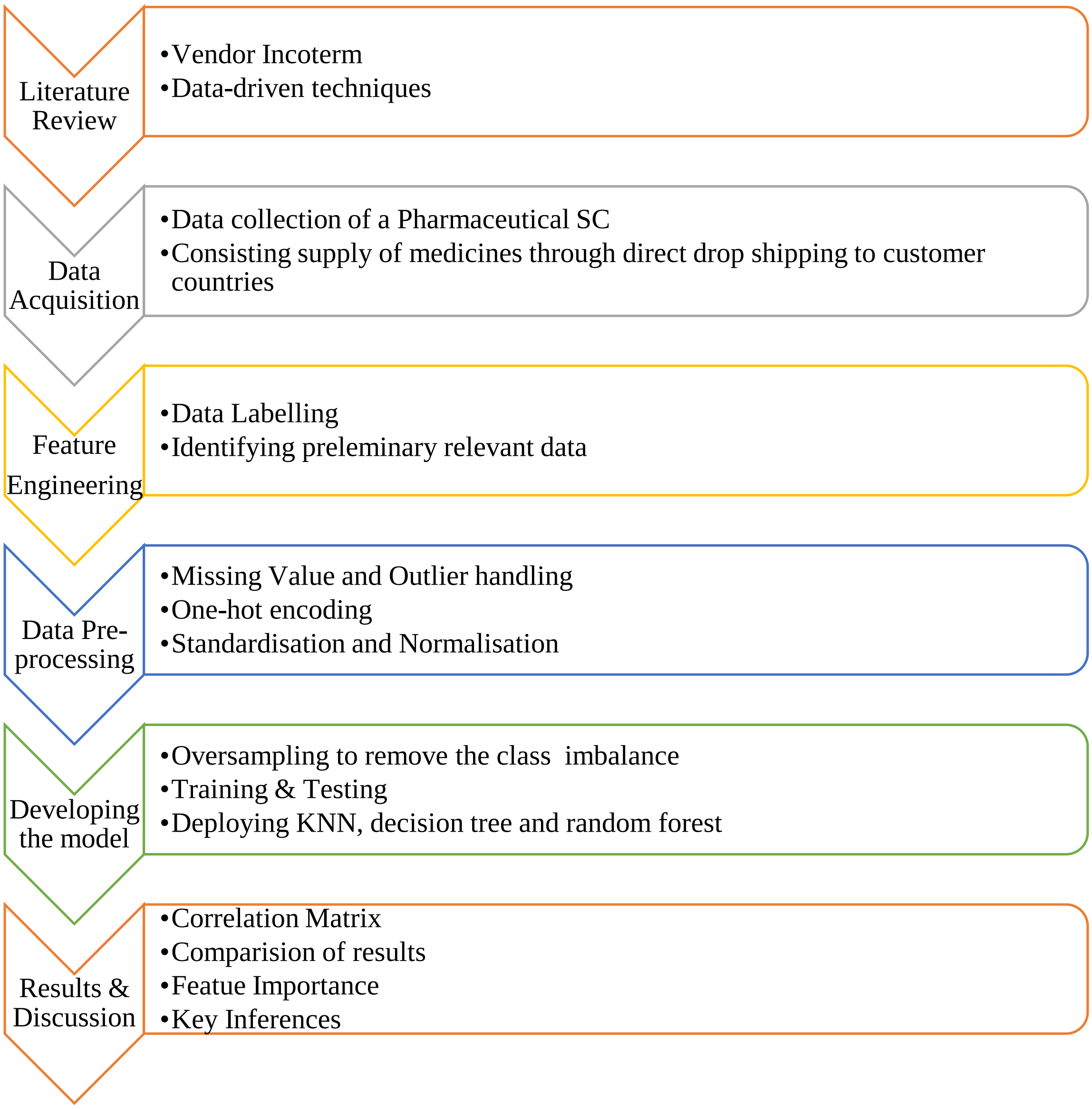

Figure 5 Proposed Methodology

Out of these 2612 data points, more than 50% of the data points were from a single category of vendor incoterm. This bias is called 'class imbalance' in the machine learning language. To negate the effect of class imbalance and to avoid overfitting of the data, we employed the "synthetic minority oversampling technique (SMOTE)" (Kulkarni et al., 2020) on our dataset. This oversampling technique duplicates the values of minority classes until the count becomes equal to that of the majority class with the highest data points. Although, this technique doesn't add any new information to the model.

Before the data points were supplied to the algorithms, the training and testing proportions of the data points had to be defined. For training and testing, we maintained a 70:30 ratio. This means that each of the three machine learning algorithms will use 70% of all data points to train the algorithm and the remaining 30% for prediction.

We used three machine learning algorithms on our dataset after determining this ratio: KNN, decision tree, and random forest. As explained in subsection 2.1 of the literature review, no single algorithm can work well with all data sets and for all sorts of requirements. We picked these three algorithms because they are easy to interpret and give good results with large datasets (refer to Table 1). Moreover, we plan to suggest the best algorithm to predict the vendor incoterm by comparing results obtained through these three algorithms.

Following the deployment of algorithms, the next step was to investigate the results and make appropriate interpretations based on the metrics acquired. We employed a correlation matrix to examine the relationships between the dataset's features. We used indicators like Accuracy, f-score, Precision, and recall comparing ML algorithms' performance. Afterward, we constructed a plot for feature importance to ascertain factors' influence on vendor incoterm selection.

Finally, we highlighted the managerial implications of our research based on our findings. This section discusses the potential impact of this strategy on the pharmaceutical industry and other stakeholders. In the last half of our essay, we discussed the limitations of this data-driven model's applicability and future research opportunities.

5. Case of Predicting Vendor Incoterm (contract)

This section will closely analyze the composition of SC for the pharmaceutical industry, the variables we have used to develop our model, the distribution of the data, and the dynamics of the data-driven model.

5.1 Composition of pharmaceutical supply chain

The case is represented by a dataset that contains logistics records of the shipping of medicine goods between the years 2006 to 2015. The data is obtained from a global repository of pharmaceutical supply chain records. PMO-US, a business consulting firm in the United States of America, oversees these shipments. There are 34 attributes, with 10324 data points in the dataset. The values of features mix practically all data kinds, with integer, floating-point, and string values. The medications are sent to 43 customer nations across four continents from records. There are mainly two supply chains: direct drop shipping and regional distribution centers (RDCs)-based shipment, which relies on a distribution network (Zhu et al., 2021) to provide pharmaceutical items.

5.2 Information about variables

We split the 34 accessible attributes into two categories: qualitative and quantitative. Country, Fulfil Via, Vendor Incoterm, Shipment Mode, Product Group, Sub Classification, Vendor, Manufacturing

Site, First Line Designation, medical information linked factors, and Identification associated variables are qualitative variables. Lead Time, Line-Item Quantity, Item Value, Pack Price, Unit of Measure, Pack Price, Unit Price, Weight, Freight Cost, and Line-Item Insurance are the quantitative variables. Table 2 lists all the variables present in the obtained dataset.

Figure 6 shows a statistical report for per kilogram freight costs expressed in US dollars (USD). We evaluated this parameter by dividing the total freight cost by the weight of the shipment. A-squared is the test statistic for the Anderson-Darling Normality Test. It is a measure of how strictly a dataset follows the normal distribution. The null hypothesis for this test is that the data is normal.

Table 2 Classification of available variables

Qualitative/Categorical Variables	**Quantitative/Numerical Variables**
Country (Customer)	PO Sent Date
Fulfill via (Direct Drop/RDC)	Scheduled Delivery Date
Vendor Incoterms (EXW, DDP FCA, etc.)	Delivered to client date
Shipment Mode (Air, Truck, etc.)	Delivery recorded date
Product Group (HRDT, ARV)	Unit of measure (per pack)
Subgroup (Pediatric/Adult)	Line-item Quantity
Vendor	Line-item value
Item Description	Pack price
Brand	Unit price
Dosage Form	Weights (Kgs.)
Manufacturing Site	Freight cost (USD)
First Line Designation	Line-item insurance (USD)

So, if we get a rationally large A-squared, you will get a small p-value and thus reject the null hypothesis. 'A-squared value' for per kilogram freight cost data points is 776.63. The P-value, less than 0.005 in this case, is very low and indicates a significant result. The smaller the P-value, the more evidence that the null hypothesis is probably wrong

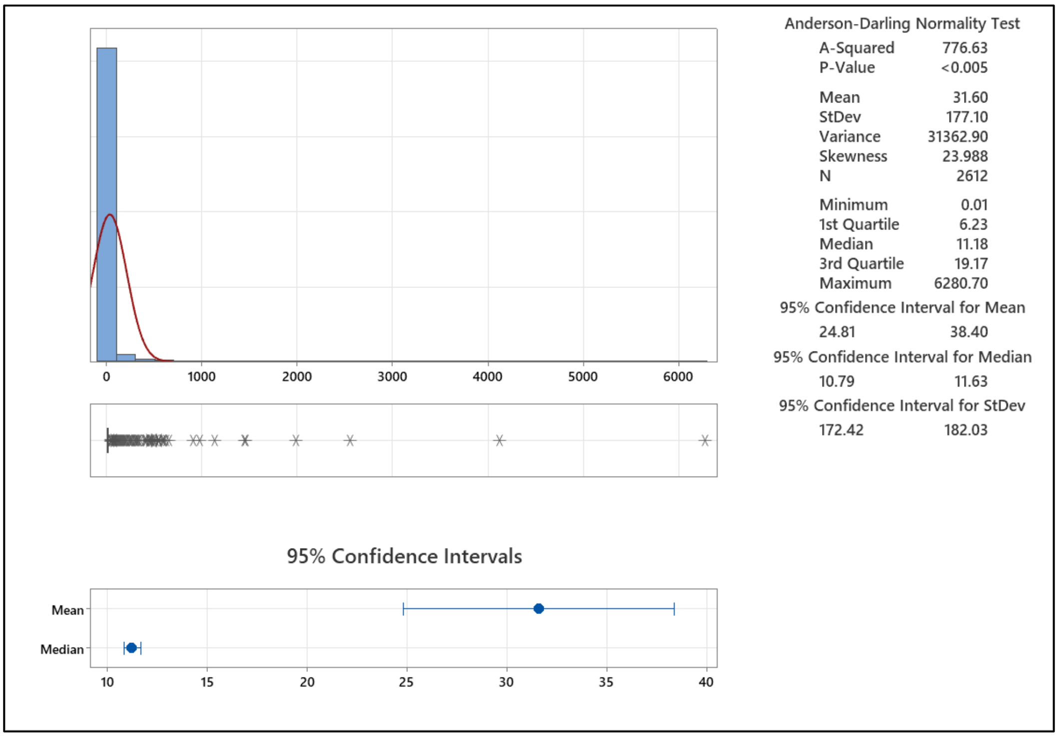

Figure 6 Summary Report for Per Kilogram Freight Cost (in USD)

On the other hand, small A-squared values imply large p-values; thus, we cannot reject the null hypothesis. The mean per kilogram freight cost came out to be 31.6 USD, while the maximum value was 6280.7 USD among 2612 data points. The variance, standard deviation, and skewness values are high, representing high variation in freight cost data points. Statistical analysis includes minimum, interquartile, and maximum values per kilogram freight cost (USD). In the end, lower and upper limit values of mean, median, and standard deviation are listed with a 95% confidence interval.

Figure 7 shows a statistical summary report for Lead Time expressed in the number of days. We have calculated the lead time by taking the difference between the 'date of sending the PO to vendor' and the 'date of delivery to the client.' The 'lead time' function shows the time between placing an order for a product and receiving it. In our data, the lead time spans from 0 to 88 weeks. A large A-squared value (67.5) and low p-value (<0.005) means data points for lead time do not follow the normal distribution (fails the Andersen-Darling normality test). Although, the skewness value is comparatively lower for lead time data points. All the lead time values lie from 12 days to 616 days. The report shown in figure 7 also depicts the box plot (including interquartile range values) variation and 95% confidence interval values of mean, median, and standard deviation for lead time data points.

Since both the quantitative variables, per kilogram freight cost and lead time, fail the normality test, we opted for standardization and normalization at the data preprocessing stage of our model development to account for this variation, as shown in the methodology flow chart (refer to figure 5).

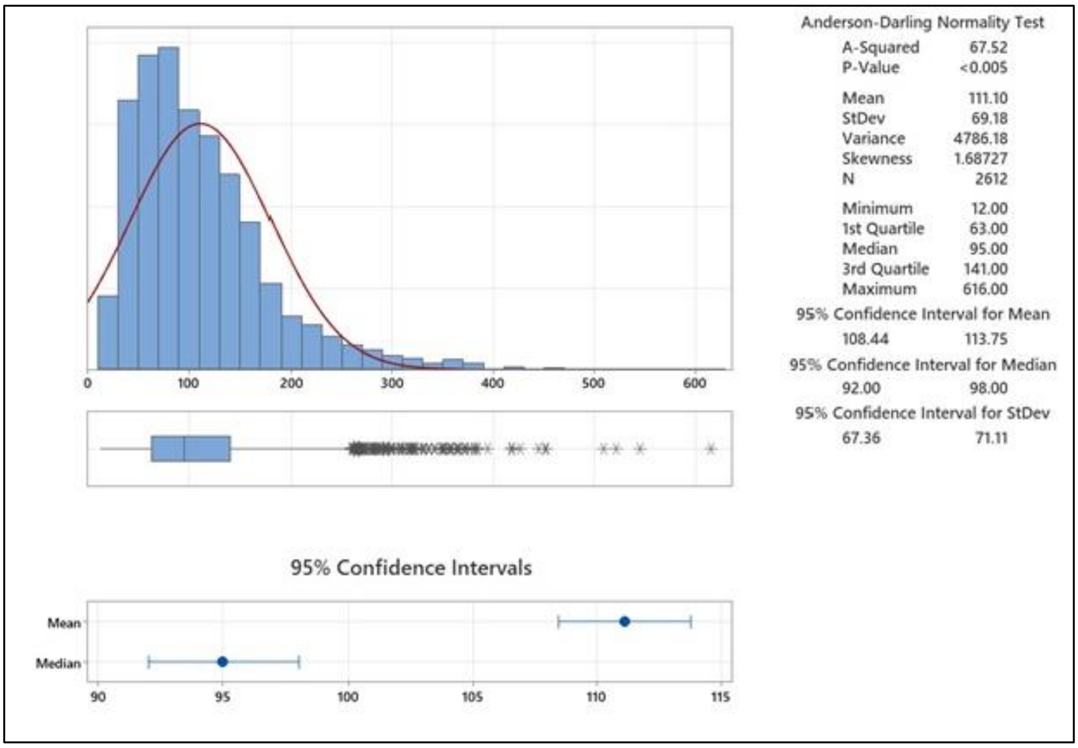

Figure 7 Summary Report for Lead Time (in days)

5.3 The Model

We have constructed a model around this diversified range of attributes and accompanying data points that can predict a suitable vendor incoterm for a given set of input parameters. This study plans to use the most common data attributes for each firm. In our proposed model, we chose 'vendor incoterm' as our response variable (label) and 'product group, sub-classification, vendor, per kg freight cost (in USD), and lead time' as our predictive features. As established before, we deployed three ML algorithms to develop the model- K-nearest neighbor (KNN), decision tree, and random forest algorithm.

After understanding the available attributes, their characteristics, the variation in data values, and the composition of our model, we will now discuss the results obtained from this ML-based prediction model in the next section. This will be followed by an evaluation of the model using various performance matrices. The forthcoming section will also quantify the parameters that affect the selection of vendor incoterm in a direct drop shipping environment.

6. Results and discussion

First, we developed a correlation matrix to check the applicability of linear regression models. Since any of the values in the correlation matrix (shown in table 3) is not close to either 1 or -1, we can say that no direct or inverse relationship exists between any two variables. A linear regression model would not lead to a conclusive model as the relationship between any two sets of variables is weak or inconsequential. Hence to evaluate the overall effect of all five input variables on the evaluation of Vendor Incoterm, we deployed three distinct ML algorithms (on the preprocessed dataset) falling under the category of supervised learning.

Table 3 Correlation matrix

	Per Kg Freight Cost (in USD)	Lead time (days)	Vendor Incoterm	Product Group	Sub-classification	Vendor
Per Kg Freight Cost (in USD)	1.000000	-0.009022	0.010154	0.055173	0.005956	0.051812
Lead time (days)	-0.009022	1.000000	-0.091539	-0.111822	0.038450	-0.059865
Vendor Incoterm	0.010154	-0.091539	1.000000	0.124100	0.085799	-0.013019
Product Group	0.055173	-0.111822	0.124100	1.000000	0.006676	0.626020
Sub-classification	0.005956	0.038450	0.085799	0.006676	1.000000	-0.141805
Vendor	0.051812	-0.059865	-0.013019	0.626020	-0.141805	1.000000

We further compared and analyzed the results from these algorithms to evaluate the performance of the developed model. Compared to KNN and the Decision tree, the Random Forest algorithm performs best. We first use Accuracy and F1-Score as our metrics to compare the results. F1 metric is a predominantly used metric for the case of multi-classification problems. The more the value of the F1-Score, the better the prediction effect of the model (Zhou et al., 2021). F1-Score and prediction accuracy for these three algorithms are shown in Table 4.

Table 4 Comparison based on F1 score and accuracy

S. No	ML Algorithm	F1-Score	Accuracy
1.	K-Nearest Neighbor (KNN)	0.98	98.09%
2.	Decision Tree	0.99	99.48%
3.	Random Forest	1.00	99.52%

The confusion matrix is a widely used metric for analyzing classification problems. It can be used to solve issues in both binaries and multiclass classification. Table 5 illustrates a confusion matrix for the case of binary type. Confusion matrices provide counts based on predicted and actual values (Kulkarni et al., 2020).

Table 5 Confusion matrix for binary classification.

		Predicted	
		Negative	**Positive**
	Negative	TN	FP
Actual	**Positive**	FN	TP

The output "TN" stands for True Negative and exhibits the number of correctly recognized negative cases. In the same way, "TP," also known as True Positive, signifies the number of correctly identified positive instances. "FP" is a shorter form of False Positive value representing the number of actual negative examples classified as positive, and "FN" is an abbreviation for False Negative value, meaning the number of real positive examples classified as negative. So, values on the diagonal of the matrix are true values, and it's preferable if the maximum predicted values fall on the diagonal of the matrix. The confusion matrix obtained from three ML algorithms is illustrated in Table 6, Table 7, and Table 8.

Table 6 Confusion matrix for K- Nearest Neighbor Algorithm (KNN)

[609	0	0	0	0	0	0]
[0	594	0	0	0	0	0]
[0	0	579	0	0	0	0]
[0	0	0	597	0	0	0]
[0	0	0	0	599	0	0]
[2	45	0	26	0	544	1]
[0	2	0	0	0	4	596]

Table 7 Confusion matrix for Decision Tree Algorithm

[609	0	0	0	0	0	0]
[0	594	0	0	0	0	0]
[0	0	579	0	0	0	0]
[0	0	0	597	0	0	0]
[0	0	0	0	599	0	0]
[0	11	0	4	0	596	7]
[0	0	0	0	0	0	602]

Table 8 Confusion matrix Random Forest Algorithm

[609	0	0	0	0	0	0]
[0	594	0	0	0	0	0]
[0	0	579	0	0	0	0]
[0	0	0	597	0	0	0]

$$
\begin{matrix}
[0 & 0 & 0 & 0 & 599 & 0 & 0] \\
[0 & 0 & 0 & 7 & 0 & 598 & 2] \\
[0 & 0 & 0 & 0 & 0 & 0 & 602]
\end{matrix}
$$

The confusion matrix for Random Forest Algorithm has the most significant number of values at the diagonal; hence it's most suited for prediction among all three algorithms. Accuracy is among the most widely utilized measures when solving classification problems. The following expression is used to determine the accuracy of a model with the help of a confusion matrix.

$$Accuracy = \frac{TN+TP}{TN+FP+FN+TP} \qquad \ldots\text{(i)}$$

Furthermore, precision and recall are also famous metrics in classification scenarios. Precision shows how well the model predicts positive values. As a result, it judges the accuracy with which a positive outcome is projected. The positive predictive value is an alternative name for precision. Recall is suitable for assessing the strength of a model to forecast positive results, and incidentally, it is called the sensitivity of a model (Kulkarni et al., 2020). Both measurements provide helpful information, but the goal is to enhance recall without compromising precision. Results of precision and recall matrices for three ML algorithms are compared in Table 9 below.

$$Precision = \frac{TP}{TP+FP} \qquad \ldots\text{(ii)}$$

$$Recall = \frac{TP}{TP+FN} \qquad \ldots\text{(iii)}$$

Table 9 Comparison based on Precision and Recall

S. No	ML Algorithm	Precision	Recall
1.	K-Nearest Neighbor (KNN)	0.98	0.98
2.	Decision Tree	0.99	0.99
3.	Random Forest	1.00	1.00

The above results show that the random forest algorithm can best predict the vendor incoterm for inputs of Product Group, Product Sub classification, Vendor, Per Kg. Freight Cost and Lead Time. The Random Forest algorithm-based model yielded the best f1-score, accuracy, precision, and recall results. Alternatively, we can use the decision tree or KNN-based model to get the vendor incoterm for a particular input scenario. The accuracy achieved by these two models, too, is on the higher side, and we believe the applicability of these two would be alright.

To answer our first research question, we have developed a machine learning-based model and identified the best-performing model that can predict the vendor incoterm for input parameter values. Now we will understand which among the five input factors has the maximum impact on the vendor incoterm prediction, our second research question.

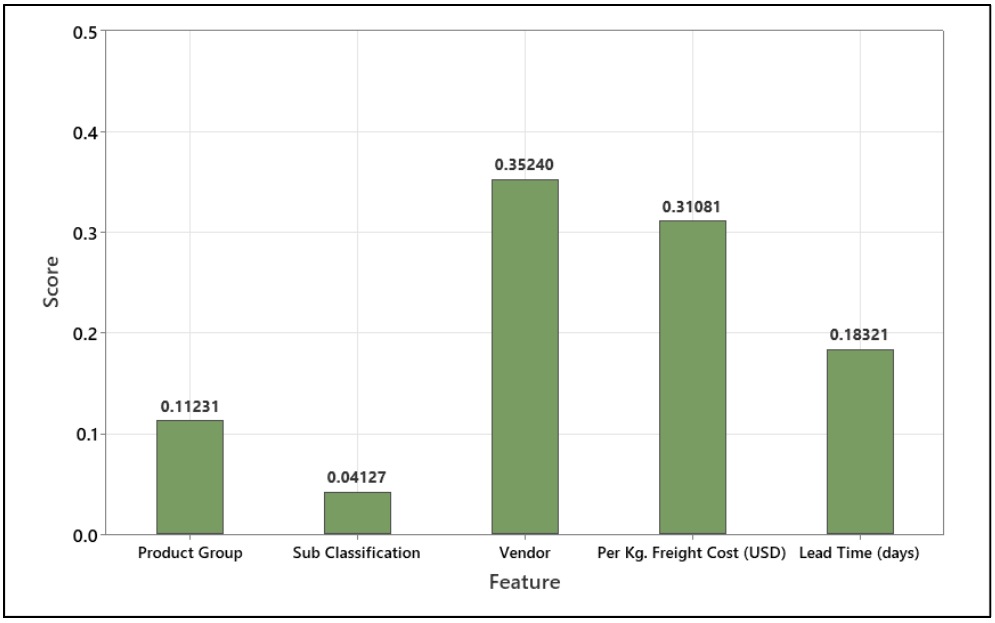

Figure 8 Plot for feature importance

For this purpose, we will analyze the importance plot of input features, as shown in figure 8. This plot shows the feature's importance, which means the higher the score, the more influential the feature is. Feature 'Vendor' has the highest score; hence it is of maximum importance concerning predicting the vendor incoterm, while feature 'Sub Classification' is of least importance among the five input attributes.

6.1 Prediction through Random Forest Algorithm- An Example Case

In this part, we'll simulate the model by comparing the produced output to a string of input variables. This sample set of input and output is also stated in terms of interpretation.

Input-

User String: [[ARV, Adult, Aurobindo Pharma Limited, 5, 60]]

Encoded String: [[2, 1, 7, 5, 60]]

Output-

Encoded String: [5]

Interpreted Output: EXW

This implies that if we use 'ARV' as the Product Group, 'Adult' as the Subclassification, 'Aurobindo Pharma Limited' as the Vendor, '5' as the Per Kg. Freight Cost (in USD), and '60' as the Lead Time (in days), we'll receive EXW or Ex-Works as our vendor incoterm. As a result, the EXW type of vendor incoterm is recommended for these input variable values.

6.2 Theoretical and Managerial Implications

Our study has contributed to the growing research sector emphasizing machine learning and data analytics in healthcare supply chains. The era of data-driven technology has enabled it to handle massive amounts of medical data. The healthcare sector has grown tremendously in recent decades because of increased awareness and technological advances (Dixit et al., 2019); other disruptions that health systems have faced include unavailability, inaccessibility, overcrowding, unaffordability, resulting in long wait times, and length of stays (Almeida & Vales, 2020; Supeekit et al., 2016).

Complex global omnichannel pharmaceutical supply chains that deal with shipping medical goods across continents can benefit from this data-driven model. We offer the following direct and indirect managerial implications that result from our study-

i. The model can expedite the entire process from selecting a suitable vendor incoterm for shipment until the material is delivered to the destination.

ii. The data-driven model has a certain effect on the global supply chains concerned with the shipment of goods through a direct drop mechanism.

iii. The proposed model can improve the overall supply chain efficiency by predicting suitable vendor incoterm for given demand and lead time requirements.

iv. We identified attributes and their relative importance that affect the choice of Vendor Incoterm selection in a direct drop shipping environment. It will help suppliers, vendors, customers, and other stakeholders make better decisions.

Using the proposed model, we not only enabled the selection of vendor incoterms in a direct drop shipping environment but also identified the factors which have a bearing on the selection of these incoterms in the pharma supply chain. This model will serve as a benchmark for future research concerning the distribution of medical supplies through direct drop type of distribution.

To our knowledge, despite the criticality of efficiency in the healthcare sector, particularly in times of crisis like a pandemic, no study has been done on employing methods like machine learning to forecast vendor incoterm in a worldwide omnichannel pharmaceutical supply chain environment. We identified the factors based on which future models can be developed and compared. Hence, this model will potentially save time and scholarly efforts for future research.

7. Conclusion

This study aims to develop a data-driven model to select the Vendor incoterm for a direct drop type shipping scheme in a global omnichannel pharmaceutical supply chain. Here we obtained the dataset containing shipment data of therapeutic goods across various countries from 2006 to 2015. We thoroughly analyzed the nature and variation of available attributes. After the data preprocessing stage, we developed our model utilizing three ML algorithms- KNN, decision tree, and random forest algorithm. We then compared the results of these algorithms using performance matrices such as f1 score, accuracy, precision, and recall. After comparison, we found that the model based on the random forest algorithm gave the best results among all three algorithms. We then plotted the feature importance to evaluate input factors' effect on vendor incoterm prediction. The 'Vendor' feature returned the highest importance score, while the 'Product sub-classification feature was the least important. Further, we also proposed theoretical and managerial implications of our research endeavor.

Upcoming studies can focus on developing the model for industries other than pharmaceuticals, where direct drop-type shipping is involved. Researchers can also create a model to select the appropriate vendor instead of incoterm for the same input parameters in the pharmaceutical supply chain.

Like every research, there exist a few limitations in our study. We have developed our model by taking "Air" as the sole mode of shipment. The possibility of Trucks, Rails, or waterways as other shipment modes can be explored in future research. We have also limited our study to direct drop-shipping data points for pharmaceutical goods. This work can be further persuaded to develop a model for a supply chain network fulfilling orders through regional distribution centers (RDCs).

Credit authorship contribution statement

Pankaj Kumar Detwal: Writing – original draft, Conceptualization, Methodology, Data curation, Modeling. **Gunjan Soni:** Writing – review & editing, Supervision, Methodology, Modeling. **Suresh Kumar Jhakar:** Writing – review & editing, Formal analysis. **Deepak Kumar Srivastava:** Writing – review & editing, Visualization. **Jitender Madan:** Writing – review & editing, Investigation. **Yasanur Kayikci:** Writing – review & editing, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Almeida, A., & Vales, J. (2020). The impact of primary health care reform on hospital emergency department overcrowding: Evidence from the Portuguese reform. *The International Journal of Health Planning and Management*, *35*(1), 368–377. https://doi.org/10.1002/hpm.2939

Bergami, R. (2013). Managing Incoterms 2010 risks: tension with trade and banking practices. *International Journal of Economics and Business Research*, *6*(3), 324. https://doi.org/10.1504/IJEBR.2013.056126

Chang, V., Doan, L. M. T., Ariel Xu, Q., Hall, K., Anna Wang, Y., & Mustafa Kamal, M. (2023). Digitalization in omnichannel healthcare supply chain businesses: The role of smart wearable devices. *Journal of Business Research*, *156*, 113369. https://doi.org/10.1016/j.jbusres.2022.113369

Chopra, S., & Meindl, P. (2007). *Supply chain management : strategy, planning, and operation*. Pearson Prentice Hall.

Clauson, K. A., Breeden, E. A., Davidson, C., & Mackey, T. K. (2018). Leveraging Blockchain Technology to Enhance Supply Chain Management in Healthcare: *Blockchain in Healthcare Today*. https://doi.org/10.30953/bhty.v1.20

Dahl, A. J., Milne, G. R., & Peltier, J. W. (2021). Digital health information seeking in an omni-channel environment: A shared decision-making and service-dominant logic perspective. *Journal of Business Research*, *125*, 840–850. https://doi.org/10.1016/j.jbusres.2019.02.025

Dixit, A., Routroy, S., & Dubey, S. K. (2019). A systematic literature review of healthcare supply chain and implications of future research. *International Journal of Pharmaceutical and Healthcare Marketing*, *13*(4), 405–435. https://doi.org/10.1108/IJPHM-05-2018-0028

Dolgui, A., Ivanov, D., & Sokolov, B. (2018). Ripple effect in the supply chain: an analysis and recent literature. *International Journal of Production Research*, *56*(1–2), 414–430. https://doi.org/10.1080/00207543.2017.1387680

Falamarzi, A., Moridpour, S., & Nazem, M. (2019). A review of rail track degradation prediction models. *Australian Journal of Civil Engineering*, *17*(2), 152–166. https://doi.org/10.1080/14488353.2019.1667710

Garg, P., Patil, A., Soni, G., Keprate, A., & Arora, S. (n.d.). *Machine Learning-Based Abnormality Detection Approach for Vacuum Pump Assembly Line*.

Han, C., & Zhang, Q. (2021). Optimization of supply chain efficiency management based on machine learning and neural network. *Neural Computing and Applications*, *33*(5), 1419–1433. https://doi.org/10.1007/s00521-020-05023-1

Haszlinna Mustaffa, N., & Potter, A. (2009). Healthcare supply chain management in Malaysia: a case study. *Supply Chain Management: An International Journal*, *14*(3), 234–243. https://doi.org/10.1108/13598540910954575

Huang, Y.-Y., & Handfield, R. B. (2015). Measuring the benefits of ERP on supply management maturity model: a “big data” method. *International Journal of Operations & Production Management*, *35*(1), 2–25. https://doi.org/10.1108/IJOPM-07-2013-0341

INCOTERMS® 2020 INTERNATIONAL CHAMBER OF COMMERCE (ICC). (n.d.).

Islam, S., & Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting

Machine learning techniques. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00345-2

Konovalenko, I., & Ludwig, A. (2021). Comparison of machine learning classifiers: A case study of temperature alarms in a pharmaceutical supply chain. *Information Systems*, *100*, 101759. https://doi.org/10.1016/j.is.2021.101759

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In *Data Democracy* (pp. 83–106). Elsevier. https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Li, D., Zhao, Y., Zhang, L., Chen, X., & Cao, C. (2018). Impact of quality management on green innovation. *Journal of Cleaner Production*, *170*, 462–470. https://doi.org/10.1016/j.jclepro.2017.09.158

Meijboom, B., Schmidt- Bakx, S., & Westert, G. (2011). Supply chain management practices for improving patient- oriented care. *Supply Chain Management: An International Journal*, *16*(3), 166–175. https://doi.org/10.1108/13598541111127155

Micol Policarpo, L., da Silveira, D. E., da Rosa Righi, R., Antunes Stoffel, R., da Costa, C. A., Victória Barbosa, J. L., Scorsatto, R., & Arcot, T. (2021). Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Computer Science Review*, *41*, 100414. https://doi.org/10.1016/j.cosrev.2021.100414

Mohamed-Iliasse, M., Loubna, B., & Abdelaziz, B. (2020). Is Machine Learning Revolutionizing Supply Chain? *2020 5th International Conference on Logistics Operations Management (GOL)*, 1–10. https://doi.org/10.1109/GOL49479.2020.9314713

Müller, A. C., & Guido, S. (n.d.). *Introduction to Machine Learning with Python A GUIDE FOR DATA SCIENTISTS Introduction to Machine Learning with Python*.

O'Brien, N., Flott, K., Bray, O., Shaw, A., & Durkin, M. (2022). Implementation of initiatives designed to improve healthcare worker health and wellbeing during the COVID-19 pandemic: comparative case studies from 13 healthcare provider organisations globally. *Globalization and Health*, *18*(1), 24. https://doi.org/10.1186/s12992-022-00818-4

Pacella, M., & Papadia, G. (2021). Evaluation of deep learning with long short-term memory networks for time series forecasting in supply chain management. *Procedia CIRP*, *99*, 604–609. https://doi.org/10.1016/j.procir.2021.03.081

Pereira, M. M., & Frazzon, E. M. (2021). A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains.

International Journal of Information Management, *57*. https://doi.org/10.1016/j.ijinfomgt.2020.102165

Rosal, I. del. (2016). Factors influencing the choice of delivery terms used in Spanish seaborne container trade. *International Journal of Shipping and Transport Logistics*, *8*(3), 318. https://doi.org/10.1504/IJSTL.2016.076261

Roßmann, B., Canzaniello, A., von der Gracht, H., & Hartmann, E. (2018). The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study. *Technological Forecasting and Social Change*, *130*, 135–149. https://doi.org/10.1016/j.techfore.2017.10.005

Sakib, N., Ibne Hossain, N. U., Nur, F., Talluri, S., Jaradat, R., & Lawrence, J. M. (2021). An assessment of probabilistic disaster in the oil and gas supply chain leveraging Bayesian belief network. *International Journal of Production Economics*, *235*. https://doi.org/10.1016/j.ijpe.2021.108107

Schaefer, T. J. (2017). *Incoterms® Use in Buyer-Seller Relationships: a Mixed Methods Study*. https://irl.umsl.edu/dissertation/692

Shahbazi, Z., & Byun, Y.-C. (2020). A Procedure for Tracing Supply Chains for Perishable Food Based on Blockchain, Machine Learning and Fuzzy Logic. *Electronics*, *10*(1), 41. https://doi.org/10.3390/electronics10010041

Stojanović, Đ., & Ivetić, J. (2020). Possibilities of using Incoterms clauses in a country logistics performance assessment and benchmarking. *Transport Policy*, *98*, 217–228. https://doi.org/10.1016/j.tranpol.2020.03.012

Suma, S., Mehmood, R., Albugami, N., Katib, I., & Albeshri, A. (2017). Enabling Next Generation Logistics and Planning for Smarter Societies. *Procedia Computer Science*, *109*, 1122–1127. https://doi.org/10.1016/j.procs.2017.05.440

Supeekit, T., Somboonwiwat, T., & Kritchanchai, D. (2016). DEMATEL-modified ANP to evaluate internal hospital supply chain performance. *Computers & Industrial Engineering*, *102*, 318–330. https://doi.org/10.1016/j.cie.2016.07.019

Vildósola, A., & Pearson, A. (n.d.). *Machine-learning technologies in telecommunications*.

Vogt, J., & Davis, J. (2020). The State of Incoterm® Research. *Transportation Journal*, *59*(3), 304–324. https://doi.org/10.5325/transportationj.59.3.0304

Zhou, Y., Song, X., & Zhou, M. (2021). Supply Chain Fraud Prediction Based On XGBoost Method. *2021 IEEE 2nd International Conference on Big Data,* *Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 539–542. https://doi.org/10.1109/ICBAIE52039.2021.9389949

Zhu, X., Ninh, A., Zhao, H., & Liu, Z. (2021). Demand Forecasting with Supply-Chain Information and Machine Learning: Evidence in the Pharmaceutical Industry. *Production and Operations Management*, *30*(9), 3231–3252. https://doi.org/10.1111/poms.13426