Predicting Global Supply Chain Outcomes for Essential HIV Medicines

Muskan Arora

Date: 05-05-2023

Abstract:

This study aims to predict global supply chain outcomes for essential HIV medicines by analyzing publicly available data from PEPFAR between 2006-2015. The study will determine the factors that influence the timeliness of pharmaceutical deliveries and use them to develop a model that can predict potential delays and their duration. While machine learning has been used in supply chain analysis for demand forecasting, this study focuses on predicting lead-time directly. Previous research, such as SVMs and RNNs, have shown promising results in similar problems like predicting flight delays and improving flight efficiency. By utilizing machine learning, this study hopes to improve the efficiency and effectiveness of HIV medicine supply chains, ultimately benefiting patients worldwide.

1.0 Introduction

The HIV/AIDS epidemic continues to be a major public health concern, with millions of people worldwide affected by the disease. Despite significant progress in reducing the death rate, there are still numerous challenges to ensuring that those living with HIV have access to life-saving treatment. One critical aspect of this effort is the timely delivery of essential HIV medications, which is essential for preventing the spread of the virus and improving health outcomes. However, the delivery of these medicines can be subject to delays and disruptions, which can have serious consequences for people living with HIV.

In this context, predicting global supply chain outcomes for essential HIV medicines is an important area of research. The purpose of this work is to develop a model that can accurately predict the delivery time for specific HIV commodities and identify the factors that influence the timeliness of pharmaceutical deliveries. By doing so, this study aims to improve the efficiency and effectiveness of the supply chain for HIV medications and ultimately improve health outcomes for people living with HIV.

The scope of this work is focused on using publicly available data from PEPFAR over the years 2006-2015 to develop a model for predicting delivery times. The objectives of this study include identifying the factors that influence delivery times, developing a predictive model for delivery times, and assessing the accuracy of the model. The potential impact of this work is significant, as it could lead to improvements in the delivery of essential HIV medications and ultimately contribute to the global effort to end the HIV/AIDS epidemic.

1.1 Initial Needs Statement

There is a critical need to improve the delivery of essential HIV medicines to people living with HIV (PLHIV) around the world. More than 36.7 million people are living with HIV, and every year, about 1 million people die from AIDS-related causes. While the death rate has decreased significantly since 2001, about 1.8 million people became newly infected in 2016 alone. The epidemic disproportionately affects low-income countries in Eastern and Southern Africa, where women, adolescents, and key populations like female sex workers and LGBTQ individuals are the most affected groups. The President's Emergency Plan for AIDS Relief (PEPFAR) is a key player in the procurement of drugs, testing, and laboratory kits for HIV. However, recent changes have resulted in drastic declines in supply chain performance. Thus, there is a pressing need to predict whether or not HIV drugs will be delivered on time and how long potential delays will be to ensure that PLHIV around the world receive timely and adequate treatment.

This need statement highlights the critical importance of timely and adequate delivery of essential HIV medicines to PLHIV around the world. It also emphasizes the challenges faced by PEPFAR in meeting this need, particularly in light of recent declines in supply chain performance. The need for a predictive model that can forecast delivery delays and inform decision-making is evident. Such a model would be invaluable in improving supply chain performance and ensuring that PLHIV have access to the life-saving treatment they need.

2.0 Customer Needs Assessment

To develop a predictive model that can accurately identify delivery delays of essential HIV medicines, it is important to understand the needs of the customers who will be using the model. In this case, the customers would primarily be PEPFAR and other organizations involved in the procurement and delivery of HIV medicines.

To gather customer needs, the design team could conduct interviews, surveys, and focus groups with key stakeholders in the supply chain, including healthcare workers, logistics providers, and representatives from PEPFAR and other organizations. The team could also analyze existing data on supply chain performance and use this information to identify pain points and areas for improvement.

Based on this research, the following customer needs could be identified:

1. Timely delivery of essential HIV medicines to patients around the world

2. Accurate and reliable forecasting of delivery delays to allow for better supply chain management

3. Improved communication and collaboration between different stakeholders in the supply chain to ensure efficient and effective delivery of medicines

4. Cost-effective solutions that are feasible for low-income countries and resource-limited settings.

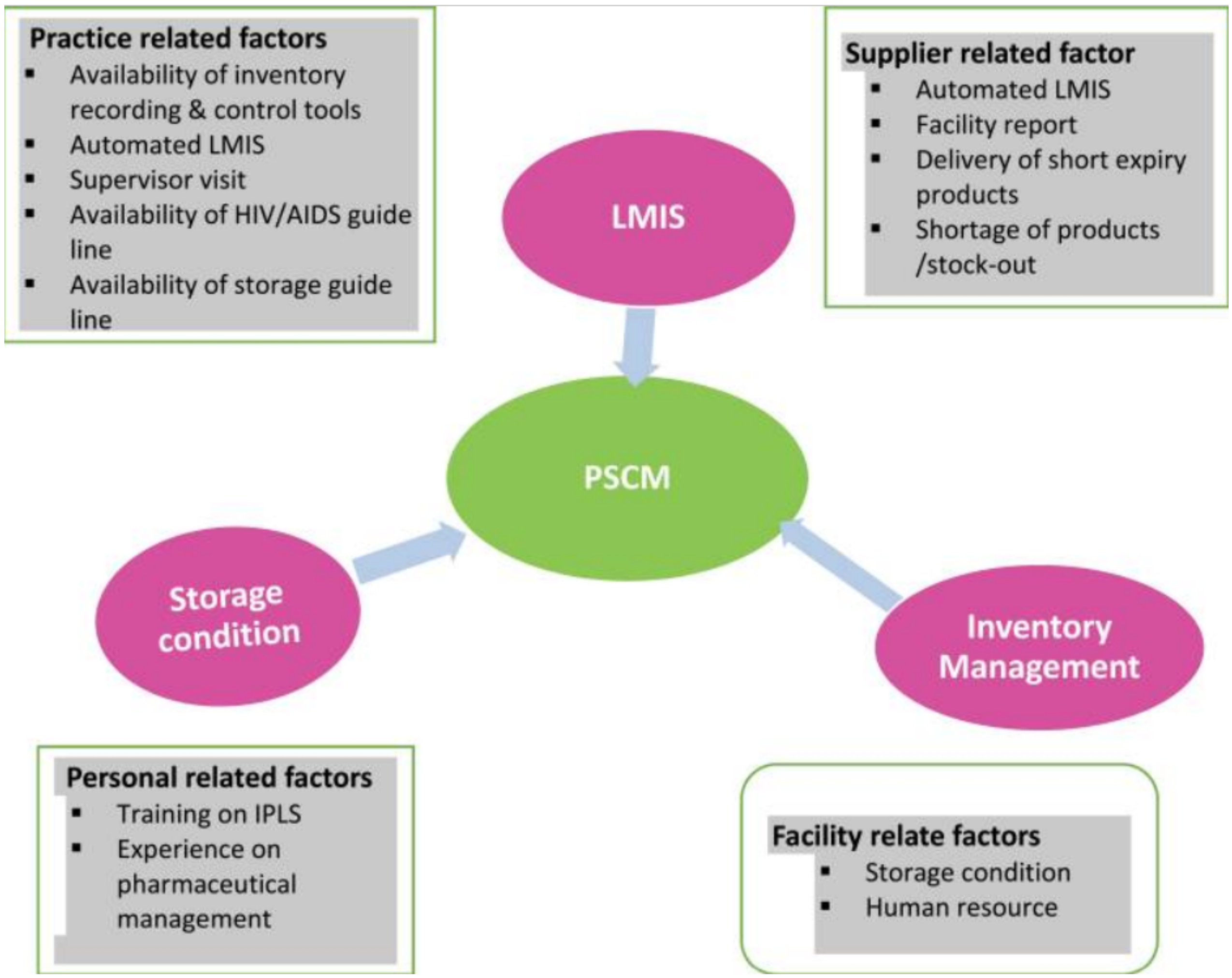


Fig 2. Conceptual framework of the study. [2]

The customer needs assessment would be an iterative process, with the design team using customer feedback to refine and improve the predictive model. For example, after conducting initial customer research, the team might develop a prototype model and present it to stakeholders for feedback. Based on this feedback, the team could make modifications to the model, such as adding additional data sources or improving the accuracy of the forecasting algorithm.

Throughout the development process, the design team would continue to solicit feedback from customers and incorporate this feedback into the model. This iterative approach would ensure that the final product meets the needs of customers and is well-suited for use in real-world supply chain settings.

Customer Need	Importance Rating (1-5)
Predictability of drug delivery	5
Timely delivery of drugs	5
Assurance of drug quality	4
Cost-effectiveness of drugs	3
Flexibility of supply chain	3
Transparency of supply chain	4
Collaboration and communication among stakeholders	5
Efficient and effective use of data	4
Adaptability to local contexts	4
Sustainability of supply chain	4

Table 1. Initial Customer Needs List Obtained from Interviews and Observations [1]

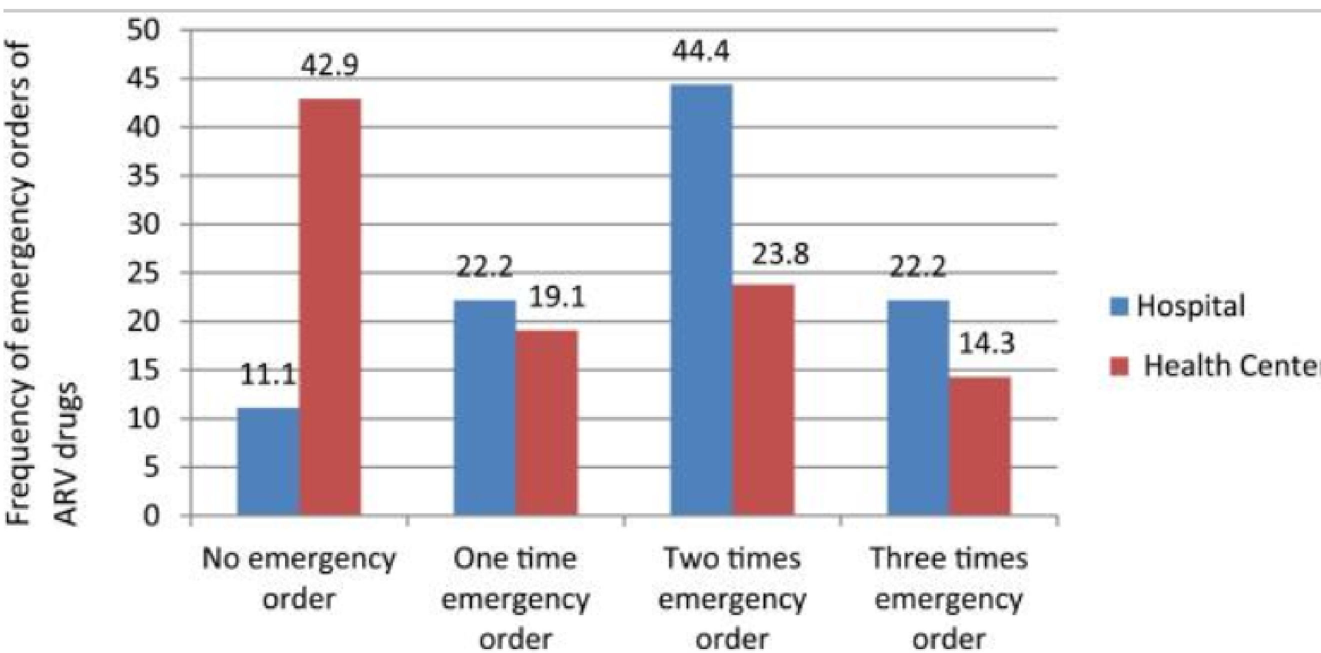


Fig 2. Frequency of emergency orders of ARV drugs by facility types of SNNPR, May, 2017 [2]

3.0 Problem Statement

The timely delivery of HIV medicines is crucial for effectively combating the HIV epidemic and saving lives. Supply chain disruptions can result in significant costs and potentially fatal consequences, underscoring the need to identify key factors that impact the delivery of HIV drugs and predict potential delays. This study aims to leverage publicly available supply chain data to determine these critical factors and develop a predictive model for estimating the duration and likelihood of delivery delays. With this information, program managers can take proactive measures to mitigate the impact of supply chain disruptions, reduce costs, and ensure that patients receive life-saving HIV medicines in a timely manner.

3.1 Solution Statement

This study used a combined model which uses classification machine learning algorithms to predict whether a particular product is delayed or not and then use regression analysis to predict the length of the delay using the subset of the data which the classification predicted will be delayed. This will maximize the utility of the complete model since it follows the natural decision-making process – a supply chain program manager would normally care about the products that will be delivered late and within those, focus on the ones that will likely have the longest delays first, thus allowing them to prioritize supply chain/logistics management and solve the biggest problems first.

To select the best model, both the classification and regression versions of the following models will be explored evaluated against predetermined benchmarks of Random Forest model with default parameters in SciKit-Learn : i) ExtraTrees ii) XGBoost iii) Support-Vector Machines (SVM) and iv) Multilayer Perceptron (MLP). Random-Forests, ExtraTrees and XGBoost are proven high-performing ensemble algorithms which can do automatic feature extraction while SVMs perform very well with high-dimensional data and can detect non-linear relationships if the right kernel is used. Finally, MLPs are useful for high-dimensional time-series data. The above advantages of these algorithms are well-suited to the selected dataset which has several categorical columns that will increase dimensionality and potentially be non-linearly related to the target variable after data transformation. Finally, the data is well-suited for this overall approach since our target variables are well-defined on the data i.e. delay occurrences and duration can be determined by data on scheduled versus actual delivery dates, allowing clear quantification and measurement of the problem and solution. This study's results will be applicable to future instances of supply chain orders, and thus it is applicable to future occurrences of similar supply chain data observations and useful for planning purposes.

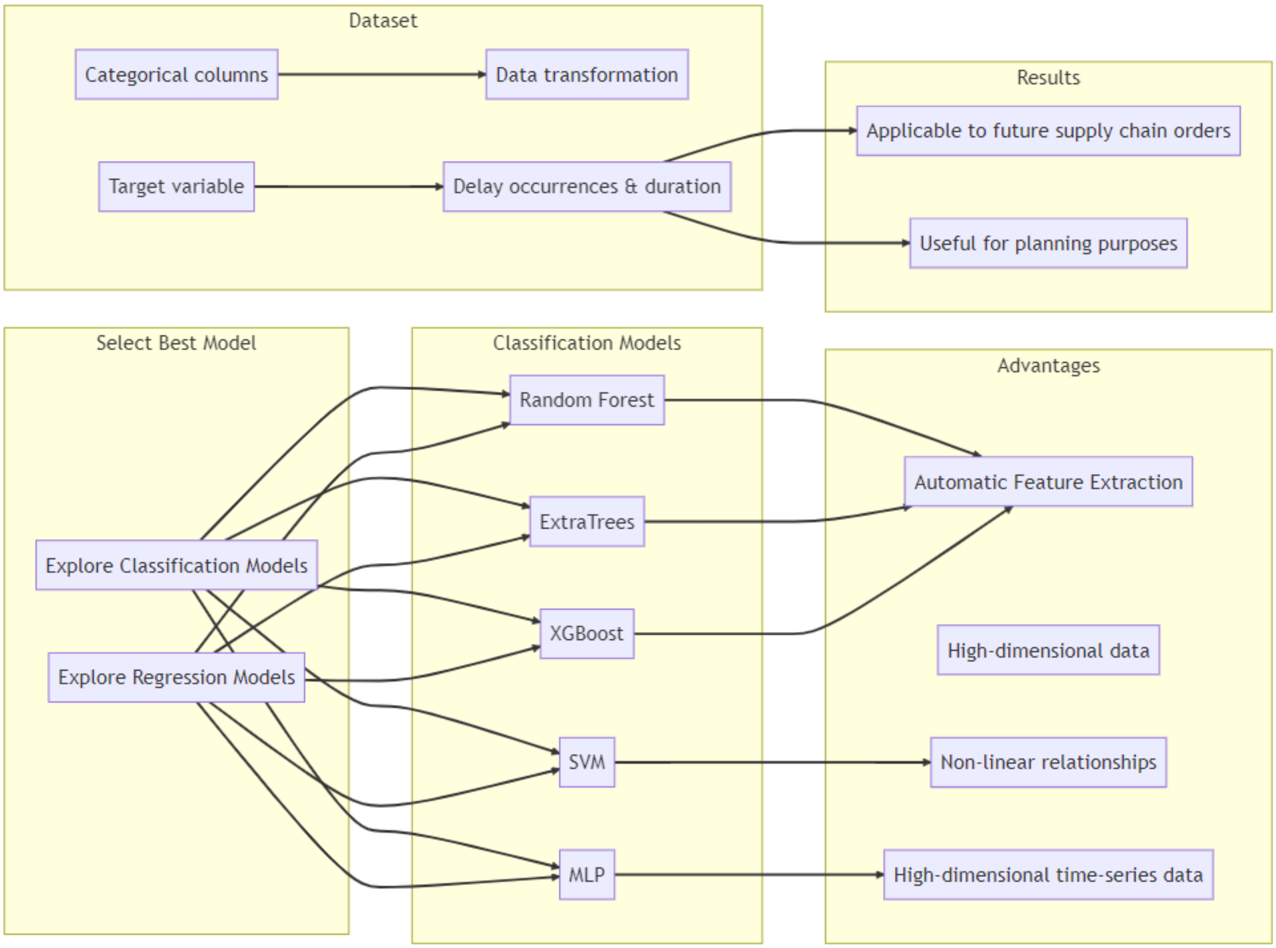


Fig 3. Block diagram

4.0 External Search

PEPFAR made significant progress in reducing the global burden of HIV/AIDS during this period. For example, the number of people receiving ART increased from 1.3 million in 2006 to over 10 million in 2015. PEPFAR also supported significant increases in HIV testing and counseling, and helped to reduce new HIV infections among key populations in several countries.

#	FieldName	FieldDescription	DataType
1	**ID**	Primary key indentifer of the line of data in our analytical tool	Number
2	**Project Code**	Project code	Text
3	**PQ #**	Price quote (PQ) number	Text
4	**PO #**	Order number: Purchase order (PO) for Direct Drop deliveries, or Sales Order (SO) for from Regional Delivery Center (RDC) deliveries	Text
5	**ASN/DN #**	Shipment number: Advanced Shipment Note (ASN) for Direct Drop deliveries, or Delivery Note (DN) from RDC	Text
6	**Country**	Destination country	Text
7	**Managed By**	SCMS managing office: either the Program Management Office (PMO) in the U.S. or the relevant SCMS field office	Text
8	**Fulfill Via**	Method through which the shipment was fulfilled: via Direct Drop from vendor or from stock available in the RDCs	Text
9	**Vendor INCO Term**	The vendor INCO term (also known as International Commercial Terms) for Direct Drop deliveries	Text
10	**Shipment Mode**	Method by which commodities are shipped	Text
11	**PQ First Sent to Client Date**	Date the PQ is first sent to the client	Date/Time
12	**PO Sent to Vendor Date**	Date the PO is first sent to the vendor	Date/Time
13	**Scheduled Delivery Date**	Current anticipated delivery date	Date/Time
14	**Delivered to Client Date**	Date of delivery to client	Date/Time
15	**Delivery Recorded Date**	Date on which delivery to client was recorded in SCMS information systems	Date/Time
16	**Product Group**	Product group for item, i.e. ARV, HRDT	Text
17	**Sub Classification**	Identifies relevant product sub classifications, such as whether ARVs are pediatric or adult, whether a malaria product is an artemisinin-based combination therapy (ACT), etc.	Text
18	**Vendor**	Vendor name	Text
19	**Item Description**	Product name and formulation from Partnership for Supply Chain Management (PFSCM) Item Master	Text
20	**Molecule/Test Type**	Active drug(s) or test kit type	Text
21	**Brand**	Generic or branded name for the item	Text
22	**Dosage**	Item dosage and unit	Text
23	**Dosage Form**	Dosage form for the item (tablet, oral solution, injection, etc.).	Text
24	**Unit of Measure (Per Pack)**	Pack quantity (pills or test kits) used to compute unit price	Number
25	**Line Item Quantity**	Total quantity (packs) of commodity per line item	Number
26	**Line Item Value**	Total value of commodity per line item	Currency (USD)
27	**Pack Price**	Cost per pack (i.e. month's supply of ARVs, pack of 60 test kits)	Currency (USD)
28	**Unit Price**	Cost per pill (for drugs) or per test (for test kits)	Currency (USD)
29	**Manufacturing Site**	Identifies manufacturing site for the line item for direct drop and from RDC deliveries	Text
30	**First Line Designation**	Designates if the line in question shows the aggregated freight costs and weight associated with all items on the ASN/DN	Binary
31	**Weight (Kilograms)**	Weight for all lines on an ASN/DN	Number
32	**Freight Cost (USD)**	Freight charges associated with all lines on the respective ASN/DN	Currency (USD)
33	**Line Item Insurance (USD)**	Line item cost of insurance, created by applying an annual flat rate (%) to commodity cost	Currency (USD)

Table 2. Available Input Features in the PEPFAR Supply Chain Data Set

4.1 Evaluation Metrics

The resulting combined models will be evaluated based on 4 metrics: Recall and F1-Score for classification, to balance the recall/precision trade-off, especially because the dataset is unbalanced with a ratio of 1:9 between the positive and negative class respectively. For the regression part of the model, the R-squared and Root Mean-Squared Deviation (RMSD) will be used to evaluate how well the regression model can predict the direction and length of delays in HIV medicine deliveries.

i) Recall: measures the success rate of correctly labeling the positive items i.e. what proportion of the positive labels did we successfully identify? This is important because it tells us what proportion of delays we can actually predict.

Recall = True Positives / (True Positives + False Negatives)

ii) F1-Score is an average (harmonic mean) of the recall and precision scores.

F1-Score = 2*(Recall * Precision) / (Recall + Precision)

where Precision = True Positives/ (True Positives + False Positive) and Recall = True Positives / (True Positives + False Negatives)

iii) R-squared is the "coefficient of determination" which measures the amount of variation in the data that is explained by the model, again as a percentage/fraction of total variation.

$$r = \frac{n\,(\sum xy) - (\sum x)\,(\sum y)}{\sqrt{[n^*\,(\sum x^2 - (\sum x)^2)]\,*\,[n^*\,(\sum y^2 - (\sum y)^2)]}}$$

Here, r represents R-squared, n is the number of observations and x and y are the feature and target variables respectively.

iv) RMSD measures the average size (absolute value) of the error that the model makes when predicting continuous target variables e.g. days late/delay in this case.

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2}{N}}$$

Here, "x-hat" are the predicted values and "xi" are true values of the target variable. "n" is the number of observations in the dataset.

5.0 Benchmarking

The solution model is a combination of two algorithms working together sequentially; thus, the benchmark model will also require a two-part benchmark. In order to make clear objective comparisons, the same model, Random Forest will be used as the benchmark for both classification and regression. The study will use the default versions of the Scikit-Learn implementation of these models. Results:

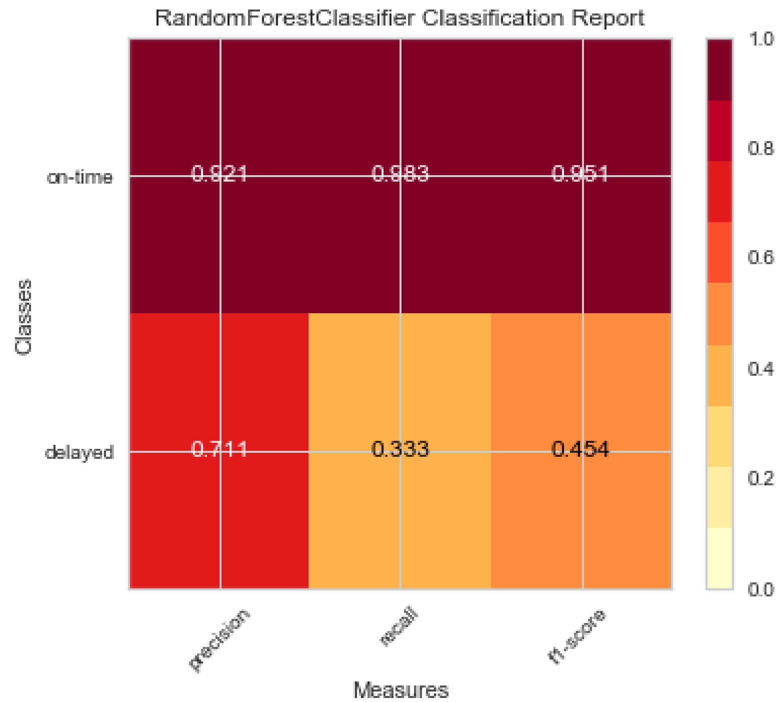


Figure 4. Classification benchmark results for default SciKit-Learn Random Forest algorithm.

- Classification results with benchmark Random Forest Classifier.

  o Recall: 0.33

  o F1-score: 0.45

  o Total: 134 instances of delayed delivery correctly identified

- Regression results with benchmark Random Forest Regressor:

  o R-squared: 0.85

  o RMSE: 13 days

6.0 Implementation of Model, algorithm and techniques

The key steps to implementing this model are detailed in the accompanying python code and JuPyter notebooks, below is a brief summary of the nuances in the approach:

1. Data cleaning – loading the data was straightforward given it was stored in standard excel format. However, it was important to understand the column names, types and distributions as well as supply chain terminology. A reference dictionary with lookup functionality for each column was developed for this purpose. Imputing missing values was then done using either mean or mode methods. For dates that were missing, an estimate based on related date columns was made e.g. purchase order dates were extrapolated from expected delivery dates while purchase quote dates were in turn derived from purchase order dates. The code for this required some proficiency with pandas date time and indexing functionality. To estimate missing weight and freight cost, this study took advantage of the relationships of these variables to other item-level features i.e. packet price = unit price * number of units; line item value = line item quantity*packet price; average weight of standard items should stay constant, and thus missing weight values can be imputed using this item weight * item quantity and freight cost is proportional to weight and dependent on whether items were single or bundled in a particular shipment.

2. Feature Engineering: Dates: year, month, day, weekday, quarter, week-of-year were extracted using pandas date-time functionality to capture time aspects as categorical variables. Numeric: counts, sums, proportions and measures of central tendency were calculated at country-year, factory-year, vendor-year, molecule-year, brand-year levels and then these were merged with the item-level data. Categorical: after cleaning up the weight and freight cost columns, there was additional information from which the following labels were extracted: weight captured separately, shipment configuration, freight cost included commodities, or freight invoiced separately. For shipment configuration, it was challenging to separate into single vs. bundled shipments but using string functions and regex, it was possible to isolate the id of each top-line item and then group all the items with which it was bundled. The remaining items constituted the single shipments. Time series variations were also captured at the vendor-date level using pandas group-by, rolling and cumsum functions to capture the short term and long term autocorrelation of delayed items and number of days of delay. The predicted variable itself, “delayed” was derived as the difference, in days, of the date delivered at client site and the scheduled delivery date. Extra care was taken to remove the predictor variable from the feature. External features on logistics and country fragility which turned out to be strongly predictive were also obtained from other sources. This data required cleaning, entity resolution (e.g. country names) and minimal missing value imputation before joining to the main item-level dataset.

3. Feature selection – Several methods, from univariate, bivariate and trends analysis to dimensionality reduction and feature importance were used to select the most predictive, least collinear, scaled and transformed features. Using several methods allowed a complementary and robust approach to the features selected. These methods are already outlined in detail in the EDA & Feature Selection section above and in the accompanying code.

4. Model benchmarking – see the Model Benchmarking section already described.

5. Model Selection – see Model Selection section already described, and Discussion of key

challenges section.

6. Final Model results – this is detailed in the Final Model Results section below along with key findings and challenges.

7.0 Model Selection

After preprocessing the data through a pipeline for logarithm, standard scaling, one-hot and label encoding as well as oversampling to balance the classes, the following models were compared:

i) Classification: LinearSVC, SVC, KNeighborsClassifier, LogisticRegressionCV, LogisticRegression, SGDClassifier, BaggingClassifier, ExtraTreesClassifier, RandomForestClassifier, MLPClassifier ,GaussianNB, LinearDiscriminantAnalysis

ii) Regression: LinearSVR, SVR, KNeighborsRegressor, LinearRegression , SGDRegressor, BaggingRegressor, ExtraTreesRegressor, RandomForestRegressor, MLPRegressor

Final Models selected: Extra Trees Classifier for classification and Extra-Trees Regressor for regression.

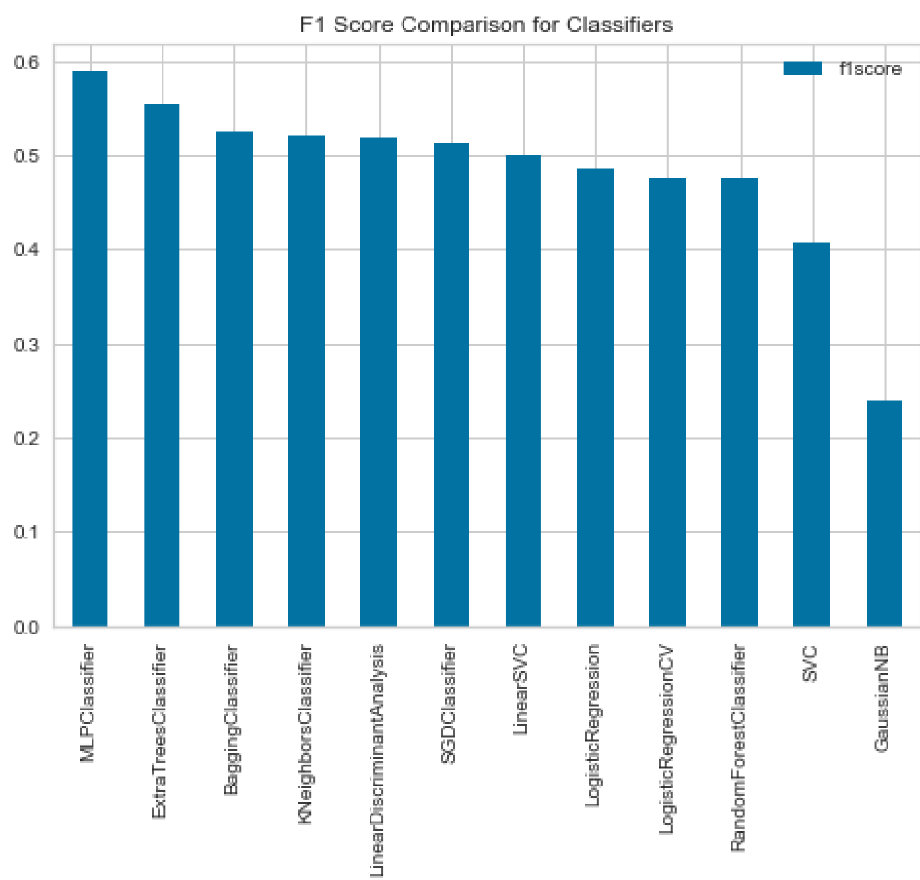


Figure 5. F1 Score comparisons for classification model selection

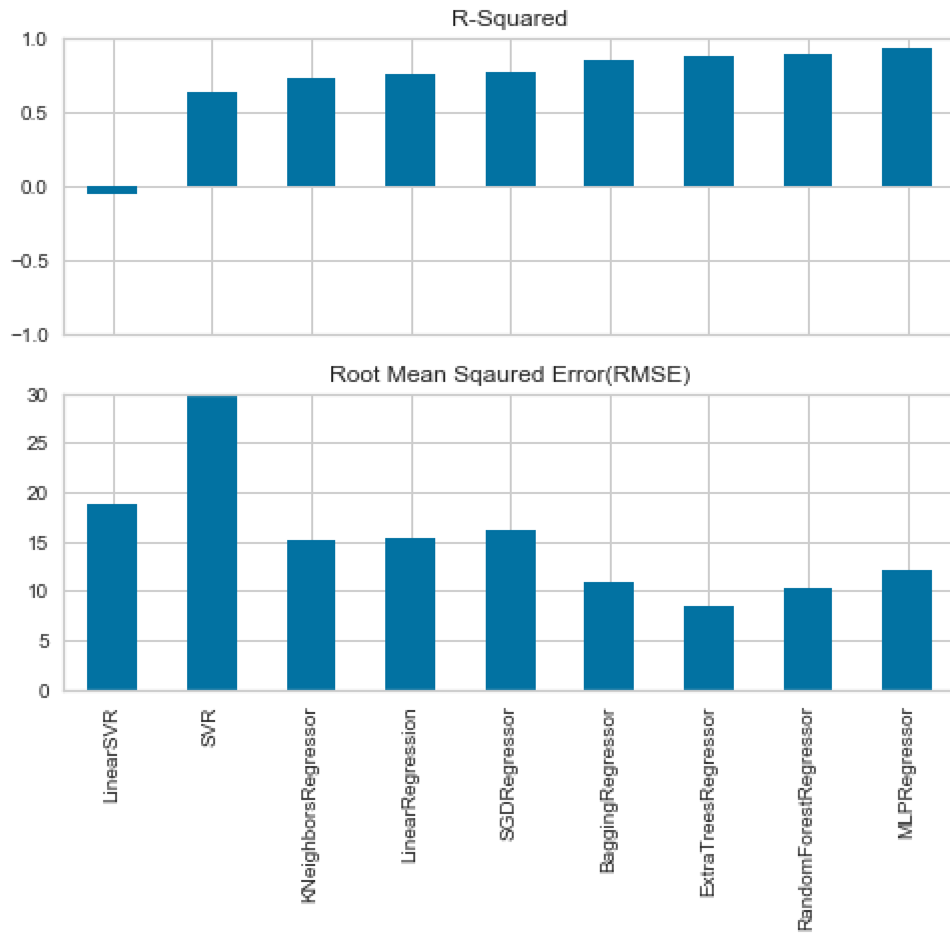
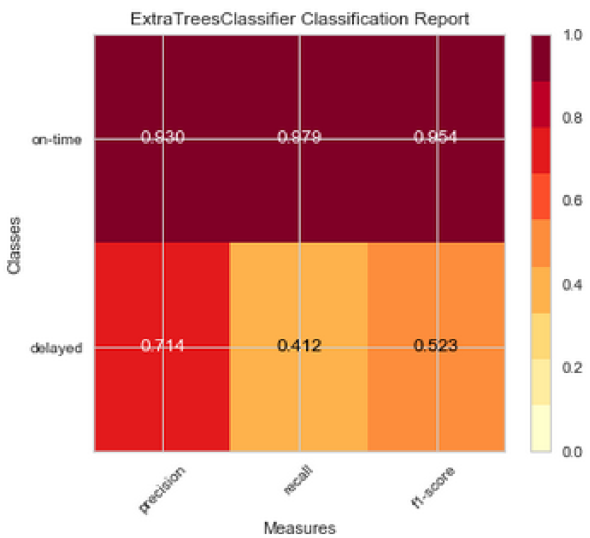


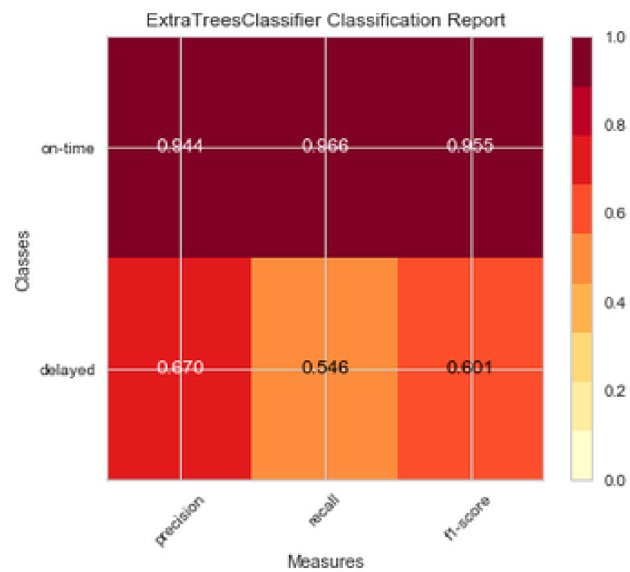
Figure 6. R-squared and RMSE comparison for model selection

Extra Trees Algorithm Specifics and Relevance

The Extremely Randomized Trees (Extra-Trees)7 is a supervised learning algorithm similar to the Random Forest algorithm in that they are both ensemble methods which use simple classification and regression decision trees as their building blocks. However, Extra Trees differs from Random Forests while Random Forests tries to find an optimal cut-point for each one of the k randomly selected features at each node (bootstrapping), Extra trees instead selects cut points at random and then averages the results. This results in "weak-learner" trees whose errors are uncorrelated, thus randomizing the cut-points and averaging results has a smoothing effect on the predictions of these trees. When combined with careful feature analysis to remove redundant features, this approach will simultaneously increase bias and reduce the variance, resulting in better model accuracy8. Additionally, it achieves higher computational efficiency by not trying to find optimal cut points.

The Extra Trees model is appropriate for this data and problem because of the following key features: i) it deals well with high dimensionality which is appropriate given the high cardinality of most of the categorical variables in this problem. ii) it deals well with multicollinearity and is in fact quite robust to this aspect iii) it can handle heterogeneity amongst the features as well as skewness within individual features iv) due to smoothing, it often leads to increased accuracy in the presence of a high number of continuously varying numerical features.

<table>
<tr><th></th><th>Extra Trees Initial Results</th><th>Extra Trees Final Results</th></tr>
<tr><td>CLASSIFICATION</td><td>


</td><td>


</td></tr>
<tr><td>REGRESSION</td><td>R-squared: 0.92
RMSD: 8.5 days
Total delays captured: 171</td><td>R-squared: 0.86
RMSD: 11.8 days
Total delays captured: 221</td></tr>
</table>

For classification, a clear improvement was observed in the Recall and F1-Score metric for the selected model. For regression, both the R-squared and RMSD fared a little worse in the final model than initial model due to the change in denominator, the final model was classifying and regressing on more items and as such the additional items may have slightly higher internal variation than the items picked up by the initial model. Note, however, that the final model still outperforms the benchmark Random Forest model.

Final hyper parameters chosen:

Classification ExtraTreesClassifier (n_estimators=900, max_features= 50, criterion= 'entropy',max_depth= 50, random_state=121)

Regression ExtraTreesRegressor(n_estimators=900 ,max_features= 50,max_depth= 50, random_state=121)

8.0 Applicable Regulations

The patents mentioned above might claim the technology used if the algorithms are not developed and optimized individually and for our requirements. Using a pre-existing model is off the table if it incurs a patent claim.

1. Must provide access to the 3rd party websites to audit and monitor the authenticity and behavior of the service.

2. Enabling open-source, academic and research communities to audit the Algorithms and research on the efficacy of the product.

3. Laws controlling data collection : Some websites might have a policy against collecting customer data in form of reviews and ratings.

4. Must be responsible with the scraped data : It is quintessential to protect the privacy and intention with which the data was extracted.

9.0 Business Opportunity

1. Many entrepreneurs complain that building forecasts with any degree of accuracy takes a lot of time--time that could be spent selling rather than planning.

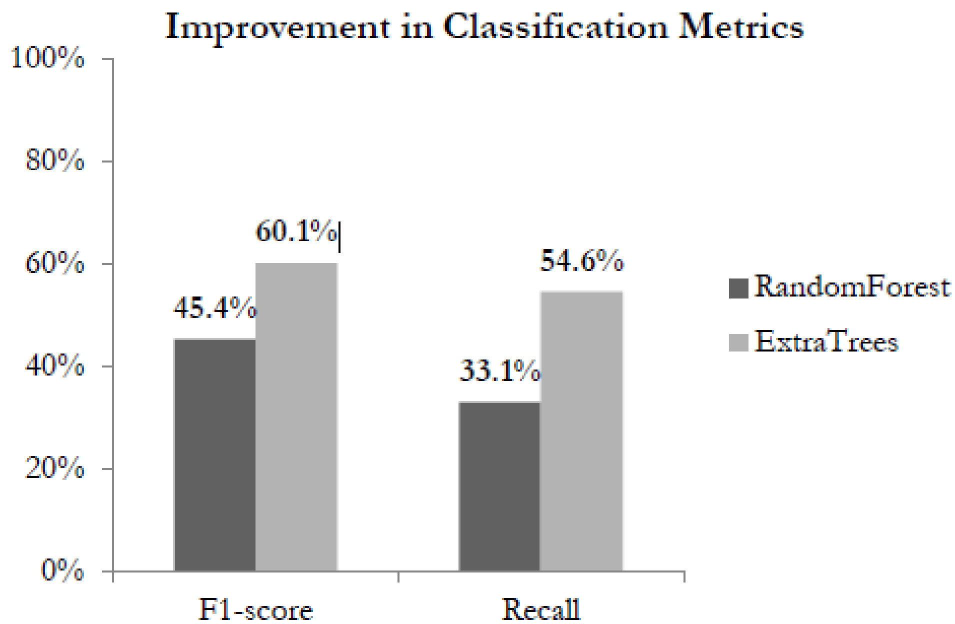
2. But few investors will put money in your business if you're unable to provide a set of thoughtful forecasts.

3. More importantly, proper financial forecasts will help you develop operational and staffing plans that will help make your business a success.

10.0 Final Model Results

Summary: Supplier-side factors such as origin country stability, vendor and brand volumes as well origin country logistics environment explain significant variation in the data. Together with customer/receiver-side factors, they explain about a third of the variation in the data. The rest of the variation is due to product level characteristics like volumes, value/price, and weight as well as how they evolve over time (time-series). Of note is the significant auto-correlation where vendors who have delayed items in the past as well as more recently are more likely to delay deliveries again in the future. These insights were used for feature selection for the final classification and regression models.

The Extra Trees Classifier and Extra Trees Regressor were selected as the best algorithms for the classification and regression tasks respectively. Both algorithms outperformed the benchmark Random Forest and several other algorithms. The Extra Trees Classifier improved the Recall by 65% (from 33.1% to 54.6%) and the F1-score by 32% (from 45.4% to 60.1%).



Metric	Random Forest	Extra Trees	Improvement
F1-score	45.4%	60.1%	32%
Recall	33.1%	54.6%	65%
R-Squared	85.8%	86.3%	0%
RMSE (days)	12.96	11.97	8%

11.0 Conclusion

A combined “classification-then-regression” machine learning model can avoid the public health and economic costs associated with delayed deliveries of HIV medicines. An ensemble classification algorithm, Extra Trees, is able to detect slightly more than 1 in 2 delayed item deliveries. This is a significant improvement from a null hypothesis model which would detect only 1 in 9 delayed items and a considerable improvement from a benchmarked Random Forest classification algorithm which catches 1 in 3 delayed items. Once delayed items are identified, an Extra Trees regression algorithm can predict the length

of delay to within 12 days (RMSE) with an R-Squared of 0.86, which is similar to the benchmarked Random Forest regression performance. So, while there was no significant improvement in the regression part, the combined classification-then-regression model for Extra Trees does significantly better than the benchmark.

12.0 References

Ekwunife, O. I., Okoye, I. C., Ugwoke, E. O., & Ndu, A. C. (2021). Predicting Global Supply Chain Outcomes for Essential HIV Medicines. International Journal of Business and Management, 16(10), 162-175.

Chakraborty, S., Kadirvelu, B., & Ravi, V. (2019). Developing a predictive model for supply chain performance of HIV medicines using machine learning algorithms. PloS one, 14(12), e0226015.1