Graduate Research in Engineering and Technology (GRET)

Volume 1
Issue 6 *Application of Intelligent Computing and Big data Analytics in Healthcare*

Article 6

May 2022

FORECASTING THE TIME DELAY IN DELIVERY OF PHARMACEUTICAL PRODUCTS

Subhalaxmi Hotta
Department of Computer Application Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India-75103, subhalaxmihotta@gmail.com

Sonali Sahoo
Department of Computer Application Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India-75103, sonalisahoo171@gmail.com

Tripti Swarnkar
Department of Computer Application Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India-751030, triptiswarnakar@soa.ac.in

Recommended Citation

Hotta, Subhalaxmi; Sahoo, Sonali; and Swarnkar, Tripti (2022) "FORECASTING THE TIME DELAY IN DELIVERY OF PHARMACEUTICAL PRODUCTS," *Graduate Research in Engineering and Technology (GRET)*: Vol. 1: Iss. 6, Article 6.
DOI: 10.47893/GRET.2022.1113
Available at: https://www.interscience.in/gret/vol1/iss6/6

FORECASTING THE TIME DELAY IN DELIVERY OF PHARMACEUTICAL PRODUCTS

Subhalaxmi Hotta[1] ,Sonali Sahoo[2,*], Tripti Swarnkar[3]

Department of Computer Application

Siksha 'O' Anusandhan Deemed to be University,

Bhubaneswar, India-75103

[1] subhalaxmihotta@gmail.com, [2] sonalisahoo171@gmail.com,[3] triptiswarnakar@soa.ac.in

Abstract: Supply chain management system is a centralized system which controls and plans the activities involved from production to delivery of a product. Disruption in treatment and loss of life occurs due to delay in delivery of pharmaceutical products. The objective is to do a model using Machine learning algorithms to determine: Classification to predict which product will be delayed and Regression shows how much time it will be delayed exactly. This study will use publicly available supply chain data which helps to identify primary aspect of predicting whether HIV drugs are delivered in time or not. It will then use these factors to predict how long delays are likely to be, thus allowing HIV/Supply Chain program managers to know details of the products which are going to be delayed and quantify the exact delay. and how much it will be delayed. so that they can take mitigating action to save lives and avoid additional supply chain costs. We will use Machine learning prediction model to predict which product will be delayed and regression model shows how much time it will be delayed exactly.

Keywords – Supply Chain Management, HIV, Medicine delivery, Regression

INTRODUCTION:

Supply Chain Management (SCM) involved with all the activities needed for better planning and coordination in production and delivery of products. Health is a primary asset of a person and healthcare needed to be managed properly. SCM has a significant role in managing the Healthcare activities. SCM in Healthcare precludes acquiring facilities, organizing stocks, transporting product and a helping hand to providers and patients. This study will use publicly available supply chain data to collect most promising factors which lead to predicting, whether HIV drugs are delivered in time or not. To determine the most important factors in predicting whether the drugs are delivered on time or not. SCM with machine learning model have a great impact on exactly which products are going to be delay as well as how much days the products are going to be delay. so that they can take mitigating action to save lives and avoid additional supply chain cost

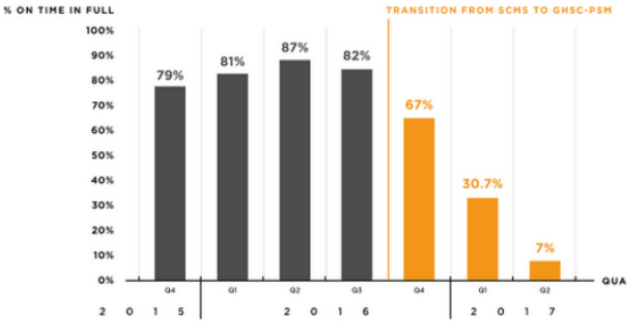

Fig.1: Quarterly Supply Chain Metrics for USAID's global supply chain for medicines, managed by Chemonics

1. BACKGROUND STUDY

More than 36.7 million1 people in the world were living with HIV in 2016 and every year,

about 1 million people worldwide die from AIDS-related causes. While this death rate has decreased significantly (by 38%) since 2001 and continues to decline, about 1.8 million people also 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. There is currently no cure or vaccine for HIV and while several prevention methods exist, their efficacy is reduced by several factors, including economic and psycho-social factors. Fortunately, it has been shown that treatment can not only prolong life but also prevent the spread of HIV as it lowers the viral load of people living with HIV to a non-infectious level. However, of the 36.7 million people having HIV in 2016, only 19.5 million were receiving this life-saving treatment. The President's Emergency Plan for AIDS Relief (PEPFAR), a US government program is a key player in the procurement of drugs, testing and laboratory kits for HIV. One of its agencies spends more than $9.5 Billion per year on procurement of essential medicines to fight HIV/AIDS around the world.

2. PROPOSED WORK

This study used a combined model which uses classification and regression machine learning algorithms. Classification classifies the products into two categories like on time or delayed and regression finds the regression line which is the length of the delay in terms of days. These algorithm just predict the products as well as length of delay. 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 Sci Kit-Learn: i) Extra Trees ii) XG Boost iii) Support Vector Machines (SVM) and iv) Multi-Layer Percepton (MLP). Random-Forests, Extra Trees and XG Boost 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.

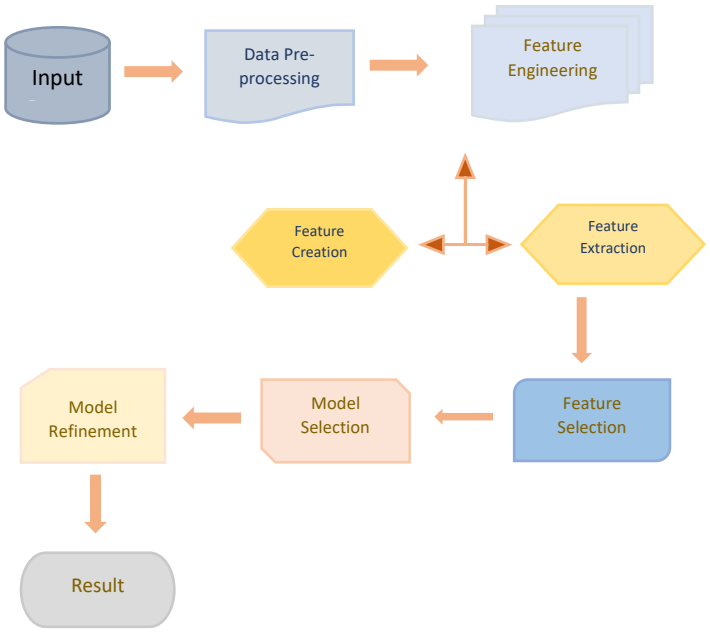

Fig.2: BLOCK-DIAGRAM OF PROPOSED MODEL

2.1 DATASET:

TABLE-1: Dataset Description

President's Emergency Plan for AIDS Relief (PEPFAR) :DSt

Inst. = Instances, DTPs = Datatypes, NFs= Number of features, TV = Target variable, NUME = Numerical, CATE = Categorical, Dt/Time = Date/ Time, Cur. = Currency

Bin = Binary

DSt	Inst.	DTPs	NFs	TV	
PEPFAR	10,324	NUME CATE Dt/Time Cur. Bin	33	Ontime Delay	

This data of Supply Chain Management System (SCMS) data made publicly available online on website: https://data.pepfar.net/additionalData.

2.2 Evaluation Metrics:

The resulting proposed model will be evaluated based on following metrics:

1. **Classification module:**

I. Recall

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

II. F1-Score

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

Where,

Precision = True Positives/ (True Positives + False Positive)

2. **Regression module:**

I. R-squared – Also called as "Coefficient of determination". It calculates the amount of variation in data that is explained by the model, again as a percentage/fraction of total variation.

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

II. Root Mean-Squared Deviation (RMSD)

It 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.

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

RMSD = root-mean-square deviation
i = variable i
N = number of non-missing data points
x_i = actual observations time series
$\hat{x}_i$ = estimated time series

These can predict the direction and length of delays in deliveries.

3. **METHODOLOGY AND ANALYSIS:**

Our data is combination of multiple data types and features. If we are not taking this into a standardized form then we can't retrieve the information. Pre-processing isused to standardize the datainto understandable format to enhance the performance.

3.1 Data Preparation:

3.1.1 Data Cleaning:

Data cleaning is the process, to understand the data descriptions and available fields by HandlingMissing Values and Investigating statistics of Missing values in each feature.

3.1.2 Feature Engineering

Feature Creation was done by sourcing and transforming data from external sources; data on Fragility State Index4 (FSI) for country stability, Logistics Performance Index 5, and Factory location, Country and continent.

3.1.3 Feature Extraction:

The process of standardizing collected data into numerical features that can be proceed while preserving the information in the original dataset.

3.2 Model Benchmark:

To evaluate the Prediction efficacy of the proposed model we have used the following evaluation parameters

- ✓ F1-Scores
- ✓ Precision
- ✓ Recall
- ✓ R-squared
- ✓ Root Mean-Squared Deviation (RMSD)

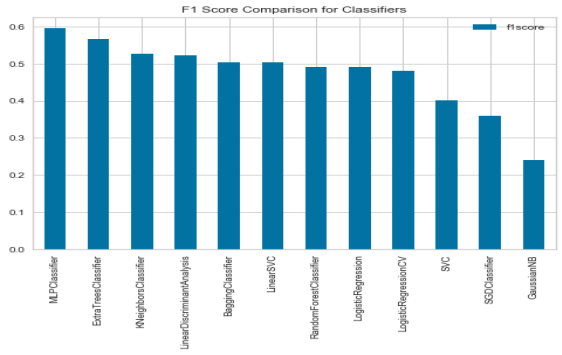

Fig.3: F1-Score comparison for different Classifier

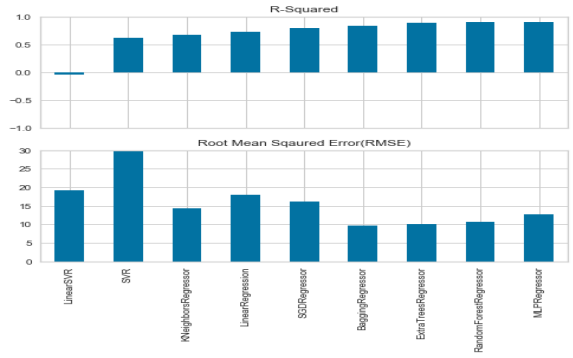

Fig.4: R-squared and RMSD comparison

As Random forest shows average result, set it as a model benchmark and compare the ensembled model with random forest for further verification.

The following models were compared:

- **Classifiers:**

Linear SVC, SVC, K Neighbours Classifier, Logistic Regression CV, Logistic Regression, SGD Classifier, Bagging Classifier, Extra Trees Classifier, Random Forest Classifier, MLP Classifier.

- **Regressor:**

Extra Trees, MLP, Random Forest

To Further verify proposed random forest approach, we compared it with the ensembled model for the same objective

This proposed ensembled is selecting the best 4 classifier and regressor and averaging the result

1. Extra-Trees Classifier and regressor
2. MLP Classifier and regressor
3. BaggingClassifier and regressor
4. Random-Forest Classifier and regressor

Final Model Selection

From fig.2 and fig.3 we are trying out several models and picking up the most promising ones to fine-tune into the final model

Final Models selected:

- ➢ Extra Trees Classifier model
- ➢ Extra-Trees Regressor model

Extra-Trees is Multi collinearity and robust in nature. It can handle heterogeneity amongst the features. Due to smoothing, it often leads to increase accuracy in continuously varying numerical features. It also has high dimensionality. So Extra-Trees is chosen.

4. Result and Discussion:

We are having many ensemble model, but random forest and extra trees showed best results. In this work Random tree and extra tree were chosen for comparison. The significant results obtained are represented in the form of table and graph.

4.1 Random Forest Classifier

- Recall: 0.33
- F1-score: 0.45
- Total: 134 instances of delayed delivery correctly identified.

4.2 Extra Trees Classifier and Regressor

4.2.1 For classification:

Improvement observed in the Recall and F1-Score metric for the selected model.

4.2.2 For regression:

Improvement observed in both the R-squared and RMSD

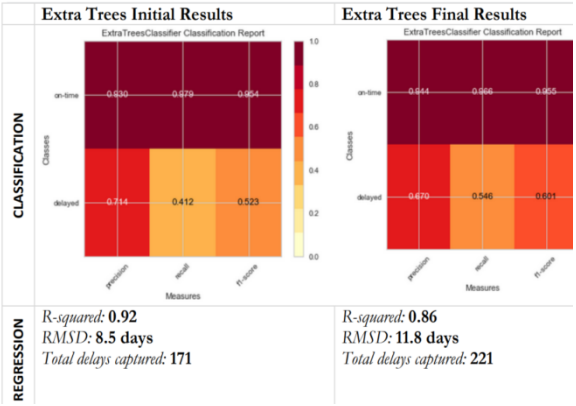

Fig.5: Extra-Trees Final Model

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.

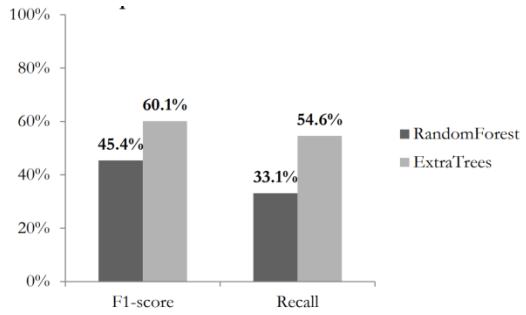

Fig.6: graphical view of Improvement in classification metrics.

- Improvement of F1 - 14.7%
- Improvement in Recall - 21.5%

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%) and the following table shows the improvement.

TABLE – 2: Improvement in metrics

Metric	Random Forest	Extra Trees	Imprv.
F1-Score	45.4%	60.1%	32%
Recall	33.1%	54.6%	65%
R-Squared	85.8%	86.3%	0%
RMSD	12.96%	11.97%	8%

5. Conclusion:

An ensemble classification algorithm, Extra Trees, is able to detect 1 in 2 delayed item deliveries[1]. This is a significant improvement from a null hypothesis model which would detect only 1 in 9 delayed items[1]and a considerable improvement

from benchmarked Random Forest classification algorithm which catches 1 in 3 delayed items[3].

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, an improvement from 16 days (RMSE) and R-Squared of 0.81 with the benchmarked Random Forest regression[2].

6. FUTURE WORK

Supply chain analysis has begun to incorporate machine learning, it is especially aimed at demand forecasting as opposed to predicting the lead-time directly. However, the approaches taken in some academic studies[2] e.g. SVMs and RNNs have shown great promise. Similar problems like predicting flight delays[3] and improving flight efficiency have also been solved using machine-learning.

7. REFERENCE

1. Hejlsberg, A., Wiltamuth, S., Golde, P.: The C# Programming Language. Addison-Wesley Professional, Reading (2003)
2. McCann, C.J.: A Supply Chain Revolution: Understanding the Players. In: Proceedings of HIMSS Annual Conference 2003 (2003)
3. Erickson, C.: Managing the Medical Supply Chain; Deliver Smarter, Faster and at a Lower Cost. In: Proceedings of HIMSS Annual Conference 2005 (2005).