International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Enhancing Efficiency in Healthcare Supply Chains: Leveraging Machine Learning for Optimized Operations

Soumit Roy[1], Mainak Mitra[2]

[1]Data Analytics Presales Lead, TCS
[2]TPM, Google Inc

Abstract

In this complex strings of healthcare supply chain, optimal efficiency is the most supreme to ensure cost and time effectiveness of medical resources. This paper scrabbles through Machine Learning (ML) classifiers such as, Na¨ıve Bayes, K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM) and Linear Regression to intensify the performance of healthcare supply chain management. This study focuses on five classes: 'Inspection Results', 'Defect Rate', 'Transportation Modes', 'Routes' and 'Cost'. According to my findings, Random Forest classifier indicated 87% accuracy in the 'Inspection Results' and 'Transportation Modes' classifications whereas KNN classifier signified an impressive accuracy of 86% in the 'Routes' classification. These findings underline the need for ML approaches within different classes. The diversified performances of classifiers within classes highlight the importance of selecting the most suitable algorithm based on supply chain aspect. This research not only shows the effectiveness of ML classifiers in healthcare supply chain optimization but also highlights the possibility of automation for different surface of supply chain management. It provides as a fundamental step for optimizing healthcare supply chain.

Keywords: Machine Learning, Classification, Supply Chain, Healthcare

I. INTRODUCTION

Artificial intelligence (AI) associates creating computer systems that can think and make decisions like a human by focusing on computer programs. [1] Predicting from data, it helps to improve production and supply chain management. Overall, AI is modifying industries for better efficiency through machine learning (ML).

With a focus on five key classes – 'Inspection Results', 'Defect Rates', 'Transportation Modes', 'Routes' and 'Costs' – I am aiming to address the cost and time effectiveness of medical resources according to the classification and regression classes. I am adopting ML classifiers including Naive Bayes, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machines (SVM), and Linear Regres- sion which forms the backbone of my investigation.The primary goal is to increase the performance of healthcare supply chain management, a critical aspect in the delivery of quality healthcare services. This study addresses the broader prospect of automation in healthcare supply chain management. The effectiveness of ML classifiers not only in boosting accuracy but also in expediting decision-making processes hints at the feasibility of automating across various supply chain management surface. [2] The COVID-19 pandemic has globally impacted healthcare and society. Instead of controlling the virus, many countries failed to provide healthcare supplies. To ensure vaccine distribution, reduce risk, ensure safety, and provide traceability, it addressed variations and storage facility in forecasting for COVID-19 vaccines. By correlating between drug supply management and predicting parameters in disease, this study addresses a connection between predictive parameters in machine learning and drug supply management.

Studies underlines the need for application of advanced analytic to tackle challenges for ensuring medicine supply during crises. My findings reveal variations in the performance of ML classifiers across different classes. The separation in classifier performances highlight the healthcare supply chain challenges and the importance of selecting the most appropriate algorithm for each specific parts of the supply chain. [3] In today's world, global events have shown how important it is to have strong and adaptable supply chains. Researches reviewed academic works for exploring how AI and ML frameworks increase supply chain resilience [4]. This research sets the stage for a complete investigation into the domain of healthcare supply chain optimization using ML classifiers. By shedding light on the unique challenges within each class, I aim to contribute valuable intuitions that inform the development of data-driven strategies for healthcare supply chain management.

II. RELATED WORKS

In [5], the researchers have examined the usage of artificial intelligence (AI) in Supply Chain Management (SCM) and identified top 5 key search areas like supply chain network design, selection of supplier, inventory and budget planning and green uses of SCM. They offered a framework to guide people to understand AI trends in SCM to improve the process with AI's authorization.

During COVID-19 pandemic, researchers [6] addressed the prediction of delivery times for medical supplies using AI and ML (Machine Learning). By analyzing data, AI models improved its prediction and reached accuracy up to 93.5% with their introduced method. Here, authors offered a reliable way to acknowledge the arrival of medical supplies.

In recent years [7] data analytic and Artificial Intelligence (AI) on healthcare highlighted healthcare professionals struggle to understand Al and it's basic. Authors proposed a framework of two combined group, key research areas and questions for future studies. They offered to improve collaboration among researchers, analysts and healthcare professionals in healthcare sector.

[8]Combination of blockchain and artificial intelligence is getting very popular because of its security, efficiency, productiveness in business force. In this paper, authors re- viewed the running state using AI in supply chains to see the activities of the technologies and future work exploration.

Fellow researchers [9] aimed to use AI and ML to predict the longevity to process and deliver medical supply from e-pharmacy back in COVID- 19 pandemic. By combining multiple ML algorithms they improved their accuracy 25% more within 3 days. The aim of this paper is to establish that AI and ML can make medical supply chain more reliable.

Song [10] addressed to improve the healthcare supply chains using technologies like deep learning, artificial intelligence and blockchain. Researchers introduced a model to find the best connections and make healthcare supply chains more efficient.

Many healthcare organizations struggle to manage medical supplies and patient care efficiently because of the lacking of coordination. [11] In this study it shows how AI is used in e-Healthcare supply chains and the privacy issues. It's about using AI to make healthcare supply chains work smoothly while ensuring security and privacy concerns.

In this studies [12] authors addressed the challenge of managing Pharmaceutical Supply chains (PSC) and created a math model for designing networks to minimize costs and delivery delays. For improving forecasts and reduce shortages, they suggested Linear Quadratic Regression of Machine Learning and this approach leads more efficient and cost effective PSC. [13]Authors aimed on a global crisis call vaccine supply chain for distribution by using math model to balance cost as well. This model also considers factors like unpredictable demand and vaccine shelf-life. They showed that one method outperforms the other and can help to create an effective vaccine supply chain.
[14]Technology is improving healthcare and customer experiences which combine online and offline channels for real time information. This study explored the role of wearable in hospital supply chains and how ethical concerns affect their adoption by proposing a framework that combined traditional statistics and machine learning.

III. METHODOLOGY

The objective of this research is to optimize the Healthcare Supply Chain (HCS) using Machine Learning (ML). To achieve this goal, the research process began with data collection and preprocessing, which is briefly discussed in Section III-A
As HCS comprises several areas, I have selected five highly relevant areas, including 'Inspection Results,' 'Defect Rates,' 'Transportation Modes', 'Routes', and 'Costs'. Among these, defect rates and costs pose regression problems, while the rest are classification tasks. The selection of ML models is based on the specific requirements, as outlined in Section III-B.
I have divided the dataset into two categories, and the entire process is illustrated in Figure 1. During the data preprocessing step, the dataset was imbalanced. Addition- ally, categorical columns required conversion into numerical format using one-hot encoding techniques, which is detailed in Section III-C.

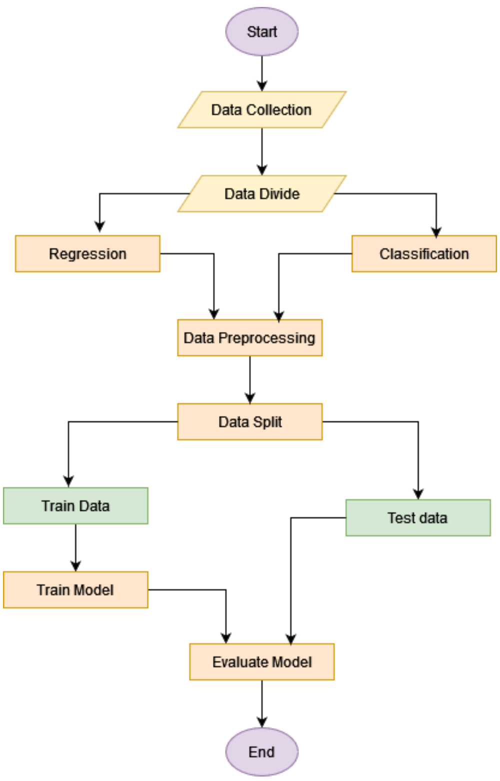


Fig. 1: High level overview of the process.

A. *Data Collection and Preprocessing*

The dataset [15] consists of 100 data objects with 22 columns. Out of these 22 columns, 5 are classes and 17 are features. The 'Inspection Results' column contains three labels: 41% pending results, 36% pass, and 23% fail. The distribution of this column is visualized in Figure 2. Sim- ilarly, the 'Transportation Modes' column contains 17% sea mode, 29% road, 28% rail, and 26% air. In addition, the classification column 'Routes' contains three dummy routes: Route-A, Route-B, and Route-C, with the following distribution: 43%, 37%, and 20%.

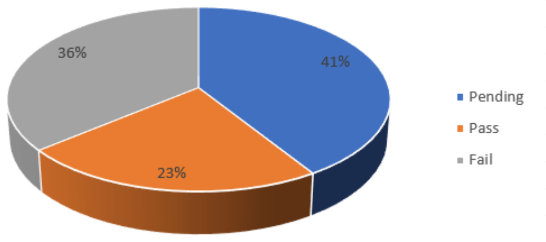
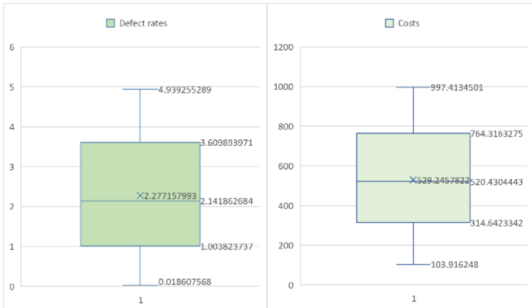


Fig. 2: Inspection Results Label Distribution.

The dataset includes two additional regression columns, 'Defect Rates' and 'Costs,' which are visualized in Figure 3a and Figure 3b, respectively.



(a) Defect Rate Distribution. (b) Costs Distribution.

Fig. 3: Defect Rate and Costs Distributions.

B. *Algorithms*

As the dataset [15] comprises both regression and classification columns, the following algorithms have been chosen to address each class.

1) **Linear Regression:** Linear regression models the relationship between variables by fitting a linear equation to observed data. It is selected for regression tasks due to its simplicity and interpretability.
2) **Naive Bayes Classifier:** The Naive Bayes classifier is a probabilistic algorithm based on Bayes' theorem. It
3) **K-nearest Neighbors (KNN):** K-nearest neighbors assigns a new data point's class or value based on its nearest neighbors. It is adaptable for both classification and regression tasks, suitable for various types of data
4) **Random Forest:** Random Forest is an ensemble method combining decision trees. It is a robust choice for classification and regression, effective in handling complex relationships.

5) **Support Vector Machine (SVM):** Support Vector Ma- chines are versatile algorithms used for classification and regression tasks. They find optimal hyperplanes for data separation and are known

for their strong predictive capabilities.

C. *Categorical Data Transformation and Encoding*

One-hot encoding is a technique that converts categorical data into numerical vectors. For each categorical feature, a new dimension is created in the feature vector, with the value 1 representing the presence of the current category and 0 representing the absence of the current category. This ensures that only one dimension of the feature vector is active for a given state. The dataset [15] consists of 6 categorical columns that need to be transformed into numerical vectors using the one-hot encoding technique.

IV. RESULTS AND DISCUSSION

D. *Experimental Setup*

The experiment was conducted within a Jupiter Notebook environment, utilizing a machine equipped with an Intel Core i5 5600G CPU. This CPU is notable for its integrated Graphics Processing Unit (GPU), which facilitated the execution of machine learning tasks. Additionally, the machine featured 8 GB of RAM. The experiment was implemented using Python as the programming language, complemented by four libraries, namely, Numpy, Pandas, and SKlearn.

E. *Evaluation Matrices*

Evaluation matrices are essential for measuring the performance of the model, offering distinct perspectives on the results. In this study, four classification metrics were employed:

- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$
- $F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$

For regression tasks, three metrics were utilized:

- $MeanSquareError(MSE) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$
- $MeanAbsoluteError(MAE) = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$
- $R - squared(R) = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$

Here, TP is True Positive, TN is True Negative, FP is False Positive, FN is False Negative, Yi is the observed value, ˆ Yi is the predicted value, and ¯ Y is the mean of observed values.

D. Experimental Results

I conducted an extensive evaluation of various machine learning classifiers and models, exploring their performance on classification & regression tasks. My experimentation involved four classifiers, including NaiveBayes, KNN (K- Nearest Neighbors), RandomForest, and SVM (Support Vector Machine). Additionally, I considered three distinct classification classes: ’Inspection Results’, ’Transportation Modes’, and ’Routes’. The selected classifiers and models were configured and trained to assess their accuracy in classifying the different inspection results, transportation modes, and routes. Table I provides a detailed analysis of the classification accuracy achieved by each model for the three classification classes. The results showcase the performance metrics for NaiveBayes, KNN, RandomForest, and SVM. Figure 4 presents a visual comparison of the classification accuracy across the different models. Notably, RandomForest demonstrates superior accuracy, particularly in the ’Inspec- tion Results’ & ’Transportation Modes’ classification with

Models	**Inspection Results**	**Transportation Modes**	**Routes**
NaiveBayes	0.77	0.67	0.69
KNN	0.85	0.8	0.86
RandomForest	0.87	0.86	0.35
SVM	0.7	0.65	0.85

TABLE I: Classification accuracy of classification classes in different algorithms

Models	**Defect Rates**			**Costs**		
	MSE	**MAE**	**R^2**	**MSE**	**MAE**	**R^2**
Linear Regression	5.06	1.82	-3.6	117269.33	281.3	-0.8
KNN Regression	1.56	1.04	-0.42	95576.52	272.97	-0.47
RF Regression	2.13	1.11	-0.94	114238.03	299.96	-0.75
SVM Regression	1.75	1.09	-0.59	85675.38	258.15	-0.31

TABLE II: MSE, MAE and R^2 value of regression class in different algorithms

an accuracy of 87% . On the other hand, KNN performed well in ’Route’ with an accuracy of 86%, while SVM and Naive Bayes did not perform well for any class expect SVM in the ’Routes’ classification with an accuracy of 85%.

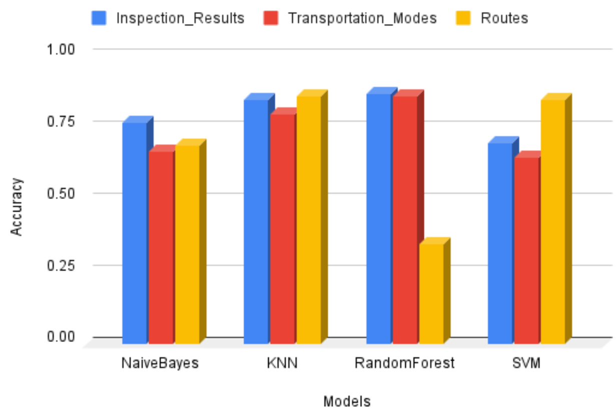


Fig. 4: Accuracy of different classifiers

Among the regression models, the Linear Regression and KNN Regression models performed relatively well for predicting ’Defect Rates’ as table II shows the results. The Linear Regression model achieved an MSE of 5.06 and an MAE of 1.82, while KNN Regression demonstrated an MSE of 1.56 and an MAE of 1.04. In contrast, the Random Forest Regression and SVM Regression models showed poorer performance with higher MSE and lower R^2 values. For predicting ’Costs,’ the RF Regression model excelled with the lowest MSE of 114238.03 and an MAE of 299.96. The SVM Regression model also performed reasonably well. However, the Linear Regression and KNN Regression models struggled, as indicated by their higher MSE and lower R^2 values. Careful consideration of these outcomes is crucial for selecting models that balance predictive accuracy and generalization.

V. CONCLUSION AND FUTURE WORK

The use of machine learning to optimize the healthcare supply chain has yielded promising results. The Random Forest classifier, in particular, demonstrated significant improvement in classification tasks, achieving an accuracy of 87% after hyperparameter tuning and Principal Component Analysis (PCA). This highlights the critical role of data quality in the effectiveness of machine learning models, emphasizing the importance of rigorous data preprocessing and feature selection techniques. Future research will focus on collecting real-world data from healthcare clinics to enhance the model’s robustness and validate its performance in practical healthcare settings. Additionally, ongoing exploration of advanced preprocessing techniques and feature engineering methodologies will be essential for refining model performance and ensuring its adaptability to evolving healthcare scenarios. The findings from this research lay the foundation for continued investigations, addressing the dynamic challenges within the healthcare supply chain. Incorporating real-world data and further refining machine learning techniques will pave the way for more accurate and adaptable predictive models, contributing to the ongo- ing improvement of healthcare logistics and supply chain management.

REFERENCES

1. I. First, H. Pallathadka, M. Mustafa, D. Sanchez, G. Sajja, S. Gour,M. Naved, and M. Jawarneh, “Impact of machine learning on man- agement, healthcare and agriculture,” *Materials Today: Proceedings*, 07 2021.
2. T. I. Meghla, M. Rahman, A. A. Biswas, J. T. Hossain, and T. Khatun, “Supply chain management with demand forecasting of covid-19 vaccine using blockchain and machine learning,” 07 2021, pp. 01– 07.
3. G. Emmanuel, J. Wieczorek Michałand Siłka, D. Połap, M. Woźniak, and R. Damaševičius, “Real-time neural network based predictor for cov19 virus spread,” PLOS ONE, vol. 15, no. 12, 2020, doi: 10.1371/journal.pone.0243189
4. *and*
5. Y. Han, W. K. Chong and D. Li, "A systematic literature review of the capabilities and performance metrics of supply chain resilience", International Journal of Production Research, vol. 58, no. 15, pp. 4541-4566, 2020.
6. Dwivedi, A., Wickramasinghe, N., & Bali, R. K. (2017). "Designing a machine learning-based predictive decision support system for healthcare supply chain management." Industrial Management & Data Systems.
7. Paul SK, Sarker R, Essam D (2017) A quantitative model for disruption mitigation in a supply chain.

Eur. J. Oper. Res. 257(3):881–895. .
8. Ong JBS, Wang Z, Goh RSM, Yin XF, Xin X, Fu X (2015) Understanding natural disasters as risks in supply chain management through web data analysis. Internat. J. Comput. Comm. Engrg. 4(2):126–133
9. Abbas, K., Afaq, M., Khan, T. A., & Song, W.-C. (2020). A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. Electronics (switzerland), 9(5), 852.
10. Singh H, Majumdar A, Malviya N (2020) E-pharmacy impacts on society and pharma sector in economical pandemic situation: a review. J Drug Deliv Ther 10(3-s):335–340.
11. Martin Beaulieu, Omar Bentahar,Digitalization of the healthcare supply chain: A roadmap to generate benefits and effectively support healthcare delivery,Technological Forecasting and Social Change,Volume 167,2021,120717,ISSN 0040-1625.
12. Dimitrios Bechtsis, Naoum Tsolakis, Dimitrios Vlachos, Eleftherios Iakovou,Sustainable supply chain management in the digitalisation era: The impact of Automated Guided Vehicles,Journal of Cleaner Production,Volume 142, Part 4,2017
13. F. Goodarzian, A.A. Taleizadeh, P. Ghasemi, A. Abraham, An integrated sustainable medical supply chain network during COVID-19, Eng. Appl. Artif. Intell.100 (2021) 104188, http://dx.doi.org/10.1016/J.ENGAPPAI.2021.104188.
14. L.M. West, L. Diack, M. Cordina, D. Stewart, Applying the delphi technique to define ‘medication wastage’, Eur. J. Hosp. Pharm. 22 (5) (2015) 274–279,http://dx.doi.org/10.1136/ejhpharm-2014-000593
15. S. Priyan, R. Uthayakumar, Optimal inventory management strategies for pharmaceutical company and hospital supply chain in a fuzzy–stochastic environment, Oper. Res. Heal. Care 3 (4) (2014) 177–190, http://dx.doi.org/10.
16. 1016/J.ORHC.2014.08.001.W.H. Organization, World Health Statistics 2010, World Health Organization,Geneva PP - Geneva, 2010, [Online]. Available: https://apps.who.int/iris/handle/10665/44292..