Procurement Timeframe Analysis

[1]Sumedh Kharolkar, [2]Siddharth Koli, [3]Koushal Bhat, [4]Sourav Sharma, [5]Dr. Shyam Deshmukh

Department of Information Technology
Pune Institute of Computer Technology
Pune, Maharashtra

***Abstract*- The essential part for the procurement timeframe analysis to be successful is procuring the right material in acceptable timing. Procurement in DRDO may be internal or external. DRDO has established and well-documented procurement procedures. For any procurement case sanction, there are specific steps involved with recommendations, concurrences, and approval at various levels. Based on this analysis, we would try to find or locate various bottlenecks using algorithms like Apriori that happen during order, which hinder the efficiency and create delays in the entire process. We would also work on creating models that would help us appropriately predict the expected time for the approval process, based on the historical data fed. If we can predict it properly, it would be helpful for the project managers to manage their timeline and resources appropriately.**

Keywords: **Procurement, Bottleneck, Aproiri, Predictive analysis,**

I. INTRODUCTION

Defence Research and Development Organisation (DRDO) is the premier agency under the Department of Defence Research and Development in the Ministry of Defence of the Government of India, charged with the military's research and development, headquartered in Delhi, India. For any project to be successful, the essential part is procuring the right material in acceptable timing. Procurement in DRDO may be internal or external. For external procurement, the organization generally involves outsourcing the requirement to outside vendors.Outside vendors submit proposals in the form of tenders for the requirement. These proposals are subject to negotiations in order to get the best agreement possible. DRDO has established and well-documented procurement procedures. For any procurement case sanction, there are specific steps involved with recommendations, concurrences, and approval at various levels. For procurement and inventory management, DRDO follows a general pipeline. This pipeline typically includes the following steps in the following order.

i. Demand Order	iv. Supply Order
ii.Request for Proposal	v. Receiving the order
iii.Negotiation	vi. Payment

This general pipeline may vary depending on various parameters such as cost, nature of the item, and so on.For example, if an item costs between Rs 5 Crore and Rs 50 Crore, this pipeline would also include several approvals from department Directors.Any procurement process can be broadly divided into 2 parts:
i.Order
ii.Delivery

In procurement timeframe analysis ,we would be focusing on the order part of this process where we would analyze the whole process using the data from various DRDO labs regarding their procurement history.

II. MOTIVATION

This analysis will help the organization's project managers to detect which stage takes the most time for completion and try to eliminate or, at the very least, minimize these delays. The prediction model will allow the project managers to estimate the approximate time for procurement based on object attributes.

III. OBJECTIVES

- To apply an analytical algorithm which will help us to detect bottlenecks in procurement timeline.
- Based on historical data (Traces), apply a machine learning model which can handle large scale data.
- It would be able to predict the procurement time based on input parameters.

IV. SCOPE

The scope of procurement timeframe analysis is to identify bottlenecks and pinpoint them for further investigation, which will allow them to derive solutions to these bottlenecks. The prediction model will be accompanied by a web interface for easy use by project managers allowing them to better plan their project cycles.

V. Literature Review

The authors of paper [1] compared the accuracy performance of six ML algorithms - Decision Tree, Support Vector Machine, Naive Bayes, K-Nearest Neighbor, Logistic Regression, and Random Forest using 1282 real student's course grade dataset. However, one of the main limitations is that the predictions are being made based on the past year's academic records, which undermines the possibility of severe change of grades in the current year.

In paper [2], the authors proposed a complex system with many internal and external variables is modeled using system dynamics model to predict the performance of the procurement process. However, even though it can capture a wide range of changing values in its variables, a system dynamics model can only run one version of a situation at a time.

In paper [3], the supply chain has been represented as a directed graph that describes the relationship between its attributes. Using this representation, bottlenecks can be identified. However, the risk structure for hazard events is not taken into account, and the overall approach is highly time complex.

In paper [4], the authors describe detailed steps for the identification of bottlenecks in a production system. This includes the identification of flow sequence, investigation of unit processes, and investigation of process cycle times. However, it is a manual process, and a flaw in identification in one stage can greatly disrupt the supply chain, leading to loss.

In paper [5], the management of supply and demand, including distribution, is discussed. Raw ingredients are delivered by suppliers to producers, who then transform them into finished goods. This process is mirrored in how manufacturers deliver their goods to customers. The importance of distribution to trade and business is highlighted, and any barriers to it may impair the functioning of the supply chain, which includes producers as one of its constituents. The interviews conducted suggest that issues with loading and unloading at ports may have negative consequences, but a more realistic scenario should be considered.

In paper [6], the Build-to-Order (BTO) supply chain is discussed as an effective pattern for meeting end-user requirements. However, it is challenging to implement current product control procedures and techniques, such as Just-In-Time (JIT), in a BTO supply chain.

This paper introduces a new theoretical method for identifying bottlenecks in the BTO supply chain. The strategy is combined with input/output control and the theory of constraints. This method is used to define bottlenecks and explore how they affect production control.

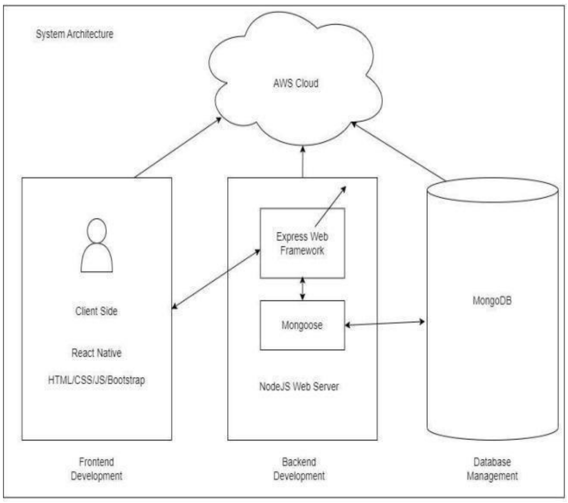
Next, an analysis and development of a production control model based on the bottlenecks takes place. However, a more detailed algorithm which considers different parameters to optimize the whole supply chain should be developed, e.g., a negotiation mechanism with parameters like discount, price, etc.

In paper [7], authors examine how the issue of an unbalanced class distribution occurs when there are significantly fewer samples in one class of a dataset than there are in the other classes. This is a widespread issue in many real-world datasets, including those used for fraud detection, anomaly identification, and medical diagnosis, when the proportion of positive examples is substantially smaller than that of negative cases. In order to balance the class distribution, the SMOTE algorithm, which is suggested in the paper, generates synthetic examples for the minority class. By interpolating between minority class samples, the method generates synthetic cases.

In particular, SMOTE develops new examples by randomly choosing one or more of the minority class examples and one or more of the majority class examples.

The authors of study [8] contend that optimizing hyperparameters is essential for getting high performance in machine learning models. They contrast GridSearchCV and RandomizedSearchCV, two well-known methods for hyperparameter optimization, in terms of effectiveness and efficiency.

A brute-force method called GridSearchCV does an exhaustive search over a given set of hyperparameters. Contrarily, RandomizedSearchCV selects hyperparameters at random from a given distribution.The authors contend that GridSearchCV becomes unworkable when searching over a large number of alternatives, and that RandomizedSearchCV is better suited for high-dimensional hyperparameter spaces.



The authors of study [9] present the experimental setup and results of their evaluation. They use 10 multi-label datasets and train four different types of classifiers: binary relevance, classifier chains, label powerset, and random k-label sets.

The results show that there is no single evaluation metric that is universally best for all datasets and classifiers. Instead, the optimal evaluation metric depends on the specific characteristics of the dataset and the classifier. The authors find that macro F1 score and micro F1 score are generally the most informative and reliable evaluation metrics for multi-label classification tasks.
The study insights into the performance of different evaluation metrics.

VI. DATASET DESCRIPTION

Our Procurement cycle dataset consists of data from over 20+ DRDO labs across the country.It includes following 48 columns which are:

SUPPLY ORDER NO	APPROVAL AMOUNT
SUPPLY ORDER DATE	SANCTION DATE
MODE OF TENDER SUPPLY ORDER NO	CFA CODE CFA DESCRIPTION
SUPPLY ORDER DATE	CFA
MODE OF TENDER	CFA APPROVAL DATE
MODE OF TENDER CODE	CFA APPROVAL REFERENCE
DEMAND NO	CONCURRENCE LEVEL CODE
SANCTION DATE	CONCURRED BY
UNITCODE	CONCURRENCE DATE
UNITCODE DESCRIPTION	CONCURRENCE BY
RECEIPT DETAILS	CONCURRENCE DATE 1
DEMAND DATE	CONCURRENCE AMOUNT
DIVISION NO	PROCUREMENT MODE CODE
DIVISION NAME	PROCUREMENT MODE DESC
METHOD OF PURCHASE	PROCUREMENT MODE DESC 1
TOTAL COST	SANCTION DATE
RFP NO	AGAINST SANCTION NO
FINANCIAL POWER CODE	SUB MAJOR HEAD
FINANCIAL POWER SERIAL NO	MINOR HEAD
GEM CONTRACT NO	HEAD CODE
UO NO	IS PAC
GEM CONTRACT DATE	DEMAND NO 1
BUDGET HEAD DESCRIPTION	ESTIMATED COST

Out of these 48 attributes we selected 21 attributes. Below is a brief description of few of them:

MODE OF TENDER CODE : It is used to categorize on the basis of different modes of tender.

FINANCIAL POWER CODE : It is the Financial code associated with a procurement.

CFA CODE : It is the code for competent financial authority.

CONCURRENCE BY : It is the code for the authority responsible for concurrence.
BUDGET HEAD CODE : It is used to categorize on the basis of the budget.
TOTAL COST : Gives the total cost of the order

FINANCIAL POWER SERIAL NUMBER : It is used to categorize the order based on the serial number of financial power

UO NO : It is used to categorize based on the Urban Operations number.

GEM CONTRACT NO : It is used to categorize the tender based on the gem contract number.

ESTIMATED COST : Provides an estimated cost required for the procurement.

DIVISION NO : It is used to categorize the order based on the division number.

APPROVAL AMOUNT : Provides the approval amount required for the tender to be passed.

SUPPLY ORDER DATE YEAR : It is used to categorize based on the year of the supply order date.

SUPPLY ORDER DATE DAY : It is used to categorize based on the day of the supply order date.

Here we created a function for EST COST RANGE and TOTAL DAYS RANGE which converts the columns EST COST and TOTAL DAYS into a range value. This method is better for procurement timeline prediction since it compresses the values into a fixed range.

VII. PROPOSED SYSTEM

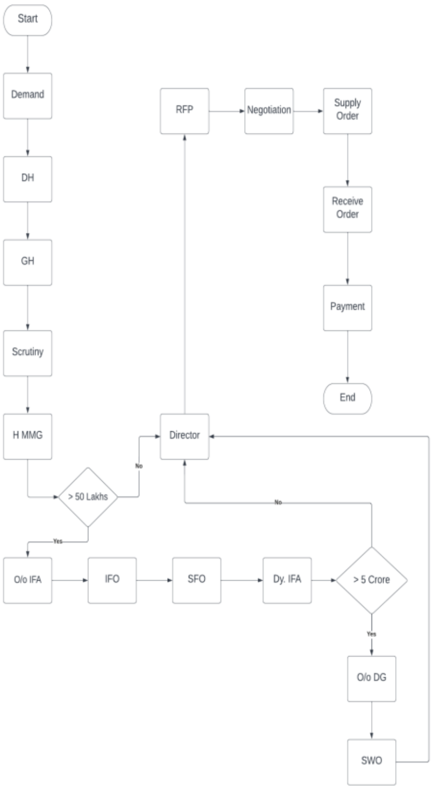


Fig.2. Flow Diagram of DRDO Procurement Cycle

The general procurement cycle defines the steps that are involved in the procurement of a proposed demand until the final delivery of the demand raised. The procurement cycle for the transactions below 50 lakhs of amount follow the below steps :

1. The demand is proposed.
2. The approval of Divisional Head and General Head is taken.
3. The proposal is scrutinized and further further approval of Higher management and Director is taken.
4. The last steps involved are negotiation, receiving and payment for the demand.

The procurement cycle for the transaction above 50 lakhs and below 5 Cr of amount follow the below steps :

1. The demand is proposed.
2. The approval of Divisional Head and General Head is taken.
3. The proposal is scrutinized and further further approval of Higher management is taken.
4. The proposal also needs to be approved from the offices of IFO, SFO ,Dp. IFA and Director.
5. The last steps involved are negotiation, receiving and payment for the demand.

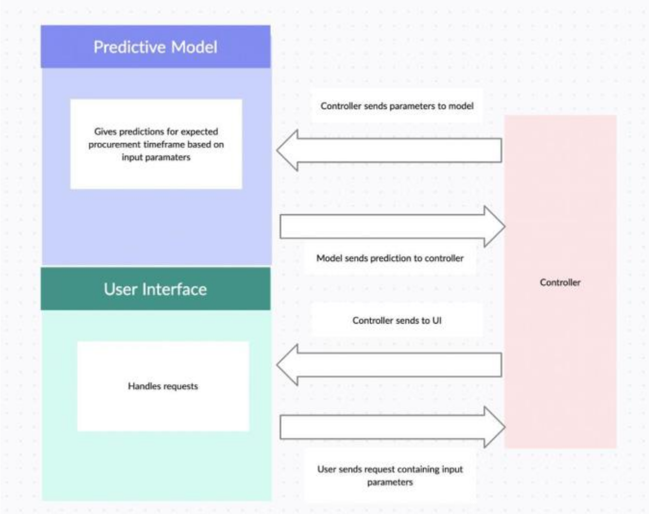


Fig. 1. Software Architecture for Procurement timeframe analysis

A user interface, controller, and predictive model make up the software architecture. The part of the system with which the user will engage in order to use it is the user interface. It includes every parameter that the user must supply as input to the model and will display the output, which is the model's estimate of the period. It transmits the user's request to the controller.The binding part of the system is the controller.

The user interface for this system requires input to be provided for six different columns of data. These columns include "EST_COST," which represents the estimated cost of the project, "MODE OF TENDER" which describes how the tender process will be conducted, and "METHOD OF PURCHASE," which indicates the specific method by which the goods or services will be acquired. Additionally, the "FINANCIAL POWER CODE" is used to identify the level of financial authority needed to approve the purchase, while the "CFA CODE" is used to track funding sources for the project. The "CONCURRENCE BY" column indicates the individuals or groups that must provide their approval before the purchase can be made, and the "BUDGET HEAD CODE" is used to identify the specific budget allocation for the purchase. The "IS PAC" column denotes whether or not a pre-audit clearance is required before the purchase can proceed. By providing accurate and complete data for each of these columns, users can ensure that their procurement processes are compliant and efficient.

Once the user has entered the required data into the twenty one columns of the user interface and clicks the submit button, the system will process this information and generate a prediction for the "TOTAL DAYS RANGE" column. This prediction is based on a variety of factors such as the estimated cost of the project, the method of purchase, and the level of approval required, among others. The system uses algorithms and statistical models to analyze the input data and calculate the estimated timeline for the procurement process. This predicted timeline can be used to help the user plan and manage their project timeline more effectively, ensuring that they have sufficient time to complete all necessary steps in the procurement process.

The process of generating a prediction for the "TOTAL DAYS RANGE" column involves several components working together. The user interface captures user input data, which is then passed on to the controller. The controller acts as an intermediary between the user interface and the model, and it sends the input parameters to the model for processing. The model then uses its algorithms and statistical models to analyze the input data and generate a prediction for the "TOTAL DAYS RANGE" column. Once the prediction is generated, the model sends it back to the controller, which in turn sends it to the user interface. The user interface then displays the predicted total days range to the user. This process of passing data between the controller, model, and user interface allows for a streamlined and efficient prediction process, ensuring that the user receives accurate and timely predictions for their procurement project timeline.

VIII. IMPLEMENTATION

Bottleneck Identification using Apriori Algorithm:

To identify the bottleneck in our procurement process, we used the Apriori algorithm. We started by collecting transactional data on the procurement process of a large organization. The dataset contained information on various attributes of the procurement process such as the estimated cost, mode of tender, method of purchase, financial power code, CFA code, concurrence by, budget head code, and the time taken to procure each product.We then applied the Apriori algorithm to the dataset to identify frequent patterns or sets of items. The algorithm works by generating frequent itemsets, i.e., sets of items that appear together in a minimum number of transactions. It then generates association rules that represent the relationship between the items in these frequent itemsets.

We used the support and confidence metrics to evaluate the generated association rules. Support measures the frequency of occurrence of the itemsets in the transactions, while confidence measures the likelihood of an itemset to appear in the transaction, given that another itemset is present. Based on the generated association rules, we identified the most frequent sets of items that were associated with longer procurement times. These sets of items represented the bottleneck in our procurement process. By identifying these bottleneck items, we were able to streamline our procurement process and reduce the procurement time for these items.

In conclusion, the Apriori algorithm proved to be an effective technique in identifying the bottleneck in our procurement process. By applying this algorithm, we were able to streamline our procurement process and reduce the time taken to procure items.

Data Preparation:

We started by collecting data on the procurement process of a large organization. The dataset contained information on various attributes of the procurement process such as the estimated cost, mode of tender, method of purchase, financial power code, CFA code, concurrence by, budget head code, and whether the procurement required PAC (Prior Approval Committee) clearance or not. We also included the total number of days it took to procure each product as the target variable.To prepare the data for analysis, we performed several preprocessing steps. We first identified and removed highly correlated features using a correlation matrix and removing columns with a correlation coefficient of greater than or equal to 0.9. We then used the backward elimination technique with a significance level of 0.05 to select the most relevant features for our model.

Model Selection and Training:

After preparing the data, we selected the XGBoost algorithm to predict the time it would take to procure a particular product based on its attributes. We used the Python library xgboost to build our model, and experimented with different hyperparameters such as the number of estimators and the maximum depth of each tree. We trained the model on 70% of the data and used the remaining 30% as a holdout set for evaluation.To address the issue of class imbalance in the dataset, we used the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples of the minority class. This helped us to balance the dataset and improve the performance of the model.

Model Evaluation:

We evaluated the performance of our model using several metrics such as accuracy, precision, recall, and F1 score. We also used a confusion matrix to visualize the distribution of the actual and predicted classes. Our final model achieved an accuracy of 0.86 and an F1 score of 0.86 on the holdout set, indicating that it was able to predict the time it would take to procure a particular product based on its attributes with reasonable accuracy.

Our approach shows promising results in predicting the time it would take to procure a particular product based on its attributes. However, there are several limitations to our study that should be considered. First, our dataset was limited to the procurement process of a single organization, and may not generalize well to other contexts. Second, our model was trained on a relatively small dataset, which may affect its performance on larger and more diverse datasets. Finally, our approach did not take into account external factors such as market conditions and supplier performance, which may have a significant impact on the procurement timeframe. Despite these limitations, our study provides a foundation for further research in this area.

To implement our machine learning model, we created a simple frontend using Angular, a popular web framework for building single-page applications. The frontend communicates with our backend, which is built using Node.js, a lightweight and efficient JavaScript runtime. The backend runs a Python script that generates results based on our trained .sav model files. The frontend design is intuitive and user-friendly, allowing users to easily input their data and receive results. The Angular components were designed to work seamlessly with our backend APIs, allowing for efficient data transfer and processing. Overall, the combination of Angular, Node.js, and Python provided a robust and efficient platform for implementing and deploying our machine learning model.

IX. RESULTS

The performance of our model was evaluated using accuracy and F1 score. The accuracy of the model was found to be 0.86, indicating that it correctly classified 86% of the data points. The F1 score, which takes into account both precision and recall, was found to be 0.86 . This suggests that the model has a good balance between identifying true positives and avoiding false positives, as well as identifying true negatives and avoiding false negatives. Overall, these results demonstrate the effectiveness of our model in accurately predicting the outcomes.

X. CONCLUSION AND FUTURE SCOPE

The web app would be developed which would help the organization in predicting the time frame for the particular procurement process which would be based on majorly contributing parameters. In the procurement cycle it becomes crucial to understand the steps that involve higher complexities which may lead to delay in the procurement cycle.

The system can be extended further to support many different platforms and cycles.The model can be further improved and fine-tuned to improve accuracy and thus enhance the reliability of the system. With the possibility of acquiring even more data from DRDO labs, the model can be improved to account for subtle nuances in the cycle and improve prediction.

REFERENCES:

1. Anitha K, Bhoomika C, J Andrea Kagoo, Kruthika K, Aruna, Dept of CSE, M S Engineering College, Bangalore, Karnataka (2022) "Student Grade Prediction Using Multiclass Model", August 2022| IJIRT | Volume 9 Issue 3 | ISSN: 2349-6002, https://www.ijirt.org/master/publishedpaper/IJIRT156211_PAPER.pdf
2. M. H. Abolbashari(), A. Zakeri, and E. Chang, E. (2018), "Using System Dynamics for Predicting an Organization's Procurement Performance", Springer Nature Switzerland AG 2018 https://doi.org/10.1007/978-3-030-00828-4_29
3. Kamil J. Mizgier a, Matthias P. Jüttner b & Stephan M. Wagner a (2018), "Bottleneck identification in supply chain networks", International Journal of Production Research, Vol. 51, No. 5, 1 March 2013, 1477–1490. https://doi.org/10.1080/00207543.2012.695878
4. Y.R.Perdana, R. Dwitasari, Maria M., Marsono Y. and A.S.Kuswati, "The Bottleneck of Distribution in Supply Chain's Manufacturers", 2021 IEEE 8th (ICIEA), https://ieeexplore.ieee.org/document/9436726
5. Wiselaw Urban, Patrycja Rogowska, "Case Study of bottleneck identification to the theory of constraints", MAPE 2018, volume 1, issue 1, pp. 399-405, https://www.researchgate.net/publication/328337894_The_Case_Study_of_Bottlenecks_Identification_for_Practical_Implementation_to_the_Theory_of_Constraints
6. William J. Christensen, Richard Germain Ph.D Laura Birou (2005) " Build-to-Order and Just-in-Time as Predictors of Applied Supply Chain Knowledge and Market Performance" .July 2005 Journal of Operations Management 23(5):470-481 , 17(3):37-54 https://dx.doi.org/10.1016/j.jom.2004.10.007
7. "N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer",SMOTE: Synthetic Minority Over-sampling Technique,https://arxiv.org/abs/1106.1813
8. " James Bergstra, Yoshua Bengio ",Comparing Random Search and Grid Search https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf
9. Grigorios Tsoumakas and Ioannis Vlahavas ", Comparing Evaluation Measures Multi-Label Learning,https://link.springer.com/article/10.1007/s10994-007-0054-7
10. DRDO Procurement Manual (2020) issued vide No. DRDO/DFMM/PM/0000120/M/801/D(R&D) dated 23-10-2020
11. Delegation of Financial Powers issued vide No. DRDO/DFMM/PL/83226/M/01/1976/D(R&D) dated 18-12-2019