INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

Machine learning based supplier defect prediction with real time risk alerts for uplift operations

1st.D.Jeyablessy
Department of Advanced Computing Science
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
Blessydavisha03@gmail.com

2nd DR.R.Thailambal
Department of Advanced Computing Science
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
Thaila.scs@vistas.ac.in

ABSTRACT-Efficient supply management is regarded as an essential aspect in ensuring business success, particularly in uplift operations where the timely nature of supply and quality of products are regarded as critical factors in determining business success. The research focuses on presenting a proposal on how machine learning can be utilized in predicting defects in supply and sending timely signals. The timely nature of supply is critical in ensuring timely decisions are made in supply management. Data collected from suppliers on the quantity of ordered supplies, delay in supply, and past defects in supply were utilized in building a random forest classifier in predicting defects in supply. The suggested framework is critical in ensuring timely efficiency in uplift operations, thus ensuring any potential losses are averted by defective supply. The research has shown how machine learning can be utilized in optimizing uplift operations. Moreover, the incorporation of predictive analytics ensures timely decisions are made, thus enhancing supply chain transparency. The suggested solution is deemed to be suitable in ensuring organizations remain consistent in quality and efficiency.

KEYWORDSMachinelearning,RandomRorest,Supplier defect,Uplift operations, Risk managemnt,Supply predictive analytics.

1. INTRODUCTION

Supplier performance is an essential factor in ensuring that the business activities of an organization are running efficiently. Uplift operations are the supply of raw materials as well as the quality of products supplied by various suppliers. In case the quality of raw materials supplied by various suppliers is low, there are high chances for an organization to be affected in terms of productivity and performance regarding the supply of raw materials.

In the past, various organizations have been using the monitoring method for the evaluation of the performance of various suppliers in various fields. The conventional method of monitoring for the evaluation of the performance of various suppliers involves the inspection of the quality of raw materials as well as the performance review of various suppliers. Although the monitoring method for the evaluation of the performance of various suppliers is essential, there are high chances for an organization to be affected in terms of performance in an attempt to identify the problems after they have occurred.

Recent developments in data analytics and machine learning also bring new opportunities for improving supplier monitoring and predicting risks more effectively.Sometimes, data from suppliers can be analyzed and used to predict patterns associated with defective supplier deliveries using machine learning models.This helps in identifying potential risks more effectively, which can then be addressed to avoid operational issues in the future.

This study aims to develop a machine learning model for predicting supplier risks associated with defective supplier deliveries in real-time.A Random Forest model is developed for classifying supplier risks based on data from suppliers such as order quantity, supplier delay, and previous rates of defects from suppliers.The system can automatically alert the operational team if a supplier is found to be beyond a certain level of risk as determined by the machine learning model developed in this study.

2. LITERATURE SURVEY

Supplier performance evaluation is a vital aspect of supply chain management. This is because the quality of the products is directly affected. In the past, the supplier performance evaluation was carried out through the manual inspection of the supplier's performance. However, the traditional method of supplier performance evaluation was found to be inadequate.

The literature on supplier performance evaluation using machine learning algorithms was focused on the prediction of supplier-related risks using the historical data. Decision Tree and Random Forest algorithms were found to be efficient in supplier performance evaluation. In addition, the literature on predictive analytics was focused on the exploration of the potential of the data-driven models in improving the decision-making process through the early detection of supplier-related risks.

However, the recent research was mostly based on offline prediction and is not applicable for real-time purposes.
Although there have been recent developments in the field of supplier performance evaluation, there still exists a big gap between the integration of prediction models and real-time alert systems. In order to bridge this gap, the proposed system was designed and implemented by integrating machine learning-based defect prediction and real-time risk alert systems.

3.PROPOSED METHODALOGY

In this particular piece of work, a machine learning-based approach has been used for the prediction of defects in the suppliers during the uplift operations. The basic idea is to get a clear understanding of what exactly supplier defects are by taking a look at the historical data of the suppliers, which will eventually provide a general idea of the supplier that may end up providing defects in the material supplied.

A. Data Collection

The first step is to collect information related to suppliers. Generally, this information is obtained from the company's records, including the type of material purchased from a supplier, any delays by the supplier, and the amount of defective material obtained. These factors, as a whole, provide a wide view of the quality of a supplier.

B. Data Preprocessing

Data is Preprocessed for improvement. This process involves the removal of duplicate data, handling missing data, and organizing the raw data. Feature scaling is done by utilizing the Min-Max normalization method.

FORMULA :

$$Xnorm = \frac{Xmax \quad -Xmin}{X-Xmin}$$

C. Feature Selection

However, not all data is necessarily important. We are interested in significant data. We are particularly interested In delivery delay,defect percentage, and order quantity in this problem. The defect percentage is defined by:

Defect Rate=Defective Item\total item

D. Model Building (Random Forest)

The algorithm for making predictions is the Random Forest algorithm. This algorithm works by creating a number of decision trees and combining their outcomes. It does not rely on a single decision tree; rather, it relies on the outcome of a number of decision trees.

E. Training and Testing

The dataset is divided into two parts: one for training and the other for testing. The model learns from the training data and is then tested using new data to see how well it performs. This helps in checking whether the model is actually useful or not.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

E.Model Evaluation

After the training, the model can be evaluated based on some basic parameters like accuracy, precision, recall, and F1 score. These parameters can give an idea of how well the model is predicting the supplier's defects.

G. Prediction

Once the model has been built, it can be used for predicting the chances of a supplier delivering defective products.

H. Visualization

Finally, the results are presented in a simple way by using graphs and charts, which makes it easier to understand which suppliers are a risk and how the defects are progressing over a period of time.

4 .ARCHITECTURE DIAGRAM

Fig. 1 Architecture of Machine Learning-Based Supplier Defect Prediction System

The architecture diagram shows a system that predicting defects by a Machine Learning algorithm starts from the collection of the data to the final prediction. Once the data is processed, a random forest algorithm is used for analyzing the data to detect defective suppliers. The results are then visualized for decision-making purposes..

5.INPUT

The system begins with a pre-structured set of supplier performance records. Each record contains information about how a supplier runs their business and how they perform in terms of quality during a transaction.

The main fields of information contained within these records are the Supplier ID, Order Quantity, any Delivery Delays, Lead Time, Defect Rate, Past Defects, Supplier Rating, Reliability Score, etc. The data is formatted into a CSV file, which is then preprocessed before it is sent into the machine learning model .

Supplier_ID	Order_Quantity	Delivery_Delay (days)	Lead Time (days)	Defect_Rate (%)	Past_Defects	Supplier_Rating	Reliability_Score	Risk_Level
S001	500	2	7	2.5	10	4.5	0.90	Low
S002	750	5	10	6.8	25	3.8	0.70	Medium
S003	300	8	12	10.2	40	3.2	0.55	High
S004	600	1	6	1.5	5	4.8	0.95	Low
S005	900	6	11	7.5	30	3.5	0.65	Medium
S006	450	9	13	12.0	50	3.0	0.50	High
S007	700	3	8	3.0	15	4.2	0.85	Low
S008	850	7	12	8.9	35	3.6	0.60	Medium
S009	400	10	14	13.5	60	2.8	0.45	High
S010	650	2	7	2.0	8	4.6	0.92	Low

6.PSEUDO CODE

START

Load dataset

Remove duplicate records

Handle missing values

Select features: delay, defect_rate, quantity

Set target: defect

Normalize data

Split data into training and testing sets

Train Random Forest model using training data

Predict defects using testing data

Calculate accuracy

Calculate precision

Calculate recall

Display results and graphs

END

7.OUTPUT

```
Top 10 Risky Suppliers:
     Supplier_ID  Risk_Score
301            5    0.992085
334           15    0.990155
34             4    0.988275
59            10    0.984000
331           15    0.975307
49            10    0.960490
302            8    0.966659
240            4    0.958424
287            1    0.957783
37             9    0.957888

ALERT: High Risk Suppliers Found
     Supplier_ID  Risk_Score
4              8    0.904799
6             19    0.870450
8             11    0.911526
11             3    0.788251
12             2    0.745131
..           ...         ...
484            5    0.763618
492           10    0.824550
493            7    0.843240
497           11    0.804722
499           18    0.761672
```

```
ALERT: High Risk Suppliers Found
     Supplier_ID  Risk_Score
4              8    0.904799
6             19    0.870458
8             11    0.911526
11             3    0.788251
12             2    0.745131
..           ...         ...
484            5    0.763618
492           10    0.824550
493            7    0.843240
497           11    0.804722
499           18    0.761672

[150 rows x 2 columns]
Email alert sent!

New Supplier Prediction: High Risk
Risk Score: 0.63
New supplier is safe
```

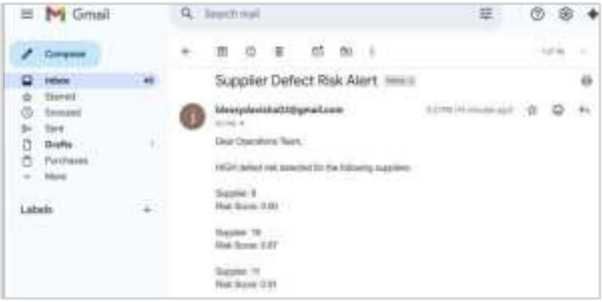
8.RESULT AND DISCUSSION

However, it was also determined that the accuracy level of the model, which was based on the data provided for the suppliers, order quantity, delivery delays, and the prior defect rate, was around 0.46, based on an 80/20 split for training and testing the data. This is a fair level of accuracy for a model that simply needs to make a prediction with respect to a supplier that could cause a defect.

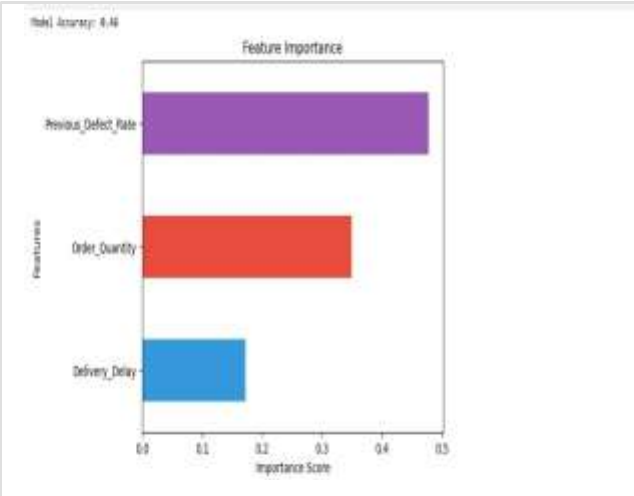
Analysis of the data showed that the Previous Defect Rate was the most important factor in determining which suppliers could produce a defect, followed by Delivery Delay, and then Order Quantity. This suggests that how a supplier has performed in the past, and whether they deliver on time, is a major factor in determining whether they could produce a defect

A risk score was calculated for all the suppliers. Suppliers whose risk score was greater than 0.70 were marked as risky.

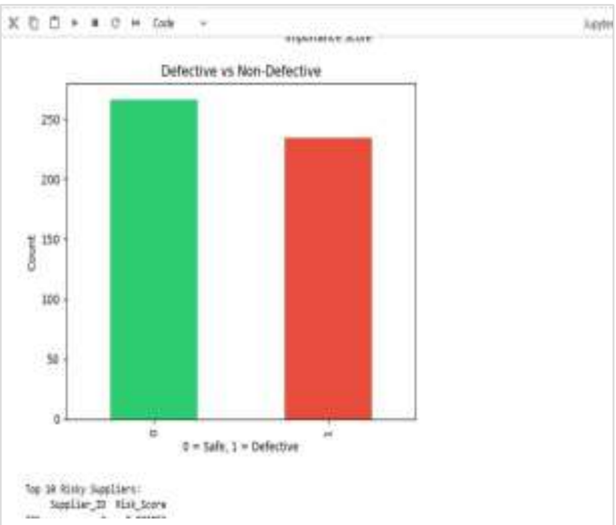
The model also effectively identified the highest risk suppliers, thus preventing possible issues.



This system is also capable of evaluating a new set of suppliers and provide a real-time risk assessment, which is either Safe or High Risk, with a risk score. Furthermore, a dynamic dashboard with a bar chart was developed for visualizing the output. It also indicates feature importance and the distribution of defective and non-defective items, which is easy to analyze and comprehend.



Overall, the project demonstrates how machine learning can be effectively used for defect prediction, risk assessment, automated alerting, and data visualization, improving decision-making in supplier management



9. CONCLUSION

It is also capable of forecasting the risk level of the new supplier. The paper is aimed at exploring the machine learning technique to identify the defective suppliers in the supply chain with special emphasis on Random Forests. The factors like order amount, number of late orders, etc., are considered to correctly identify the defective suppliers.

It has been observed that the complex, nonlinear data related to suppliers is effectively handled by the Random Forest, which is an ensemble machine learning technique. The addition of a visualization component to the system would make it even more practical, helping in timely and informed decisions.

If the bigger picture is considered, it is observed that this paper is important in the sense that it highlights the importance of having a data-driven system to transform the traditional supplier assessment mechanism to a more proactive mechanism. This system not only removes the defective products but also saves money, makes the processes more efficient, and improves the quality of products.

The accuracy of this system depends on the quality of the input data. To further enhance this study, it is recommended that it include real-time data, features, and algorithms, which will surely enhance the performance of the system. In conclusion, this study has created a good foundation for intelligent systems that could be used to detect defective suppliers. It has created a solution that could be used by different industries to produce better quality products.

10. REFERENCE

[1] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[2] G. Baryannis, S. Dani, and G. Antoniou, “Predicting supply chain risks using machine learning: The trade-off between performance and interpretability,” Future Generation Computer Systems, vol. 101, pp. 993–1004, 2019.

[3] S. Heydarbakian and M. Spehri, “Interpretable Machine Learning to Improve Supply Chain Resilience, An Industry 4.0 Recipe,” IFAC-PapersOnLine, vol. 55, no. 10, pp. 2834–2839, 2022.

[4] S. Haman, “Machine Learning Techniques in Supply Chain Management: An Exploratory Literature Review,” in Artificial Intelligence and Transforming Digital Marketing, Springer, 2022.

[5] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2016.

[6] A. El Filali, E. H. Ben Lahmer, and S. El Filali Hassan, “Machine Learning Techniques for Supply Chain Management: A Systematic Literature Review,” 2022.

[7] J. Wang, H. Yang, R. Wang, and J. Shou, "Data-driven supply chain management: A review," *International Journal of Production Research*, vol. 54, no. 23, pp. 7060–7080, 2016.

[8] S. Min, Y. Lim, and S. Yoo, "Machine learning in supply chain management: A review," *International Journal of Production Research*, vol. 57, no. 15–16, pp. 4707–4726, 2019.

[9] T. Choi, "Innovative 'bring-service-near-your-home' operations under Corona-virus (COVID-19/SARS-CoV-2) outbreak: Can logistics become the Messiah?" *Transportation Research Part E*, vol. 140, 2020.

[10] A. Kusiak, "Smart manufacturing," *International Journal of Production Research*, vol. 56, no. 1–2, pp. 508– 517, 2018.

[11] D. Ivanov and A. Dolgui, "Viability of intertwined supply networks: Extending the supply chain resilience angles towards survivability," *International Journal of Production Research*, vol. 58, no. 10, pp. 2904–2915, 2020.

[12] D. Ivanov, "Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS- CoV-2)," *Transportation Research Part E*, vol. 136, 2020.

[13] M. Christopher and H. Peck, "Building the resilient supply chain," *International Journal of Logistics Management*, vol. 15, no. 2, pp. 1–14, 2004.