ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Backorder Prediction in Inventory Management: Classification Techniques and Cost Considerations

Sarit Maitra[1] and Sukanya Kundu[2]

ABSTRACT

This article introduces an advanced analytical approach for predicting backorders in inventory management. Backorder refers to an order that cannot fulfilled immediately due to stock depletion. Multiple classification techniques, including Balanced Bagging classifiers, Fuzzy Logic, Variational Autoencoder (VAE) - Generative Adversarial Networks, and Multilayer Perceptron classifiers, are assessed in this work using performance evaluation metrics such as ROC-AUC and PR-AUC. Moreover, this work incorporates a profit function and misclassification costs, considering the financial implications and costs associated with inventory management and backorder handling. The study suggests a hybrid modelling approach, which includes ensemble techniques and VAE, which effectively addresses imbalanced datasets in inventory management. This approach emphasizes interpretability and reduces false positives and false negatives. This research contributes to the advancement of predictive analytics and offers valuable insights for future investigations in backorder forecasting and inventory control optimization for decision-making.

Article information:

Keywords: Backorder Forecasting, Cost Sensitive, Decision Science, Inventory Management, Machine Learning, Predictive Analytics

Article history:
Received: July 22, 2023
Revised: October 12, 2023
Accepted: November 30, 2023
Published: December 16, 2023 (Online)

DOI: 10.37936/ecti-cit.2023174.253571

1. INTRODUCTION

Backorders in inventory management refer to a customer's order for a product that is temporarily out of stock, resulting in a delay in fulfilment and delivery. Backorders can be both beneficial and detrimental. Orders might be delayed due to high demand or because of poor inventory planning. Organizations are continually striving for a balance in the management of backorders. It's a tight line to walk too much supply raises inventory expenses, while too little supply raises the danger of customers cancelling purchases, and excess stock increases inventory costs significantly.

The challenge is to classify and forecast severely imbalanced classes and assess their propensity to have backorders. Despite the considerable research efforts ([1]; [2]; [3]; [4]) and researchers addressing backorder forecasting using new age technologies, e.g., artificial intelligence (AI) and machine learning (ML)-based prediction, there are still open questions and challenges in this field. There is a lack of comprehensive studies that compare different AI-ML-based prediction systems in terms of their performance, advantages, limitations, and suitability for backorder forecasting. Moreover, while some studies look at financial consequences, the integration of cost and profit issues, such as misclassification costs and the profit function, is not thoroughly examined in the current literature. Many AI-based prediction models lack interpretability, making it difficult for decision-makers to comprehend the underlying issues driving backorders and take appropriate action.

This study addresses these gaps by employing advanced generative AI and ML techniques to bridge some of the existing gaps. The theoretical foundation of this article lies in the intersection of supply chain management, predictive analytics, and advanced AI-ML techniques, with a specific focus on backorder management and inventory system optimization. This study considers several metrics as cost-sensitive approaches to creating a final model, including misclassification cost, macro F1-score (harmony between precision and recall), precision-recall area under the curve (PRAUC), and receiver operating characteristics area under the curve (ROCAUC). This comprehensive evaluation helps to assess the accuracy, robustness, and cost-effectiveness

[1,2] The authors are with Alliance Business School, Alliance University, Bengaluru, India, E-mail: sarit.maitra@gmail.com and sukanya.kundu@alliance.edu.in

of the prediction model and offers important insights for decision-makers and practitioners in selecting the best applicable solutions for their specific supply chain contexts. Organizations can save costs, improve customer satisfaction through on-time delivery, mitigate the bullwhip impact, and enable proactive decision-making in changing market conditions by enhancing forecasting accuracy.

2. LITERATURE REVIEW

Before delving into demand and inventory management, we attempted to comprehend the key aspects influencing supply chain performance. In the past, researchers found empirical evidence that supply chain structure, inventory management policy, information interchange, customer demand, forecasting method, lead time, and review period duration are important contributors to inventory management [5]. Our study incorporates all of these and eventually narrows down to a forecasting method that demonstrates interdependence with inventory management policy, customer demand, lead time, and review period distribution.

The stochastic demand has attracted the attention of several scholars in the last decade. Researchers experimented with ML to improve the precision of backorder forecasts (e.g., [6], [7], [8], [9], etc.). An empirical study [6] reported an accuracy gain of nearly 20%. Each of these studies demonstrated the importance of carrying out more research to investigate various algorithms, build a cost-sensitive learning framework, and verify performance improvements. A recent study investigated the uses of AI and ML inside supply chains, opening new options for further research into how AI and ML can be used in SCM [8].

Building upon this foundation, the existing literature has delved into various aspects, notably focusing on the economic order quantity (EOQ) and economic production quantity (EPQ) models, even accounting for scenarios involving backorders ([1], [10]). This intersection offers a rich landscape for further investigation and synthesis, opening avenues for innovative approaches that blend traditional inventory models with cutting-edge AI and ML techniques. Moreover, these studies contribute to forecasting by integrating backorders into traditional inventory models and enhancing our understanding of managing backorders in inventory management. In a different approach, researchers combined ARIMA and ANN for backorder prediction ([11], [12]). The development of a Bayesian method for demand forecasting based on compound Poisson distributions is a significant contribution that outperforms other current methods [13]. This method improves the accuracy and dependability of demand forecasts by using the probabilistic framework of Bayesian inference and the diversity of compound Poisson distributions. The effectiveness of this strategy not only demonstrates its potential to develop forecasting methodologies but also emphasizes the need to apply Bayesian principles and novel distribution models for demand forecasting. Researchers also explored the use of artificial neural networks for backorder prediction ([3], [12]). The combination of all the above methods highlights the advancement of backorder forecasting.

Several authors (e.g., [14], [6], [1], [15], [3], [8]) have applied ML techniques to the same dataset to determine whether advanced ML techniques can improve the effectiveness of backorder forecasting in the early stages of the supply chain. All these studies enhance our understanding of managing backorders, optimizing inventory systems, improving customer service levels, and aiding decision-making. Adaptable and resilient models are required to handle the complexity of massive inventory data and deliver the correct insights for decision-making ([16], [17]). However, despite the progress made, the field of AI and ML in supply chain management, including backorder forecasting, is still in its early stages [18]. The challenges of working with big data in inventory management, such as data volume, variety, heterogeneity, and statistical biases, need to be addressed to develop adaptable and resilient models for accurate decision-making.

This literature review serves as a theoretical foundation for developing the conceptual model for our work. Fig. 1 outlines a flow diagram with the different steps and measures that were taken in our study to explore and validate the effectiveness of AI and ML techniques in backorder forecasting.

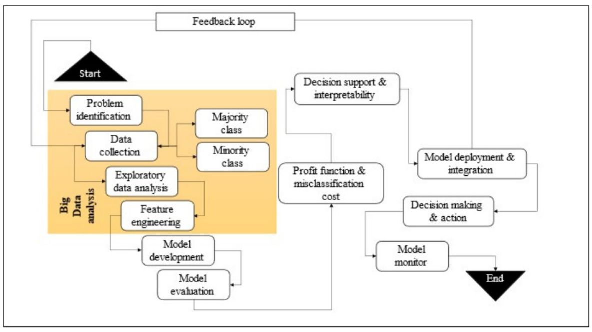

Fig.1: *Conceptual framework.*

By operationalizing the conceptual model into a logical framework, we aim to provide a structured framework for conducting empirical research, analyzing data, and deriving actionable insights for inventory management and supply chain decision-making. Furthermore, the logical model considers cost-sensitive learning frameworks, validation in real-world implementations, and considerations of contextual factors in backorder forecasting. Profit functions and misclassification costs are incorporated into the model to address the financial consequences and costs related to inventory management and backorder.

Table 1: Summary Statistics.

Sl no.	Variables	Description	Data type	Mean	Median	Std Dev	Maxima	Value
2	**nationalInv**	Current inventory level	int64	489.42	15.00	28595.83	12334400	unit
3	**leadTime**	Transit time for product	float64	7.84	8.00	7.04	52	weeks
4	**inTransitQty**	Product in transit	int64	45.36	0.00	1390.53	489408	
5	**forecast3Month**	Forecast for next 3 months	int64	185.22	0.00	5032.30	1218328	unit
6	**forecast6Month**	Forecast for next 6 months	int64	360.88	0.00	10067.64	2461360	unit
7	**forecast9Month**	Forecast for next 9 months	int64	528.91	0.00	14895.45	3777304	unit
8	**sales1Month**	Sales revenue prior 1 month	int64	57.30	0.00	2067.93	741774	unit
9	**sales3Month**	Sales revenue prior 3 months	int64	180.46	0.00	5263.48	1094112	unit
10	**sales6Month**	Sales revenue prior 6 months	int64	352.46	0.00	9773.35	2146625	unit
11	**sales9Month**	Sales revenue prior 9 months	int64	544.33	0.00	15195.65	3201035	unit
12	**minBank**	Minimum amount to stock	int64	54.14	0.00	1244.24	313319	unit
13	**potentialIssue**	Past overdue	int64	3.28	0.00	299.43	146496	unit
14	**piecesPastDue**	Performance last 6 months	float64	-7.05	0.82	26.84	1	N/A
15	**perf6MonthAvg**	Performance last 12 months	float64	-6.62	0.80	26.14	1	N/A
16	**perf12MonthAvg**	Amount of stock orders overdue	int64	0.63	0.00	35,18	12530	unit
17	**localBoQty**	Issue identified	object					
18	**deckRisk**	Risk flag	object					
19	**oeConstraint**	Risk flag	object					
20	**ppapRisk**	Risk flag	object					
21	**stopAutoBuy**	Risk flag	object					
22	**revAtop**	Risk flag	object					
23	**wentOnBackorder**	Product went on backorder	object					

To this end, the problem statement can be hypothesized as follows: A hypothetical manufacturer has a data set that indicates whether a backorder has happened. The goal is to accurately identify future backorder risk using machine learning and predictive analytics and to determine the optimal way to stockpile high-risk products.

3. DATA ANALYSIS & MODEL DEVELOPMENT

The inventory dataset used in this study has 1,04,8575 entries with 8 categorical and 15 numerical variables. A unique identification SKU (stock-keeping unit) assigns each data point to a distinct product. This label provides no additional value and is thus eliminated from further analysis. Table 1 shows a summary of the dataset's statistics. We identified the categorical columns (7 columns with object data type, which excludes SKU). The data file includes eight weeks of historical data, which comes before the week we aim to forecast.

We used SMOTE (synthetic minority oversampling technique) and generative AI-based unsupervised variational auto-encoder (VAE) to deal with imbalanced classes, which increases modelling accuracy and efficiency. SMOTE is used for balancing imbalanced datasets. It works by generating synthetic samples for the minority class to balance class distribution. VAE learns a probabilistic mapping between the input data and a latent space, allowing for the generation of new samples that resemble the training data.

The second task is to optimize for the business case. To do this, we employ:

- Profit maximization via classification models which entails developing a profit function, optimizing the decision threshold, feature engineering, cost-sensitive learning, model selection, and continuous monitoring to ensure that the model's performance aligns with the business's financial objectives. It is a data-driven and business-focused strategy to increase profitability while considering any trade-offs and risks.
- Misclassification costs, which is a practical and business-oriented approach to optimization. It allows for informed decisions regarding model performance, model selection, and decision thresholds while considering the financial and operational aspects of the business case. The problem is viewed as a cross-sectional problem.

3.1 Exploratory Data Analysis

A thorough data mining was essential for this dataset. The has helped in understanding the characteristics, patterns, and relationships within the data. The relevant attributes are:

- *nationalInv* indicates the current inventory levels. The mean inventory is 489.42 units, with a wide range from 15 units to a maximum of 12,334,400 units. The median value is 15 units, indicating a possible right-skewed distribution.
- *leadTime* represents the transit time for the product. The mean transit time is 7.84 weeks, with a median of 8 weeks. The standard deviation is 7.04 weeks, indicating some variability in transit times.
- *inTransitQty* indicates the number of products currently in transit. The high standard deviation and maximum value indicate significant variation in transit quantities.
- Sales 1, 3, 6, and 9 months include variables representing sales revenue for different periods as well as forecasts for the next 3, 6, and 9 months. The mean and median values indicate the average and middle values of sales and forecasts. The high standard deviations suggest variability in sales performance.
- *minBank* represents the minimum amount required to stock. The mean and median values indicate the average minimum stocking requirement, while the high maximum value suggests that some products may require a significantly higher minimum stock level.
- *piecesPastDue* represents the number of products that are past due. The mean and median values suggest a low average quantity of past-due items.
- *perf6MonthAvg* and *perf12Monthavg* represent the performance of products over the last 6 and 12 months, respectively. The negative mean values for both variables indicate below-average performance, while the standard deviations suggest variation in performance.
- *localBoQty* indicates the number of stock orders overdue. The low mean and median values suggest a relatively low average amount of overdue stock orders.
- Other attributes, such as *potentialIssue*, *deckRisk*, *oeConstraint*, *ppapRisk*, *stopAutoBuy*, and *revStop*, represent risk flags associated with the products. These variables are categorical and indicate the presence or absence of specific risks.
- *wentOnBackorder* is the target variable that represents whether a product went on backorder.

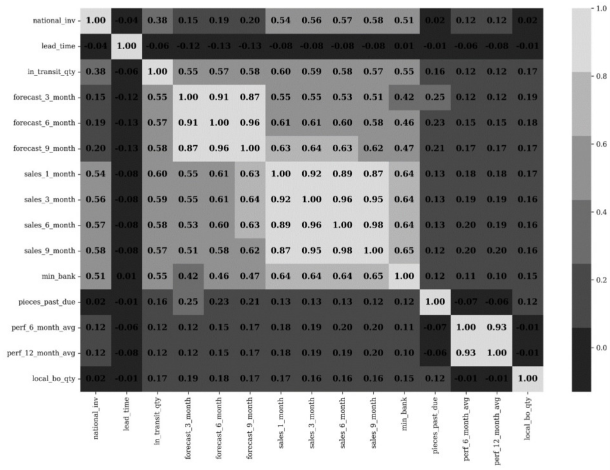

Fig.2: *Correlation heatmap of numerical features.*

Fig. 2 displays the Spearman correlation matrix of numerical attributes. Spearman correlation does not assume a linear relationship between variables. It measures the monotonic relationship, which means it is suitable for detecting both linear and non-linear associations. All the significant correlations observed are positive.

- *forecast3Month, forecast6Month* and *forecast9Month* are strongly correlated (coefficient = 0.99).
- *sales1Month, sales3Month, sales6Month,* and *sales9Month* are strongly correlated with each other with a degree varying from 0.82 to 0.98.
- *perf6MonthAvg* and *perf12MonthAvg* are very highly correlated with each other (coefficient = 0.97).
- *minBank* is highly correlated with sales and forecast columns as stock in inventory is directly proportional to sales.
- *inTransitQty* is highly correlated with sales, forecast, and *minBank* columns. This is obvious because high sales of a product => more of that product in transport for inventory replenishing high sales of a product => high forecast.
- *piecesPastDue* is meekly correlated with sales

and forecast columns *nationalInv* is meekly correlated with *minBank* and weekly correlated with sale columns.

Overall, the correlation matrix indicates that the number of features used to forecast whether an item will be placed in backorder may be less than the number of features in the data set. Fig. 3 displays the correlation matrix of categorical attributes. Based on the chi-squared values, there is no strong association between any pairs of variables. There might be a relationship between *oeConstraint* and *potentialIssue*. *RevStop* and *oeConstraint* may be related in some way. There might be a relationship between *potentialIssue* and *revStop*. Even if the coefficients are high, the features mentioned above have the lowest scores in comparison.

We employed several data pre-processing steps to get the raw data ready for modelling. This includes replacing -99.0 in performance columns with nan for imputing and employing an iterative imputer to fill in the missing values. Moreover, as some of the features in the numerical columns are correlated, statistical models which are based on statistical assumptions, may not perform as well as the coefficients of separating plane change.

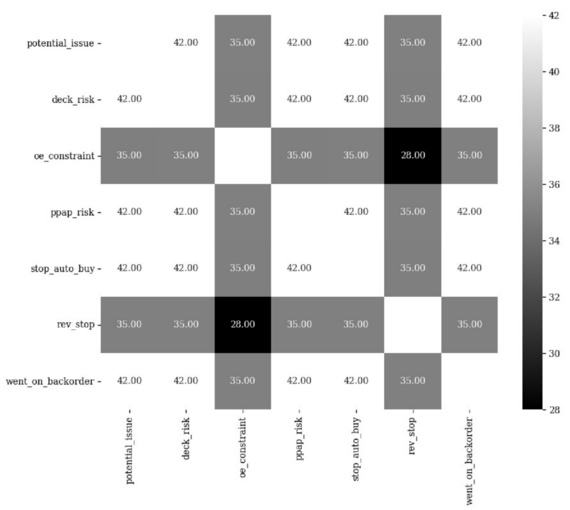

Fig.3: *Chi-squared test heatmap of categorical features.*

The values in the category columns are encoded with "Yes" as 1 and "No" as 0. To fill in the missing variables, we employed model-based imputation (iterative-imputer). Moreover, we treated the positive skewness (possible outliers) in the data in two ways, resulting in two distinct datasets to apply.

- Values = (value - median) / (75% value - 25% value) to standardize the data while scaling it without taking outliers into account.
- Applied the log transformation followed by the Standard Scaler to the columns in the dataset with positive skewness.

We employed Principal Component Analysis (PCA) to identify major features explaining 99% variance in the dataset. Fig. 4 displays the top three features in our PCA: *nationalInv, forecast9Month,* and *sales9Month*.

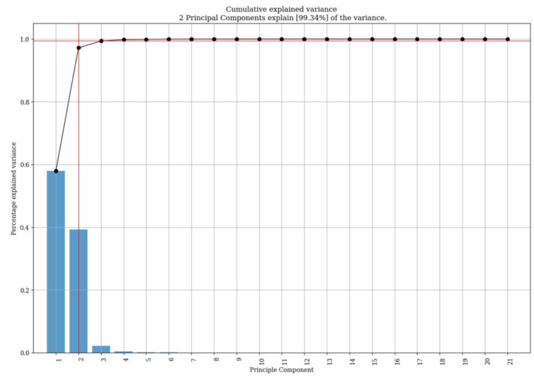

Fig.4: *Feature importance using PCA.*

The performance metrics used in this work are ROCAUC, (differentiates between positive and negative classes), PRAUC (to select an appropriate threshold that balances the trade-off between precision and recall), Macro F1-Score (average of the F1 scores of both the positive and negative classes). Besides these, we also report Mathews Correlation Coefficient (MCC), precision, recall, and Brier scores for better comparison. MCC considers true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values to provide a balanced measure of classification accuracy. Considering these, our work aims to provide a comprehensive evaluation of the predictive models in terms of their performance, interpretability, and latency constraints. By not having latency constraints, the models can potentially make more accurate and informed predictions by leveraging a broader range of data and capturing any patterns or trends that may emerge over longer periods.

3.2 Hypotheses test

We employed non-parametric Mann-Whitney and Chi-Square tests to help assess the association between each attribute and the target variable, or their differences between different groups. The results are displayed in Table 2. The results indicate significant associations between certain attributes and backorders. Low and significant p-values for the attributes *deck_risk*, *oe_constraint*, and *ppapRisk* indicated strong relationships. All the numerical attributes displayed significant relationships with backorders, based on the Mann-Whitney U tests. These findings suggest that these attributes may play a crucial role in predicting backorders.

3.3 Summary of data analysis and feature engineering

We have a severely imbalanced dataset with a 99.99% majority class with a binary classification problem where we have to predict whether or not a product will go to backorder.

The numerical attributes are highly skewed.

- *leadTime* attributes come with missing values.
- *perf6MonthAvg* and *perf12MonthAvg* are heavily left skewed rest all the features are right skewed.
- The numerical attributes have a small interquartile range, and some have negative values.
- The attributes of sales, forecast, and performance are correlated.
- The dataset contains 8 categorical attributes, including *wentToBackorder* as the target variable.
- Encoded categorical attributes to numerical attributes.
- The dataset is split into train and test datasets with 80:20.
- Iterative imputer is used for missing value imputation.
- Performed PCA to determine feature importance.
- Feature transformation was applied to the dataset (Robust scaler, Log transform, and Standard scaler).

3.4 Modeling and Evaluation

We experimented with 5 different algorithms to address the complexities and challenges of inventory management systems. There is no one size fits all and researchers in the past have experimented with different techniques in their work (e.g., [6], [3], etc.) We used a dummy model as the base model, and subsequently Balanced Bagging classifier (BBC), the BBC with VAE, Fuzzy logic with the BBC, Random Forest, and Multilayer Perceptron. Each model was implemented with four transformed datasets (Robust Scaling, log Transform Standard Scaling, Quantile Transform). Thus, a total of 20 models were trained and tested.

To start with, a grid search cross-validation (GridSearchCV) was employed to fine-tune the Balanced Bagging Classifier (BBC) by exploring different hyperparameters. The optimum hyperparameters are chosen depending on the ROC AUC. Table 3 displays the pseudocode for hyperparameter search.

The purpose of controlling the maximum number of features during grid search is to introduce randomness and diversity into the individual base estimators of the ensemble. It helps to prevent overfitting and improve the generalization of the ensemble. By restricting the number of features considered at each split, the base estimators are forced to make more independent decisions, which can lead to a more robust and accurate ensemble model.

Table 2: Statistical hypotheses tests.

Sl. no.	Attributes	Statics with Significance
0	nationalInv	U statistic: 1834036383.5, p-value: 0.00
1	leadTime	U statistic: 4239446959.0, p-value: 0.00
2	inTransitQty	U statistic: 7002852514.5, p-value: 0.00
3	forecast3Month	U statistic: 6905403097.0, p-value: 0.00
4	forecast6Month	U statistic: 6827462574.5, p-value: 0.00
5	forecast9Month	U statistic: 5869355615.0, p-value: 0.00
6	sales1Month	U statistic: 5910651312.5, p-value: 0.00
7	sales3Month	U statistic: 5809789939.0, p-value: 0.00
8	sales6Month	U statistic: 5740565459.0, p-value: 0.00
9	sales9Month	U statistic: 4688282621.5, p-value: 0.019
10	minBank	U statistic: 5062287539.5, p-value: 0.00
11	potentialIssue	U statistic: 4193563375.0, p-value: 0.00
12	piecesPastDue	U statistic: 4174380247.5, P-value: 0.00
13	perf6MonthAvg	U statistic: 5145229552.0, P-value: 0.00
14	perf12MonthAvg	U statistic: 1834036383.5, p-value: 0.00
15	localBoQty	U statistic: 4239446959.0, p-value: 0.00
16	deckRisk	Test statistic: 219.903, p-value: 0.00
17	oeConstraint	Test statistic: 28.274, p-value: 0.00
18	ppapRisk	Test statistic: 110.458, p-value: 0.00
19	stopAutoBuy	Test statistic: 4.564, p-value: 0.032
20	revAtop	Test statistic: 2.978, p-value: 0.084

The BBC combines the advantages of bagging and sampling techniques to address the issue of imbalanced datasets [20]. Several researchers in recent times and the past have recommended fuzzy logic (e.g., [11], [20], [21], [22], [23], [24], [25]). The objective is to add human-centric design along with advanced machine-learning algorithms. Therefore, fuzzy logic was integrated into the BBC to handle imprecise information that is frequently encountered in inventory management systems.

Furthermore, we leveraged the power of the Generative Adversarial Network—Variational Auto Encoder (VAE) to experiment with our modelling approach. Researchers claimed the superior performance of unsupervised VAE compared to supervised

Table 3: *Pseudocode - exhaustive hyperparameter search.*

```
# Hyperparameters and the possible values
parameters = {
   'n_estimators': [20, 50, 100, 200, 300, 400, 500, 600,
700, 800, 900, 1000, 1200],
   'max_features': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8,
0.85, 0.90, 0.92, 0.95, 1.0],
   'bootstrap': [True, False],
   'bootstrap_features': [True, False],
}

# BBC Instance
classifier = Balanced Bagging Classifier ()

# Grid Search cross-validation instance
gridSearch = GridSearchCV(estimator = classifier,
param_grid = parameters, cross validation = 5, scoring
= 'roc_auc')

# Fit GridSearchCV to the training data
gridSearch.fit(X_train, y_train)

# Visualize the results
gridSearchPlot(grid_search)
gridSearchTable(grid_search)
```

ML techniques for imbalanced data (e.g., [26], [27]). We combined supervised BBC with unsupervised VAE to develop a powerful approach to address class imbalance and capture complex feature representations. VAE generates meaningful latent representations of the input data, which are then combined with the original features to improve the performance of the BBC. Table 4 presents the pseudocode for VAE implementation.

The VAE model consists of an encoder and a decoder. The encoder converts the input data x to a latent space representation z that is parameterized by mean and variance:

$$q(z|x) = N(z|(x), 2(x)) \tag{1}$$

This latent space representation is then used as input for the decoder. Based on the learned latent representations, the decoder provides a meaningful and accurate reconstruction of the original data. This reconstruction method aids in capturing the key aspects and qualities of the data, allowing the model to generate correct predictions and classify instances. $p(x|z) = N(x|f(z))$, where $f(z)$ denotes the mapping from the latent space to the original feature space. The VAE aims to maximize the ELBO objective, which is a trade-off between reconstruction accuracy and regularization of the latent space:

$$ELBO = E[\log p(x|z)] - D\_KL(q(z|x)||p(z)) \tag{2}$$

Table 4: *Pseudocode for handling imbalanced data with VAE and BBC model training.*

```
X, y = SplitData(df);

# Data Preprocessing
numeric_features = ['nationalInv', 'leadTime',
'inTransitQty', 'piecesPastDue', 'localBoQty'];
categorical_features = ['potentialIssue', 'deckRisk',
'oeConstraint', 'ppapRisk', 'revStop'];

# VAE model
input_dim = GetNumberOfFeatures(X);
latent_dim = 10;

# Encoder and decoder layers for the VAE
"""Encoder and decoder layers of the VAE are defined
with activation functions. These layers are part of the
neural network architecture used in the VAE"""
input_layer = DefineInputLayer(input_dim);
encoded =
DefineDenseLayerWithActivation(input_layer, 32,
'relu');
encoded = DefineDenseLayerWithActivation(encoded,
latent_dim, 'relu');
decoded = DefineDenseLayerWithActivation(encoded,
32, 'relu');
decoded = DefineDenseLayerWithActivation(decoded,
input_dim, 'linear');

# VAE model with custom loss
vae = CreateVAEModel(input_layer, decoded,
CustomLossFunction);

# Train VAE on the imbalanced data
TrainVAEModel(vae, X_train_processed,
X_train_processed, epochs=20, batch_size=32);

# Encode the input data using the VAE
encoded_X_train = EncodeDataWithVAE(encoder,
X_train_processed);

# Combine the latent representations with the original
features
combined_X_train =
ConcatenateFeatures(X_train_processed,
encoded_X_train);

# Train BBC on the combined feature set
bbc =
CreateBalancedBaggingClassifier(n_estimators=1000)
TrainBalancedBaggingClassifier(bbc,
combined_X_train, y_train);

# Preprocess the test data
X_test_processed = PreprocessTestData(X_test);
encoded_X_test = EncodeDataWithVAE(encoder,
X_test_processed);
combined_X_test =
ConcatenateFeatures(X_test_processed,
encoded_X_test);
y_pred = PredictWithBalancedBaggingClassifier(bbc,
combined_X_test);

# ROC AUC score
roc_auc = CalculateROCAUCScore(y_test, y_pred)
```

where $E[\log p(x|z)]$ represents the reconstruction term and $D\_KL(q(z|x)||p(z))$ is the KL divergence between the approximate posterior $q(z|x)$ and the prior distribution $p(z)$.

BBC is further integrated with the decoder to improve the overall performance. This approach ensures that the final prediction considers the opinions of multiple classifiers.

Lastly, we experimented with Artificial Neural Network (ANN) based Multilayer Perceptron (MLP). MLP provides efficient computing, lowering computational time and memory utilization, making it a desirable tool for real-world inventory control systems ([13], [27]).

4. RESULTS & DISCUSSIONS

Table 5 presents a consolidated report of all the classifier experimented.

- *ROCAUC*
  a) *on non-normal data:* BBC (0.9081) > VAE_BBC (0.9003) > FL_BBC (0.8759) > Dummy (0.5074).
  b) *on Log-transformed and Normalized data:* BBC (0.9073) > VAE_BBC (0.9007) > FL_BBC (0.8615) > Dummy (0.4897).
- *PRAUC*
  a) *on non-normal data:* BBC (0.4925) > VAE_BBC (0.4841) > FL_BBC (0.4646) > Dummy (0.2640).
  b) *PRAUC on Log-transformed and Normalized data:* BB (0.4917) > VAE_BBC (0.4847) > FL_BBC (0.4515) > Dummy (0.2456).
- *Macro F1-Score*
  a) *on non-normal data:* BB (0.5545) > VAE_BBC (0.5532) > FL_BBC (0.5251) > Dummy (0.3407).
  b) *on Log-transformed and Normalized data:* BB (0.5544) > VAE_BBC (0.5524) > FL_BBC (0.5195) > Dummy (0.0160).
- *Precision*
  a) *on non-normal data:* BB (0.0838) > VAE_BBC (0.0824) > FL_BBC (0.0601) > Dummy (0.0087).
  b) *on Log-transformed and Normalized data:* BB (0.0840) > VAE_BBC (0.0818) > FL_BBC (0.0555) > Dummy (0.0081).
- *Recall*
  a) *on non-normal data:* BBC (0.9005) > VAE_BBC (0.8848) > FL_BBC (0.8679) > Dummy (0.5151).
  b) *on Log-transformed and Normalized data:* BB (0.9017) > VAE_BBC (0.8865) > FL_BBC (0.8460) > Dummy (0.4786).
- *Mathew's correlation coefficient*
  a) *on non-normal data:* BB (0.2600) > VAE_BBC (0.2552) > FL_BBC (0.2099) > Dummy (0.0027).
  b) *on Log-transformed and Normalized data:* BB (0.2609) > VAE_BBC (0.2544) > FL_BBC (0.1976) > Dummy (-0.0037).
- *Brier score*
  a) *on non-normal data:* Dummy (0.2500) > VAE_BBC (0.0626) > BB (0.0623) > FL_BBC (0.0815).
  b) *on Log-transformed and Normalized data:* Dummy (0.2500) > VAE_BBC (0.0629) > BB (0.0622) > FL_BBC (0.1310).

Area Under the Precision-Recall Curve (PRAUC), F1-Score, Matthews Correlation Coefficient (MCC), ROC AUC, Precision at a Given Recall, and Geometric Mean is commonly used by researchers to report model accuracy (e.g., [28], [29], etc.). We used a combination of multiple metrics, which provides

Table 5: *Results & discussions.*

Retained original dimensions							
	ROCAUC	**PRAUC**	**Macro F1-Score**	**Precision**	**Recall**	**MCC**	**Brier**
Models on non-normal data							
BBC	0.9081	0.4925	0.5545	0.0838	0.9005	0.2600	0.0623
FL_BBC	0.8759	0.4646	0.5251	0.0601	0.8679	0.2099	0.0815
VAE_BBC	0.9003	0.4841	0.5532	0.0824	0.8848	0.2552	0.0626
Models on Log transformed + Normalized dataset							
Dummy	0.4897	0.2456	0.0160	0.0081	0.4786	-0.0037	0.2500
BBC	0.9073	0.4917	0.5544	0.0840	0.9017	0.2609	0.0622
FL_BBC	0.8615	0.4515	0.5195	0.0555	0.8460	0.1976	0.1310
VAE_BBC	0.9007	0.4847	0.5524	0.0818	0.8865	0.2544	0.0629
Random undersampler with	0.9604	0.2428	0.5483	0.0785	0.9005	0.2507	0.0641

a more comprehensive assessment of model performance. Based on the metrics, the BBC model on the log-transformed and normalized dataset performed well across multiple metrics, indicating its effectiveness in predicting the target variable. Additionally, the VAE + BBC model on non-normal data also showed promising performance. The Brier score measures the accuracy of probabilistic predictions, and a lower score indicates better calibration and accuracy.

The confusion matrix in Fig. 5 allows us to examine the model's predictions in detail. It shows the number of true negatives (TN) and false positives (FP) for backorder predictions. In this case, the high number of TNs (190,429) indicates that the model performs well in accurately identifying non-backorders. This means that the model correctly identifies a significant portion of the instances where a backorder is not present, which is crucial for efficient inventory management. The number of FP (17,506), however, indicates that there are some situations in which the model forecasts non-backorders as backorders. This can lead to unnecessary actions or costs associated with backorder handling for those specific instances.

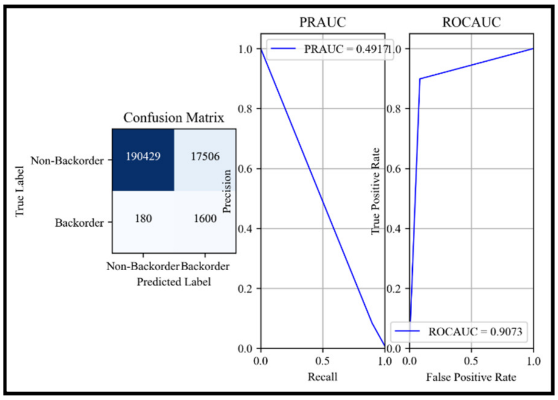

Fig.5: *VAE + BBC model - confusion matric and roc curves.*

The presence of false negatives (180) indicates the model occasionally fails to predict backorders when they occur. This means that there are instances where the model incorrectly classifies a backorder as a non-backorder. This can result in missed opportunities to take proactive measures and prevent backorders, leading to potential disruptions in the supply chain and customer dissatisfaction. On the other hand, the presence of true positives (1600) suggests that the model can identify backorders to some extent. These instances represent the correct classification of backorders by the model.

Conversely, Fig. 6 shows that a high number of TN indicates that the model identified a quite a few samples (190,235) as non-backorders, indicating that it works well in reliably identifying non-backorders. A moderate number of FP means the model misclassified a considerable number of instances as backorders when they were non-backorders (17,700). This suggests that the model tends to predict false positives, which means it may sometimes predict an item as a backorder when it is not.

A low number of FPs means the model misclassified a relatively small number of instances as non-backorders when they were backorders (202). A moderate number of TPs means the model correctly identified a reasonable number of instances as backorders (1,578). This indicates that the model can predict backorders accurately. Overall, the model shows good performance in correctly identifying non-backorders (high TN) but has some limitations in accurately predicting backorders (moderate TP and FN). Reducing false positives and false negatives would be valuable for improving the overall accuracy of backorder predictions.

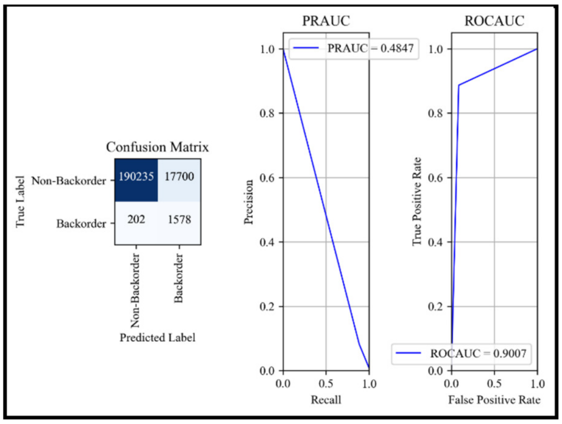

Fig.6: *BBC-A model (non-normal data).*

Matrix 2 (Fig. 6) performs slightly better than Matrix 5, as it has a lower number of FP and FN and a slightly higher number of TP. However, the differences between the two matrices are relatively small. This means that for Matrix 1 (Fig. 5), around 88.64% of products identified as potential backorders went on backorder, while for Matrix 2, around 90.02% of products identified as potential backorders went on backorder.

We justify using supervised learning under the presumption that the memory that is accessible is limited by the size of the dataset. The context of big data, distributed databases, and embedded systems all apply to this fundamental paradigm. We propose a simple yet powerful ensemble framework in which each ensemble model is constructed from a random patch of data generated by selecting random subsets of features and instances from the entire dataset. These results demonstrate that the suggested method performs on par with popular ensemble methods in terms of accuracy while simultaneously lowering memory requirements and achieving much superior performance when memory is severely limited.

4.1 Profit function

By measuring the overall profit generated by the forecasts, we can gain insights into the financial impact of the model's predictions and make informed decisions about inventory levels, ordering quantities, and backorder management strategies. This evaluation goes beyond traditional accuracy metrics and provides a more comprehensive assessment of the model's performance from a business perspective. The profit calculation in the given formula involves the following parameters: *SalesRevenue (1,3,6,and 9 months), holdingCost, inventoryLevel, backorderCost, leadTimeCost, potentialIssueCost, deckRiskCost.*

$$Revenue = \sum(sales1Month + sales3Mont + sales6Month + sales9Month) \quad (3)$$

$$\begin{aligned} Profit = Revenue – (&holdingCost \\ &* inventoryLevel) \\ &- (backorderLost \\ &* backorders) – (leadTimeCost \\ &* leadTime) – (potentialIssueCost \\ &* potentialIssue) – (deckRiskCost \\ &* deckRisk \end{aligned} \quad (4)$$

We performed optimization to maximize the profit by finding the optimal values for the decision variables: *holdingCost*, *backorderCost*, *leadTimeCost*, *potentialIssueCost*, and *deckRiskCost*. The initial guess for the decision variable values was set to x_0. Here, we defined a constraint as $holdingCost \geq 0$. The goal is to find the optimal values for the decision variables (costs) that maximize the profit. Results are displayed in Table 6. We used the numerical optimization method by employing Sequential Least Squares Quadratic Programming (SLSQP).

4.2 Cost-Sensitive Learning

The overall cost of misclassification is computed as the sum of the costs of false positives and false negatives. It calculates the entire cost of inaccurate predictions, taking into consideration the expenses associated with each form of misclassification.

- $FPcost = 10$ (cost of misclassifying a backorder item as non-backorder)
- $FNcost = 1$ (cost of misclassifying a non-backorder item as backorder)

$$Total\ Cost = \{FP * FPcost\} + \{FN * FNcost\} \quad (5)$$

Table 4 displays the cost-sensitive factors. Considering both average profit and misclassification cost, the VAE + BBC model appears to perform well in terms of optimal profit, while the BBC-A model performs well in terms of misclassification cost for non-normal data.

Table 6: *Cost perspective.*

Classifier	Optimal Profit	Misclassification cost
Non-normal data		
BBC-A	98,37,422.94	1,75,443.00
FuzzyBBC-A	74,51,418.43	2,42,238.00
VAE + BBC	98,37,424.40	1,78,353.00
Log normalized data		
BBC-B	98,37,510.94	1,75,590.00
FuzzyBBC-B	74,51,422.38	2,42,233.00
VAE + BBC	98,37,423.06	1,78,060.00
MLP	98,36,319.84	1,80,041.00

4.3 Interpretability

Interpretability is a critical aspect of the adoption of AI and ML in business ([30], [31]). It enables supply chain professionals to choose the best ML algorithms and comprehend the justification for the forecasts and decisions made by the model. This understanding empowers practitioners to optimize supply chain operations, improve decision-making, enhance visibility, mitigate risks, and drive overall efficiency and effectiveness ([32], [33]). By embracing interpretability, practitioners can make informed adjustments, ensure regulatory compliance, and establish trust with stakeholders, thus harnessing the full potential of AI and ML in transforming supply chain management.

We employed permutation importance (PI) to understand the relative importance of each attribute while making decisions. PI measures the importance of each feature by randomly shuffling its values and observing the impact on the model's performance This helps to understand the attributes affecting backorder predictions and sheds light on the contribution of each feature in the model. Fig. 7 displays the relative weights of each attribute in predicting backorders. These weights, along with their standard deviations, provide insights into how each attribute influences the model's output. We can draw several key conclusions from this analysis:

- *nationalInv* attribute carries a high weight, indicating its significant impact on the model's predictions. Changes in this attribute (national inventory) values have a substantial effect on anticipating backorders. This underscores the pivotal role that inventory levels play in predicting backorders.
- *sales1Month* is the second most influential at-

tribute. It captures recent demand patterns and market dynamics, allowing the model to make informed predictions about backorders. Its significance lies in its ability to reflect immediate sales trends and enable timely responses to changes in customer demand, ultimately contributing to customer satisfaction and operational efficiency.

- $forecast3Month$ and $Sales9Month$ attributes also exhibit substantial importance, emphasizing the value of both short-term (3-month) forecasts and the history of sales over 9 months in predicting backorders. $forecast3Month$ attribute provides forward-looking insights into expected demand, aiding in inventory planning and procurement decisions.

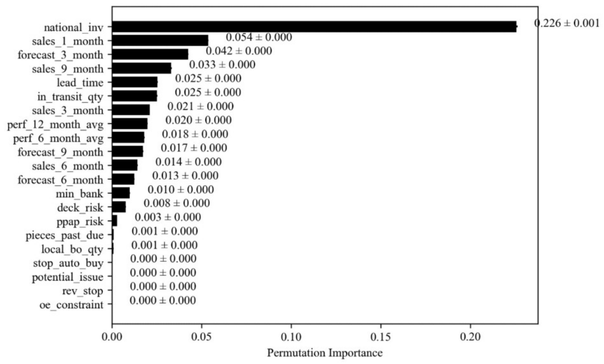

Fig.7: *Permutation Importance (BBC model).*

Some of the features, such as StopAutoBuy, PpapRisk, revStop, and OeConstraint, carry nearly zero weights. This indicates that changes in these attributes have minimal impact on backorder predictions and are less significant in this context.

4.4 Limitations

This study relies heavily on data quality and availability, which could be improved by overcoming data-related challenges. The real-world supply chain dynamics are complex and ever-changing, and future research could explore adaptive models that can adapt to dynamic environments in real time. The findings for the current work are promising, but their applicability across diverse industries and supply chain structures may vary. Future research directions include developing explainable AI models, real-time supply chain optimization, hybrid models, cost modelling, and cross-domain applications. The study paves the way for a new era of supply chain management by integrating state-of-the-art artificial intelligence and generative model with the financial considerations. However, we encourage researchers and practitioners to embrace these limitations as opportunities for further innovation. The future of supply chain analytics is bright, with groundbreaking discoveries and transformative applications.

5. CONCLUSIONS

This paper presented a novel perspective on the impact of advanced predictive analytics techniques on backorder management. By combining generative AI-based unsupervised VAE with the supervised BBC model, the study offers valuable insights into proactive strategies, stockout mitigation, and supply chain optimization. The research demonstrates the trade-off between profit and misclassification costs, providing a comprehensive evaluation of the model's performance through various metrics such as ROCAUC, PRAUC, Macro F1-Score, Precision, Recall, MCC, and Brier. By incorporating a profit function and considering misclassification costs along with these metrics, this work addresses the financial implications and costs associated with inventory management and backorder handling. Additionally, the inclusion of permutation importance enhances the interpretability of the model by identifying the relative importance of input features. This provides valuable insights into the factors influencing backorder forecasting and inventory control, contributing to future investigations in the field. The potential real-world applications of our research are vast. Businesses operating in dynamic supply chain environments can leverage our insights to fine-tune their inventory management strategies, minimize stockouts, and enhance customer satisfaction. Moreover, the findings have the potential to revolutionize the way businesses make decisions by incorporating cost-sensitive predictive analytics into their daily operations.

References

[1] C. Ntakolia, C. Kokkotis, P. Karlsson and S. Moustakidis, "An explainable machine learning model for material backorder prediction in inventory management," *Sensors*, vol. 21, no. 23:7926, 2021.

[2] U. Praveen, G. Farnaz and G. Hatim, "Inventory management and cost reduction of supply chain processes using AI based time-series forecasting and ANN modeling," *Procedia Manufacturing*, vol. 38, pp. 256-263, 2019.

[3] M. Shajalal, P. Hajek and M. Z. Abedin, "Product backorder prediction using deep neural network on imbalanced data," *International Journal of Production Research*, vol. 61, no. 1, pp. 302-319, 2023.

[4] M. Shokouhifar and M. Ranjbarimesan, "Multivariate time-series blood donation/demand forecasting for resilient supply chain management during COVID-19 pandemic," *Cleaner Logistics and Supply Chain*, vol. 5, 100078, 2022.

[5] J. George and V . M. Pillai, "A study of factors affecting supply chain performance," in *Journal of Physics: Conference Series*, vol. 1355, No. 1, p. 012018, 2019.

[6] S. Islam and S. H. Amin, "Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques," *Journal of Big Data*, vol. 7, pp. 1-22, 2020.

[7] R. B. de Santis, E. P. de Aguiar and L. Goliatt, "Predicting material backorders in inventory management using machine learning," *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, Arequipa, Peru, pp. 1-6, 2017.

[8] P. Hajek and M. Z. Abedin, "A Profit Function-Maximizing Inventory Backorder Prediction System Using Big Data Analytics," in *IEEE Access*, vol. 8, pp. 58982-58994, 2020.

[9] H. Younis, B. Sundarakani and M. Alsharairi, "Applications of artificial intelligence and machine learning within supply chains: systematic review and future research directions," *Journal of Modelling in Management*, vol. 17, no. 3, pp. 916-940, 2022.

[10] X.-R. Luo, "A detailed examination of Sphicas (2014), generalized EOQ formula using a new parameter: Coefficient of backorder attractiveness," *Symmetry*, vol. 11, no. 7, no. 931, 2019.

[11] F. Xu, H. Zhao, W. Zhou and Y. Zhou, "Cost-sensitive regression learning on small dataset through intra-cluster product favoured feature selection," *Connection Science*, vol. 34, no. 1, pp. 104-123, 2022.

[12] R. W. Tatko, C. R. White, G. M. Williams and R. K. Myers, "Backorder Prediction," *IBM Corporation*, 2021.

[13] M. Z. Babai, H. Chen, A. A. Syntetos and D. Lengu, "A compound-Poisson Bayesian approach for spare parts inventory forecasting," *International Journal of Production Economics*, vol. 232, p. 107954, 2021.

[14] Y. Li, "Backorder prediction using machine learning for Danish craft beer breweries," PhD dissertation, Aalborg University, 2017.

[15] F. Ahmed, M. Hasan, MS. Hossain and K. Andersson, "Comparative Performance of Tree Based Machine Learning Classifiers in Product Backorder Prediction," in *Intelligent Computing & Optimization: Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022)* , pp. 572-584, 2022.

[16] T. M. Choi, S. W. Wallace and Y. Wang, "Big data analytics in operations management," *Production and Operations Management*, vol. 27, no. 10, pp. 1868-1883, 2018.

[17] A. L'Heureux, K. Grolinger, H. F. Elyamany and M. A. M. Capretz, "Machine Learning With Big Data: Challenges and Approaches," in *IEEE Access*, vol. 5, pp. 7776-7797, 2017.

[18] G. Baryannis, S. Validi, S. Dani and G. Antoniou, "Supply chain risk management and artificial intelligence: state of the art and future research directions," *International Journal of Production Research*, vol. 57, no. 7, pp. 2179-2202, 2019.

[19] B. Kaya and G. Ulutagay, "Inventory and Maintenance Optimization of Conditional Based Maintenance Using Fuzzy Inference System," in *Industrial Engineering in the Covid-19 Era: Selected Papers from the Hybrid Global Joint Conference on Industrial Engineering and Its Application Areas, GJCIE 2022*, pp. 188-199, 2023.

[20] W. Rosenberg-Vitorica, T. E. Salais-Fierro, J. A. Marmolejo-Saucedo and R. Rodriguez-Aguilar, "Machine Learning Applications in the Supply Chain, a Literature Review," *Smart Applications with Advanced Machine Learning and Human-Centred Problem Design*, pp. 753-761, 2023.

[21] K. L. Yung, G. T. S. Ho, Y. M. Tang and W. H. Ip, "Inventory classification system in space mission component replenishment using multi-attribute fuzzy ABC classification," *Industrial Management & Data Systems*, vol. 121, no. 3, pp. 637-656, 2021.

[22] Md. H. Rahman, Md. A. Rahman and S. Talapatra, "The bullwhip effect: causes, intensity, and mitigation," *Production & Manufacturing Research*, vol. 8, no. 1, pp. 406-426, 2020.

[23] Z. Ashraf and M. Shahid, "Multi-objective vendor managed inventory system with interval type-2 fuzzy demand and order quantities," *International Journal of Intelligent Computing and Cybernetics*, vol. 14, no. 3, pp. 439-466, 2021.

[24] A. Aksoy, N. Ozturk and E. Sucky, "A decision support system for demand forecasting in the clothing industry," *International Journal of Clothing Science and Technology*, vol. 24, no. 4, pp. 221-236, 2012.

[25] M. Díaz-Madroñero, J. Mula and M. Jiménez, "Fuzzy goal programming for material requirements planning under uncertainty and integrity conditions," *International Journal of production research*, vol. 52, no. 23, pp. 6971-6988, 2014.

[26] G. Saraogi, D. Gupta, L. Sharma and A. Rana, "An un-supervised approach for backorder prediction using deep autoencoder," *Recent Advances in Computer Science and Communications*, vol. 14, no. 2, pp. 500-511. 2021.

[27] Y. Xu, A. Bisi and M. Dada, "A finite-horizon inventory system with partial backorders and inventory holdback," *Operations Research Letters*, vol. 45, no. 4, pp. 315-322, 2017.

[28] G. Canbek, T. Taskaya Temizel and S. Sagiroglu, "PToPI: A comprehensive review, analysis, and knowledge representation of binary classification performance measures/metrics," *SN Computer Science*, vol. 4, no. 13, 2022.

[29] D. Chicco and G. Jurman, "The advantages of

the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, no. 1, pp. 1-13, 2020.

[30] I. Lee and Y. J. Shin. "Machine learning for enterprises: Applications, algorithm selection, and challenges," *Business Horizons*, vol. 63, no. 2, pp. 157-170, 2020.

[31] H. Younis, B. Sundarakani and M. Alsharairi, "Applications of artificial intelligence and machine learning within supply chains: systematic review and future research directions," *Journal of Modelling in Management*, vol. 17, no. 3, pp. 916-940, 2022.

[32] F.-M. Jaenichen, C. J. Liepold, A. Ismail, M. Schiffer and H. Ehm, "Disruption evaluation in end-to-end semiconductor supply chains via interpretable machine learning," *IFAC-PapersoOline*, vol. 55, no. 10, pp. 661-666, 2022.

[33] G. Rebala, A. Ravi, S. Churiwala, G. Rebala, A. Ravi and S. Churiwala, "Machine Learning Definition and Basics," in *An Introduction to Machine Learning*, 2019, pp. 1-17.

Sarit Maitra is currently a professor in the Department of Information Systems and Operations Management at Alliance University's Business School. His primary focus has been on data-driven business decision making, and he has worked globally in the big data, BI, AI, ML, and mathematical optimization models etc. in various industries. Sarit's educational background includes a Ph.D. and M.S. in information technology from the Universiti Teknologi Petronas, Malaysia. Passionate about advancing AI and ML approaches, Sarit focuses on enhancing forecasting models and devising strategic roadmaps that leverage data and analytics to inform business decisions and drive monetization. His research interests involve optimization models, AI and ML approaches for outlier detection, and predictive modeling.

Sukanya Kundu is an Associate Professor at Alliance School of Business, Alliance University, Bangalore, in the area of Information Systems and Operations Management. She holds a Ph.D. degree in Management, Masters of Computer Applications and has completed a Post-graduate Diploma in Business Management. She has a bachelor's degree in Economics from the University of Calcutta. She worked as a Software Engineer in mainframe technology and was involved in maintenance and testing projects in the financial domain. She has great interest in operations management, business process models, reengineering of the business processes, digital transformation etc. She takes great interest in research related to recent advances in information systems and technologies like AI, IoT, and Blockchain. She is interested in inter-disciplinary research, especially in technology applications in banking, e-commerce, education, customer experience, etc..