Leveraging Machine Learning Classifiers for Backorder Prediction: A Comprehensive Framework for Enhancing Supply Chain Efficiency and Inventory Management Addressing Class Imbalance issue in Backorder Prediction

Amin Sepehrnia

A Thesis
In

The Department

Of

Supply Chain & Business Technology Management

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Supply Chain Management at

Concordia University
Montreal, Quebec, Canada

January 2025

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared
By: Amin Sepehrnia

Entitled: Leveraging Machine Learning Classifiers for Backorder Prediction: A Comprehensive Framework for Enhancing Supply Chain Efficiency and Inventory Management - Addressing Class Imbalance issue in Backorder Prediction

and submitted in partial fulfillment of the requirements for the degree of
Master of Supply Chain Management

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Pr. Xiaodan Pan	Chair
Pr. Anton Shevchenko	Examiner
Pr. Danielle Morin	Examiner
Pr. Salim Lahmiri	Supervisor

Approved by: Dr. Satyaveer S. Chauhan, Graduate Program Director

Date: 31st January 2025

Dean: Dr. Anne-Marie Croteau

ABSTRACT

Leveraging Machine Learning Classifiers for Backorder Prediction: A Comprehensive Framework for Enhancing Supply Chain Efficiency and Inventory Management Addressing Class Imbalance issue in Backorder Prediction

Amin Sepehrnia

The current research explores the application of advanced machine learning and ensemble learning techniques to address the challenges of backorder prediction in supply chain management, specifically when the dataset is severely imbalanced. Considering the critical importance of accurate forecasting in supply chains, this study evaluates the performance of five resampling techniques (Random Under Sampling, ADASYN, SMOTE-ENN, Borderline-SMOTE, and SMOTE-SVM), combined with hyperparameter tuning (Randomized Search CV) and two cross-validation methods (5-fold and 10-fold). The research methodology involved training 98 combinations of two machine learning and five ensemble learning models, incorporating feature selection with SHAP and dimensionality reduction using PCA, alongside sophisticated data preprocessing techniques such as MICE for handling missing values. The primary evaluation metric is AUC-ROC, complemented by secondary metrics including balanced accuracy, F1 Score, and AUC-PR, ensuring a holistic assessment of model performance. Key findings demonstrate that ensemble learning models, particularly XGBoost, outperforms classical machine learning models in terms of robustness and being accurate in backorder prediction. Resampling techniques such as SMOTE-ENN and Random Under Sampling significantly enhance model performance, with SMOTE-ENN proving especially effective due to its noise reduction capabilities. Interestingly, dimensionality reduction using PCA was found to have little benefit, whereas feature selection using SHAP consistently improved efficiency and accuracy. The insights derived from this study provide a comprehensive framework for improving predictive performance in supply chain management applications, specifically backorder prediction. By addressing class imbalance, optimizing preprocessing techniques, and rigorously evaluating resampling methods, this research establishes best practices for tackling forecasting challenges in imbalanced, high-dimensional data environments.

Keywords: Supply Chain Management, Backorder Prediction, Machine Learning, Demand Forecasting, Inventory Management, Imbalanced Class

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof Salim Lahmiri, for his insightful guidance and encouragement throughout the course of my research. His expertise and constant feedback were instrumental in shaping this work, and I am truly grateful for their mentorship.

And a special thank you to the examiner committee, Prof. Anton Shevchenko, Prof. Danielle Morin, and Prof. Xiaodan Pan, for their valuable time, constructive criticism, and support during my thesis defense. Their expertise and insightful comments have greatly enhanced the quality of this research.

I am also profoundly grateful to all the professors in the MSCM program over the past two years. Their dedication, knowledge, and passion for teaching have significantly contributed to my academic growth and success.

Thank you to all who have contributed, directly or indirectly, to the success of this research.

Table of Contents

Table of Contents v
List of Figures vii
List of Tables viii
Acronyms ix
1. Introduction 1
1.1. Supply Chain Management, Backorder Prediction and Machin Learning Classifiers 1
1.2. Background 5
1.2.1. Supply Chain and Demand Forecasting 5
1.2.2. Inventory Management 6
1.2.3. Backorder 7
1.2.4. Imbalanced Dataset 8
1.2.5. Bullwhip Effect 9
1.2.6. Artificial Intelligence and Machine Learning 10
1.3. The main objective and motivation 11
1.4. The main research questions 13
1.5. The contributions of this research 13
1.6. Theoretical Framework and Relevance 14
1.7. Thesis Structure 15
2. Literature Review 16
2.1. The Role of Machine Learning in Addressing Demand Uncertainty and Backorder prediction 16
2.2. Machine Learning Approaches to Backorder Prediction in Supply Chain Management 17
2.3. Review of Related Research on Backorder Prediction and Supply Chain Management 18
2.4. Imbalance Class 29
3. Dataset 31
4. Research Methodology 35
4.1. Data Preprocessing and Handling Missing Values 37
4.1.1. MICE: Multiple Imputation by Chained Equations 37
4.2. Addressing Imbalanced Data 38
4.2.1. SMOTE-ENN: Synthetic Minority Over-Sampling Technique - Edited Nearest Neighbor 38
4.2.2. SMOTE-SVM: Synthetic Minority Over-Sampling Technique-Support Vector Machine 40

4.2.3. Borderline-SMOTE: Borderline-Synthetic Minority Over-Sampling Technique (BSM) 40
4.2.4. ADASYN: Adaptive Synthetic Oversampling Technique 40
4.2.5. Random Under Sampling (RUS) 42
4.3. Feature Selection Using SHAP (SHapley Additive ExPlanations) 42
4.4. Feature Engineering and Dimensionality Reduction 44
4.5. Machine Learning Models 45
4.5.1. Neural Networks 45
4.5.2. K Nearest Neighbours 48
4.6. Ensemble Learning 48
4.6.1. Random Forest 49
4.6.2. Gradient Boosting 49
4.6.3. XGBoost Extreme Gradient Boosting 50
4.6.4. LightGBM 51
4.6.5. Stacking 51
4.7. Overview of hyperparameter tuning: RandomizedSearchCV and GridSearchCV 52
4.8. K-Fold Cross Validation 53
4.9. Leave-One-Out Cross-Validation (LOOCV) 54
5. Evaluation Metrics 54
6. Results Analysis 55
6.1. Comparison of Results from Resampling Techniques 56
6.2. Analysis and Comparison of Resampling Techniques Combined with SHAP and PCA 57
6.3. Adjusted Feature Selection 64
6.4. Evaluation Metrics Analysis 64
7. Discussion of the results 75
8. Conclusion 79
8.1. Remarks 79
8.2. Managerial Implications 80
8.2.1. Focus on the Most Important Variables for Backorder Prediction 80
8.2.2. Choose the Right Machine Learning Model 81
8.2.3. Implementing and using machine learning models in supply chain operations 81
8.3. Limitation 82
8.4. Future research 82
References 84

List of Figures

Figure 1.1: The role of Machine Learning classifiers in Supply Chain Management 4
Figure 3.1: Class Distribution (Percentage%) - Backorder vs non-backorder - A comparison between original dataset and stratified dataset 33
Figure 3.2: Class Distribution (count) - Backorder vs non-backorder - A comparison between original dataset and stratified dataset 33
Figure 3.3: Correlation matrix heatmap 34
Figure 4.1: The methodology flow chart of the thesis 36
Figure 4.2: General Neural Network structure (Shaik et al, 2021) 46
Figure 4.3: The basic function of a neural network (Shaik et al., 2021) 46
Figure 4.4: General layout of FNN (Li and Zhuang, 2020) 48
Figure 6.1: Distribution of SHAP values for individual features 58
Figure 6.2: SHAP Feature Importance Bar Chart (ADASYN) 59
Figure 6.3: SHAP Feature Importance Bar Chart (Borderline-SMOTE) 59
Figure 6.4: SHAP Feature Importance Bar Chart (SMOTE-ENN) 60
Figure 6.5: SHAP Feature Importance Bar Chart (SMOTE-SVM) 60
Figure 6.6: SHAP Feature Importance Bar Chart (RUS) 61
Figure 6.7: Transformed Data Scatter Plot (ADASYN) 61
Figure 6.8: Transformed Data Scatter Plot (Borderline-SMOTE) 62
Figure 6.9: Transformed Data Scatter Plot (SMOTE-SVM) 62
Figure 6.10: Transformed Data Scatter Plot (SMOTE-ENN) 63
Figure 6.11: Transformed Data Scatter Plot (RUS) 63
Figure 6.12: ROC Curve - SMOTEENN 10-Fold Cross Validation 68
Figure 6.13: ROC Curve - SMOTEENN 5-Fold Cross Validation 68
Figure 6.14: ROC Curve - SMOTE-SVM 10-Fold Cross Validation 69
Figure 6.15: ROC Curve - SMOTE-SVM 5-Fold Cross Validation 69
Figure 6.16: ROC Curve - ADASYN-10 Fold Cross Validation 70
Figure 6.17: ROC Curve - ADASYN– 5-Fold Cross Validation 70
Figure 6.18: ROC Curve - Borderline-SMOTE-10 Fold Cross Validation 71
Figure 6.19: ROC Curve - Borderline-SMOTE 5-Fold Cross Validation 71
Figure 6.20: ROC Curve - Random Under Sampling-10 Fold Cross Validation (Without PCA) 72
Figure 6.21: ROC Curve - Random Under Sampling-10 Fold Cross Validation 72
Figure 6.22: ROC Curve - ADASYN -10-Fold Cross Validation (Without PCA) 73
Figure 6.23: ROC Curve - SMOTEENN -10-Fold Cross Validation (Without PCA) 73
Figure 6.24: ROC Curve - Borderline-SMOTEENN -10-Fold Cross Validation (Without PCA) 74
Figure 6.25: ROC Curve - SMOTESVM -10-Fold Cross Validation (Without PCA) 74

List of Tables

Table 2-1 Previous studies focused on backorder prediction by using machine learning techniques 28
Table 3-1 The dataset details 32
Table 3-2 Variables description 32
Table 3-3: Summary statistics of dataset 34
Table 6-1: Comparison of Results from Resampling Techniques 56
Table 6-2: Comprehensive Evaluation Metrics of Model Training: Integration of Resampling Techniques, ML and Ensemble Methods, Feature Selection (SHAP and PCA), and 5-Fold CV 65
Table 6-3: Comprehensive Evaluation Metrics of Model Training: Integration of Resampling Techniques, ML and Ensemble Methods, Feature Selection (SHAP and PCA), and 10-Fold CV 66
Table 6-4: Comprehensive Evaluation Metrics of Model Training: Integration of Resampling Techniques, ML and Ensemble Methods, Feature Selection (SHAP), and 10-Fold CV 67

Acronyms

AAC - Average Accuracy
ADASYN – Adaptive Synthetic Sampling
ALT – Administrative Lead Time
AUC – Area Under the Curve
BBC – Balanced Bagging Classifier
BSM - Borderline-SMOTE
CART - Classification and Regression Trees
CBUS – Clustering-Based Under Sampling
CNN – Convolutional Neural Network
CV – Cross-Validation
CI - Class imbalance
DNN – Deep Neural Network
ENN – Edited Nearest Neighbors
ERP – Enterprise Resource Planning
F1-Score – Harmonic Mean of Precision and Recall
FPR – False Positive Rate
GBM – Gradient Boosting Machine
LightGBM - Light Gradient Boosting Machine
KNN – K-Nearest Neighbors
MCC – Matthews Correlation Coefficient
MICE – Multiple Imputation by Chained Equations
ML – Machine Learning
MLP – Multilayer Perceptron
NN – Neural Networks
PCA – Principal Component Analysis
PLT – Production Lead Time
PR-AUC – Precision-Recall Area Under the Curve
RF – Random Forest
DFS - Deep Feature Synthesis
RFE – Recursive Feature Elimination
RNN – Recurrent Neural Network
ROC – Receiver Operating Characteristic
ROC-AUC – Receiver Operating Characteristic Area Under the Curve
RUS – Random Under Sampling
SHAP – Shapley Additive Explanations
SMOTE – Synthetic Minority Oversampling Technique
SMOTE-ENN – SMOTE with Edited Nearest Neighbors
SMOTE-SVM: Synthetic Minority Oversampling Technique - Support Vector Machine
SCM - Supply Chain Management
SVM – Support Vector Machine
SVR - Support Vector Regression
TPR – True Positive Rate
VAE – Variational Autoencoder
XAI – Explainable Artificial Intelligence
XGB – Extreme Gradient Boosting

1. Introduction

This research examines the interaction between supply chain management, backorder prediction, and machine learning classifiers to address crucial challenges in modern supply chains. A total of 98 models, resulting from the combination of multiple resampling strategies, feature engineering, machine learning and ensemble learning have been trained on a severely imbalanced dataset. The introduction investigates the foundational aspects of supply chain operations, including demand forecasting, inventory management, and the impacts of backorders. It also explores the implications of imbalanced datasets, the bullwhip effect, and the integration of artificial intelligence and machine learning in improving operational efficiencies. Key objectives and motivations are highlighted, alongside research questions and contributions, to underscore the significance and relevance of this study. The theoretical framework and its practical implications provide a solid foundation, followed by an outline of the thesis structure to guide the discourse.

1.1. Supply Chain Management, Backorder Prediction and Machin Learning Classifiers

In the supply chain management, inventory planning often depends on accurate demand assessment, which can be complex due to the interdependence of manipulated variables and their outcomes. On the one hand, effective demand forecasting serves as a proxy for future sales estimates, making it essential for aligning supply levels with anticipated demand and ensuring smooth inventory flow. On the other hand, handling backordered products is a frequently examined subject in supply chain management and inventory planning, as it plays a critical role in boosting company profitability (Wang and Tang, 2014; Rodger, 2014; Adana, Cevikparmak, Celik, and Uvet, 2019; de Santis, Aguiar, and Goliatt, 2017; Islam and Amin, 2020; Ntakolia, Kokkotis, Moustakidis, and Papageorgiou, 2022; Iqbal, Rosenberger, Ha, Gregory, and Anoruo, 2023; Lawal and Akintola, 2021; Hajek and Abedin, 2020; Shajalal, Boden, and Stevens, 2022, 2023; Maitra and Kundu, 2023; Gao, Ren, and Lv, 2022; Lopez, Panduro, and Pumayauri, 2022; Ahmed, Hasan, Hossain, and Andersson, 2022; Ali, Jayaraman, Azar, and Maalouf, 2024). These studies have had a specific focus on examining and analyzing the critical role of classification techniques in SCM. The above-mentioned studies demonstrated the effectiveness of Machine Learning (ML) classifiers in inventory management and backorder prediction, producing promising results. In addition, numerous studies demonstrate the effective application of machine learning to enhance solutions within SCM. ML algorithms offer powerful predictive capabilities across various supply chain metrics, including profit margins, operational costs, credit assessments, backordered items, and market demand (Wang and Zhang, 2020 and Abbas et al., 2020). Moreover, machine learning techniques, through nonlinear analysis, enhance the predictive accuracy of demand, inventory requirements and sales, offering valuable insights to manage inventory precisely (Aguilar-Palacios et al., 2020).

There are critical challenges, such as forecasting demand, managing backorders, and replenishing inventory effectively, which are increasingly complicated and exacerbated by the growing complexity and disruptions within global supply chains (Tang and Ge, 2021). Modern supply chains, which involve various suppliers, numerous stakeholders, and intricate distribution channels and logistics networks, face significant challenges in maintaining balanced inventory levels. (Shajalal et al., 2022). Disruptions such as geopolitical tensions, natural disasters, and global crises like the COVID-19 pandemic can cause substantial delays and material shortages, further complicating inventory planning (Liu and Wang, 2007). Additionally, the dynamic and unpredictable nature of customer demand, influenced by market trends, seasonal fluctuations, and economic shifts, adds to the difficulty of demand forecasting (Carbonneau et al., 2008).

Backorder management is a critical metric in inventory analysis and supply chain management. Backorders occur when an order is accepted despite the supplier not having the product in stock or the manufacturer not yet producing it. Common causes of backorders include delays in order placement, warehouse discrepancies, human errors, manufacturer shortages, incorrect reorder points, and unexpected demand (Banik et al., 2023). Backorders have been considered as one of the causes of the bullwhip effect in SC performance (Pillai and Chinna Pamulety, 2013).

The balance between demand and supply can be disrupted by the bullwhip effect, leading to either excess inventory or stockouts, both of which directly harm a company's profitability (Iqbal et al., 2023). In backorder prediction "Accuracy" plays a crucial role for reducing the costs related to production and also improving inventory services. While precise predictions are required for effective inventory control, modelers must consider not only decreasing prediction errors but also assessing the economic benefits of these predictions (Hajek and Abedin, 2020) Additionally, incorrect predictions of material backorders can adversely affect inventory management and production systems. Thus, reliable material predictions are vital for minimizing the risk of inventory backorders (Hajek and Abedin, 2020). Accurate forecasting of client backorders is critical in sectors that rely on excellent inventory management and supply chain operations for profitability. It enables businesses to optimize inventory levels, reduce operational expenses, and improve customer happiness. Backorder prediction in the supply chain comprises creating algorithms based on past data to anticipate the likelihood of a product running out of stock as well as the expected time it will take to fulfill back orders. This information is crucial for optimizing inventory levels, adjusting production schedules, and managing customer expectations across the supply chain. In improving the accuracy of backorder predictions, many factors are involved such as supplier reliability, demand fluctuations, lead times, and market trends which must be carefully considered precisely (Ali et al., 2024).

In backorder prediction numerous significant challenges are involved like the uneven distribution of data across classes, commonly referred to as class imbalance. Simply put, this occurs when one class contains a disproportionately higher number of instances, such as having substantially more non-backordered products compared to backordered ones (de Santis, Aguiar, and Goliatt, 2017; Hajek and Abedin, 2020; Ntakolia, Kokkotis, Moustakidis, and Papageorgiou, 2022; Islam and Amin, 2020; Lawal and Akintola, 2021; Shajalal, Boden, and Stevens, 2022, 2023; Maitra and Kundu, 2023; ; Iqbal, Rosenberger, Ha, Gregory, and Anoruo, 2023). In the recent years, to deal with this issue, a number of techniques have been considered as solutions such as resampling techniques and machine learning models, particularly ensemble learning. Examining and analyzing the critical role of classification techniques in SCM has been one of the most important topics in this field in recent years. Previous research has focused on various aspects of this relationship, providing foundational insights and practical applications of classification prediction within supply chains, from inventory management to demand forecasting and backorder management. For example, Hajek and Abedin (2020), in their paper focus on integrating predictive classification techniques to maximize the profitability of inventory management by addressing backorders. The authors discuss how accurate prediction models allow for better decision-making in supply chain management by classifying high-risk stock items and optimizing order levels accordingly. In one of the most recent studies, authors emphasize the need for predictive classification to anticipate backorders and manage supply chain disruptions (Ali et al., 2024). They explore how accurate classification models help in stock optimization, thereby enhancing customer satisfaction and reducing operational costs. Lawal and Akintola (2021), investigate RNN-based classification models for predicting backorders in supply chains. They focus on the importance of handling imbalanced data in backorder prediction, emphasizing how accurate classifications improve overall inventory and supply chain efficiency. Shajalal et al., (2022) emphasize the importance of interpretability in machine learning classifiers for backorder prediction, particularly when understanding the sources of backorders in complicated supply chains. By introducing CNN models, the study illustrates the role of classification accuracy in ensuring responsive and adaptive inventory management. Carbonneau et al., (2008) in their study highlights the importance of using classification in managing supply chain complexities and mitigating the bullwhip effect. This work compares traditional forecasting methods with machine learning models, including classification approaches such as SVM and neural networks, for demand prediction. Hong and Ping (2007), provide a detailed examination of the bullwhip effect within supply chains, underscoring the role of accurate demand classification and prediction to prevent inventory mismatches. This foundational article highlights how accurate classification impacts inventory control, planning, scheduling and whole supply chain performance. In another research, tang and Ge (2021) in their research demonstrated the application of Convolutional Neural Networks (CNN) for material demand forecasting. By optimizing predictive classifications, the study shows how machine learning models improve demand accuracy, crucial for effective inventory management and timely fulfillment of orders. Integrating machine learning classification models can significantly enhance supply chain performance, increase customer satisfaction, and reduce costs, as shown in these studies. Traditional models often struggle to handle this volatility, leading to either stock shortages or excess inventory—both of which negatively affect operational efficiency and customer satisfaction (Ahmed et al., 2022) Among these challenges, backorders play a significant role, as they arise when customer demand exceeds available inventory (Lawal and Akintola, 2021). Accurately predicting backorders is essential for businesses aiming to optimize inventory management and minimize the disruptions caused by stock shortages (Ali et al., 2024). Precise forecasting would enable companies and manufacturers to manage inventory more effectively, adjust their production schedules, and determine optimal order quantities and timings (Maitra and Kundu, 2023). This, in turn, helps businesses mitigate risks associated with delayed supplies, minimize financial losses, and maintain customer trust through timely responses to demand fluctuations (Shajalal et al., 2023). Quick and accurate backorder management not only prevents customer dissatisfaction and canceled orders but also reduces the broader operational risks posed by supply chain disruptions (Hajek and Abedin, 2020). To address these complexities, the adoption of machine learning (ML) techniques has become crucial in enhancing the precision of backorder predictions (Dehghan-Bonari et al., 2021). ML models, capable of analyzing large and complex datasets, offer significant improvements in supply chain forecasting by considering key factors such as demand variability, inventory levels, sales history, lead times, and production workflows (Adana et al., 2019). These models allow companies to proactively manage backorder risks, maintaining balanced inventory levels, optimize decision-making, and improve overall SC performance (Santis et al., 2017). Methods such as supervised learning, ensemble classifiers, and explainability tools enable businesses to refine backorder predictions, identify critical factors driving stock shortages, and evaluate both traditional and modern forecasting approaches (Ntakolia et al., 2021; Shajalal et al., 2022). Ultimately, the use of ML not only enhances operational efficiency but also supports revenue growth by providing actionable insights into inventory management and customer service strategies (Ali et al., 2024). The goal of this research is to construct and assess a variety of machine learning models, such as Neural Networks, KNN, and sophisticated ensemble approaches, in order to determine the most effective and precise model for backorder prediction inside supply chains.

In summary, the integration of ML classifiers into supply chain and inventory management addresses critical challenges, including demand forecasting, backorder prediction, and the bullwhip effect, by leveraging advanced predictive capabilities and data-driven insights. These classifiers, when paired with novel preprocessing, feature selection, and ensemble learning methods, improve operational efficiency, customer happiness, and overall decision-making. Figure 1.1. illustrates the central role of machine learning classifiers in supply chain management, depicting their connections to key processes, techniques, and outcomes, thereby providing a comprehensive framework for understanding their impact on addressing supply chain complexities.

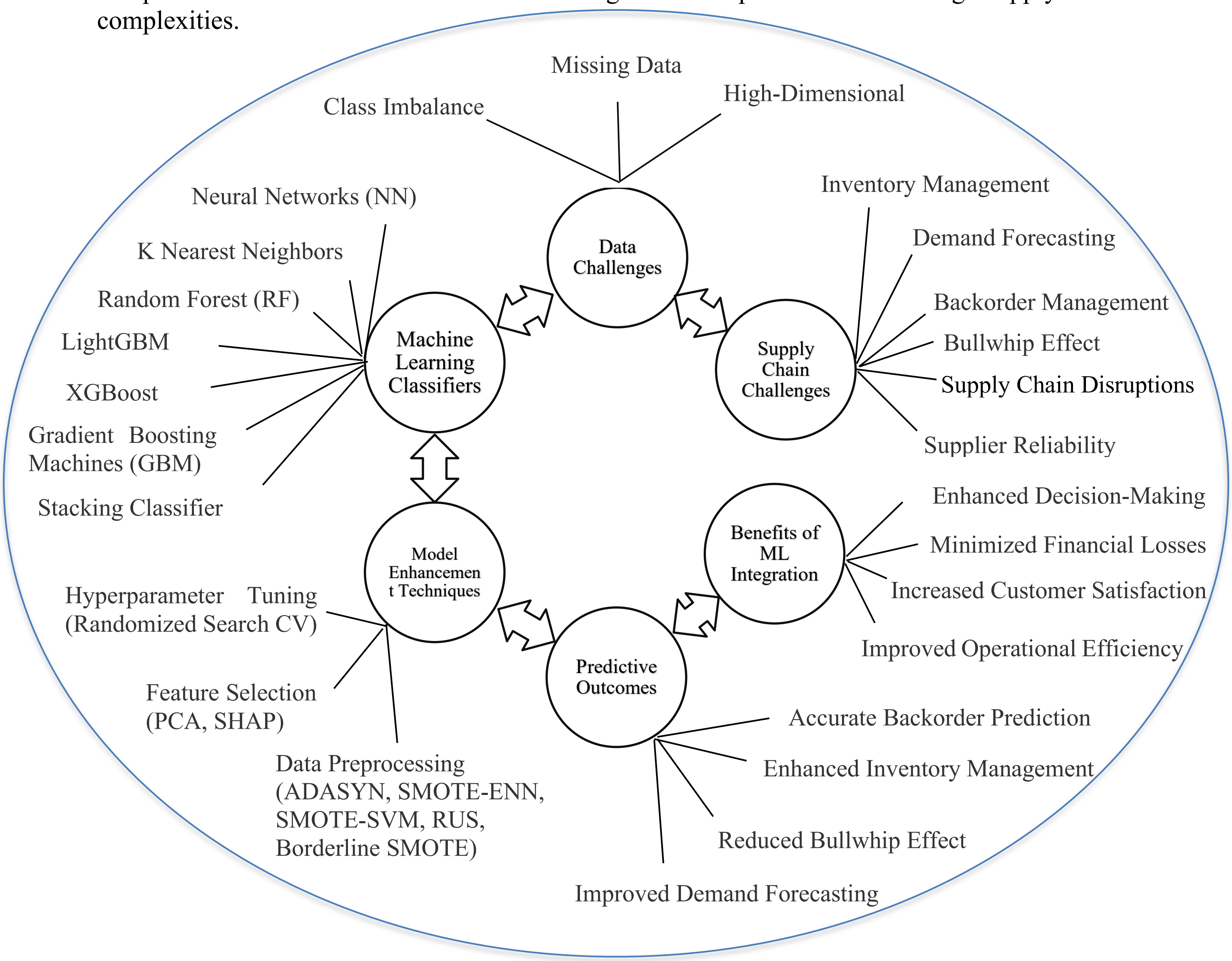


Figure 1.1: The role of Machine Learning classifiers in Supply Chain Management

1.2. Background

The existing body of research on predicting backorders in SCM emphasizes various advanced machine learning (ML) techniques, addressing the challenges of class imbalance, prediction accuracy, and interpretability. A common focus across many studies is improving predictive accuracy by leveraging ML models, particularly in dealing with the complex, non-linear relationships inherent in supply chain data. For instance, Adana et al. (2019) and Carbonneau et al. (2008) demonstrated that neural networks (NNs) and support vector machines (SVMs) outperform traditional logistic regression models in backorder forecasting. Similarly, Ahmed et al. (2022) evaluated multiple ML algorithms to assess their effectiveness in predicting backorders, with a particular emphasis on handling imbalanced datasets and complex patterns in inventory data. Both Lawal and Akintola (2021) and Shajalal et al. (2023) focused on addressing the challenge of class imbalance by utilizing techniques such as SMOTE, ADASYN, and under-sampling to ensure more accurate backorder predictions. Several studies have also explored the interpretability of these models. For instance, Shajalal et al. (2022) highlighted the "black box" nature of deep learning models and proposed integrating explainable AI (XAI) techniques to enhance transparency, a concern shared by Rodger (2014), who emphasized the complexity of backorder aging prediction in supply chains. Other research, such as that by Iqbal et al. (2023), has adopted more interpretable methods like Classification and Regression Trees (CART), aiming to balance predictive performance with clarity in decision-making. Maitra and Kundu (2023) incorporated cost-sensitive learning approaches to minimize financial losses due to stockouts, an approach also explored by Hajek and Abedin (2020), who integrated profit-maximization into their model development. Furthermore, studies like Ali et al. (2024) and Santis et al. (2017) examine the trade-offs between model complexity and computational cost. Ali et al. (2024) showed that reducing the number of predictors could maintain acceptable accuracy while lowering computation time. In contrast, Garcia et al. (2022) targeted the practical implications of backorders within specific sectors, such as cross-docking in HomeCenter services, by optimizing fulfillment processes using ML models. Overall, this diverse range of approaches, from big data analytics (Hajek and Abedin, 2020) to novel architectures like convolutional neural networks (CNNs) (Shajalal et al., 2022), highlights a multi-faceted effort to improve supply chain resilience and predict backorders more accurately.

1.2.1. Supply Chain and Demand Forecasting

Demand forecasting plays a critical role in supply chain management, serving as a foundational input for key decision-making processes in retail, including inventory management, network planning, pricing strategies, and revenue management (Ge et al., 2019). As products, services, and data traverse organizational boundaries across diverse system platforms in the expanding global marketplace, supply chain management (SCM) faces significant challenges. At a broader scale, the global economic landscape and political dynamics contribute to disruptions within supply chains, generating uncertainties, delays in delivery schedules, and complications with regulatory compliance. The lasting effects of the COVID-19 pandemic serve as a clear example of these challenges. (Iqbal et al., 2023). In every sector of the supply chain, demand prediction is vital not only for profitability but also for ensuring that products are available in the right quantities when needed. This makes it a crucial element in planning and decision-making within SCM and enterprise operations. Accurate demand forecasting is essential for making informed decisions regarding capacity expansion, resource allocation, and strategies for both forward and backward integration (Adhikari, 2017). In today's extremely competitive economic world, organizations are increasingly relying on precision marketing to sustain or increase profit margins. Consequently, forecasting models have become integral to precision marketing by enabling businesses to better understand and meet customer needs and expectations (Seyedan and Mafakheri, 2020). In SCM, it is often assumed that factors such as capacity, demand, and costs are fixed and known. However, in reality, these factors are subject to significant uncertainties due to fluctuations in customer demand, supply chain disruptions, organizational risks, and lead times. Among these, demand uncertainty has the most profound impact on supply chain performance, affecting production scheduling, inventory management, and transportation. Thus, precise demand forecasting is crucial for mitigating these uncertainties and optimizing SCM operations (Seyedan and Mafakheri, 2020).

Businesses must forecast future demand to prepare for unpredictable needs and streamline inventory management costs. Given the lead time between ordering and delivery, companies often must order in bulk before actual demand materializes (Seyedan et al., 2023). The unpredictability of customer demand complicates demand forecasting, which can render traditional SCM systems less effective. This can lead to problems such as inaccurate demand predictions and misclassification of back-ordered products (Islam and Amin, 2020). Predictive models that forecast the likelihood of backorders for specific products enable companies to proactively optimize inventory management strategies, thereby mitigating stockout risks and facilitating better planning and decision-making (Shajalal et al., 2022).

1.2.2. Inventory Management

Supply chain management (SCM) encompasses the flow of goods, services, and information from origin points to customers, involving a network of interconnected entities and activities (Ge et al., 2019). Inventory management, a critical component of SCM, involves making strategic decisions about the quantity and placement of stocked goods within a multi-layered system that operates across various locations and facilities (Ge et al., 2019). It covers the entire process of inventory control, including the monitoring of goods as they move into and out of warehouses and distribution centers, and the reconciliation of inventory balances (Ge et al., 2019). Inventory control specifically supports operational decisions about when and how much to replenish for each stock-keeping unit (SKU), including parts and materials used in production (Goltsos and Syntetos, 2022). Effective inventory management plays a crucial role in maintaining a steady flow of raw materials and finished products, ensuring seamless business operations across both manufacturing and production activities. Effective inventory management helps maintain smooth operations by coordinating purchases, sales, and logistics activities (Oluwaseyi et al., 2017). The primary goals of inventory management systems are to ensure the seamless operation of production processes, minimize ordering costs, capitalize on quantity discounts, and avoid sales opportunity losses (Akintola and Lawal, 2021). As a key element of the supply chain, inventory management encompasses responsibilities such as overseeing and tracking purchases, managing stock storage, controlling available product quantities, and ensuring order fulfillment. To properly satisfy customer demand, organizations with complicated supply chains and manufacturing processes must balance inventory levels. (Banik et al., 2023). Decisions regarding the timing and quantities of SKUs and their associated materials and components are central to inventory management. The main objective is to meet customer demand at a specified service level (Seyedan et al., 2023). Retailers, including medium and small-sized businesses, must regularly manage a wide range of distinct items to minimize operational costs and optimize sales (Seaman, 2018). A crucial aspect of inventory management involves determining the need for and timing of placing orders for specific items, as well as the appropriate quantity for each order (Theodorou et al., 2023). Proper inventory management guarantees the timely and cost-efficient procurement of essential supplies from vendors, supporting both the production process and the delivery of finished goods. Inventory management is essential for meeting demand, enhancing profitability, and ensuring the smooth operation of a company, making it a critical function for maintaining competitiveness (Namir et al., 2021). Inventory management profoundly influences the overall performance of manufacturing and trading enterprises. It requires a delicate balance between storage costs and the potential losses from insufficient inventory levels and customer dissatisfaction. Optimal inventory management involves maintaining satisfactory service levels while avoiding excessive stockpiles that disproportionately increase storage costs (Sustrova, 2016). Inventory control has long been a significant concern in industrial engineering and operational research, focusing on decisions about when to place orders and how much to order through various control mechanisms (Tsou, 2008). On a micro level, managing inventory effectively involves addressing issues such as customer demand, transportation logistics, and technology platform communication, with product backorder being one of the most common challenges (Iqbal et al., 2023). Among the various types of inventory costs, stock-out costs are often the most significant, arising from fluctuating customer demand, forecast inaccuracies, and variability in lead times. To mitigate these risks, businesses may hold sufficient stock to cope with unexpected or excess demand, a strategy known as maintaining safety stock (Ge, 2019). Safety stock is crucial for addressing uncertainties in vendor lead-time predictions and is the foundation of broader inventory replenishment challenges. The core challenge in supply chain and inventory management lies in inventory replenishment planning, which involves determining order quantities (lot sizing) and establishing replenishment schedules (Boctor and Bolduc, 2015). The increasing complexity of global supply chains, with numerous stakeholders and diverse flows of goods and information, adds to the challenge of managing inventory effectively. Poorly managed inventory can lead to stock-outs, disrupting production and increasing costs, or to excess supplies, which inflate costs and disrupt cash flow (Namir et al., 2021). In global production systems, inventory management plays a pivotal role in balancing costs and service objectives while serving as a buffer against uncertainties in demand and supply (Chinelloa et al., 2020). Effective inventory management is crucial for any company seeking to remain competitive in today's business environment (Moshtagh and Taleizadeh, 2017). A critical aspect of inventory models is the prediction of backorders, which detects out-of-stock items and enables organizations to promptly replenish their inventory (Zhang et al., 2016).

1.2.3. Backorder

In inventory management and supply chains, a backorder occurs when a customer purchases an item that is currently unavailable or out of stock. Instead of selecting a substitute, the consumer opts to wait for the item to be restocked, resulting in delays in order fulfillment and delivery. (Maitra and Kundu, 2023; Ntakolia et al., 2022; Islam and Amin, 2020; Shajalal et al., 2022). Backorders are a frequent feature of inventory management systems, emphasizing the need for efficient handling of customer expectations and stock levels. They offer critical insights into consumer demand, inventory precision, and the overall performance of supply chains, enabling businesses to refine inventory management, enhance customer satisfaction, and minimize lost sales opportunities (Maitra and Kundu, 2023). A backorder is essentially a contingent order based on future inventory availability with flexible delivery timelines (Shajalal et al., 2022). Stockouts can elicit varied responses from customers; some choose to delay their purchase until the item becomes available, whereas others opt to cancel their order or seek alternatives from competing providers.

These reactions result in either postponed sales through backorders or revenue losses due to diverted purchases. (Wang and Tang, 2014). Predicting backordered products before customers place orders is crucial for regulating production to reduce lead times and enhance profitability (Hajek and Abedin, 2020). Over the years, inventory models have been designed to identify the ideal inventory levels needed for efficient production, regulate the timing and frequency of orders, establish appropriate quantities of goods or raw materials to maintain in stock, and monitor supply chains to guarantee seamless customer service without disruptions in delivery (Akintola and Lawal, 2021). Many customers tend to avoid companies with a history of frequent backorders, as it signals inefficiencies and can lead to lost sales (Chan et al., 2017). The manner in which a company handles backorders can significantly affect customer trust, satisfaction, and the company's overall market performance. Promptly addressing backorders can enhance a company's reputation, while delays can lead to dissatisfied customers, order cancellations, loss of revenue, and supply chain disruptions (Ntakolia et al., 2022; Islam and Amin, 2020; Santis et al., 2017). While some studies have assumed that excess demand results in either lost sales or backorders, more realistic scenarios involve partial backordering, where some excess demand leads to lost sales and the remainder results in backorders (Chen et al., 2015). Backordering offers organizations advantages such as maintaining customer retention, fostering a responsive supply chain, and supporting effective risk management. Conversely, ineffective backorder management can lead to increased expenses, including financial and operational costs related to procurement, production, and distribution, alongside intangible drawbacks like diminished customer satisfaction and the risk of losing customers to competitors (Akintola and Lawal, 2021). Backorders are challenging to predict due to various factors, including unusual customer demands, forecast complexity and inaccuracy, material or product shortages, logistics issues, and data inaccuracies (Iqbal et al., 2023). Although backorders may indicate strong demand for a product, they also pose the risk of losing customers if not managed properly (Iqbal et al., 2023). Optimizing inventory policies, such as review periods, lead times, and service levels, is crucial to managing the trade-off between maintaining product availability and minimizing inventory costs (Theodorou et al., 2023). Backorders often occur for highly demanded products, but for other items, demand can be less predictable. Large orders based on backorders can expose retailers to reputational risks if expected delivery dates are not met. Customers may cancel orders due to long wait times or purchase from competitors with available stock, leaving companies with excess inventory and potential financial losses (Shajalal et al., 2022). The handling of backorders affects a company's market position, customer satisfaction, and overall sales performance. Prompt responses to backorders enhance reputation, while delays can result in dissatisfied customers, lost revenue, and supply chain disruptions. Accurate backorder prediction is crucial for devising mitigation strategies and adjusting production processes accordingly (Ntakolia et al., 2022). If backorders are not addressed promptly, they can lead to lost customers, decreased revenue, and a decline in market share. Conversely, quickly fulfilling backorders can strain supply chain processes, increasing labor, production, and shipping costs (Islam and Amin, 2020). Stock-outs result in lost sales when no backorders are allowed, leading to reduced sales gains from higher product variety as consumers opt for alternatives in the market (Wan et al., 2020).

1.2.4. Imbalanced Dataset

Identifying backordered products in inventory management is particularly challenging because products are far more often available than backordered, leading to a common consequence: an imbalanced binary classification problem. Classifying a product as a backorder is difficult due to its infrequent occurrence compared to availability, creating an imbalanced binary classification issue (Ntakolia et al., 2022). Class imbalance (CI) in classification tasks occurs when one class contains significantly fewer observations compared to another (Khan et al., 2024, (Shoeibi et al., 2023; Kaur and Singh, 2023; Osama et al., 2023; Makki et al., 2023). In typical supply chain management scenarios, the occurrence of goods in backorder (positive or minority class) is relatively rare compared to the number of available or non-backorder items (negative or majority class). This situation is referred to as class imbalance. Such class-imbalanced scenarios frequently occur in various real-world predictive applications, including loan approval analysis, corporate bankruptcy forecasting, and credit card fraud detection (Kim and Hwang, 2022). Class imbalance is a prevalent challenge across various fields of study and research, including medical images for the diagnosis of brain diseases (Shoeibi et al., 2023), image segmentation and classification (Kaur and Singh, 2023), medical diagnostics (Zhu et al., 2018), disease diagnosis (Kim and Hwang, 2022) and prediction methods in biotechnology and medicine (Osama et al., 2023), fraud detection (Makki et al., 2023). In these scenarios, minority positive instances typically hold greater importance than the majority class examples. Addressing class imbalance primarily involves achieving an optimal balance between majority and minority instances to enhance the impact of the positive class. Furthermore, the differing costs of misclassifying backorder and non-backorder items introduce an additional complexity that must be effectively managed (Hajek and Abedin, 2020). Shoeibi et al. (2023), identified imbalanced data as a critical challenge in data fusion for diagnosing brain diseases. To address this issue, they proposed resampling techniques as a viable solution. Specifically, they implemented three strategies: under-sampling the majority class, over-sampling the minority class, and employing the Synthetic Minority Over-sampling Technique (SMOTE) to create artificial samples. These methods were effective in increasing the representation of the minority class, thereby improving data balance. Bader et al. (2019) emphasized data-level strategies to mitigate the effects of class imbalance by either generating additional minority class instances (oversampling) or removing some majority class instances (undersampling). These methods are typically applied during the data preprocessing stage and operate independently of subsequent learning algorithms (Kaur & Singh, 2023). Bader et al. (2019) emphasized the effectiveness of ensemble learning methods, which improve predictive accuracy by aggregating the outputs of several base classifiers into a unified decision. By combining multiple models, ensemble learning provides a more resilient solution and has been extensively utilized alongside data augmentation strategies to address challenges associated with class imbalance (Shoeibi et al., 2023; Kaur and Singh, 2023; Osama et al., 2023; Makki et al., 2023).

1.2.5. Bullwhip Effect

The term "bullwhip effect" refers to a phenomenon in supply chain management where the variability of outgoing orders tends to be greater than the variability of incoming demands at each level of the supply chain (Gaalman et al., 2022). Carbonneau et al., (2008) in their research concluded that the bullwhip effect refers to the amplification of demand variability as information moves upstream in the supply chain, causing distortions in demand signals. This distortion often results in excessive inventory or stockouts, leading to inefficiencies in supply chain operations. An essential concept in supply chain management, the bullwhip effect, indicates that demand variability intensifies as it progresses up the supply chain. When a supply chain experiences the bullwhip effect, it can result in misguided capacity planning and missed production schedules due to a lack of visibility into product sales at the distribution channel stage. This phenomenon also contributes to numerous inefficiencies, including insufficient or excessive production capacities, excessive inventory investments, poor customer service due to unavailable products or extended backlogs, lost revenue opportunities, uncertainty in production planning (leading to frequent revisions), and increased correction costs (Hong and Ping, 2007). Backorders, as a key driver of supply chain disruptions, play a crucial role in amplifying the bullwhip effect by creating delays in fulfilling customer orders, which further complicates demand forecasting and inventory management across the supply chain. Supply shortages drive up costs for companies due to limited availability, leading to higher prices for goods and services. This impact is particularly evident in last-mile delivery, where escalating fuel expenses translate into increased shipping charges, further inflating the prices of products. Such price surges can impose financial strain on consumers and may hinder overall economic growth (Iqbal et al., 2023). The bullwhip effect, first identified by logistics executives at Procter and Gamble (P&G), refers to the amplification of demand or order variability as it moves from downstream to upstream stages in the supply chain (Lee et al., 1997). This dynamic can result in overstocked inventories, subpar customer service, revenue declines, inefficient transportation utilization, flawed capacity planning, disruptions to production schedules (Lee et al., 1997), diminished market share (Wright and Yuan, 2008), and avoidable costs, including those associated with stockouts or excess inventory obsolescence (Shukla et al., 2009), all of which negatively affect supply chain performance (Nienhaus et al., 2006). Backorders exacerbate the bullwhip effect by disrupting inventory, warehousing, production, and transportation processes across the supply chain. The bullwhip effect, a significant challenge in managing supply chain uncertainty, arises when minor variations in consumer demand trigger increasingly amplified fluctuations upstream, affecting wholesalers and manufacturers. Contributing factors include delays in information flow, order batching, price variability, and inventory management practices. These dynamics often result in imbalanced inventory levels and elevated carrying costs for organizations. Pillai and Pamulety (2013) examined the bullwhip effect as a performance measure in supply chains, specifically studying the impact of backorders under short lead times. Their research, conducted through experiments involving lost sales and backorder scenarios, demonstrated that backorders are a key factor contributing to the bullwhip effect. The traditional forecasting methods tend to struggle with mitigating the bullwhip effect, as they are not designed to handle the non-linear distortions caused by the phenomenon. Machine learning models, such as SVMs and neural networks, offer a potential solution by improving the accuracy of forecasting and helping to reduce the impact of the bullwhip effect, though their effectiveness depends on the specific supply chain context (Carbonneau et al., 2008). Zhang et al. (2021) utilized the Support Vector Machine (SVM) model to mitigate the bullwhip effect, a phenomenon of demand estimation inaccuracies. They trained a series of observations to classify outcomes based on assigned variables, emphasizing that larger datasets yield more stable results for this type of processing.

1.2.6. Artificial Intelligence and Machine Learning

Big data analytics provide competitive advantages by extracting valuable insights from vast databases, helping enterprises make informed business decisions, enhance strategies, improve operational efficiency, and boost supply chain sustainability and economic performance (Dubey, 2019). These applications also deepen understanding of enterprise dynamics, increase customer engagement, optimize routine operations, and generate new profit streams (Wang et al., 2016). These benefits have led to increasing attention on big data analytics within supply chain management (SCM) (Wang et al., 2016). As SCM focuses on satisfying customer demand while minimizing total supply costs, the application of machine learning algorithms facilitates precise, data-driven demand forecasts, aligning supply chain activities with these predictions to improve efficiency and satisfaction (Seyedan and Mafakheri, 2020). Although predicting demand accurately is challenging due to market uncertainties, the use of extensive historical data and big data analytics has improved the accuracy of demand forecasting (Seyedan et al., 2023). Advanced machine learning (ML) and deep neural network models have garnered widespread adoption in various industries, including business, healthcare, and bioinformatics, due to their superior predictive capabilities. These technologies have diverse applications, such as optimizing SCM, predicting credit risk, detecting credit card fraud, and informing retail banking strategies (Shajalal et al., 2022). The integration of artificial intelligence and big data into SCM has been transformative, especially as barriers to their implementation—such as costs, computing power, and access to open-source platforms—have diminished. Machine learning is now used to design and develop predictive models that evaluate all aspects of management, providing crucial insights that enable companies to respond effectively to operational changes (Santis et al., 2017). ML algorithms can comprehend non-linear processes by analyzing large datasets, making predictions or recommendations based on observed patterns rather than assumptions about data generation. These algorithms have shown outstanding performance across many SCM domains, including substituting statistical methods in demand forecasting, classifying inventory items, and predicting optimal transportation routes (Theodorou et al., 2023). Neural networks, one of the most complex AI processes, require significant processing power to collect and convert data into actionable insights quickly, aiding in prediction and decision-making (Papernot et al., 2017). A decision tree is a predictive tool that assists in forecasting outcomes based on a series of decisions. It autonomously selects variables from a dataset to create subdivisions, helping guide the decision-making process (Namazkhan et al., 2019). Support Vector Machines (SVM) are supervised ML algorithms used for classification and regression, analyzing binary variables and seeking maximum separation between observations. SVM is particularly suitable for complex, small to medium-sized datasets (Khan et al., 2021). SVM aims to draw a hyperplane in an "n" dimensional vector space to separate data into distinct patterns representing respective classes. Random forests, an ensemble ML technique, are highly adaptable to data and capable of identifying correlations and interactions between variables. Random forest models can be more effective than neural networks, especially when dealing with tabular data where variables are individually significant and lack temporal or spatial structure (Lundberg et al., 2020).

1.3. The main objective and motivation

The primary objective of this research is to develop a highly accurate and efficient machine learning-based predictive model to identify products or materials at risk of backorders within supply chain systems. By leveraging advanced machine learning techniques, this study aims to enhance backorder forecasting accuracy, enabling businesses to proactively manage inventory levels, mitigate the impacts of stock shortages, and optimize supply chain operations. This goal will be accomplished by implementing and comparing a total of 98 configurations consisting of various machine learning models, including K-Nearest Neighbors (KNN), Neural Networks, Random Forest (RF), ensemble methods and model stacking, and numerous resampling techniques to identify the most effective predictive approaches. As discussed in the background, one of the critical challenges in backorder prediction is data imbalance. This research addresses this issue through the use of data augmentation methods and ensemble learning techniques. These efforts are supported by innovative data preprocessing, feature selection, and dimensionality reduction methods, along with the application of a wide range of machine learning and ensemble learning techniques. Unlike some prior studies that opted to remove records with missing values (e.g., Ntakolia et al., 2022; Ali et al., 2024), this research acknowledges the critical value such records may hold. While other studies have explored missing value imputation methods (e.g., Santis et al., 2017; Hajek and Abedin, 2020), and Adana et al. (2019) employed the impute data node in SAS Enterprise Miner, this study emphasizes the importance of more sophisticated approaches. For instance, Iqbal et al. (2023) assumed missing data to be random, using linear and logistic regression for imputation, while Maitra and Kundu (2023) implemented model-based imputation. However, traditional methods, such as mean, mode, or median replacements proposed by Gao et al. (2022), can result in information loss and reduced model accuracy. To address these issues, this research adopts advanced imputation techniques, such as Multiple Imputation by Chained Equations (MICE), ensuring valuable insights from incomplete data are preserved. Additionally, model explainability is prioritized through SHAP analysis, offering transparent insights into feature contributions and enhancing decision-making processes. By focusing on a robust and scalable solution, this study aims to improve decision-making, reduce operational costs, enhance supply chain efficiency, minimize risks of stockouts or excess inventory, and elevate customer satisfaction. Building on this foundation, this research leverages SMOTEENN, an advanced hybrid technique combining SMOTE (Synthetic Minority Over-sampling Technique) with Edited Nearest Neighbors (ENN), representing a significant improvement over traditional resampling methods such as SMOTE or ADASYN. While prior studies have emphasized the need to address class imbalance (e.g., Ntakolia et al., 2022; Shajalal et al., 2022, 2023; Islam and Amin, 2020; Ali et al., 2024; Adana et al., 2019; Kaur and Singh, 2023), none have specifically explored the SMOTEENN approach alongside advanced machine learning algorithms for backorder prediction. Studies, including Santis et al. (2017) and Shajalal et al. (2023), have underscored the challenge posed by the class imbalance between backordered and non-backordered items, which impedes accurate predictions. Previous research has applied various strategies to mitigate this issue, such as Random Under Sampling, ADASYN, Weighted Samples, SMOTE, Oversampling, Random Down Sampling, and Stratified Holdout (e.g., Shajalal et al., 2023; Islam and Amin, 2020; Adana et al., 2019). By incorporating SMOTEENN, this study seeks to achieve a more optimal balance between precision and recall, particularly improving recall for the minority class representing backorders. The SMOTEENN technique's combination of oversampling (SMOTE) and noise reduction (ENN) results in a cleaner, more balanced dataset, significantly enhancing model accuracy in detecting backorders. This hybrid approach is especially effective in addressing the minority class of "Backorder," a persistent challenge in backorder prediction. By addressing this gap, the research improves the reliability of backorder predictions and establishes a robust framework for handling imbalanced datasets, contributing to more precise and actionable insights in supply chain management.

Previous studies, such as Adana et al. (2019) and Santis et al. (2017), have acknowledged the importance of feature selection but often relied on traditional or manual selection methods. In this research, we integrate Principal Component Analysis (PCA) into our feature engineering strategy to effectively manage high-dimensional data while preserving critical information for accurate predictions. PCA achieves dimensionality reduction by transforming the original features into uncorrelated principal components that retain the maximum variance of the dataset (Momeni et al., 2020). This process mitigates issues like multicollinearity, which can compromise the accuracy and stability of machine learning models and reduces the risk of overfitting. By applying PCA before model training, we enhance computational efficiency, particularly for resource-intensive algorithms such as Neural Networks. This step ensures faster training and improves the models' ability to generalize to new data. These advancements enable us to develop scalable and high-performing predictive solutions for backorder forecasting in complex supply chain systems. Through the integration of PCA, our research modernizes feature engineering practices, establishing it as a pivotal component of our strategy to create a precise, efficient, and reliable backorder prediction model.

Building on prior research, this study expands the evaluation of machine learning algorithms to identify the most effective model for predicting backorders. While earlier studies, such as Ali et al. (2024), have tested models like Random Forest, Gradient Boosting, and Neural Networks, our research introduces additional methods, including K-Nearest Neighbors (KNN), to enhance model diversity and robustness. This broader comparison aims to provide a comprehensive understanding of each model's performance, identifying their respective strengths and limitations within the context of supply chain data. Furthermore, this research advances the field by incorporating ensemble techniques, such as Gradient Boosting Machines (GBM), XGBoost, and LightGBM. These boosting methods combine the predictive strengths of multiple models, enabling the capture of intricate patterns and interactions within the dataset. Additionally, a key focus of this study is to leverage ensemble learning techniques specifically to address class imbalance issues, ensuring more reliable and balanced backorder predictions. Through this comprehensive evaluation and the application of ensemble methods, our research offers a novel and practical approach to predictive modeling in supply chain management.

1.4. The main research questions

The main research questions of this research are as follows:

Question 1: How do advanced machine learning models, such as Neural Networks (NN) and K-Nearest Neighbors (KNN), perform compared to traditional forecasting methods in predicting backorders in supply chain management, particularly when dealing with imbalanced data?

Question 2: Which ensemble learning techniques—such as XGBoost, LightGBM, and Stacking Models—are most effective in improving the accuracy and reliability of backorder prediction models?

Question 3: How do different data preprocessing techniques, such as resampling methods (e.g., SMOTEENN and Random Under Sampling) and advanced imputation methods, affect the performance of machine learning models in backorder prediction?

Question 4: Does feature selection (e.g., using SHAP) improve the accuracy of backorder prediction models, and why do dimensionality reduction techniques like PCA fail to deliver similar improvements?

1.5. The contributions of this research

Based on the identified gaps in the existing literature, the contribution of the current research can be summarized as follows:

This study aims to enhance the accuracy and robustness of machine learning models for backorder prediction within supply chain management by addressing several overlooked areas in existing research. First, this research systematically compares five key imbalanced data handling techniques, Random Under Sampling (RUS), ADASYN, SMOTE-ENN, SMOTESVM, and

Borderline-SMOTE, which are rarely assessed side-by-side. This comparison will provide valuable insights into their respective impacts on model performance, particularly within the context of backorder forecasting, where class imbalance is often a significant challenge. In addition, the current study uniquely incorporates an evaluation of different sampling strategies, specifically comparing 5-fold and 10-fold cross-validation methods. By assessing how these approaches influence model generalizability and predictive accuracy, the study will shed light on how sampling techniques affect the reliability of machine learning models in supply chain applications. Another significant contribution lies in the examination of hyperparameter optimization methods. While previous studies often neglect or inconsistently apply optimization methods, this research rigorously evaluates Randomized Search CV to determine the most efficient technique for optimizing machine learning models tailored to supply chain backorder prediction. Additionally, unlike previous research that often overlooks systematic approaches to handling missing values, this study employs MICE (Multiple Imputation by Chained Equations), a sophisticated technique particularly suited to supply chain datasets, where missing data can heavily impact model reliability and interpretability. A key focus of this research is resolving class imbalance issues not only through resampling techniques but also by leveraging ensemble learning methods such as Gradient Boosting Machines (GBM), XGBoost, and LightGBM. These ensemble techniques are specifically designed to handle skewed datasets while enhancing the robustness and predictive accuracy of backorder classification models. By implementing a suite of preprocessing, sampling, and optimization techniques and evaluating their influence on machine learning models, this research will provide a comprehensive framework for improving predictive performance in supply chain management applications. This work will also inform best practices for future studies aiming to tackle similar forecasting challenges within imbalanced, high-dimensional data environments.

1.6. Theoretical Framework and Relevance

This research is grounded in the theoretical framework of predictive analytics and machine learning theory, focusing on backorder prediction within supply chain management. By utilizing advanced machine learning techniques such as ensemble learning, and feature engineering, the study builds on principles of supervised learning and data-driven decision-making. The methodology integrates core concepts from data preprocessing, class imbalance handling, feature selection, and dimensionality reduction, establishing a comprehensive machine learning workflow. Techniques like Principal Component Analysis (PCA) and advanced imputation align with theoretical foundations in data transformation and dimensionality reduction, aiming to enhance model accuracy and interpretability. This research holds particular relevance for industries where supply chain management is essential, including manufacturing, retail, and e-commerce. Timely and accurate backorder prediction is crucial for optimizing inventory levels, reducing operational costs, and maintaining high levels of customer satisfaction. The study’s focus on addressing challenges such as data imbalances, missing values, and feature selection is particularly beneficial for businesses seeking to improve supply chain efficiency in a data-rich environment. This work not only identifies effective predictive models but also provides actionable strategies for optimizing inventory management and mitigating risks associated with backorders, thereby strengthening the resilience and competitiveness of contemporary supply chains.

1.7. Thesis Structure

This thesis is structured into eight chapters, each addressing key aspects of supply chain management, demand forecasting, and the integration of Machine Learning (ML) models. Chapter one introduces the background, research problem, objectives, contributions, and the significance of addressing inefficiencies in inventory management and backorder issues through predictive models. Chapter two provides a comprehensive review of the relevant literature, identifying the research gaps that this study aims to address. In chapter three, the dataset is described in detail. Chapter four discusses the research methodology, including data preprocessing, handling imbalanced data, feature selection, and the application of various ML models for prediction. Chapter five explains the evaluation metrics used in this research. Chapter six presents the results of the analysis, comparing model performances. This is followed by a thorough discussion in chapter seven, which interprets the findings and explores their practical implications. Finally, chapter eight concludes the research by summarizing key insights, addressing the study's limitations, and offering directions for future research in supply chain management, backorder prediction, and inventory planning and replenishment.

2. Literature Review

This chapter provides a comprehensive review of prior research on the integration of machine learning techniques in supply chain management, with a specific focus on addressing demand uncertainty and predicting backorders. The current chapter examines the role of machine learning in mitigating the challenges of demand fluctuations and imbalanced datasets, highlighting its potential to enhance backorder prediction accuracy and operational efficiency. Key aspects of previous studies, including the methodologies, machine learning classifiers, model enhancement techniques, and the challenges identified in earlier works, are analyzed. The discussion is organized into subsections that explore the contributions of machine learning to supply chain management, the specific applications of machine learning in backorder prediction, a critical review of related research, and the implications of data imbalance on predictive modeling.

2.1. The Role of Machine Learning in Addressing Demand Uncertainty and Backorder prediction

In today's rapidly evolving business landscape, supply chains are facing unprecedented levels of complexity, encompassing challenges in production planning, operations, inventory management, demand forecasting, and backorder fulfillment. The integration and globalization of supply chains have amplified this complexity, leading to a heightened focus on efficient supply chain management among both academic and industry stakeholders. This interest is driven by the escalating costs of manufacturing and transportation, coupled with the dynamics of global markets. These trends have led to the emergence of intricate and dynamic supply networks, which in turn have intensified and redistributed uncertainties throughout the supply chain. Consequently, companies are now required to allocate more resources to anticipate and manage uncertainties related to demand, supply, and internal operations, all of which are crucial for enhancing the sustainability of their supply chains. Notably, this surge in uncertainty is not solely due to external factors but is also attributed to the increasing complexity of supply chain structures and the diverse mechanisms employed within supply chain operations (Shin et al., 2012). In practice, demand uncertainties stem from fluctuations in customer demand, transportation issues, organizational risks, and variable lead times. Among these, demand fluctuations have the most profound impact on supply chain performance, influencing production scheduling, inventory planning, and transportation strategies (Seyedan and Mafakheri, 2020). Inventory planning, a critical aspect of supply chain management, involves making strategic decisions about when and how much to order, employing various control mechanisms to ensure efficiency (Santis et al., 2017). Backorders occur when customers order products that are temporarily unavailable. In such situations, companies must decide whether to manufacture or source the backordered items, while customers may opt to cancel their orders if the delay is too long, possibly leaving the company with excess inventory. As a result, making strategic inventory management decisions is essential, and incorporating AI-driven insights can improve these decision-making processes (Shajalal et al., 2022). A wide range of research has investigated the causes and effects of backorders on various aspects of supply chain management, production, and inventory control (Bao et al., 2018; Hajek and Abedin, 2020; ElHafsi et al., 2021; Umakanta et al., 2021; Thinakaran et al., 2019; Shin et al., 2012). Some studies have specifically focused on using machine learning models to predict backorders in inventory management systems (Islam and Amin, 2020; Hajek and Abedin, 2020; Shajalal et al., 2023; Santis et al., 2017). For instance, one study developed a machine learning pipeline incorporating explainability analysis to pinpoint key features in backorder prediction

(Ntakolia et al., 2022), while another used a supervised learning model with sampling techniques and classifier ensembles for improved backorder forecasting (Santis et al., 2017). Additionally, a study introduced a range-based approach to determine various predictive features, accounting for real-time data anomalies caused by human or machine errors, offering a flexible solution for predicting backorder scenarios (Islam and Amin, 2020). Machine learning models have been widely used to accurately forecast different aspects of supply chain management, such as demand, sales, revenue, production, and backorders. These techniques have been applied to predict unpredictable manufacturer demands, with some studies comparing ML-based approaches to traditional forecasting methods to evaluate their predictive accuracy (Carbonneau et al., 2008). Machine learning is effective for forecasting backorders in the supply chain when relevant data is available. The decision to use statistical methods or machine learning depends on the organization's specific needs and resources, along with the complexity and variability of the data being examined (Tirkolaee et al., 2021). The competition between different ML techniques enhances forecasting precision, enabling better decision-making that can ultimately increase revenue.

2.2. Machine Learning Approaches to Backorder Prediction in Supply Chain Management

The literature on backorder prediction and supply chain management increasingly emphasizes the role of machine learning models to improve forecasting accuracy, reduce costs, and optimize inventory management. Backorders occur when customer demand surpasses the available inventory, resulting in higher operational costs, delays in production, and lower customer satisfaction. As such, various studies have explored machine learning approaches to address this issue. Ali et al. (2024) emphasized the importance of balancing model complexity and computational efficiency in backorder prediction, demonstrating that simplified ML models like Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB) can reduce computational costs while maintaining high predictive accuracy. Similarly, Islam and Amin (2020) utilized Distributed Random Forest and Gradient Boosting Machines (GBM) to handle the large-scale datasets and complexity inherent in predicting backorders. Their research employed a distributed computing framework to improve scalability and efficiency, making it well-suited for real-time applications in global supply chains. Other studies have focused on integrating advanced ML techniques with economic considerations. Hajek and Abedin (2020) proposed a Random Forest classifier combined with a Clustering-Based Under sampling (CBUS) technique to improve backorder prediction while optimizing profitability. Their innovation lies in the application of big data analytics to handle large datasets, offering a more practical solution by incorporating economic factors into the prediction model. Santis et al. (2017) tackled the issue of class imbalance in backorder datasets by employing ensemble learning models like Logistic Regression, Classification and Regression Tree (CART), and Gradient Boosting, using techniques like SMOTE and Random Under-Sampling (RUS) to improve predictive accuracy. The integration of these methods demonstrates the critical importance of addressing the imbalance between backordered and non-backordered items to enhance model performance. The use of deep learning models has gained traction in recent studies. Shajalal et al. (2023) employed Deep Neural Networks (DNN), combining data balancing techniques such as SMOTE and under sampling to address the significant class imbalance in supply chain datasets. This approach significantly improved the accuracy of backorder prediction, demonstrating the effectiveness of DNNs in handling complex data patterns. Furthermore, Shajalal et al. (2022) introduced Explainable Artificial Intelligence (XAI) techniques by integrating Convolutional Neural Networks (CNN) with SHAP and LIME, providing greater transparency and interpretability in backorder predictions. This innovation is critical for fostering trust in ML models among business stakeholders, especially in contexts where decision-making transparency is essential. Additional studies have explored the role of recurrent neural networks (RNNs) in backorder prediction. Lawal and Akintola (2021) developed an RNN-based model to predict backorders, addressing the challenge of imbalanced datasets through the application of SMOTE, ADASYN, and RUS techniques. Their research showed that RNNs, when combined with advanced data balancing methods, provide a robust solution for backorder prediction. Similarly, Iqbal et al. (2023) applied a Classification and Regression Tree (CART) model to predict backorders, leveraging decision trees for their interpretability and ability to handle both continuous and categorical variables. This approach, while simpler than deep learning models, offers a practical solution for businesses looking for accessible and efficient ways to manage backorders. Lastly, Maitra and Kundu (2023) examined the financial implications of backorder misclassification, integrating cost-sensitive learning into their model. By employing techniques like Balanced Bagging Classifiers and Variational Autoencoder-GAN (VAE-GAN) models, they demonstrated that ML models can not only improve backorder prediction but also optimize inventory management by minimizing financial losses. This focus on profit functions highlights the practical relevance of backorder prediction models in real-world supply chains, where financial and operational efficiency is paramount. These studies collectively demonstrate the critical role of machine learning in improving the prediction of backorders in supply chain management. The methods range from classical models like logistic regression to more complex deep learning architectures, each addressing key challenges such as class imbalance, computational efficiency, and interpretability. By leveraging machine learning, businesses can optimize inventory management, reduce operational costs, and enhance customer satisfaction, making it a vital tool in modern supply chain operations. Table 2.1. provides an overview of studies that have utilized machine learning algorithms to predict backorders.

2.3. Review of Related Research on Backorder Prediction and Supply Chain Management

Islam and Amin (2020) focused on improving the accuracy and efficiency of predicting backorder scenarios within supply chains by utilizing advanced machine learning techniques. The main concern addressed by the paper is the need for robust predictive models that can handle the complexity and scale of modern supply chains. The authors have emphasized that traditional prediction methods often fall short in accurately forecasting backorders due to the dynamic nature of supply chain variables and the vast amount of data involved. This research aims to bridge this gap by applying Distributed Random Forest and Gradient Boosting Machine (GBM) techniques, which are well-suited for managing large datasets and capturing intricate patterns in supply chain data. To achieve its objectives, the paper employs a methodology that integrates Distributed Random Forest and Gradient Boosting Machine learning techniques within a distributed computing framework. This approach allows for the processing and analysis of extensive supply chain data in parallel, significantly reducing computation time and enhancing scalability. The study meticulously compares these advanced machine learning models against traditional methods, demonstrating superior performance in terms of prediction accuracy and reliability. One of the key innovations of this research is the implementation of distributed computing, which enables the models to efficiently handle real-time data and adapt to the evolving complexities of global supply networks. Additionally, the paper introduces novel feature engineering strategies tailored to supply chain data, enhancing the models' ability to identify critical factors contributing to backorders. This combination of distributed machine learning and customized feature engineering sets the study apart from previous research, offering a more effective and scalable solution for backorder prediction in contemporary supply chains. The authors have employed two machine learning models to predict backorder scenarios: Distributed Random Forest (DRF) and Gradient Boosting Machine (GBM). These models were chosen for their ability to handle large datasets and provide robust predictions in the context of supply chain management. The performance of the prediction models was evaluated using several metrics, including 1. Area Under the Curve (AUC) which is a common measure for assessing the accuracy of classification models, reflecting the model's ability to distinguish between classes (Kaur and Singh, 2023). 2.LogLoss which is a measure of the model's prediction error, capturing the probability-based classification accuracy. 3.Mean Per Class Error: This metric evaluates the average error across different classes, giving insight into how well the model performs for each class individually. 4.ROC (Receiver Operating Characteristic) Curve, which visualizes the performance of the models by plotting the true positive rate against the false positive rate at various threshold settings, with AUC summarizing the performance in a single value (Kaur and Singh, 2023).

Hajek and Abedin (2020) considered the material backorder prediction based on big data characteristics and its profitability occurring from misclassification. The have concentrated on predicting inventory backorders using a data-driven approach aimed at maximizing profitability. Their research introduced a profit function maximization method within a backorder prediction system designed to optimize the economic impacts of backorder decisions. The approach involved four key steps. First, they developed a modified version of the CBUS method, utilizing the k-means algorithm to balance instances of inventory backorders. Next, they established a profit-based classification metric to weigh the trade-offs between the benefits and costs associated with backorders. Following this, machine learning algorithms were trained on the balanced dataset, or cluster-specific classifiers were employed to enhance the accuracy of data subsets. Finally, they implemented a genetic algorithm-based search procedure to optimize the profitability metric. Hajek and Abedin (2020) focused on enhancing inventory backorder prediction by integrating big data analytics with a profit-maximization approach. The study addresses the challenge of predicting backorders, which is critical for optimizing inventory management and improving economic outcomes for businesses. The main concern of the paper is the limitation of traditional backorder prediction methods that rely heavily on stochastic approximations, which often fail to leverage the full potential of historical inventory data. The authors propose a machine learning model equipped with an under-sampling technique that aims to maximize the expected profit from backorder decisions, setting it apart from previous research that did not incorporate profit-based measures into backorder prediction systems. The methodology used in the paper involves modifying the Clustering-Based Under sampling (CBUS) approach to balance the class distribution in the dataset, which is highly imbalanced with a minority of backordered items. The modified CBUS method is combined with a Random Forest classifier, which is enhanced with a profit-based classification measure to optimize economic outcomes. The innovation of this paper lies in its focus on maximizing profitability rather than merely improving prediction accuracy, as well as its use of big data analytics to handle large and complex datasets. Compared to earlier studies, this approach offers a more practical and economically beneficial solution for inventory backorder prediction, especially in the context of big data. Researchers in this study have employed the Random Forest classifier as the primary machine learning model. This model is integrated into the modified CBUS (Clustering-Based Under sampling) technique to handle the imbalanced dataset and optimize the profit function in backorder prediction. The prediction performance has been evaluated by following measures: Area Under the ROC Curve (AUC) and Profit-Based

Classification Measure (π), which this novel metric is designed to evaluate the economic impact of prediction decisions, incorporating both the benefits and costs associated with correct and incorrect classifications.

In their paper, Santis et al. (2017) suggest using a supervised learning model to predict backorders in inventory control. Authors in their research primarily focuses on addressing the challenge of predicting material backorders in supply chain management. The main concern in their paper is the class imbalance problem, where the number of items that go into backorder (positive class) is significantly lower than those that do not (negative class). This imbalance poses a significant challenge in accurately predicting backorders, which can lead to inefficiencies in inventory management and overall supply chain performance. The methodology used in this study involves applying supervised learning models, specifically focusing on the combination of sampling methods and ensemble learning classifiers to tackle the class imbalance issue. The paper compares various learning classifiers, such as Logistic Regression, Classification Trees, Random Forest, and Gradient Boosting, in conjunction with techniques like Random Under-Sampling (RUS) and Synthetic Minority Over-sampling Technique (SMOTE). The innovation of this paper lies in its approach to integrating sampling techniques with ensemble learning to improve the predictive accuracy of backorder events, setting it apart from previous studies that did not adequately address the class imbalance in such a comprehensive manner. Several machine learning models have been used in this research, including Logistic Regression (LOGIST), Classification and Regression Tree (CART), Random Forest (FOREST) Gradient Tree Boosting (GBOOST) and Blagging (BLAG), which is a combination of under-sampling and tree ensemble methods.in addition the performance of the prediction models was evaluated using Area Under the ROC Curve (AUC) and Precision-Recall Curves.

Ntakolia et al. (2022) focused on interpreting and explaining the importance of key features in their dataset's prediction model. They developed a machine learning pipeline that included: (i) data preprocessing; (ii) feature selection using Recursive Feature Elimination (RFE) with Random Forest; (iii) building a classification model with Random Forest and under-sampling to handle imbalanced data; and (iv) applying the SHAP (SHapley Additive exPlanations) approach for post-hoc model interpretation. In this study, the RFE technique was used with Random Forest as the prediction model to identify significant features. The classification task was performed using the Random Forest (RF) algorithm, an ensemble method that constructs multiple decision trees based on randomly selected subsets of training data and features. The RF model trains on in-bag samples, which consist of about two-thirds of the dataset, while the remaining samples (out-of-bag) are used for internal cross-validation, yielding an out-of-bag error estimate. The recursive feature elimination process, combined with 5-fold cross-validation, was employed to determine feature importance and finalize the feature set for training the predictive models. The study demonstrated the superiority of the RF classifier by comparing it with two other well-known classifiers. Logistic Regression (LR), a regression model that extends linear regression to handle classification problems (binary outcomes 0 and 1), and Support Vector Machines (SVMs), were used in the analysis to benchmark the performance of the RF classifier.

Iqbal et al. (2023) centers on developing a predictive model for determining the likelihood of product backorders within the context of supply chain management. The main concern of the study is the challenge posed by supply chain uncertainties, which can lead to either a surplus or shortage of products, with backorders being a critical issue. The unpredictability of product backorders, influenced by factors such as fluctuating demand, logistical delays, and supply chain disruptions, necessitates an effective prediction model to minimize operational inefficiencies and improve customer satisfaction. To address this concern, the authors employed the Classification and Regression Tree (CART) model, a decision tree technique, to predict backorders based on various independent variables, such as inventory levels, transit times, sales forecasts, and risk factors. The CART model was chosen for its ability to handle both continuous and categorical data, providing interpretable results that highlight the most influential factors in predicting backorders. The innovation of this paper lies in its application of the CART model to the specific problem of backorder prediction, which is relatively novel compared to previous approaches that often relied on more complex machine learning models. By using CART, the study aims to offer a practical and accessible solution that can be readily implemented by businesses facing similar supply chain challenges. The study primarily uses the Classification and Regression Tree (CART) model to predict the probability of a product being backordered. The performance of the CART model was evaluated using several metrics: Accuracy: Measures the overall correctness of the model's predictions. Sensitivity: Evaluates the model's ability to correctly identify actual backorders (true positives). Specificity: Assesses the model's ability to correctly identify non-backorders (true negatives). R-squared (R^2): Used to measure the goodness-of-fit of the model, indicating how well the independent variables explain the variance in the dependent variable (backorder occurrence).

Lawal and Akintola (2021), in their research focused on developing a predictive model for inventory backorders using Recurrent Neural Networks (RNNs). The main concern addressed by this paper is the significant challenge of accurately predicting product backorders due to the imbalanced nature of the dataset, where backordered items are significantly fewer than non-backordered items. This imbalance often leads to inaccurate predictions, which can result in delays, increased costs, and reduced customer satisfaction. To address this issue, the paper proposes a deep learning model based on RNNs, which are particularly suited for processing sequential data and capturing temporal dependencies. The methodology includes several steps: preprocessing the data, including missing value imputation, feature conversion, and normalization; applying data balancing techniques such as SMOTE, ADASYN, and Random Under Sampling (RUS); and finally, training the RNN on the balanced dataset. The innovation of this study lies in the combination of RNN with advanced data balancing techniques to enhance the predictive performance of the model. Compared to previous studies, this approach provides a more robust solution to the backorder prediction problem, particularly in the context of large and imbalanced datasets. Recurrent Neural Network (RNN) is the primary machine learning model used in this study. The study also incorporates data balancing techniques such as: SMOTE (Synthetic Minority Over-sampling Technique) ADASYN (Adaptive Synthetic Sampling) and Random Under Sampling (RUS). The prediction performance of the models is evaluated using the following metrics: Precision, Recall, F1-Score and Area Under the Curve (AUC).

Shajalal et al. (2022), in their paper primarily focuses on improving the transparency and interpretability of machine learning models used for product backorder prediction in inventory management systems. The main concern addressed in this study is the "black-box" nature of complex machine learning models, which can hinder trust and adoption among business stakeholders who rely on these predictions for critical decision-making. To overcome this challenge, the authors propose a convolutional neural network (CNN)-based model for predicting product backorders, enhanced with explainable artificial intelligence (XAI) techniques. These techniques, including SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), provide both global and local interpretability, allowing stakeholders to understand the decision-making process of the predictive model. The methodology involves several key steps. First, the dataset is preprocessed to handle missing values, normalize features, and address class imbalance using the ADASYN (Adaptive Synthetic Oversampling) technique. The CNN model is then trained on this balanced dataset. The innovation in this paper lies in the integration of XAI techniques with the CNN model, which not only improves the predictive performance but also makes the model's decisions transparent and understandable to non-expert users. Compared to previous studies, which often focused solely on improving prediction accuracy, this approach offers a significant advancement by making the model's decision-making process more accessible and actionable for business stakeholders. The primary machine learning model used in this paper is a Convolutional Neural Network (CNN). The study also compares the performance of the CNN model with classical machine learning models, such as: Decision Tree, Support Vector Machine (SVM), Gradient Boosting. The prediction performance of the models is evaluated using several metrics: Accuracy, Area Under the ROC Curve (AUC) Receiver Operating Characteristic (ROC) Curves.

Shajalal et al. (2023), primarily addresses the challenge of predicting product backorders in inventory and supply chain management. The main concern is the significant class imbalance between backordered and non-backordered items, which complicates accurate prediction. The study proposes a novel methodological framework that leverages deep neural networks (DNNs) combined with data balancing techniques such as SMOTE (Synthetic Minority Oversampling Technique) and under sampling to handle this imbalance. This approach aims to improve the accuracy and reliability of backorder predictions, which is crucial for optimizing inventory levels, minimizing lost sales, and enhancing supplier-customer relationships. The methodology involves preprocessing the dataset, which is highly imbalanced, with a ratio of 137:1 between non-backordered and backordered items. The authors use a combination of oversampling and under sampling techniques to balance the dataset. They then implement four different DNN models: Weighted_DNN, Ran_Over_DNN, SMOTE_Over_DNN, and Com_SMOTE_Under_DNN. These models are trained on the balanced data to predict backorders. The innovation of this paper lies in its integration of DNNs with advanced data balancing techniques, which contrasts with previous studies that often relied on traditional machine learning models or simpler balancing methods. This approach demonstrates superior performance in predicting backorders, as evidenced by the new state-of-the-art results achieved. The study uses four variations of deep neural networks (DNNs): Weighted_DNN, Ran_Over_DNN, SMOTE_Over_DNN, Com_SMOTE_Under_DNN. The performance of the models is evaluated using several metrics: Area Under the ROC Curve (AUC), Precision and Recall, Expected Profit Measure.

Maitra and Kundu (2023), in their paper centers on improving backorder forecasting in supply chain management by utilizing advanced machine learning techniques. The study aims to reduce the negative impact of stockouts and backorders on inventory systems, customer satisfaction, and overall operational performance. This research introduces a comparative analysis of multiple machine learning classification techniques, including Balanced Bagging Classifiers, Fuzzy Logic, and Variational Autoencoder (VAE) integrated with Generative Adversarial Networks (GAN). The authors also account for the financial implications of backorder misclassification by incorporating profit functions and misclassification costs into their models. This work is significant in demonstrating how backorder prediction models can enhance inventory management, leading to improved customer satisfaction and organizational efficiency. The paper addresses several gaps in the literature, including the limited focus on cost-sensitive learning and profit considerations in backorder prediction models. While previous studies have developed machine learning models for backorder forecasting, few have incorporated a comprehensive analysis of both financial impacts and classification performance. Additionally, the study highlights the lack of real-world implementations that integrate machine learning interpretability for decision-making in inventory management. The authors employ multiple machine learning techniques, such as Balanced Bagging Classifiers (BBC), Fuzzy Logic, and VAE-GAN models. These are compared based on their ability to handle imbalanced datasets and their classification performance. The models are trained and tested on a large inventory dataset using key metrics like ROC-AUC, PRAUC, Macro F1-Score, and profit maximization. A notable innovation of the paper is the integration of cost-sensitive learning and profit functions into machine learning models for backorder prediction. By combining traditional machine learning techniques with VAE-GAN models, the authors aim to improve not only predictive accuracy but also the financial implications of backorders. The use of permutation importance to interpret model features also distinguishes this study from previous research, as it enhances transparency for decision-makers. The study found that the Balanced Bagging Classifier (BBC) outperformed other models across multiple performance metrics, including ROC-AUC and profit maximization. The integration of VAE with BBC also demonstrated strong performance, especially in dealing with class imbalance. The results suggest that advanced machine learning techniques can significantly improve backorder prediction accuracy and minimize financial losses caused by misclassification. The paper builds on existing machine learning and inventory management theories, employing both unsupervised learning (VAE) and supervised classification techniques (BBC) within a cost-sensitive learning framework. This combination allows for both predictive accuracy and cost minimization, positioning the study within the broader field of supply chain optimization. This study is highly relevant to the field of supply chain management, particularly in industries where backorders and stockouts can cause significant disruptions. The integration of machine learning techniques with financial and cost considerations makes this research applicable for real-world inventory systems, offering a pathway for businesses to enhance their decision-making processes and operational efficiency.

Rodger (2014), in his paper focuses on the use of advanced statistical and machine learning techniques to predict and mitigate backorder aging in complex supply chain systems. He emphasizes that the main concern is how uncertainties in supply chain management—such as demand variability, production lead time (PLT), and administrative lead time (ALT)—affect backorder creation and customer wait time. The authors propose using a Bayesian network-based approach, combined with fuzzy clustering and stochastic simulation, to forecast the probability of backorders and assess their impact on supply chain performance. This method allows supply chain managers to anticipate backorder risks and optimize decision-making to reduce customer wait time and improve inventory management efficiency. In terms of innovation, this paper introduces a novel application of Bayesian probabilistic estimation integrated with fuzzy logic and Markov blankets, which is different from traditional backorder prediction models. The use of Bayesian networks provides a structured approach to understanding the relationships between various supply chain variables, while fuzzy clustering offers flexibility in dealing with uncertainties and imprecise data. The study's key contribution lies in its ability to predict backorders by dynamically adjusting trigger points—such as the Acquisition Advice Code, Acquisition Method Suffix Code, and other supply chain metrics—based on changes in lead times, unit prices, and stock levels. This approach differs from earlier models that were often static and less adaptable to real-time supply chain fluctuations. Two machine learning models have been used in this research including Bayesian Network and Fuzzy Clustering. The study does not specify common accuracy measures such as precision or recall but focuses on the probabilistic outcomes of Bayesian networks which includes calculating posterior probabilities for backorder occurrence and comparing the impact of different supply chain variables on backorder aging.

Wang and Tang (2014), in their research focused on developing optimal inventory rationing policies for systems that handle multiple demand classes with both backorders and lost sales. The main concern of this study is the complexity involved in prioritizing different demand classes over time, where customers may react differently to stockouts—either by waiting for backorders or by switching to alternative sources, leading to lost sales. The research introduces a dynamic rationing policy that adapts to the fluctuating priority of demand classes, which contrasts with traditional static policies that do not account for changes in customer behavior over time. The study uses a Markov decision model to determine optimal rationing levels for different demand classes, helping to balance the cost of holding inventory against potential penalties from backorders and lost sales. One of the key innovations in this paper is the introduction of dynamic rationing in systems with mixed backorders and lost sales, which had previously been studied in isolation. The research extends existing models by allowing for time-dependent changes in the priority of demand classes, based on their respective penalty costs. This dynamic approach results in a more nuanced policy where rationing levels for different classes change as time progresses toward the next replenishment cycle. The paper also proposes a heuristic policy to simplify the computational complexity of the Markov model, making it easier to implement in real-world systems. The findings demonstrate that this dynamic rationing policy outperforms static policies, leading to reduced costs in inventory management. This study does not explicitly use machine learning models. Instead, it relies on Markov decision processes for optimization and a heuristic algorithm for simplifying the dynamic rationing process. The evaluation of performance focuses on the cost gap between dynamic and non-rationing policies. Cost is evaluated in terms of inventory holding costs and penalty costs for backorders and lost sales.

Gao et al. (2022), in their research focused on improving backorder prediction in supply chain management by using machine learning algorithms, specifically neural networks and Naive Bayes, to forecast product backorders. The research aims to give decision-makers increased flexibility, clarity, and accuracy in predicting when products will be on backorder, thus enabling more efficient inventory management. The study contributes by proposing a machine learning-based approach that integrates neural networks and Naive Bayes to anticipate product backorders before actual sales occur. This allows companies to manage stock shortages proactively. The approach is tested using real-life data from a well-known e-commerce business, demonstrating how these algorithms can be effectively applied to predict future backorders, contributing to improvements in supply chain efficiency. The paper addresses a gap in the literature where few studies have focused on using machine learning algorithms specifically for backorder prediction. While machine learning has been applied extensively in inventory management and demand forecasting, its application to predicting backorders, particularly with an emphasis on integrating neural networks and Naive Bayes, has not been thoroughly explored. In this study, machine learning models specifically neural networks and Naive Bayes—are applied to predict backorders. The performance of these models is then evaluated using various performance metrics on the dataset. The innovative aspect of this study is the use of both neural networks and Naive Bayes for backorder prediction, combined with a comprehensive data preprocessing approach that ensures model accuracy. Additionally, this research is applied to a real-world case study, offering practical insights into how these machine learning models can be implemented in e-commerce to anticipate backorders. The research is grounded in machine learning theory, particularly in supervised learning, and uses neural networks and Naive Bayes as the primary algorithms. These algorithms are employed to classify whether a product will be backordered based on inventory data, customer demand, and sales forecasts.

Ahmed et al., (2022) in their research assess and compare the effectiveness of various machine learning (ML) techniques in predicting backorders within supply chains. By examining different algorithms, the study seeks to improve the accuracy of backorder predictions and enhance supply chain management processes. The research explores how ML models can handle complex datasets and imbalanced classes, which are common challenges in inventory management. They also demonstrate that machine learning models, when applied correctly, can offer substantial improvements in predicting backorders in a supply chain. A key contribution of the study is its comparative analysis of traditional and advanced ML techniques, including Random Forest (RF), Support Vector Machine (SVM), Neural Networks, and Gradient Boosting (GB). The paper highlights the strengths and limitations of these models in predicting product shortages and delays. Furthermore, the study emphasizes the importance of feature selection and model optimization in improving prediction accuracy. The paper is grounded in the theoretical framework of predictive analytics, where machine learning models are used to forecast future outcomes based on historical data.

Garcia and Panduro (2022) in their research focused on applying machine learning models to reduce backorders in the cross-docking sales process within the homecenter order service. They seek to identify the most effective machine learning algorithm to enhance order fulfillment, minimize stockouts, and streamline logistics. The paper contributes to the field by proposing a solution to predict and mitigate backorders in inventory management using machine learning. Through the application of models like Neural Networks, Random Forest, Decision Tree, and Support Vector Machine (SVM), the study highlights the model that best predicts backorders and improves decision-making for supply chain efficiency. The key contribution lies in optimizing inventory management to reduce pending orders and enhance service quality. A quantitative and explanatory-correlational approach is used, where historical data from 2018 to 2020 are analyzed. Variables such as demand projection, inventory levels, and backorder status are used as input for machine learning models. The models are evaluated using Orange Software, and performance is measured using indicators like accuracy, AUC (Area Under the Curve), and the confusion matrix. The innovation of this paper lies in its focus on applying multiple machine learning models to the specific problem of backorder prediction in cross-docking sales processes. The study offers a novel comparative analysis, showcasing the superior performance of neural networks in reducing backorders compared to other traditional models like SVM and decision trees. The results reveal that the neural network model performed best, achieving an accuracy of 99.5% and the highest ROC curve performance. This model significantly outperformed others, such as SVM (79.7%) and Random Forest (96.4%), making it the most suitable for predicting and managing backorders in the given context.

Ali et al., (2024), in their paper provide a comprehensive analysis of using machine learning (ML) techniques for backorder prediction in supply chains, focusing on optimizing both predictive performance and computational efficiency. It evaluates the trade-offs between using simplified machine learning models with fewer predictors versus more complex models with a larger set of predictors. By reducing the number of input variables, the study seeks to decrease computational costs while maintaining an acceptable level of predictive accuracy. The paper's key contribution is its demonstration that simplified machine learning models with a limited number of high-impact features can significantly reduce computational costs with only marginal reductions in predictive accuracy. The study provides a detailed comparison of traditional statistical methods like logistic regression with advanced machine learning algorithms such as Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB) for predicting customer backorders. Previous research has often focused on using large, complex models for backorder prediction, but there has been little exploration into how simplifying these models by reducing the number of predictors could impact both accuracy and computational efficiency. The paper fills this gap by evaluating the performance of simplified models using a subset of important predictors, addressing the balance between model complexity and computational cost. The novelty of this study lies in its focus on balancing predictive performance and computational efficiency. While previous studies have concentrated on achieving high accuracy, this research emphasizes the importance of reducing computational costs without sacrificing much accuracy. It also adds value by demonstrating that simpler models with fewer features can still provide competitive performance, which makes them more feasible for practical implementation. The findings showed that reducing the number of predictors from 22 to 5 resulted in only a marginal reduction in accuracy (0.6% to 4.2%), while significantly reducing computational costs (30% to 98%). Random Forest and Gradient Boosting were identified as the best-performing algorithms, particularly in the simpler models with fewer predictors. The most important predictors for backorder prediction were identified as inventory levels and sales forecasts.

Saban Adana et al. (2019) in their research offer a comprehensive analysis of using advanced machine learning (ML) techniques to predict backorders in a supply chain setting. The focus of the paper is on demonstrating the utility of advanced machine learning models for predicting backorders, which are a significant issue in supply chain management. The study seeks to show that ML models, such as neural networks, auto neural networks, and decision trees, provide more precise forecasts than traditional methods like logistic regression. The main contribution of this research is its empirical comparison of machine learning models against traditional methods for forecasting backorders. In addition, the authors in their study highlight the effectiveness of advanced models in improving prediction accuracy, which can help companies optimize inventory management and minimize costs. By implementing these models, businesses can better predict when backorders might occur, enabling them to take preventive measures that improve customer satisfaction.

Carbonneau et al. (2008) in their research emphasized on applying Machine Learning (ML) Techniques to improve demand forecasting in the supply chain, specifically targeting the distorted demand signals that occur due to the bullwhip effect. This research evaluates the efficacy of advanced ML methods like neural networks (NN), recurrent neural networks (RNN), and support vector machines (SVM) compared to traditional forecasting methods. The paper as a main idea addresses the demand distortion problem that arises as demand signals move through the supply chain. In addition, the study provides a comparative analysis of traditional forecasting models (like moving average and regression) against ML models (NN, RNN, SVM).

Seyedan and Mafakheri (2020), in their research has emphasized on the application of big data analytics (BDA) in demand forecasting for supply chain management (SCM). The primary concern is the unpredictability and uncertainty in supply chains due to fluctuating customer demand, logistical delays, and the complexities introduced by global supply chains. Traditional demand forecasting methods, such as statistical models, struggle to cope with the sheer volume, velocity, and variety of data in modern supply chains. This research highlights how BDA, through advanced machine learning and data analytics techniques, can provide more accurate, data-driven forecasts that help businesses manage their supply chains more efficiently, improve customer satisfaction, and reduce operational costs. The methodology involves a thorough review of existing literature, categorizing the BDA techniques used in SCM demand forecasting, including time-series forecasting, clustering, regression, support vector machines (SVM), and artificial neural networks (ANN). The authors identify gaps in the literature, particularly regarding the application of BDA in closed-loop supply chains (CLSCs), which deal with the reverse flow of materials like returns and recycling. The innovation of the paper lies in its comparative analysis of machine learning models and BDA techniques, as well as its focus on the future potential of BDA in SCM. This paper differs from previous studies by offering a comprehensive classification of BDA methods and highlighting opportunities for integrating prescriptive analytics to optimize decision-making. The paper reviews several machine learning models that are widely used for demand forecasting in supply chains: Neural Networks (ANN), Support Vector Machines (SVM), Time-Series Forecasting (ARIMA), and Decision Trees. To evaluate the performance of the predictive models, the paper references several common metrics: Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), Root Mean Square Error (RMSE), Area Under the ROC Curve (AUC).

Tang and Ge (2021) in their paper focused on improving the accuracy of material demand forecasting in manufacturing enterprises using deep learning algorithms. Their goal was to optimize Enterprise Resource Planning (ERP) systems, reduce the time spent on calculations, and enhance the response times for both manufacturing enterprises and suppliers, ultimately lowering costs and preventing production disruptions due to material shortages. The paper introduces a Backpropagation (BP) neural network model for material demand forecasting, which integrates historical sales demand and material consumption data to predict future demand more accurately. By utilizing deep learning algorithms, the study achieves greater precision in demand forecasting and optimizes lead times, allowing suppliers more time to prepare materials. The research's key contribution is the development of an auxiliary method that supports ERP systems, reducing human intervention in the forecasting process while also lowering material costs for manufacturing companies. The innovative aspect of this research is the application of deep learning algorithms, particularly the BP neural network, to enhance material demand forecasting in manufacturing enterprises. The paper introduces an auxiliary method that extends ERP system functionality, allowing enterprises to increase forecast lead times, reduce costs from production plan changes, and streamline the forecasting process. Additionally, the model focuses on optimizing both sales demand forecasting and material consumption variables, a significant departure from traditional approaches that primarily rely on consumption data. The results demonstrate that the deep learning model significantly improves forecast accuracy when compared to traditional models. By incorporating both sales demand forecasts and material consumption data, the BP neural network model achieves lower forecast errors. The extended lead time allows suppliers to better prepare, reducing the costs associated with part claims and last-minute adjustments in production plans. The model's success in accurately predicting material demand supports the utility of deep learning in enhancing ERP systems.

Table 2-1 Previous studies focused on backorder prediction by using machine learning techniques

Study (year)	Dataset	Technique to deal with CI Problem / Sampling Techniques	Feature Selection / Feature Engineering Techniques	Machine Learning Models Ensemble Learnings	Dataset Split	Main Findings Performance Obtained
The current study	Highly Imbalanced	RUS, SMOTE-ENN, SMOTE-SVM, ADASYN, Borderline-SMOTE	SHapely Additive. PCA	Neural Network, KNN and Random Forest, Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBM), LightGBM, Stacking	Standard Splits: 70% Training, 30% Test Cross-Validation Methods: 5-Fold 10-Fold - RandomizedSearchCV	balanced accuracy, AUC-ROC, Specificity, F1 Score, AUC-PR
PETR HAJEK AND MOHAMMAD ZOYNUL ABEDIN (2020)	Highly Imbalanced	SMOTE and clustering-based under-sampling (CBUS) using the k-means algorithm	NO / Not Specified	Random Forest	10-fold cross-validation.	RF: ROC 91.57 SVM: ROC: 78.32 XGBoost
Samiul Islam and Saman Hassanzadeh Amin (2020)	Highly Imbalanced	SMOTE	NO / Not Specified	Distributed Random Forest (DRF) Gradient Boosting Machine (GBM)	Not Specified	DFR: AUC 0.959 GBM: AUC 0.946
Rodrigo Barbosa de Santis, Eduardo Pestana de Aguiar, Leonardo Goliatt (2017)	Highly Imbalanced	Random Under Sampling (RUS) SMOTE and Ensemble Learnings	No / Not Specified	Logistic Regression (LOGIST), Classification and Regression Tree (CART), Random Forest (FOREST) Gradient Tree Boosting (GBOOST), Blagging (BLAG)	Training Set: 85 % Test Set: 15 %	GBOOST: AUC 0.9482, Random Forest: AUC 0.9441, BLAG: AUC 0.9478 Precision-Recall Curves
Charis, Ntakolia Christos, Kokkotis Serafeim, Moustakidis Elpiniki, Papageorgiou (2022)	Highly Imbalanced	Random Under Sampling technique	RFE-RF Recursive Feature Elimination – Random Forest	Machine Learning Classifiers: Support vector machines (SVMs) Logistic regression (LR) Random Forest (RF)	70% - 30 % Training-Test Set Random Data Split	RF: AUC 0.95, Accuracy: 87.87 % LR: AUC 0.80 - Accuracy: 71.08 % SVM: AUC 0.84, Accuracy: 74.72 %
Iqbal, G. M. D., Rosenberger, M., Ha, L., Gregory, S., and Anoruo, E (2023)	Highly Imbalanced	No technique and solution to deal with imbalance issue	Not Specified	Classification and Regression Tree (CART) model	Not Specified	Accuracy 0.98, Sensitivity, 0.14, Specificity: 0.96 R-squared (R^2): 0.03
Lawal and Akintola (2021)	Highly Imbalanced	ADASYN, SMOTE, Random Under Sampling	Not Specified	Recurrent Neural Networks RNN	70% 30% Training and Test Set	Precision: 0.901, Recall: 0.879 F1-Score: 0.889
Shajalal, M., Boden, A., and Stevens, G (2022)	Highly Imbalanced	ADASYN SMOTE	SHapley Additive exPlanation (SHAP)	Convolutional Neural Network (CNN)	Not Specified	AUC 0.9489 Accuracy 0.894
Shajalal, M., Boden, A., and Stevens, G (2021)	Highly Imbalanced	minority class weight boosting, randomised oversampling, random under sampling, SMOTE over- sampling,	No / Not Specified	Weighted_DNN Ran_Over_DNN SMOTE_Over_DNN Com_SMOTE_Under_DNN	85 % 15 % Training and Test Set 10-Fold Cross Validation	AUC: 0.9427 Precision and Recall, Expected Profit Measure.
Sarit Maitra And Sukanya Kundu (2023)	Highly Imbalanced	Random Under Sampling	Permutation importance (PI)	Balanced Bagging Classifiers, Fuzzy Logic, and Variational Autoencoder (VAE) integrated with Generative Adversarial Networks (GAN)	Not Specified	ROC-AUC: 0.9604, PRAUC, 0.2428, Macro F1-Score: 0.5483
Hui Gao, Quanhui Ren and Chunfeng Lv (2022)	Highly Imbalanced	No / Not Necessary	No / Not Specified	Neural networks and Naive Bayes	Not Specified	Naive bayes: ACC 0.99, Precision 0.97 - Neural network: ACC 0.99, Precision 0.99
Garcia Lopez, Y. J., Panduro, J., and Pumayauri, S. (2022).	Small and Normal Dataset	No / Not Necessary	No / Not Specified	Neural Networks, Random Forest, Decision Tree, and Support Vector Machine (SVM)	Not Specified	SVM: AUC 0.704, RF: AUC 0.967, DT: AUC 0.864, NN: AUC 0.994
Ali, A., Jayaraman, R., Azar, E., and Maalouf, M. (2024).	Highly Imbalanced	Random down sampling	Feature Importance method, also	Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB)	a splitting ratio of 80:20	Accuracy of 0.88 by RF accuracy, F1-score, and feature importance

			called Gini importance			
Ahmed, F., Hasan, M., Hossain, M. S., and Andersson, K (2022)	Highly Imbalanced	Random undersampling method	Statistical Hypothesis Test	Decision tree, Random Forest, adaptive boosting and gradient boosting	the dataset is divided into training and test sets with a ratio of 8:2.	Rand Forest outperforms: Accuracy 0.86, precision 0.88, recall 0.88, f1-score 0.88, AUCROC: 0.9458 and AUCPRC: 0.9383
Adana, S., Cevikparmak, S., Celik, H., and Uvet, H (2019)	Highly Imbalanced	stratified hold-out method and randomly under sampled the frequent event	PCA	Decision Trees, Neural Networks, and Logistic Regression, and Auto Neural Models	Not Specified	Accuracy, Sensitivity, and Precision Autoneural: 87.2%, 88.49%, 85.95%, 86.29%.

2.4. Imbalance Class

To address this imbalance, Santis et al. (2017) compared various learning classifier algorithms, incorporating techniques such as sampling and ensemble methods. In their research on class imbalance, Islam and Amin (2020) applied a synthetic minority oversampling technique (SMOTE) to the target class. Ntakolia et al. (2022) used a random under sampling method to lower the majority class to the number of the minority one in order to address the issue of imbalanced data. The data set was then normalized to [0,1]. Maitra and Kundu (2023), in their paper address the challenge of imbalanced datasets, which is a significant concern in backorder prediction. In inventory management, the occurrence of backorders is often much less frequent compared to non-backordered items, leading to an imbalance in the dataset. The authors point out that standard machine learning models may perform poorly as a result of this imbalance since they may become biased in favor of the majority class (non-backorders), which would lead to low predicted accuracy for the minority class (backorders). To overcome this issue, the authors employ techniques such as Balanced Bagging Classifiers (BBC) and Variational Autoencoder (VAE) combined with Generative Adversarial Networks (GAN) to enhance model performance on imbalanced data. These methods aim to balance the dataset during training, allowing the models to better learn from both classes and improve backorder prediction accuracy. Wang and Tang (2014) don't specifically focus on imbalanced datasets, but it discusses the challenge of handling varying demand classes in inventory management systems. In scenarios where there are mixed backorders and lost sales, certain demand classes are often prioritized, leading to potential imbalance in how inventory is allocated. While not explicitly focused on dataset imbalance, the decision-making processes in the paper indirectly address the issue by dynamically adjusting inventory levels for different classes. Dehghan-Bonari et al. (2021) address the issue of imbalanced datasets explicitly. They highlight that in many inventory management systems, the number of backordered products is significantly smaller than the number of non-backordered products, leading to a highly imbalanced dataset. To overcome this challenge, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and other sampling methods are applied to balance the dataset and improve the performance of machine learning algorithms. This imbalance presents a critical problem for accurate backorder predictions and was a key concern addressed in the study. Shajala et al. (2021) in their study explicitly focuses on the issue of imbalanced datasets in backorder prediction. It acknowledges that backordered products are much rarer than non-backordered products, creating a highly imbalanced dataset. The authors address this by using oversampling techniques such as SMOTE and ADASYN to balance the dataset before applying deep learning models like neural networks. This approach helps improve the accuracy of the predictions and ensures that the models do not become biased toward the majority class. In another paper, Shajalal et al. (2022) have mentioned the class imbalance problem in the context of predicting product backorders. Since backorders occur less frequently than non-backorders, the dataset is heavily imbalanced. To counter this, the study applies the ADASYN (Adaptive Synthetic Sampling) technique, which helps to balance the minority and majority classes, ensuring that the convolutional neural network (CNN) model can accurately predict backorders. The imbalance problem and its solutions are a significant part of the methodology to improve prediction reliability. Choosing the appropriate evaluation metrics is crucial for guiding the development of a predictive model, with the confusion matrix being a key tool in binary classification for tracking the correct and incorrect classifications of each classification (Santis et al., 2017).

3. Dataset

The study's dataset is a severely imbalanced dataset from the Kaggle competition that is publicly accessible, "Predict Product Backorders. Can you predict product back orders?" (Santis, 2017), as shown in Table 3.1. It comprises 23 features, with 15 being numerical and 8 (including the target variable 'went_on_back_order') categorical, capturing various aspects of supply chain performance. The dataset comprises 1,671,569 observations, containing a mix of floating-point, integer, and string values. This dataset offers historical weekly data snapshots for the eight weeks leading up to the target week and includes attributes such as ‘national inventory’ (representing current stock levels), ‘lead time,’ ‘in transit qty’ (products currently being transported), sales forecasts for 3, 6, and 9 months, along with actual sales figures for 1, 3, 6, and 9 months. Additionally, it features supplier performance metrics over the previous 6 and 12 months, the ‘min bank’ (indicating the minimum required stock level), and multiple binary risk flags. The primary objective is to predict whether a product will be on backorder, a binary classification task. However, the dataset exhibits a significant class imbalance, as most products are not backordered, making accurate predictions challenging since models may be biased toward the majority class. The target variable (y_1: went_on_back_order) is represented as a binary label, where 0 denotes non-backorder items (‘No’ class), and 1 indicates items that will be backordered (‘Yes’ class). The 'inventory' feature reflects the stock available for each product, while the 'Lead Time' attribute represents the duration between the ordering of products and their delivery to customers, ranging from 0 to 52 weeks in this dataset. The sales data is divided into one-month, three-month, six-month, and nine-month intervals, while the forecasted sales are similarly presented in three columns, showing projections for three, six, and nine months, respectively.

To address the computational complexity and resource-intensive nature of processing a large dataset containing over 1.6 million observations, this study employs stratified sampling to create a smaller, representative subset for analysis. Specifically, a subset of 16,715 samples was selected to maintain the balance between classes and preserve the structural characteristics of the original dataset, ensuring the integrity and representativeness of the data while reducing computational demands. Implementing multiple machines learning classifiers and employing advanced preprocessing techniques such as MICE, RUS, ADASYN, SMOTE, PCA, and SHAP, alongside hyperparameter tuning techniques like RandomizedSearchCV, requires substantial computational power and time. By focusing on a representative portion of the dataset, this approach ensures feasibility while maintaining the integrity of the analysis. The smaller dataset allows for efficient experimentation with various preprocessing, feature selection, and optimization techniques without compromising the reliability or validity of the results. This choice strikes a balance between computational efficiency and the study’s objective of exploring the performance of diverse machine learning models and methodologies.

Sadaiyandi et al., (2023) in their research provide a comprehensive review of stratified sampling. They stated that stratified sampling divides data into groups (strata) based on specific values to preserve structural information. Random samples are then drawn from each stratum, ensuring representative samples and balanced class distributions. For example, in forest datasets, stratified sampling balances the number of samples per class while accounting for within-class variance, maintaining the original data structure. This method is particularly effective for imbalanced datasets, as it randomly selects examples from both positive and negative classes to create balanced training sets. The main advantage of stratified sampling is its ability to reduce estimation error by grouping similar data objects and applying random or systematic sampling within each stratum.

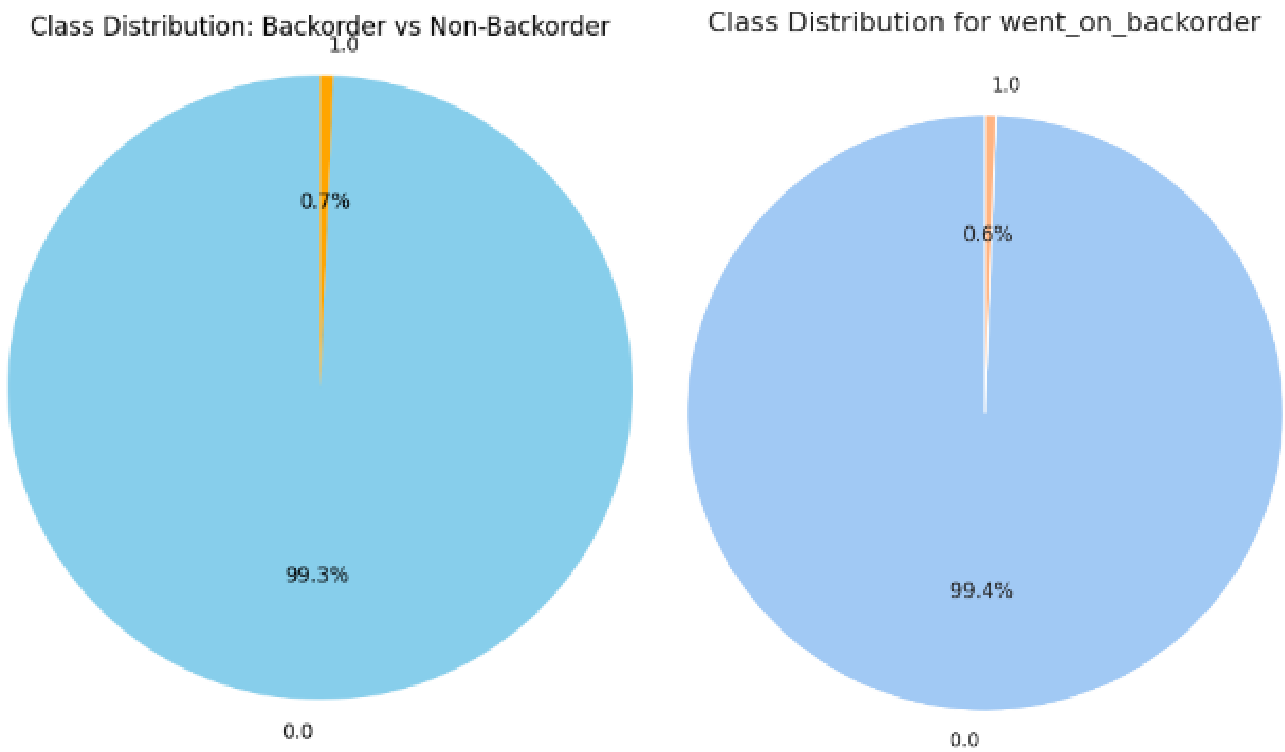
The dataset used in this study has been described comprehensively through various tables and figures, providing a detailed understanding of its structure and characteristics. Table 3.1 shows the overall dataset details, including the total number of observations and key attributes. Table 3.2 presents a detailed description of the variables, highlighting their roles in predicting backorders. To illustrate the imbalance between the "backorder" and "non-backorder" classes, Figures 3.1 and 3.2 compare the class distribution in the original dataset with the stratified dataset, emphasizing the adjustments made to address class imbalance. Figure 3.3 displays the correlation matrix heatmap, reflecting the relationships between numerical features and aiding in the identification of highly correlated variables. Additionally, Table 3.3 provides a summary of the dataset's statistical properties, including measures such as mean, median, standard deviation, and range, offering insights into the overall data distribution and variability. Together, these visualizations and summaries contribute to a comprehensive analysis of the dataset's features and suitability for backorder prediction, described through various tables and figures, providing a detailed understanding of its structure and characteristics.

Table 3-1 The dataset details

Dataset	**Independent Variables**	**Target Variable**	**Total Observation**
Product Backorder Prediction	22	Product on BO Classification 0, 1	16,715

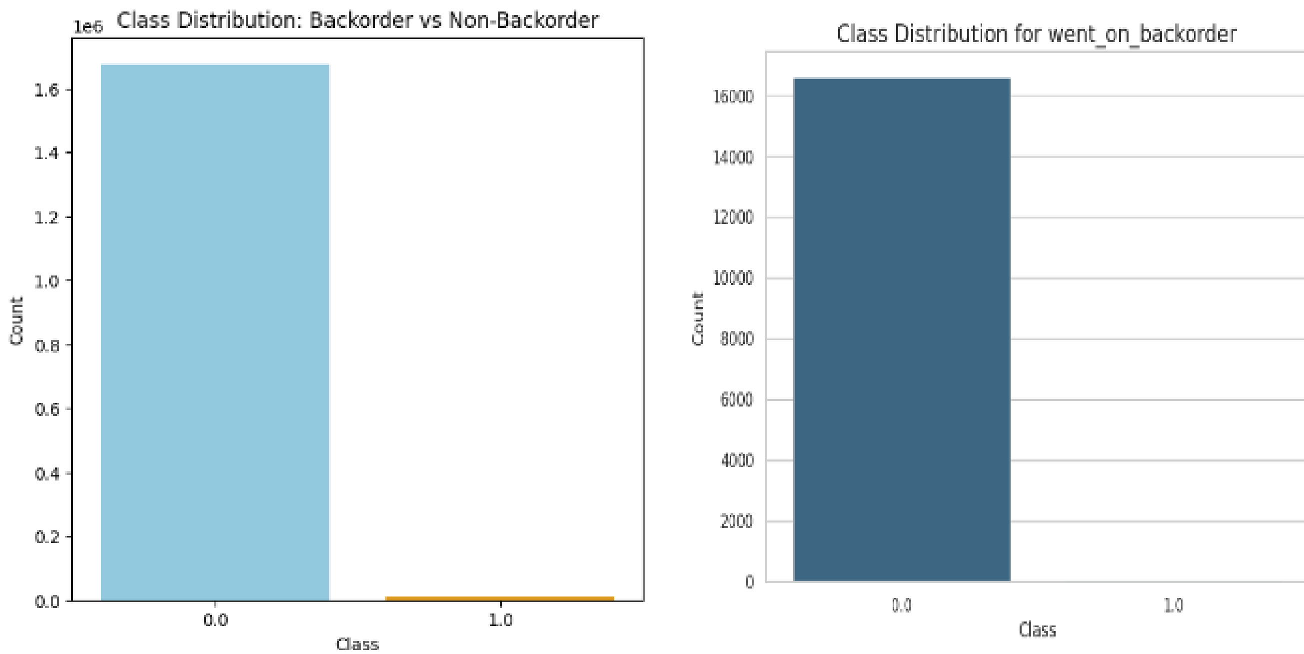
Table 3-2 Variables description

No	Variables	Description	Type
-	SKU	SKU code	-
1	x_1: National Inventory	Current inventory level of component	Numerical
2	x_2: Lead Time	Transit time	Numerical
3	x_3: In Transit Quantity	Quantity in transit	Numerical
4	x_4: Forecast 3_Month	Sales forecast - the next 3 months	Numerical
5	x_5: Forecast 6_Month	Sales forecast - the next 6 months	Numerical
6	x_6: Forecast 9_Month	Sales forecast - the next 9 months	Numerical
7	x_7: Sales 1_Month	Sales quantity - the prior 1 months	Numerical
8	x_8: Sales 3_Month	Sales quantity - the prior 3 months	Numerical
9	x_9: Sales 6_Month	Sales quantity - the prior 6 months	Numerical
10	x_{10}: Sales 9_Month	Sales quantity - the prior 9 months	Numerical
11	x_{11}: Min Bank	Minimum recommended amount in stock	Numerical
12	x_{12}: Potential_Issue	Indictor variable noting potential issue with item	Categorical
13	x_{13}: Pieces Past Due	Parts overdue from source	Numerical
14	x_{14}: Perf 6 months avg	Source performance in the last 6 months	Numerical
15	x_{15}: Perf 12 months avg	Source performance in the last 12 months	Numerical
16	x_{16}: Local BO Quantity	Amount of stock orders overdue	Numerical
17-22	x_{17-22} General Risk Flags	multiple binary risk flags	Categorical
23	y_1: Went On Back Order	Product went on backorder	Categorical



Original Stratified

Figure 3.1: Class Distribution (Percentage%) - Backorder vs non-backorder - A comparison between original dataset and stratified dataset



Original Stratified

Figure 3.2: Class Distribution (count) - Backorder vs non-backorder - A comparison between original dataset and stratified dataset

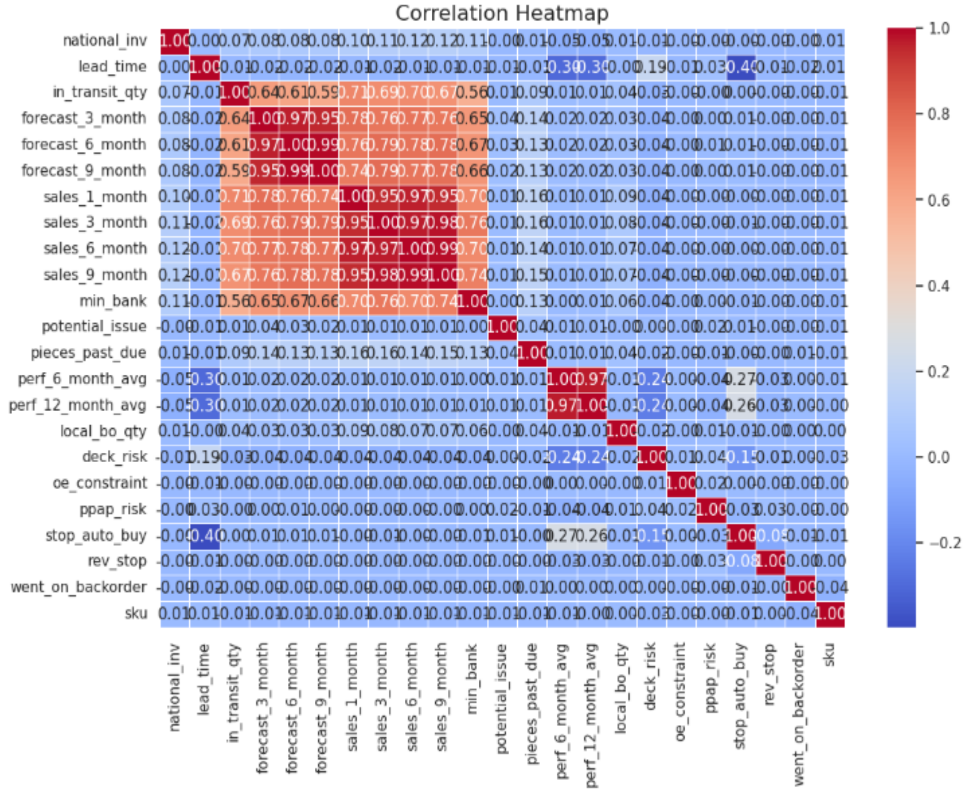


Figure 3.3: Correlation matrix heatmap

Table 3-3: Summary statistics of dataset

Variable	Feature	Non-Null Count	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
0	national_inv	16715	474.6351	13503.772	0	4	15	82	1370327
1	lead_time	16715	8.405171	7.323036	0	4	8	9	52
2	in_transit_qty	16715	37.62238	543.594257	0	0	0	0	34060
3	forecast_3_month	16715	149.3052	2004.5584	0	0	0	4	124380
4	forecast_6_month	16715	293.1877	3918.89074	0	0	0	11	236316
5	forecast_9_month	16715	433.0183	5912.67271	0	0	0	19	358428
6	sales_1_month	16715	40.2895	489.011661	0	0	0	4	30071
7	sales_3_month	16715	131.2472	1557.10325	0	0	1	15	104079
8	sales_6_month	16715	271.2694	3452.68306	0	0	2	30	212700
9	sales_9_month	16715	413.0353	4972.09648	0	0	4	45	304121
10	min_bank	16715	45.55381	565.336179	0	0	0	3	42956
11	potential_issue	16715	0.00071792	0.026785	0	0	0	0	1
12	pieces_past_due	16715	1.115704	29.16411	0	0	0	0	2100
13	perf_6_month_avg	16715	-6.756476	26.3739	-99	0.64	0.82	0.96	1
14	perf_12_month_avg	16715	-6.343496	25.689251	-99	0.66	0.81	0.95	1
15	local_bo_qty	16715	0.256895	8.010347	0	0	0	0	615
16	deck_risk	16715	0.2312893	0.42167	0	0	0	0	1
17	oe_constraint	16715	5.9827E-05	0.007735	0	0	0	0	1
18	ppap_risk	16715	0.1194735	0.324355	0	0	0	0	1
19	stop_auto_buy	16715	0.9615316	0.19233	0	1	1	1	1
20	rev_stop	16715	0.00065809	0.025646	0	0	0	0	1
21	went_on_backorder	16715	0.00616213	0.078259	0	0	0	0	1

4. Research Methodology

The methodology of this study follows a comprehensive, multi-phase approach designed to ensure accurate and reliable predictions of backorders through the application of advanced machine learning models. The process begins with data preprocessing, where the dataset is cleaned to address missing values using sophisticated imputation techniques such as Multiple Imputation by Chained Equations (MICE), ensuring critical information is preserved. To address the issue of class imbalance, a combination of oversampling and under sampling techniques, including SMOTE-ENN, ADASYN, and hybrid adaptive sampling methods, is applied, ensuring balanced representation between backordered and non-backordered items. Feature engineering and dimensionality reduction are then performed, with Principal Component Analysis (PCA) employed to transform the high-dimensional dataset into a smaller set of uncorrelated components, improving computational efficiency and enhancing model performance. Additionally, SHAP (SHapley Additive exPlanations) is utilized to interpret model predictions, identifying the contributions of individual features and enhancing the transparency and explainability of the machine learning models. The methodology continues with splitting the dataset into training and testing subsets, typically using an 70%-30% split, to ensure unbiased model evaluation. Cross-validation techniques, such as 10-fold cross-validation, has been employed during the training phase to enhance model generalizability and robustness. A range of machine learning models, including Neural Networks (NN), K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting Machines (GBM), XGBoost, LightGBM and stacking model, are then evaluated to identify the most suitable algorithms for backorder prediction. Totally 98 different configurations resulting from these techniques and method have been trained and implemented. Ensemble learning techniques, such as Random Forest and Gradient Boosting-based methods, are prioritized for their ability to combine multiple weak learners and improve predictive accuracy. Subsequently, model training is accompanied by hyperparameter optimization using methods like RandomizedSearchCV, ensuring that each model achieves optimal performance. The trained models are then rigorously evaluated using metrics such as ROC-AUC, PR-AUC, F1 Score, balanced accuracy, Sensitivity, and Specificity, providing a holistic view of performance across both majority and minority classes. Once the models are validated, predictions are generated, and a comparative analysis is conducted to determine the most effective model for backorder prediction. Finally, the results are interpreted, with actionable insights and recommendations presented for practical application in supply chain management. Figure 4.1. represents the methodology flow chart of the thesis.

Earlier studies have explored a variety of machine learning models and preprocessing techniques to address the challenges of class imbalance. For example, Adana et al. (2019) applied decision trees, neural networks, and logistic regression to a dataset with 22 variables, focusing on evaluating model performance through metrics such as accuracy, precision, sensitivity, and specificity. Similarly, Ahmed et al. (2022) compared various machine learning algorithms and employed comprehensive data preprocessing methods, including missing value imputation, normalization, and class imbalance adjustments, assessing models using F1-score, precision, recall, cross-validation, and hyperparameter tuning.

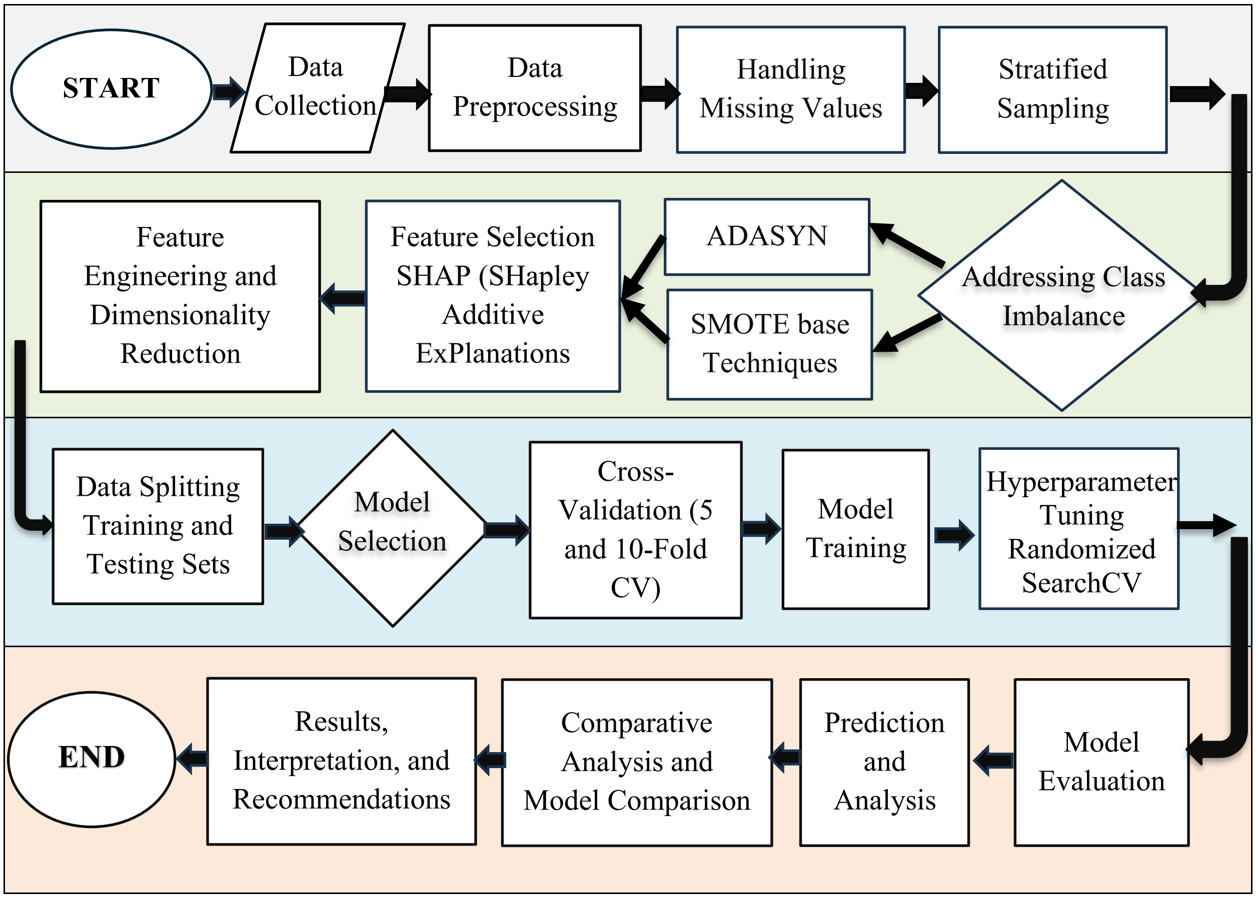


Figure 4.1: The methodology flow chart of the thesis

Lawal and Akintola (2021) introduced Recurrent Neural Networks (RNNs) alongside data balancing techniques to enhance backorder prediction accuracy in imbalanced datasets, evaluated via AUC, precision, and recall. Ali et al. (2024) conducted experiments using models built on all 22 features and a reduced top 5 feature set, incorporating preprocessing steps such as feature scaling and down sampling, with a focus on accuracy, F1-score, and computational efficiency. Garcia et al. (2022) employed a correlational approach using historical supply chain data, applying machine learning models evaluated with AUC and confusion matrices. In contrast, Carbonneau et al. (2008) compared neural networks with traditional forecasting methods, using Mean Absolute Error (MAE) as the evaluation metric. Dehghan-Bonari et al. (2021) utilized neural networks and Naive Bayes on normalized data, analyzing performance through standard machine learning evaluation metrics. Maitra and Kundu (2023) adopted advanced models like Balanced Bagging Classifiers (BBC), Fuzzy Logic, and VAE-GAN, emphasizing the management of imbalanced datasets and comparing models using metrics such as ROC-AUC and Macro F1-Score. Furthermore, Shajalal et al. (2022) and (2023) tackled class imbalance by employing techniques like ADASYN, SMOTE, and other resampling strategies to balance datasets, using Convolutional Neural Networks (CNN) and Deep Neural Networks (NN) for backorder predictions. Hajek and Abedin (2020) modified the Clustering-Based Under sampling (CBUS) approach to address highly imbalanced datasets, while Santis et al. (2017) compared ensemble learning models and sampling methods, including SMOTE and Random Under Sampling (RUS), to enhance prediction accuracy in imbalanced data scenarios.

4.1. Data Preprocessing and Handling Missing Values

In this study, we have adopted a novel approach in our methodology, particularly in the areas of data preprocessing and handling missing values. Unlike previous studies that chose to eliminate records with missing data (e.g., Ntakolia et al., 2022; Ali et al., 2024), we recognize that these records can hold valuable information. While other researchers have utilized missing value imputation techniques (e.g., Santis et al., 2017; Hajek and Abedin, 2020) and Adana et al. (2019) applied the impute data node using SAS Enterprise Miner, our approach prioritizes the use of more advanced imputation methods. For instance, Iqbal et al. (2023) treated missing data as random and employed linear and logistic regression for imputation, while Maitra and Kundu (2023) implemented model-based imputation. However, conventional approaches such as those recommended by Gao et al. (2022), which rely on straightforward mean, mode, or median replacements, may result in information loss and reduced model accuracy. Our research seeks to improve predictive performance by applying advanced imputation techniques, including Multiple Imputation by Chained Equations (MICE), which are more effective in maintaining data integrity. This approach represents a considerable advancement in preserving valuable information from records with missing values, ultimately enhancing the reliability and precision of our machine learning models.

4.1.1. MICE: Multiple Imputation by Chained Equations

Also known as "fully conditional specification" or "sequential regression multiple imputation," multivariate imputation by chained equations (MICE) is a well-known and ethical technique for dealing with missing data in statistical analysis. Unlike single imputation, MICE generate multiple imputations, allowing for the inclusion of statistical uncertainty associated with the missing values. Furthermore, the chained equations method is highly adaptable, capable of managing variables of different types, such as continuous or binary, as well as handling complex structures like bounded variables and conditional survey skip patterns (Azur et al., 2011). Multiple Imputation by Chained Equations (MICE) is a widely used statistical method for addressing missing data in analysis. It treats each missing value as a random variable and estimates it iteratively by leveraging the relationships among the other variables in the dataset. The key advantage of MICE is its capacity to generate several "complete" datasets by performing multiple rounds of regression-based imputations, effectively capturing the uncertainty related to the missing values. Samad et al. (2022) have emphasized MICE’s ability to handle various types of missing data patterns, such as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). A key role of MICE in data analysis is its capacity to estimate missing data by constructing a model that treats each missing entry iteratively, with the method cycling through each variable and imputing based on observed values. This conditional modeling allows MICE to account for relationships between variables and capture data dependencies, unlike simpler imputation methods like mean or median imputation, which ignore the relationships between variables. As highlighted in the literature, ensemble learning and deep learning approaches, such as gradient boosting within the MICE framework, have been shown to further enhance imputation accuracy (Samad et al., 2022). MICE is becoming increasingly common in machine learning, particularly for preparing datasets before training models. Machine learning algorithms typically cannot handle missing data natively, and MICE is used to impute these missing values, allowing for a completer and more usable dataset for supervised learning tasks like classification and regression. Studies have shown that MICE, when used alongside machine learning models such as random forests or support vector machines, improves the predictive accuracy by preventing loss of data due to missing values (Azur et al., 2011). Moreover, ensemble learning techniques integrated with MICE further enhance its performance by using non-linear regression methods for imputation.

4.2. Addressing Imbalanced Data

Previous studies have emphasized the considerable challenge presented by class imbalance between backordered and non-backordered items, which often impedes the accuracy of prediction models (Santis et al., 2017; Shajalal et al., 2023). Various research efforts have employed techniques like SMOTE and ADASYN to tackle this issue. Additionally, hybrid sampling techniques have been used, which involve combining both oversampling and undersampling methods, as well as adaptive sampling approaches that adjust dynamically based on dataset characteristics. Several studies have addressed class imbalance through different strategies, including Random Under Sampling (Ntakolia et al., 2021; Shajalal et al., 2023), Adaptive Synthetic Oversampling (ADASYN) (Shajalal et al., 2022), Weighted Sampling (Shajalal et al., 2023), SMOTE (Shajalal et al., 2022; Islam and Amin, 2020), Oversampling (Shajalal et al., 2023; Islam and Amin, 2020), Random Down Sampling (Ali et al., 2024; Adana et al., 2019), and Stratified Holdout (Adana et al., 2019). Our research seeks to build upon these techniques by implementing more sophisticated hybrid and adaptive sampling methods to improve the handling of class imbalance, ultimately enhancing the accuracy and reliability of backorder predictions. In this research we have used the SMOTE-ENN (Synthetic Minority Over-sampling Technique with Edited Nearest Neighbors) technique which handles class imbalance by combining oversampling of the minority class (using SMOTE) with undersampling of the majority class (using ENN). This approach refines the dataset to make it more balanced, which is essential for improving model performance on imbalanced datasets. In addition, SMOTE-ENN removes noisy data and ensures the model is not biased toward the majority class.

4.2.1. SMOTE-ENN: Synthetic Minority Over-Sampling Technique - Edited Nearest Neighbor

This over-sampling technique increases the representation of the minority class by generating synthetic samples based on the feature space similarities among existing minority class instances, utilizing the k-nearest neighbors in Euclidean space (De Santis et al., 2017). A synthetic sample is created by adding a minority sample m to a scaled difference vector d, which is multiplied by a random value i within the range [0,1], as expressed by the formula $s = m + d * i$. Here, d is calculated as $d = m - r,$ representing the positive difference between the feature vectors of a randomly chosen minority neighbor r and the minority sample m. (Lawal and Akintola, 2021). Chawla et al., (2002) in their research with the subject *SMOTE: synthetic minority over-sampling technique* have analyzed this technique completely and thoroughly, The Synthetic Minority Over-Sampling Technique (SMOTE) is a popular method for handling imbalanced datasets, especially in classification tasks where the minority class is underrepresented. Instead of merely duplicating the existing minority class samples, SMOTE generates synthetic samples to equalize the dataset. It does this by creating new data points through interpolation between current minority class instances, forming synthetic examples along the lines connecting k-nearest neighbors. This process increases diversity in the data and helps prevent overfitting, which is a common issue with straightforward replication of minority instances (Chawla et al., 2002). The author in their research have emphasizes that SMOTE is widely applied in fields where imbalanced data is common. In medical diagnosis datasets, for example, it is frequently used when one class, such as patients with a rare disease, is much smaller than the other (Shoeibi et al., 2023). In fraud detection, SMOTE is essential for identifying fraudulent transactions, which occur far less often than legitimate ones. Financial datasets, where defaults or rare events constitute the minority class, also benefit from SMOTE as it enhances the performance of predictive models. By creating synthetic data to balance the dataset, SMOTE helps machine learning models such as decision trees, support vector machines (SVMs), and neural networks better handle both majority and minority classes without bias. In data analysis, SMOTE is crucial for enhancing model performance by tackling the issue of class imbalance, which often results in biased predictions (Bader et al., 2019). Imbalanced data tends to cause machine learning models to focus on the majority class, reducing the accuracy for the minority class. SMOTE addresses this by creating synthetic samples, compelling models to recognize patterns in the minority class, which improves generalization and strengthens the classifiers. Additionally, SMOTE's ability to prevent overfitting and avoid under-sampling of the majority class makes it superior to methods relying solely on random oversampling or undersampling. Notably, using SMOTE improves not only classification accuracy but also key metrics like recall, precision, and F1 score for the minority class. Despite its effectiveness, SMOTE has certain limitations. One issue is that it may introduce noise into the dataset, since synthetic samples are generated without fully accounting for the data's underlying distribution. This can lead to overlap between classes, especially when the boundaries between them are unclear. To address this issue, we have used SMOTE-ENN in this research. SMOTE-ENN is a hybrid approach that integrates SMOTE with an undersampling method called Edited Nearest Neighbor (khan et al., 2024). The authors stated that Edited Nearest Neighbor (ENN) identifies the k-nearest neighbors for each observation and compares the majority class among these neighbors to the class of the observation. If the majority class differs from the observation's class, both the observation and its k-nearest neighbors are removed. The SMOTEENN technique combines the principles of SMOTE and Edited Nearest Neighbors (ENN) to address class imbalance. It first generates synthetic samples for the minority class using SMOTE, and then ENN is applied to remove noisy or overlapping samples from both the majority class and the newly created synthetic samples (Vukovic et al., 2024). By combining synthetic data generation with noise reduction, SMOTEENN offers a comprehensive solution to class imbalance, resulting in more robust and generalizable classification models. It is specifically aimed at addressing class imbalance in datasets by first oversampling the minority class using SMOTE, followed by ENN, which eliminates noise and borderline instances from both classes. This combination creates a more balanced and cleaner dataset for machine learning model training. ENN works as a data-cleaning method by evaluating the nearest neighbors of each sample and removing those whose class labels differ from the majority of their neighbors (Bader et al., 2019). This process helps reduce class overlap and filters out noise that may have been introduced during the oversampling process (Batista et al., 2004). SMOTEENN (SMOTE + ENN) is a two-step method that combines the Synthetic Minority Oversampling Technique (SMOTE) and Edited Nearest Neighbors (ENN) to address data imbalances (Satpathy, 2020). In the first step, SMOTE is used to create synthetic data to oversample the minority class. Then, ENN is applied to undersample the majority class by removing noisy or misclassified cases. By combining these strategies, SMOTEENN aims to create a more balanced dataset by increasing the minority class instances and eliminating noisy examples from the majority class. This approach is particularly useful in situations where there is a significant class imbalance, and both oversampling and undersampling are needed to produce a well-balanced and relevant dataset (Lanjewar and Panchbhai, 2024).

4.2.2. SMOTE-SVM: Synthetic Minority Over-Sampling Technique-Support Vector Machine

Ezziane et al. (2022) and Guo et al. (2024), in their research to address the issue of class imbalance, emphasized the use of SMOTE-SVM. In another study, Ahmad Khan et al. (2024) reviewed various papers aimed at resolving class imbalance, covering methods ranging from different data augmentation and resampling techniques to various deep learning and ensemble methods. After analyzing these papers, the authors concluded that SMOTE-SVM is one of the solutions to address class imbalance. The support vector machine (SVM) is a widely used machine learning model primarily designed for classification tasks, though it has also been adapted for regression and applied across various fields. SVMs are particularly effective in complex areas such as handwriting recognition and cancer genomics. The method works by identifying a hyperplane that separates different classes in an N-dimensional space. By utilizing only, a small number of support vectors instead of all training samples, SVM reduces computational complexity and mitigates issues related to high dimensionality. However, one drawback of SVM is that its performance significantly decreases when applied to imbalanced datasets. To address this issue, SVM has been combined with SMOTE to tackle class imbalance problems, and the resulting SMOTE-SVM approach has gained popularity in applications involving spatial datasets (Ahmad Khan et al., 2024).

4.2.3. Borderline-SMOTE: Borderline-Synthetic Minority Over-Sampling Technique (BSM)

Han et al., (2005) in their research introduced the Borderline-SMOTE (BSM) which was as an enhancement to SMOTE. This method focuses on creating synthetic samples near the class boundaries. Borderline-SMOTE targets instances on or near the borderline, as they are more susceptible to misclassification compared to those farther from the boundary. The authors emphasize that by concentrating on these critical areas, BSM can enhance classifier performance and often outperforms the original SMOTE method (Wang et al., 2015). BSM uses SMOTE to identify the k-nearest neighbors for all minority class samples and selects random instances based on the oversampling rate. New synthetic samples are then generated along the borderline to strengthen these critical minority examples. Unlike standard SMOTE, which generates synthetic samples for all minority instances, Borderline-SMOTE targets only those near the class boundary, as they are more influential for improving classification performance (Han et al., 2005; Wang et al., 2015; Smiti and Soui, 2020). In another research Smiti and Soui (2020), In Borderline-SMOTE, minority class instances that are at a higher risk of misclassification receive additional focus during training. The algorithm identifies borderline minority samples and generates synthetic instances by interpolating them with their k-nearest neighbors. When applied to bankruptcy prediction, the Borderline-SMOTE method significantly increases the size of the minority class by creating these new synthetic instances.

4.2.4. ADASYN: Adaptive Synthetic Oversampling Technique

Khan et al. (2024) concluded in their paper that the ADASYN has been used to address the CI problems in areas like modeling recreational water quality, warning systems for harmful algae blooms, and design of wireless intrusion detection systems after analyzing numerous studies in various fields of research. An extension of the SMOTE method, the adaptive synthetic (ADASYN) sampling approach aims to oversample the minority class by creating artificial examples. The ADASYN algorithm uses a weighted distribution for various minority class examples based on their learning difficulty, where more synthetic data is generated for minority class examples that are harder to learn than those that are easier to learn. This is in contrast to SMOTE, which generates an arbitrary number of synthetic minority examples to correct the imbalance in the dataset (He et al., 2008). This approach is essential as the data generated by the algorithm will not only ensure a balanced representation of class distribution, but it will also force the learning algorithm to focus on those difficult to learn examples (Lawal and Akintola, 2021). In another study, Shajalal et al., (2022) in their paper discussed on the Handling class imbalance with ADASYN. The authors have mentioned, product backorders are rare events that result in a highly imbalanced dataset. To address this, they employed ADASYN (Adaptive Synthetic Oversampling), an effective oversampling technique designed to balance datasets. ADASYN generates synthetic samples for the minority class, with a focus on samples that are more difficult to classify. This method adapts the generation of synthetic samples based on the density distribution of the minority class, ensuring that more challenging minority examples receive higher weights during sample generation) (He et al., 2008). Given a training dataset D_{train} with *N* samples, where each sample is denoted as (x,y), the vector *x* represents a *K*-dimensional feature vector of an ordered product, and *y* is the binary label (0 for non-backordered and 1 for backordered). Let m_{min} and m_{maj} represent the number of samples in the minority and majority classes, respectively, such that m_{min} and m_{maj} = *N*, and for backorder prediction, $m_{min} << m_{maj}$. The ADASYN algorithm balances the dataset by generating *G*, the number of synthetic samples required, based on the degree of imbalance *d* and a user-defined imbalance ratio β (β∈[0,1]). A value of β=1 indicates the dataset will be fully balanced. For each minority sample x_i , ADASYN calculates the difficulty ratio r_i using *K-nearest* neighbors and Euclidean distance, where Δ_i is the number of nearest neighbors from the majority class. The normalized ratio r_i^ determines the number of synthetic samples g_i to be generated for each x_i. Finally, synthetic samples are generated by adding a random perturbation to the minority sample along the distance vector to its nearest neighbors, represented as $\lambda \cdot (x_{nn} - x_i)$, where λ is a random value in [0,1] and x_{nn} is a nearest neighbor. The following steps show ADASYN algorithm.

Input:

- D_{train}: Training dataset with N samples
- m_{min}: Number of minority class examples
- m_{maj}: Number of majority class examples
- β ∈[0,1]: Desired balancing ratio (default: β=1)
- *K*: Number of nearest neighbors (default: 5)

Output:

- A balanced dataset with synthetic minority class examples added

Steps:

1. Calculate the Degree of Imbalance (d):

$$d = \frac{m_{maj} - m_{min}}{m_{min}} \tag{1}$$

2. Determine the Number of Synthetic Examples (G):

$$G = \beta \times (m_{maj} - m_{min}) \tag{2}$$

3. For Each Minority Example (x_i):
   - Identify the K-nearest neighbors using Euclidean distance:

$$d\,(i) = \|x_i - x_{nn}\| \text{ for } x_{nn} \in K \tag{3}$$

   - Count the number of neighbors (Δ_i) belonging to the majority class.
   - Compute the Difficulty Ratio (r_i) for each x_i

$$r_i = \frac{\Delta_i}{K} \tag{4}$$

4. Normalize the Difficulty Ratios:

$$r_i \wedge = \frac{r_i}{\sum_{k=0}^{m_{min}} r_j} \quad (5)$$

5. Compute Synthetic Examples (gig_igi) for Each xix_ixi:

$$g_i = r_i\wedge \times G \quad (6)$$

g_i represents the number of synthetic examples to generate for x_i.

6. Generate Synthetic Examples:
   - For each g_i, generate a synthetic sample $x_{\text{synthetic}}$

$$x_{\text{synthetic}} = x_{\text{i}} + \lambda.(x_{\text{nn}} - x_{\text{i}}) \quad (7)$$

   - x_{nn} is a randomly selected nearest neighbor of x_{i}.
   - λ is a random number sampled from [0,1].
7. Add Synthetic Examples to the Dataset:
   - Append G synthetic examples to the minority class to create the balanced dataset.

4.2.5. Random Under Sampling (RUS)

Random under-sampling (RUS) is a straightforward strategy used to address class imbalance in datasets by equalizing the class distribution. This technique involves eliminating instances from the majority class within the training data (Ahmad Khan et al., 2024). The under-sampling method achieves balance by randomly selecting a subset of examples from the majority class while preserving all instances of the minority class. However, a notable drawback of this approach is the potential loss of important information from the majority class, which could negatively impact the model's overall performance when applied to the complete dataset (De Santis et al., 2017). The random under-sampling (RUS) technique seeks to reduce the number of majority class instances in cases where one class significantly outnumbers the other in a highly imbalanced dataset. Let T represent the training dataset, N denote the majority class examples, and P signify the minority class examples. In such scenarios, RUS addresses the imbalance by decreasing the size of N, thereby mitigating the disproportion between N and P. Given the rapid expansion of datasets and their features, under-sampling often presents a more efficient alternative to oversampling methods for balancing class distributions (Hajek and Abedin, 2020).

4.3. Feature Selection Using SHAP (SHapley Additive ExPlanations)

In this research we have used SHAP method to explain the output of machine learning models. It is based on Shapley values from cooperative game theory, which attribute the contribution of each feature to the final prediction. In the context of feature selection, SHAP values indicate how much each feature contributes to the model's predictions, helping to identify irrelevant or less important features. This can be used to reduce dimensionality, improve model interpretability, and enhance model performance. Ntakolia et al. (2022) used the SHAP technique in their research. The authors stated that, “to explain the predictive model and the contribution of the most important features, Shapley Additive Explanations (SHAP) were adopted. SHAP is a game-theoretic approach that explains the output of any machine learning model. It connects optimal credit allocation with local explanations using classic Shapley values from game theory and their extensions” Ntakolia et al. (2022).

SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee (2017), provides a unified approach to explain and interpret machine learning model predictions. SHAP values, derived from Shapley values in cooperative game theory, quantify the contributions of individual features to the prediction for a particular instance x. These contributions can be seen as positive or negative impacts on the model's prediction (Shajalal et al., 2022). SHAP values are computed to create both local explanations (for individual predictions) and global explanations (for the overall model). Within Explainable AI (XAI), SHapley Additive exPlanations is a commonly used technique that provides comprehensible and consistent explanations for machine learning models. SHAP, which has its roots in cooperative game theory, gives each characteristic an importance value that represents how much it contributes to the model's predictions. By dissecting intricate and non-linear model behaviors into easily comprehended parts, this method improves transparency and is a potent method for deciphering black-box models in a variety of domains. SHAP promotes confidence in the model's judgments by measuring the impact of each characteristic, which makes it easier to make better selections. SHAP assigns each feature an importance value for a specific prediction. Its key innovations include: (1) identifying a new class of additive feature importance measures, and (2) providing theoretical results that demonstrate the existence of a unique solution within this class that satisfies a set of desirable properties (Lundberg and Lee, 2017). One of the key technical innovations of SHAP is its ability to handle complex models by decomposing predictions into additive feature contributions, which sum to the actual output of the model. This additive decomposition makes SHAP particularly versatile, allowing it to be applied to a broad range of machine learning models, including tree-based models like XGBoost and random forests, as well as deep learning architectures. Chen et al. (2020) highlights the advantages of SHAP in tree-based models, demonstrating that TreeSHAP, an optimized algorithm for tree ensemble methods, allows for efficient and consistent computation of Shapley values, overcoming the computational limitations that often-hindered earlier Shapley-based methods. In this study, we utilize SHAP to assess the contribution of features to the performance of our predictive model for the case of Backorder. In SHAP, each feature's importance is represented by the SHAP value, ϕ_j, where a higher ϕ_j indicates greater influence of feature j on the model's output. For a specific instance x, the SHAP values can be interpreted through an additive feature importance model:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z_j' \tag{8}$$

where:

- g is the explanation model that approximates the original model f by using SHAP values.
- z′ is a simplified binary vector (or coalition vector) indicating which features are present (1) or absent (0) in the instance.
- ϕ_0 represents the base value or expected prediction over all instances.
- $\phi_j$is the SHAP value for feature j, reflecting its contribution to the prediction.

To calculate ϕ_0, the Shapley value formula from game theory is applied, which assesses each feature's marginal contribution across all possible combinations of features. For a given predictive model f, the SHAP value ϕj for feature j can be computed as:

$$\phi_j = \sum_{z' \subseteq x' \setminus \{j\}}^{M} \phi_j z_j' \left(\frac{(|z'|!(M-|z'|-1)!}{M!} (f(z' \cup \{j\}) - f(z') \right) \tag{9}$$

where:

- $z' \subseteq x' \setminus \{j\}$ represents all subsets of features excluding j.
- M is the total number of features.

- $f(z' \cup \{j\}) - f(z')$ represents the change in the model's prediction when feature j is included in the subset z', providing the marginal contribution of j.

This game-theoretic approach allows SHAP to isolate each feature's impact on the model's decision by evaluating its contribution in various combinations of features, providing insights for model interpretability.

4.4. Feature Engineering and Dimensionality Reduction

In this study, Principal Component Analysis (PCA) is employed as a robust technique for feature engineering and dimensionality reduction. Previous research, such as by Adana et al. (2019) and Santis et al. (2017), mainly relied on traditional or manual feature selection methods; however, our approach leverages PCA to enhance model performance. PCA transforms the high-dimensional dataset into a smaller set of uncorrelated principal components, optimizing machine learning models like Support Vector Machines (SVM) and Neural Networks by addressing multicollinearity issues and reducing computational complexity. PCA is a widely used statistical technique that efficiently reduces the dimensionality of large datasets while preserving most of the original variability (Jolliffe and Cadima, 2016). This process not only improves model accuracy but also alleviates the "curse of dimensionality," where an increase in dataset features requires exponentially more data to achieve statistical significance (Bellman, 1961). High-dimensional data often lead to challenges such as overfitting, extended computational time, and reduced model accuracy. By implementing PCA, we reduce the feature space complexity, making clustering and classification tasks more efficient and less resource-intensive (Abdi and Williams, 2010). In this study, the application of PCA allows for a more streamlined and effective solution to backorder prediction by focusing on the most impactful features. By integrating PCA into our methodology, we ensure that our machine learning models operate with greater efficiency and predictive accuracy, ultimately enhancing the overall performance of our backorder prediction framework.

Principal Component Analysis (PCA) is a statistical method that applies an orthogonal transformation to convert a group of correlated variables into a set of uncorrelated variables. PCA is one of the most used techniques in both exploratory data analysis and machine learning, particularly for building predictive models. Principal Component Analysis (PCA) generates new features that are linear combinations of the original features. In a d-dimensional space, PCA transforms the dataset into a new k-dimensional space, where $k < d$. These new features are called principal components (PCs), and each PC captures the maximum variance in the data, with successive components capturing progressively less variance. The first principal component (PC_1) retains the largest variance, while subsequent components retain decreasing amounts of variance. Each principal component can be mathematically represented as:

$$PC_1 = a_1 X_1 + a_2 X_2 + \ldots + a_d X_d \tag{10}$$

Here, X_j represents the original features, and a_j denotes the coefficients (or weights) for the *j-th* original feature in the calculation of the *i-th* principal component. These coefficients are derived from the eigenvectors of the covariance matrix of the original dataset, ensuring orthogonality between the components and a hierarchical variance structure. As an unsupervised learning method, PCA investigates the relationships between variables, functioning similarly to a general factor analysis where regression is used to find the best-fitting line. The primary objective of PCA is to reduce the dimensionality of a dataset while retaining the key patterns or relationships among the variables, without requiring prior knowledge of the target variables (GeeksforGeeks, 2024). PCA preserves the most significant details by retaining the most representative measurements and discarding smaller, less relevant ones. It generates new features that are linear combinations of the original variables, transforming data from a d-dimensional space to a k-dimensional space, where k is less than d. These new features, known as principal components (PCs), capture the maximum variance in the data, with the first component accounting for the highest variance and each subsequent component capturing progressively less variance. PCA is traditionally used for dimensionality reduction, but more recently, it has been applied to identify and eliminate redundancies in different layers of neural networks by pruning redundant features (Sudharsan and Thailambal, 2023; Chakraborty et al., 2020).

4.5.Machine Learning Models

In this step, since the data preprocessing has been completed, the process moves to Building and Training Machine Learning Models. The dataset has been split into training and testing sets to ensure the models can be validated accurately. Various machine learning models, including Neural Networks, K-Nearest Neighbors (KNN) and Random Forest, and Modern Gradient Boosting models like XGBoost and LightGBM, Gradient Boosting, will be implemented and trained on the data. Ensemble learning technique such as Model Stacking will then be employed to combine the strengths of these individual models, resulting in more accurate and robust backorder predictions. Each model's hyperparameters will be carefully tuned to optimize performance, and the models will be evaluated using primary metrics such as ROC-AUC, PR-AUC, F1 Score and confusion matrix and secondary metrics such as balanced accuracy, Specificity (True Negative Rate). Implementing this approach will lead to a comprehensive comparative analysis, where the predictive performance of all models will be analyzed to identify the most effective ones for backorder forecasting.

4.5.1. Neural Networks

Carbonneau et al. (2008) explained that while artificial neural networks encompass various types, their study focused on the widely used feed-forward error back-propagation neural networks. In these architectures, individual components (neurons) are arranged in layers, where the output signals from neurons in one layer are transmitted to all neurons in the subsequent layer. This setup ensures that neural activations flow exclusively in one direction, moving layer by layer. The simplest configuration involves two layers: an input layer and an output layer. However, additional layers, known as hidden layers, can be inserted between the input and output layers to enhance the network's computational capabilities. With enough hidden units, a neural network has the potential to function as a "universal approximator".

Neural networks are computational models inspired by the structure and functioning of the human brain. They consist of layers of interconnected nodes (often referred to as neurons) that process input data to identify patterns and relationships. These networks simulate how the human brain learns and generalizes information, attempting to model complex relationships within data. Neural networks can refer to both biological neuron-inspired architectures and artificial implementations used in machine learning and data analysis. Figure 4.2 illustrates the general architecture of a neural network, typically comprising an input layer, one or more hidden layers, and an output layer, with each layer performing specific computations to transform the input into the desired output (Shaik et al., 2021)

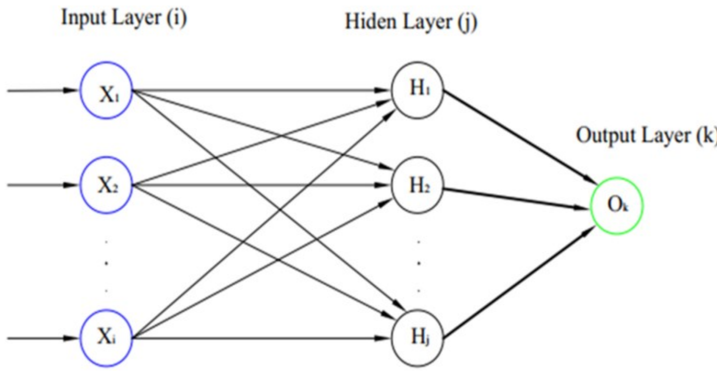


Figure 4.2: General Neural Network structure (Shaik et al, 2021)

Carbonneau et al, (2008) considered the Neural Networks (NN) as a powerful tool for handling complex, non-linear relationships in demand forecasting. They emphasize that NNs have the ability to model intricate patterns in data, making them highly suitable for forecasting in unpredictable and dynamic environments, such as supply chains. The authors highlight that neural networks are particularly effective when traditional methods fail to capture the complexity of real-world data. In their experiments, neural networks showed competitive performance compared to traditional methods. However, similar to Support Vector Machines (SVM), the authors caution that while NNs perform well, the marginal improvement over simpler models like multiple linear regression may not always justify the additional computational complexity and training required for neural networks.

The core functions of neural networks, regardless of type, include receiving data from external sources, determining whether the data is significant enough to be considered or discarded as irrelevant, minimizing errors through iterative processing, and ultimately producing an output or performance result for the trial. In an artificial neuron, the initial step involves summing various inputs (x_i) after multiplying each by its corresponding weight (w_i). These weighted inputs ($x_i w_i$) are then processed through a summation function, followed by iterative adjustments to minimize errors (Profillidis and Botzoris, 2019).

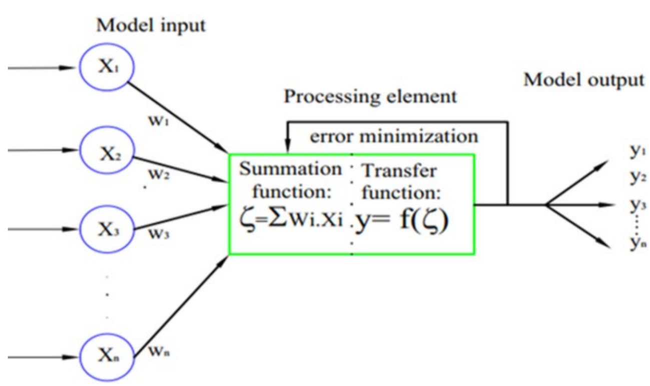


Figure 4.3: The basic function of a neural network (Shaik et al., 2021)

4.5.1.1. Back propagation

Shaik et al (2021) in their research have mentioned Back-propagation, first introduced by Paul Werbos in 1974 and later popularized by Rumelhart, Hinton, and Williams in 1986, is one of the most widely used learning algorithms for training neural networks. It is particularly common in multilayer perceptron (MLP) frameworks, which account for a significant proportion of neural network applications (Garrido, 2014). Back-propagation operates as a supervised learning algorithm that trains MLPs using gradient descent to minimize the error between the network's predicted output and the target output. During the learning process, the error is calculated at the output layer and propagated backward through the network, enabling the adjustment of weight coefficients in a manner that minimizes this error. The goal of back-propagation training is to iteratively update the weights between neurons to achieve an optimized error function (Pradhan and Sameen, 2020). The error function E_i is minimized using the following equation:

$$E_f = 0.5N \sum_i^N (p^{\wedge}{}_{\mathrm{i}} - p_{\mathrm{i}})^2 \tag{11}$$

(5)

$p_i = Target\ Value$
$p^{\wedge}{}_{\mathrm{i}}$ = Calculated Output
I = Output Layer
N = Number of Nodes (Output Layer)

4.5.1.2. Feed Forward Neural Network

The feed-forward neural network (FFNN) is among the earliest types of artificial neural networks developed. In this architecture, data flows exclusively in a forward direction across the layers: from the input layer, through the hidden layers, and finally to the output layer. This structure is referred to as a feed-forward network because it lacks feedback loops, meaning the model's output is not cycled back into the network Shaik et al (2021). A feed-forward neural network (FNN) typically comprises an input layer, one or more hidden layers, and an output layer. FNNs are capable of learning and modeling complex input-output relationships without requiring explicit mathematical formulations to describe such mappings. The learning process involves iterative adjustments to the weights and biases in the network using optimization techniques, such as the steepest descent method, commonly implemented through backpropagation (BP). This process minimizes the error between the network's predicted output values and the actual target values, culminating in the completion of training. FNNs allow only forward connections, meaning data flows in a unidirectional manner from the input layer through the hidden layers to the output layer. The activation of neuron *i* in layer *l* (a_i^l) is calculated as follows:

$$a_i^l = f(n_i^l) \tag{12}$$

$$n_i^l = \sum_{j=1}^{N_i^l} w_{ji}^l a_j^{l-1} + b_i^l \tag{13}$$

Which f represents the activation function, w_{ji}^l denotes the weights associated with each connection, and b_i^l is the bias. Given that the nonlinear response of the macroscale model depends on the loading history and the current loading increment, the macroscopic stress and strain from the previous step, along with the current incremental strain, are utilized as input parameters. The neural network's overall architecture can be outlined as follows:

$$\Delta\sigma_{ij}^{t+1} = F_{FNN}(\sigma_{ij}^t, \varepsilon_{ij}^t, \Delta\varepsilon_{ij}^{t+1}) \tag{14}$$

Where F_{FNN} is FNN that maps inputs $\sigma_{ij}^t, \varepsilon_{ij}^t, \Delta\varepsilon_{ij}^{t+1}$ and $\Delta\sigma_{ij}^{t+1}$ outputs.

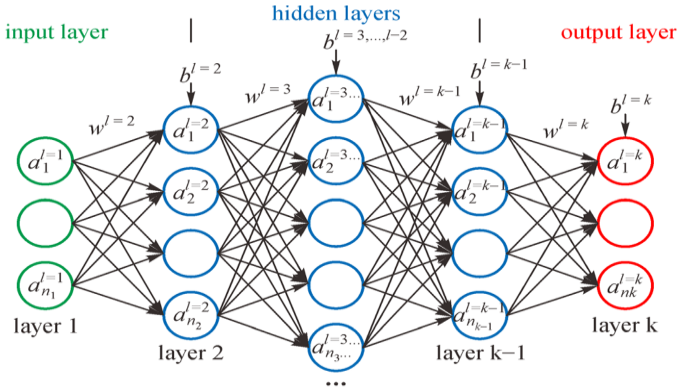


Figure 4.4: General layout of FNN (Li and Zhuang, 2020)

4.5.2. K Nearest Neighbours

KNN (K-Nearest Neighbors) is a widely utilized classification technique in pattern recognition. This algorithm determines the similarity of a given object to its surrounding objects (referred to as tuples) by creating a similarity index. Each tuple is characterized by n attributes, meaning it represents a point within an n-dimensional space. The KNN algorithm identifies the k tuples that are nearest to the target tuple, leading to the creation of clusters that group similar objects together. Additionally, KNN can be applied in regression analysis for the purpose of reducing data dimensionality (Seyedan and Mafakheri, 2020). The k-Nearest Neighbors (k-NN) algorithm is a non-parametric method widely applied in estimation and pattern recognition tasks. It predicts the output value of a new input vector by considering the outputs of its k nearest neighbors in the dataset. The similarity between data points is typically calculated using a distance function, with commonly used metrics including Euclidean distance and Mahalanobis distance. Depending on the specific implementation, the output for a new sample is determined either through a simple average (normal averaging) or by applying weights to its k nearest neighbors, where closer neighbors often have higher weights (Abbasi et al., 2020). The k-nearest neighbors (KNN) algorithm is a memory-based classification method that does not require a fitted model, making it a non-parametric technique. To classify a query point x0x_0x0, the algorithm identifies the k closest training points, $F_{(r)}$, r = 1, ..., k, based on a chosen distance metric, typically Euclidean distance in the feature space, defined as $d_i = || x^i - x_0||$. The class of x_0 is then determined by a majority vote among these k nearest neighbors. In cases of a tie, the class is chosen randomly.
When k=1, each query point is classified according to its nearest neighbor, resulting in a highly irregular decision boundary. Increasing k tends to smooth the decision boundary, as more neighbors contribute to the vote. KNN is particularly effective in classification tasks where each class has numerous prototypes, and the decision boundaries are complex. It has been applied successfully in various domains, such as classifying handwritten digits, interpreting satellite images, and analyzing EKG (electrocardiogram) patterns (Tibshirani, 2017).

4.6.Ensemble Learning

Ahmad Khan et al (2024), in their research have focused on the exploration of ensemble learning methods as a powerful approach to improving predictive performance in addressing class imbalance (CI) problems within supply chain management and backorder prediction. The authors stated that “Ensemble learning combines multiple models to construct a more robust and comprehensive predictive framework, often outperforming single-algorithm approaches. Bagging, boosting, and stacking, with prominent implementations like AdaBoost and random forests are the example of ensemble learning" (Ahmad Khan et al., 2024). In the context of our research, particular attention is given to the application of Random Forest, XGBoost, LightGBM, GBM, and Stacking methods, highlighting their effectiveness in handling complex and imbalanced datasets while enhancing model accuracy and reliability.

4.6.1. Random Forest

Using a bootstrap sample, Random Forest is a tree-based ensemble in which batches of training data are drawn with replacement. The optimum split among a random subset of the characteristics is chosen during tree construction, creating a randomness that favors the forest's performance over that of a single non-random tree. By combining the probabilistic predictions of the base classifiers, the bias increase is counterbalanced by an average variance decrease. (De Santis et al., 2017). Random Forest as an ensemble learning technique composed of multiple decision trees combines the principles of bagging (Bootstrap Aggregating) and random subspace methods to enhance predictive accuracy. In this approach, N regression trees are created, with each tree trained on a bootstrapped sample of the original dataset. Additionally, at each node of the decision trees, a random subset of the original features is selected for splitting. This dual-randomization strategy reduces correlations among the individual regression trees. By averaging the predictions of these decorrelated trees, Random Forest effectively reduces error variance, leading to more robust and accurate predictions. Umoh et al., (2022), in their research described the Random Forest as an ensemble machine learning technique used for both classification and regression tasks. The authors have mentioned it employs bagging (bootstrap aggregation), a method of creating new datasets by sampling with replacement from an existing dataset. Random Forest offers the advantages such as prevention of Overfitting, effective with smaller datasets, parallel training, and automatic feature selection. The decision tree structure within Random Forest inherently ranks and selects features, focusing on the most informative ones during training.

In the current research, Random Forest can be used to classify backorder prediction data into two categories: a negative class (non-backordered) and a positive class (Backordered). The random forest algorithm is presented below: "STEP 1: Randomly select k features from the total m features, where $k \ll m$; STEP 2: Among the "k" features, calculate the node "d" using the best split point; STEP 3: Split the node into daughter nodes using the best split; STEP 4: Repeat 1 to 3 steps until the "p" number of nodes has been reached; STEP 5: Build the forest by repeating steps 1 to 4 for "n" number of times to create "n" number of trees" (Umoh et al., 2022; Arowolo et al., 2023).

4.6.2. Gradient Boosting

Gradient Tree Boosting is a boosting-based ensemble that uses an arbitrary differentiable loss function. It is utilized in a range of applications, including web search ranking and ecology. The technique is quite competent of handling heterogeneous attributes and is robust to outliers. However, its main weakness is that it can rarely be parallelized (De Santis et al., 2017).

Khan et al., (2024) in their research have stated that the Gradient Boosting is widely recognized for its exceptional accuracy and its capability to process large datasets effectively. Gradient Boosting classifier, often referred to as Gradient Boosting Machines (GBM), minimizes the log-likelihood loss function by iteratively adding models trained on the residual errors of the preceding models. In this ensemble learning approach, M represents the total number of boosting stages, while m denotes the current stage. The final model is denoted as F_M(x), and F_m(x) represents the model obtained after incorporating mmm-stage base learners. The boosting process begins with a high-bias initial model, F_0 = γ, and progressively reduces bias by sequentially adding models from m = 1 to m = M. At each stage, the model F_{m-1}(x) is enhanced by incorporating weighted base learners, and pseudo-residuals are calculated for each training example i. The loss function L, which guides the optimization process, is computed as shown in the following equation (Khan et al., 2024; and Chen and Guestrin, 2016):

$$L = \frac{1}{n} \sum_{i=1}^{n} (y_i - F(x_i))^{2} \quad (15)$$

Which the y_i is the observed Value.

$$r_{im} = -\frac{\partial L(y,\ F_{m-1}(\mathrm{x}))}{F_{m-1}(\mathrm{x})} \,|x = x_i, y = y_i \ \forall i = 1, 2, \ldots, n \quad (16)$$

Which in this equation, each residual calculation r_{im} has been computed for ith training example to the current base learner m on the weighted sum of base learners from 1 to m-1, and the initial constant function. Then a new dataset has been generated from the original dataset and train (fit) the base learner $h_m(x)$ as shown in following equation:

$$D = \{(x_i, \gamma_{im}) : i = 1, 2\ldots, n\} \quad (17)$$

$$\gamma_{im} = argmin \sum_{i=1}^{n} L\,(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \quad (18)$$

Therefore, the $F_m(x)$ will be calculated using following equations:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x) \quad (19)$$

The function $h_m(x)$ is fitted to approximate the rate of change of the loss function L with respect to $F_{m-1}(x)$. This function $h_m(x)$ provides an estimation of the derivative of the loss function concerning$F_{m-1}(x)$, indicating the direction in which the loss decreases. To optimize this process, the parameter $\gamma_{\mathrm{optimum}}$ is determined by solving the optimization problem formulated in Eq. ().

$$\gamma_{\mathrm{optimum}} = argmin \sum_{i=1}^{n} L\left(y_i, F_m(x_i)\right)$$
$$= argmin \sum_{i=1}^{n} L\,(y_i, F_{m-1}(x) + \gamma h_m(x)) \quad (20)$$

The procedure has been run for each base model m = 1 to M and after the M iterations, the final model $F_M(x)$ has been obtain in the following equation:

$$F_m(x) = F_{m-1}(x) + \gamma_{\mathrm{optimum}} h_m(x) \quad (21)$$

(Khan et al., 2024; and Chen and Guestrin, 2016)

4.6.3. XGBoost Extreme Gradient Boosting

XGBoost (eXtreme Gradient Boosting) is a scalable, efficient, and highly optimized tree boosting system. It improves upon traditional gradient boosting methods by introducing innovations like sparsity-aware algorithms, weighted quantile sketches, and parallel computing. XGBoost scales seamlessly to billions of examples while maintaining computational efficiency, and its flexibility allows it to handle both large and sparse datasets effectively. It has been extensively applied in areas such as ad click-through rate prediction, customer behavior analysis, and high-energy physics experiments (Chen and Guestrin, 2016). The loss function and regularization for XGBoost at the t-th iteration can be expressed mathematically as shown in the Eq.

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}i^{(t-1)} + F_t(x_i)) + \Omega(f_t)\right) \quad (22)$$

$$\Omega(f) + \gamma^{T} + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

Where,

$$\hat{y}_i = \sum_{k=1}^{k} f_k(x_i),\ f_i \epsilon F \tag{23}$$

In XGBoost, *K* represents the number of trees, and *f* denotes the functional space of *F,* encompassing all possible classification and regression trees. To facilitate the use of conventional optimization methods, XGBoost uses a Taylor approximation to convert the original objective function into a form compatible with the Euclidean domain. During the t-th iteration, the goal is to train a model that maximally reduces the loss, as specified by the following equation (Khan et al., 2024; and Chen and Guestrin, 2016).

$$L^{\sim(t)}(q) = -\frac{1}{2}\sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma^T \tag{24}$$

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}i^{(t-1)}) \tag{25}$$

$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}i^{(t-1)}) \tag{26}$$

Like other ensemble learning techniques, XGBoost has certain limitations, including sensitivity to parameter selection, necessitating careful fine-tuning for optimal performance. Although XGBoost is a complex ensemble model that incorporates specialized decision trees, it remains interpretable by providing insights into feature importance. This capability makes XGBoost particularly useful for identifying the most significant features in a dataset, facilitating feature selection and improving the understanding of underlying relationships within the data (Khan et al., 2024).

4.6.4. LightGBM

LightGBM, shares similarities with XGBoost in using gradient boosting for decision trees. LightGBM differs primarily in its use of histogram-based algorithms for efficient split finding and its focus on speed and memory efficiency. Unlike XGBoost, it uses a leaf-wise growth strategy, which can lead to deeper trees and faster convergence but might be overfit in small datasets (Chen and Guestrin, 2016). To address the inherent drawbacks of GBDT, an enhanced version known as the Light Gradient Boosting Machine (LightGBM) has been introduced. LightGBM integrates Gradient-based One-Side Sampling (GOSS) with Exclusive Feature Bundling (EFB) to create a faster, distributed, high-performance, and efficient gradient boosting framework (Awe and Vance, 2020). Unlike random forests, which build trees independently for each sample, LightGBM constructs trees sequentially within the gradient boosting framework. LightGBM employs a leaf-wise tree growth algorithm, which splits the tree leaf-by-leaf rather than level-by-level, resulting in an unbalanced tree structure. This method uses information gain to determine splits at each node (Datta et al., 2022; Sai et al., 2023).

4.6.5. Stacking

Stacking is an ensemble learning technique that combines predictions from diverse base learners using a meta-learner to produce final predictions. The base learners, which can include models like SVMs, Neural Networks, or Decision Trees, are trained using k-fold cross-validation. For each fold, the dataset is split into k−1 folds for training and 1-fold for validation. During this process, each base learner generates predictions for the validation fold, creating a matrix of size $\frac{m}{k}n_i$ where m is the total number of samples, and n is the number of base learners. This matrix represents the outputs of the base learners and serves as the input for the meta-learner, which is subsequently trained to combine these predictions optimally. By leveraging the strengths of heterogeneous models and a higher-level learner, stacking enhances predictive accuracy and generalization. (Gaye et al., 2021). Belouch and Hadaj (2017) conducted a study in which they examined three ensemble learning techniques—boosting, bagging, and stacking—in an effort to increase the detection rate and lower the false alarm rate. According to their findings, bagging and boosting can both outperform single classifiers in terms of accuracy, whereas stacking outperforms other ensemble learning techniques.

Stacking, or stacked generalization, is an ensemble learning technique that combines the outputs of multiple diverse classifiers. Unlike bagging and boosting, which typically use homogeneous models, stacking integrates various types of classifiers, such as decision trees, neural networks, rule induction, naïve Bayes, and logistic regression. Stacking operates in two levels: the base learners (level-0) and the stacking model learner (level-1). The base learners are trained on the original dataset, and their predictions are aggregated to form a new dataset, where each instance is paired with the true target value it aims to predict. This new dataset is then used to train the stacking model learner, which synthesizes the base learners' outputs to produce the final prediction. By combining the strengths of different classifiers, stacking often achieves higher predictive accuracy and robustness than individual models (Belouch and Hadaj, 2017).

4.7.Overview of hyperparameter tuning: RandomizedSearchCV and GridSearchCV

A crucial stage in the practical application of machine learning is hyperparameter optimization. Hyperparameter tuning is a technique used to optimize the parameters within machine learning algorithms, facilitating the efficient development of models and improving classification accuracy (Jamaleddyn et al., 2023). Feurer et al. (2015) emphasized that model selection and hyperparameter optimization are critical when applying machine learning to novel datasets. Referring to prior research, they noted, “Hyperparameter optimization is a crucial step in the practical application of machine learning algorithms. Manually finding suitable hyperparameter settings is often a time-consuming and tedious process, requiring many ad-hoc decisions by practitioners. Consequently, recent research in machine learning has increasingly focused on developing improved hyperparameter optimization methods” (Hutter, Hoos, and Leyton-Brown, 2011; Bergstra et al., 2011; Snoek, Larochelle, and Adams, 2012; Bergstra and Bengio, 2012). Selecting an effective hyperparameter configuration for a machine learning model requires specialized knowledge, intuition, and often, trial and error. The optimization of these hyperparameters is typically formulated to maximize the predictive capability of the model. Various strategies have been proposed for hyperparameter optimization in classification algorithms, with grid search and random search being two widely used approaches. GridSearchCV and RandomizedSearchCV are essential techniques in machine learning for hyperparameter tuning, enhancing model accuracy and performance by systematically selecting optimal parameter values. GridSearchCV conducts an exhaustive search across predefined hyperparameter combinations, which, while thorough, becomes computationally expensive as hyperparameter dimensions grow. RandomizedSearchCV, in contrast, samples a subset of random hyperparameter combinations, making it more efficient in high-dimensional spaces while often yielding comparable performance to GridSearchCV (Bergstra et al., 2011; Hutter et al., 2014). *Grid search* is a tuning technique designed to identify optimal hyperparameter values through an exhaustive search of specified parameter values for the model. This traditional approach to hyperparameter optimization systematically searches a defined subset of the algorithm’s hyperparameter space. *Random Search*, by contrast, involves sampling random sets of hyperparameters to locate an optimal configuration. While similar to grid search, it has been found to yield comparatively better results in some cases. However, random search can introduce significant variance in computational outcomes. Results indicate that random search outperformed grid search in terms of accuracy, precision, recall, and F1-score, while also delivering faster execution times for some algorithms (Jamaleddyn et al., 2023). Bergstra and Bengio (2012) illustrate the efficiency of RandomizedSearchCV, especially in scenarios where only a subset of hyperparameters significantly impacts model performance. The "curse of dimensionality" associated with GridSearchCV requires a prohibitive number of trials as parameters increase, often resulting in misallocated computational resources. In contrast, RandomizedSearchCV's random sampling approach explores a larger portion of the parameter space with fewer evaluations, mitigating over-sampling of non-critical hyperparameters. This approach allows for an efficient allocation of resources without exhaustive searches, which is advantageous in models with complex interactions, like deep neural networks. In practical applications, both methods are frequently used for models such as Support Vector Machines (SVM), Decision Trees, and Neural Networks. For instance, in SVM, GridSearchCV may be employed to fine-tune parameters like regularization strength and kernel type; however, RandomizedSearchCV is preferred in high-dimensional tuning scenarios for deep neural networks (Hutter et al., 2014). Cross-validation is used in both methods to ensure that selected hyperparameters generalize well to unseen data, but for more complex settings, Bayesian optimization and Sequential Model-Based Optimization (SMBO) offer greater efficiency by concentrating on promising regions within the hyperparameter space (Bergstra and Bengio, 2012). Furthermore, RandomizedSearchCV's asynchronous and flexible nature allows it to operate effectively under constrained or variable computational resources, contrasting with the fixed grid configuration of GridSearchCV. This adaptability makes RandomizedSearchCV especially useful in distributed or cloud environments, where tasks may need to be paused or reallocated without compromising experimental integrity (Bergstra and Bengio, 2012). Both GridSearchCV and RandomizedSearchCV serve as robust baseline methods in machine learning, setting standards for more advanced hyperparameter tuning techniques (Hutter et al., 2014). Pérez-Padilla et al. (2024) utilized fine-tuned machine learning models with a randomized search cross-validation algorithm to optimize trigger day timing in minimal ovarian stimulation protocols. Their findings suggest that hyperparameter tuning through random search significantly improved model accuracy. Model training was conducted using the scikit-learn library, with hyperparameter optimization performed through a randomized search cross-validation approach. This method generated 50 random samples of parameter configurations, selecting the optimal combination based on Root Mean Squared Error (RMSE) as the scoring metric to compare different configurations. To mitigate overfitting, they employed a cross-validation value of K = 3, which allowed for model performance evaluation across multiple data subsets. By using random search within a predefined hyperparameter grid, this technique effectively reduced computational demands while identifying optimal values.

4.8. K-Fold Cross Validation

Cross-validation is a straightforward method for estimating the expected prediction error, often represented as $Err=E[L(Y, f(X))]$, which is the average generalization error when applying a method $f(X)$ to an independent test sample drawn from the joint distribution of X and Y. Ideally, cross-validation provides an estimate of the conditional error when the training set T is held fixed. In K-fold cross-validation, the dataset is split into K equally sized folds. The model is trained on $K-1$ folds and tested on the remaining fold. This process is repeated K times, with each fold used

once as a test set. The cross-validation estimate of the prediction error is calculated as the average error across all K trials.
For an indexing function k: i→1..., K that assigns each observation i to one of the K folds, let f_{-k}(x) denote the fitted function when the k-th fold is excluded from the training data. The cross-validation estimate of the prediction error can be defined as:

$$CV(f) = \frac{1}{N}\sum_{i=1}^{N} L\left(Y_i, f_{-k(i)}(X_i)\right) \quad (27)$$

where L (Y, Y^) is the loss function (e.g., squared error for regression).

4.9.Leave-One-Out Cross-Validation (LOOCV)

A special case of K-fold cross-validation is when K=N, known as leave-one-out cross-validation (LOOCV). Here, each observation i serves as a single test case, with the model trained on the remaining N−1 observations. LOOCV has the advantage of low bias in error estimation but is computationally expensive for large datasets. (Tibshirani, 2017).

5. Evaluation Metrics

Generally, the performance of any classification method is evaluated using standard metrics such as accuracy, precision, recall, and F1-score, which are derived from the confusion matrix. However, in cases where the dataset is highly imbalanced, as with the backorder prediction dataset, these metrics alone may not provide a comprehensive assessment of classifier performance. To address this, additional metrics such as AUC (Area Under the Curve) and ROC (Receiver Operating Characteristic) curves are employed to measure and visualize the classifier's ability to distinguish between classes. These metrics are particularly effective for imbalanced datasets as they provide insights into the trade-offs between true positive rates and false positive rates ((Shajalal et al., 2022; De Santis et al., 2017; Chawla et al., 2002).

The classification accuracy ACC of a model using the number of true positives, true negatives, false negatives, and false positives as shown in the following equation:
Acc= $\frac{T_p + T_n}{T_p + F_n + F_p + T_n}$, where $T_p + F_p + F_n + T_n$ denote the number of classifed samples as true positive, false positive, false negative and true negative, respectively (Shajalal et al., 2021).

- True Negative (TN): The number negative samples (non-backordered) correctly classified as negative.
- False Positive (FP): The number of negative samples incorrectly classified as positive (backordered).
- False Negative (FN): The number of positive samples incorrectly classified as negative.
- True Positive (TP): The number of positive samples correctly classified as positive (Shajalal et al., 2021).

Precision and recall are commonly used metrics to evaluate classification performance, particularly as they relate to the positive (minority) class. Precision and recall ignore true negatives, making them unaffected by class imbalance and reliable metrics for assessing a model's performance. Precision assesses the classifier's accuracy in identifying positive (backordered) instances (Shajalal et al., 2021). It can be defined as:

$$P = \frac{T_p}{T_p + F_p} \quad (28)$$

Unlike precision, recall is determined by accounting for the total number of actual positive samples (total backorders). Recall R measures the classifier's ability to correctly identify positive instances (the proportion of backorders). It is also referred to as sensitivity or the true positive rate (TPR) (Shajalal et al., 2021). It is defined as:

$$R = \frac{T_p}{T_p + F_N} \tag{29}$$

The false-positive rate (FPR) is another commonly used evaluation metric for classification techniques. In the context of backorder prediction, it quantifies the proportion of non-backordered instances incorrectly classified as backordered relative to all actual non-backordered instances (Shajalal et al., 2021). It is defined as follows:

$$F = \frac{F_p}{F_p + T_N} \tag{30}$$

The Receiver Operating Characteristic (ROC) curve is a valuable tool for evaluating the performance of a classifier, particularly when dealing with class-imbalanced datasets (Khan et al., 2024). It provides a visualization for selecting the optimal decision threshold by plotting the trade-off between the true-positive rate (TPR) and the false-positive rate (FPR). The Area Under the Curve (AUC) associated with the ROC curve measures the likelihood that the classifier ranks a randomly chosen backordered instance higher than a randomly chosen non-backordered instance. As highlighted by Chawla et al. (2002), the AUC is widely used to estimate the performance of classification techniques on imbalanced datasets (Kaur and Singh, 2023). The ROC curve offers an intuitive way to demonstrate classifier efficiency and illustrates how increasing the FPR can lead to a corresponding increase in the TPR. Moreover, like the precision-recall curve, the ROC curve helps visualize the trade-off between precision and false-positive rate, providing insights into classifier behavior under varying thresholds. AUC is one of the most efficient metrics to measure the performance of any classification model on imbalanced data (Zhu, M., et al., 2018) The AUC, derived from the ROC curve, can be mathematically expressed as follows: AUC = $\frac{1+TPR-FPR}{2}$, where TPR is the true positive rate and FPR is the false positive rate (Shajalal et al., 2022; De Santis et al., 2017; Chawla et al., 2002). A higher AUC indicates that the model performs better at distinguishing between classes. An AUC of 1 represents a perfect classifier, while an AUC of 0 means the model misclassifies all negatives as positives and all positives as negatives.

The F1 score represents the harmonic mean of precision and recall, combining them into a single unified metric and evaluates errors arising from both false positives and false negatives. as shown in following equation (Khan et al., 2024; Kaur and Singh, 2023; Zhu, M., et al., 2018):

$$F1 = 2\,\frac{P * R}{P + R}, \tag{31}$$

6. Results Analysis

This chapter presents the results from a comprehensive evaluation of 98 predictive models. These models were constructed using combinations of 5 resampling methods (Random Under Sampling, SMOTE-ENN, SMOTE-SVM, Borderline-SMOTE, and ADASYN), two levels of cross-validation (5-fold and 10-fold), and two machine learning models (K Nearest Neighbours and Neural Networks) as well as five ensemble learning techniques (Random Forest, Gradient Boosting, XGBoost, LightGBM, and Stacking). Each of these machine learning and ensemble learning models was trained and optimized using RandomizedSearchCV, a widely recognized hyperparameter tuning technique, in conjunction with the two cross-validation strategies (5-fold and 10-fold).

The performance of these 98 models was evaluated and compared based on various evaluation metrics, with a particular emphasis on handling class imbalance, the main challenge in this research. Specific metrics such as balanced accuracy, Specificity, Precision, Recall, F1-Score, ROC-AUC, and others were used to assess both the predictive accuracy and the effectiveness of the resampling methods. The objective of this comparison is to ascertain which machine learning or ensemble learning model produces the most accurate backorder predictions and which resampling technique best resolves the dataset's class imbalance.

6.1. Comparison of Results from Resampling Techniques

The analysis of resampling techniques plays a critical role in addressing the Class Imbalance in the dataset. For instance, applying SMOTE-ENN resulted in a significant transformation of the original class distribution. The used dataset had a highly imbalanced distribution, with the majority class (non-backordered products) represented by 11,629 instances and the minority class (backordered products) by only 71 instances. After applying SMOTE-ENN (as the first resampling technique) the resampled dataset achieved a more balanced distribution, with 11,568 instances of the minority class and 10,753 instances of the majority class. This balancing was achieved by combining oversampling through SMOTE, which generates synthetic samples for the minority class, and ENN, which removes noisy and borderline majority instances, resulting in a cleaner and more representative dataset. The Table 6.1 shows the performance of each resampling technique in class distribution. The results show that each technique achieved significant rebalancing of the dataset.

Table 6-1: Comparison of Results from Resampling Techniques

Original Dataset and Resampling Techniques	0: Majority Class	1: Minority Class	Total of Observations
Original dataset class distribution	11,629	71	11,700
SMOTE-ENN	10,753	11,568	22,321
SMOTE-SVM	11,629	6,427	18,056
Borderline-SMOTE	11,629	11,629	23,258
ADASYN	11,629	11,625	23,254
Random Under Sampling	71	71	142

Comparatively, other resampling techniques like SMOTE-SVM, Borderline-SMOTE, and ADASYN exhibit varying behaviors in balancing the dataset. The main goal of SMOTE-SVM is generating synthetic samples near the decision boundary, making it particularly effective for separating overlapping classes. In order to help the classifier focus on difficult areas of the feature space, Borderline-SMOTE focuses on producing synthetic samples for minority class instances near the decision boundary. ADASYN, on the other hand, adapts the synthetic sample generation based on the difficulty of classification, concentrating more on harder-to-classify minority instances. Each technique's impact on the class distribution directly influences model performance. The effectiveness of these methods in improving predictive accuracy and handling the class imbalance will be further evaluated and compared based on model-specific metrics, such as F1-Score, Recall, and Precision, across the selected machine learning models and ensemble methods.

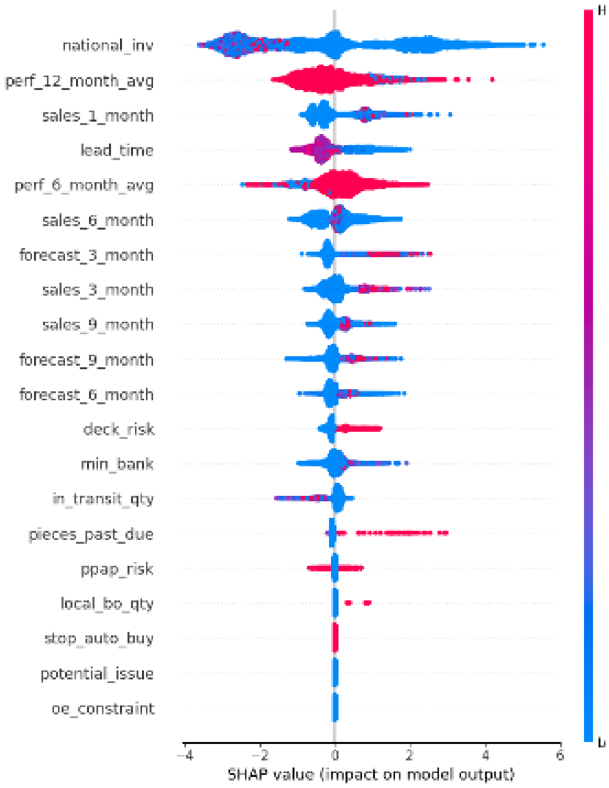
6.2. Analysis and Comparison of Resampling Techniques Combined with SHAP and PCA

The results from the five resampling techniques (SMOTE-ENN, SMOTE-SVM, Borderline-SMOTE, ADASYN, and RUS) provide valuable insights into how each method interacts with SHAP for feature selection and PCA for dimensionality reduction.

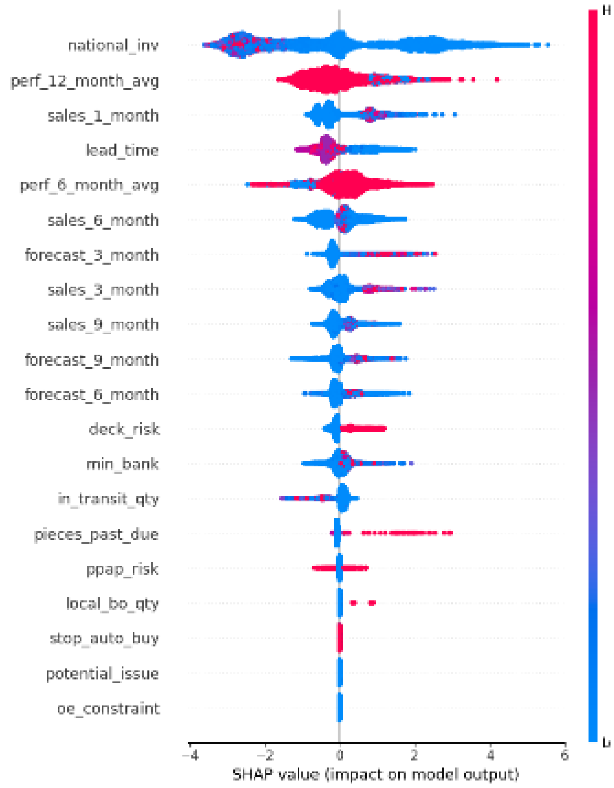
Regarding the Feature Selection, across all five resampling techniques, SHAP identified the same 16 features as the most important. This consistency reflects SHAP's robustness in selecting features that contribute significantly to backorder prediction.

Regarding Dimensionality Reduction, PCA retained 9 components for all methods, preserving a substantial portion of variance. However, the explained variance ratios vary slightly, reflecting the impact of each resampling technique on data distribution:

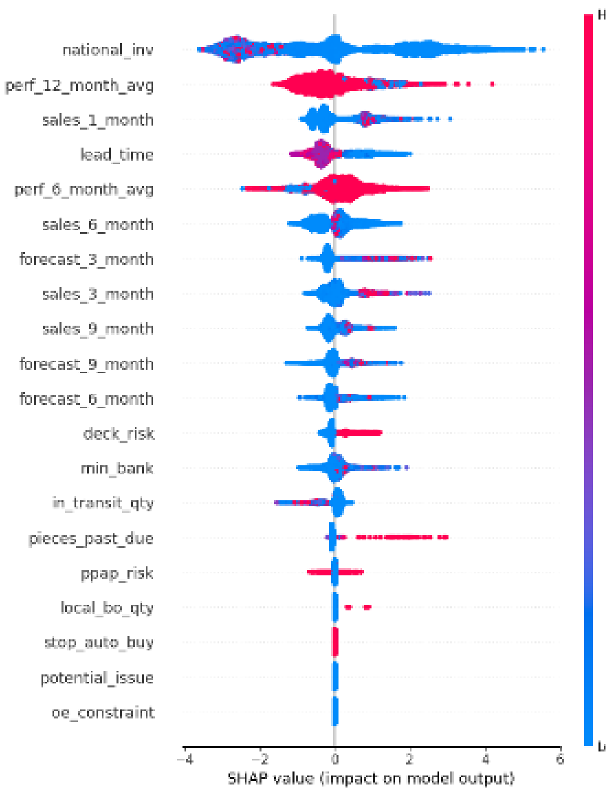
The first component explains 47.13% of the variance, resulting from SMOTE-ENN. This value from SMOTE-SVM, Borderline-SMOTE and finally ADASYN is 46.86%, 46.85% and 47.18% respectively. The results reflects that the explained variance ratios for the first component are slightly higher for ADASYN and SMOTE-ENN, suggesting that these techniques result in a dataset structure that retains more meaningful variance during dimensionality reduction. Figure 6.1 displays the distribution of SHAP values for individual features, showing their importance and impact on predictions. The combination of SHAP for feature selection and PCA for dimensionality reduction proved effective across all resampling techniques. However, the choice of resampling method significantly impacts the class balance, data structure, and variance retention. SMOTE-ENN and ADASYN appear to offer a favorable balance between variance retention and dataset size, while Borderline-SMOTE provides the most comprehensive rebalancing. The final choice of resampling technique should consider the specific requirements of the predictive models and computational resources.



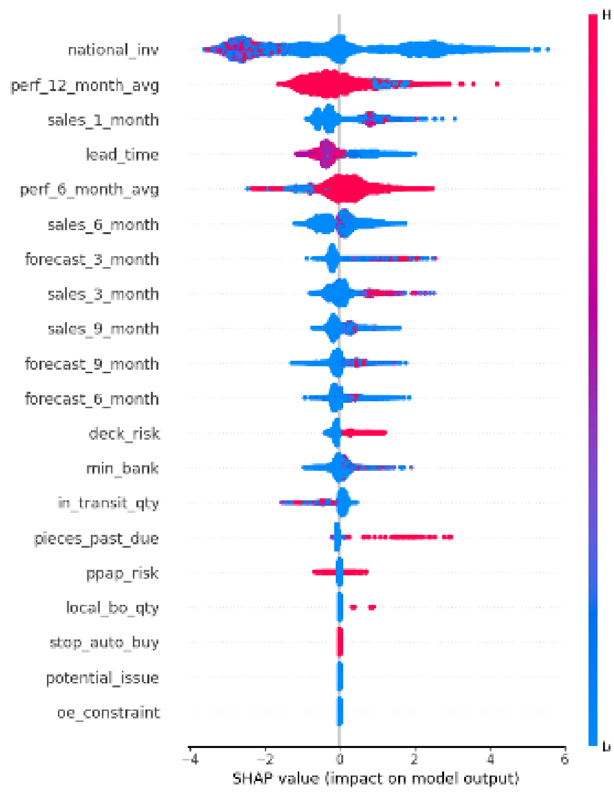
ADASYN



RUS



SMOTE-SVM



SMOTE-ENN

Figure 6.1: Distribution of SHAP values for individual features

Figure 6.2 to 6.6 visualize the importance of features as identified by SHAP values. Bar charts showing average absolute SHAP values for each variable.

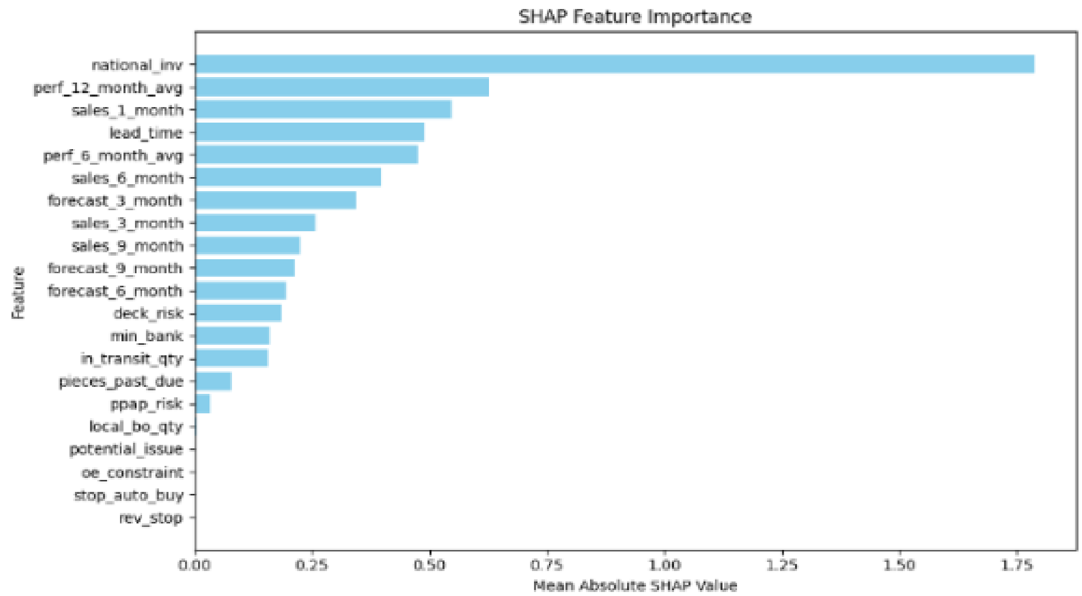


Figure 6.2: SHAP Feature Importance Bar Chart (ADASYN)

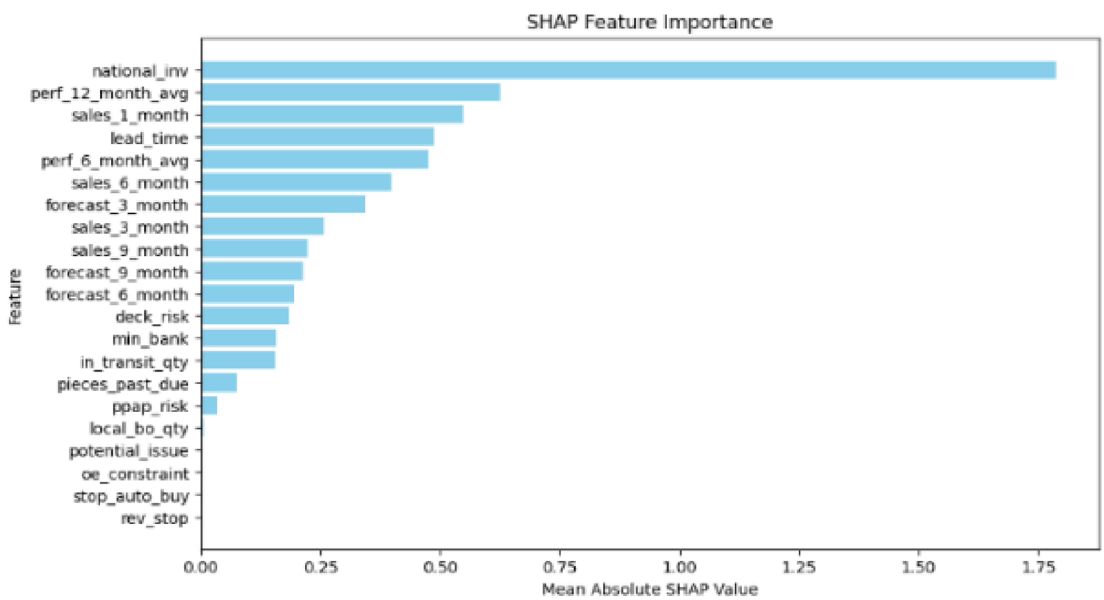


Figure 6.3: SHAP Feature Importance Bar Chart (Borderline-SMOTE)

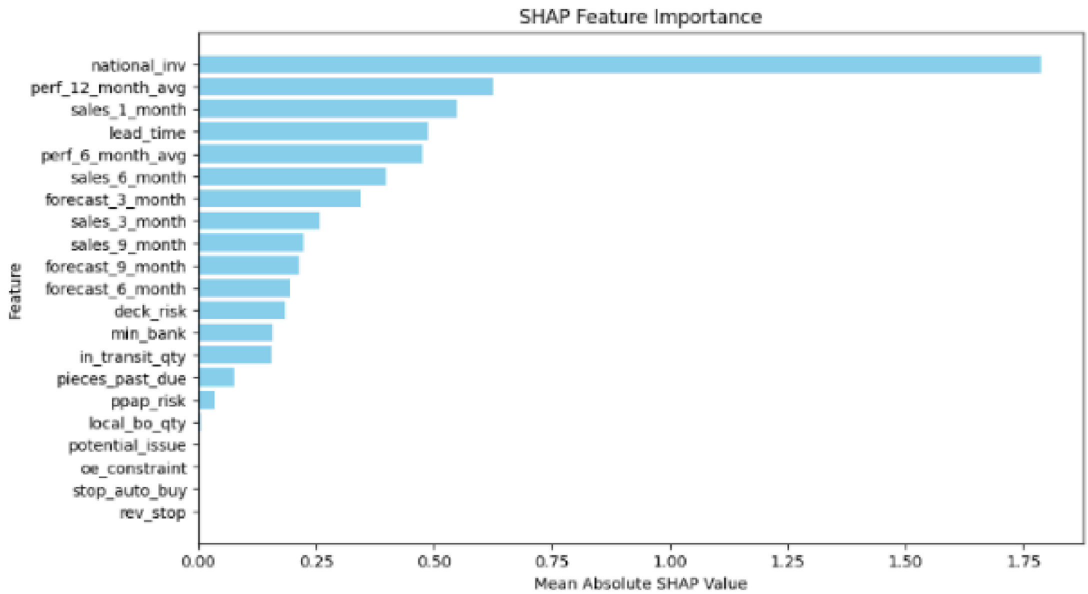


Figure 6.4: SHAP Feature Importance Bar Chart (SMOTE-ENN)

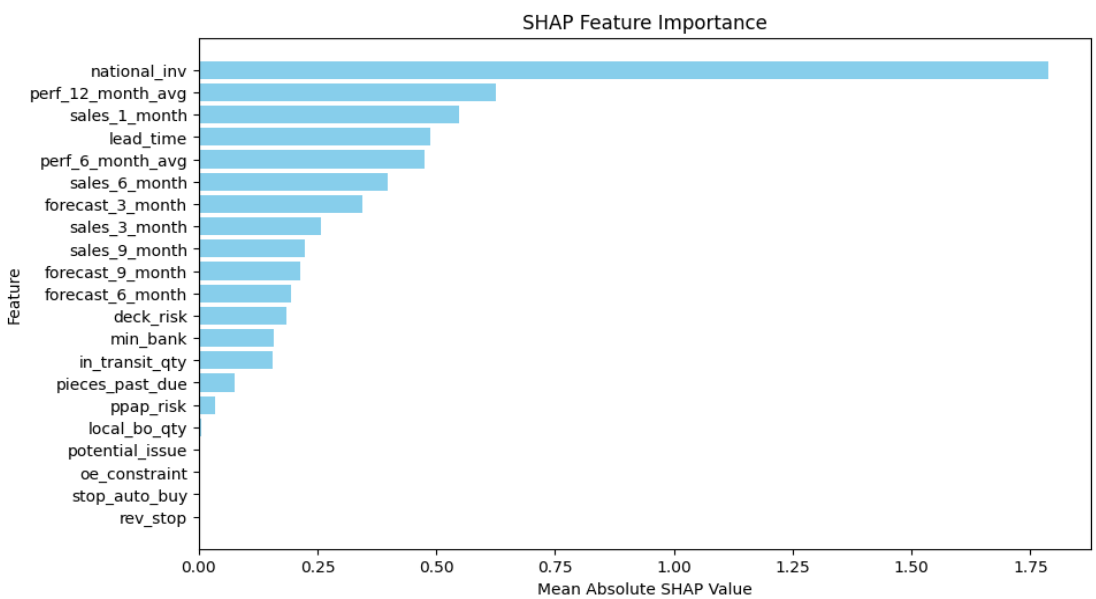


Figure 6.5: SHAP Feature Importance Bar Chart (SMOTE-SVM)

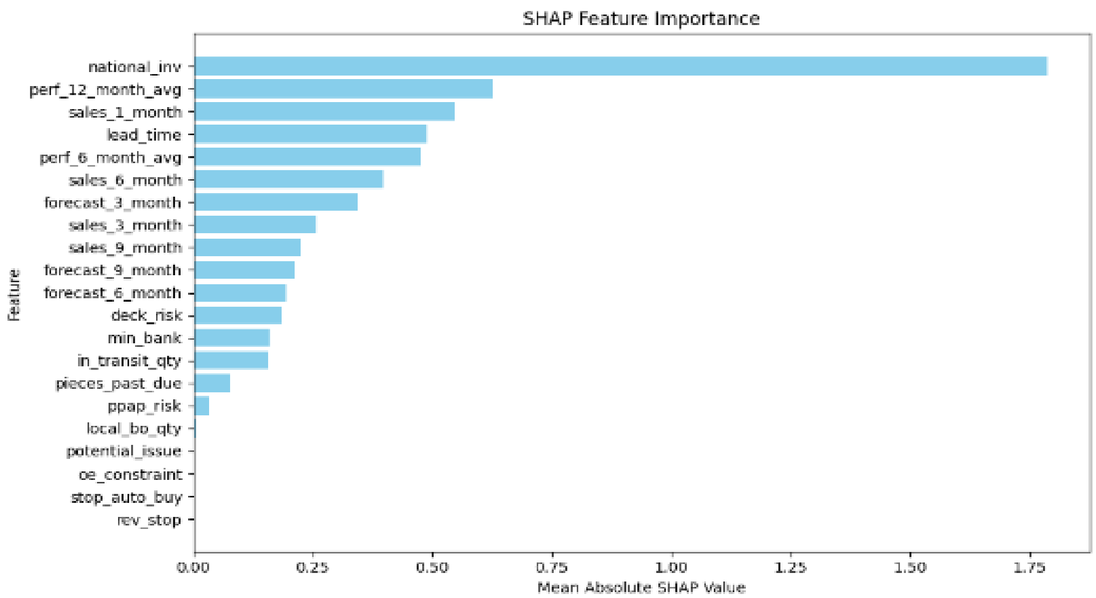


Figure 6.6: SHAP Feature Importance Bar Chart (RUS)

Figure 6.7 to 6.11 visualize the PCA-reduced data in 2D (using first two components).

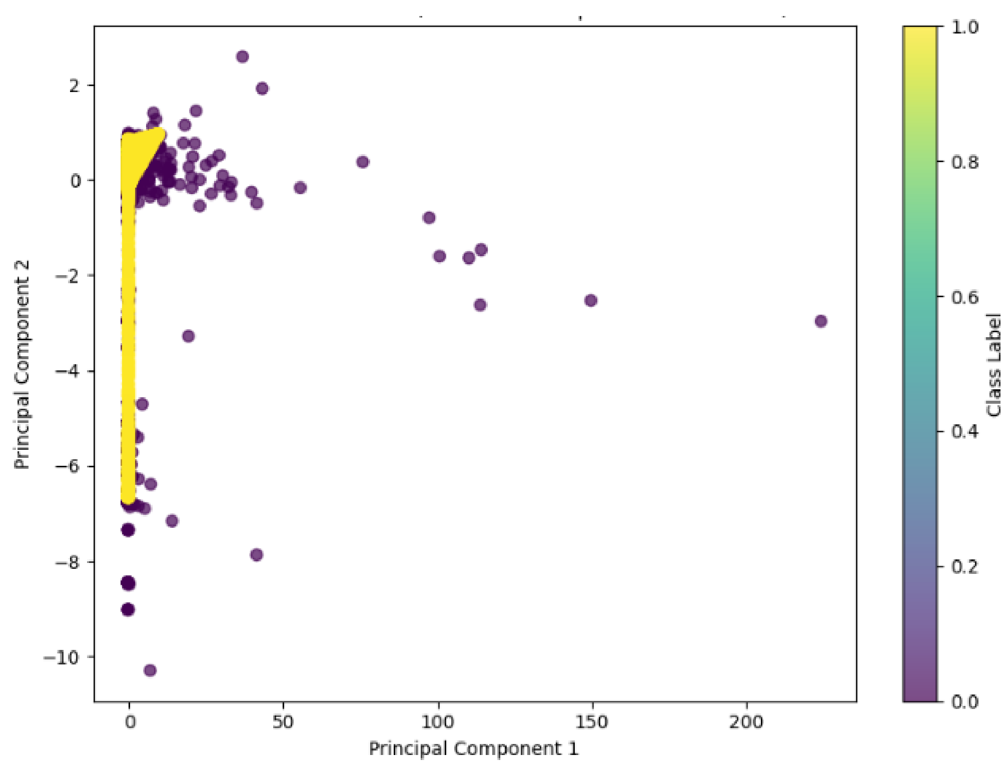


Figure 6.7: Transformed Data Scatter Plot (ADASYN)

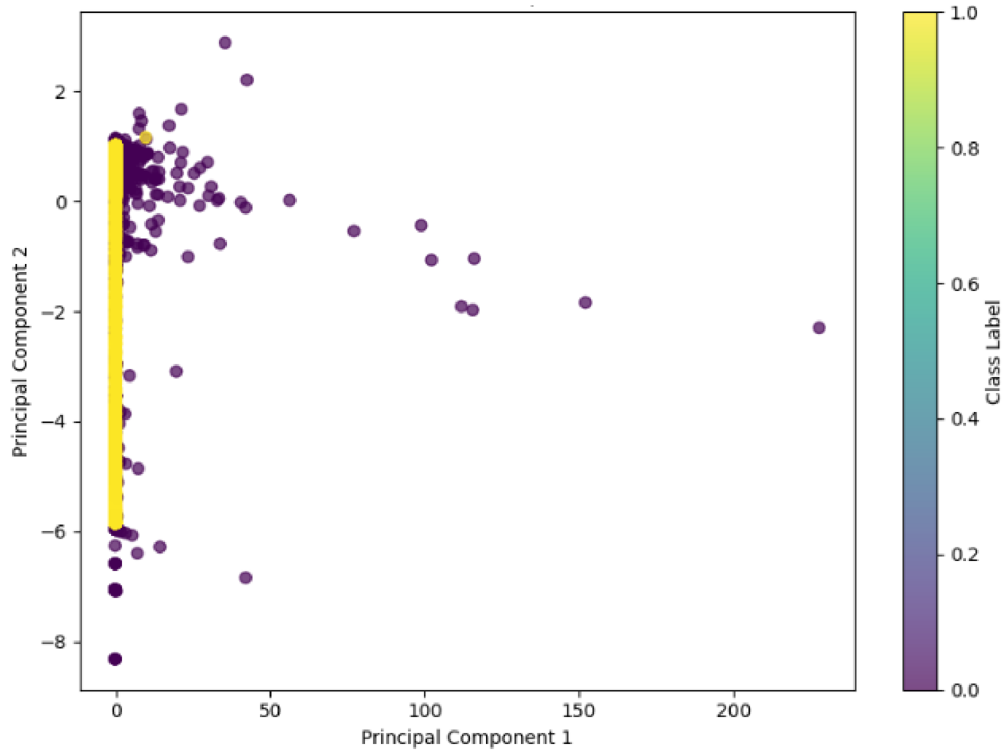


Figure 6.8: Transformed Data Scatter Plot (Borderline-SMOTE)

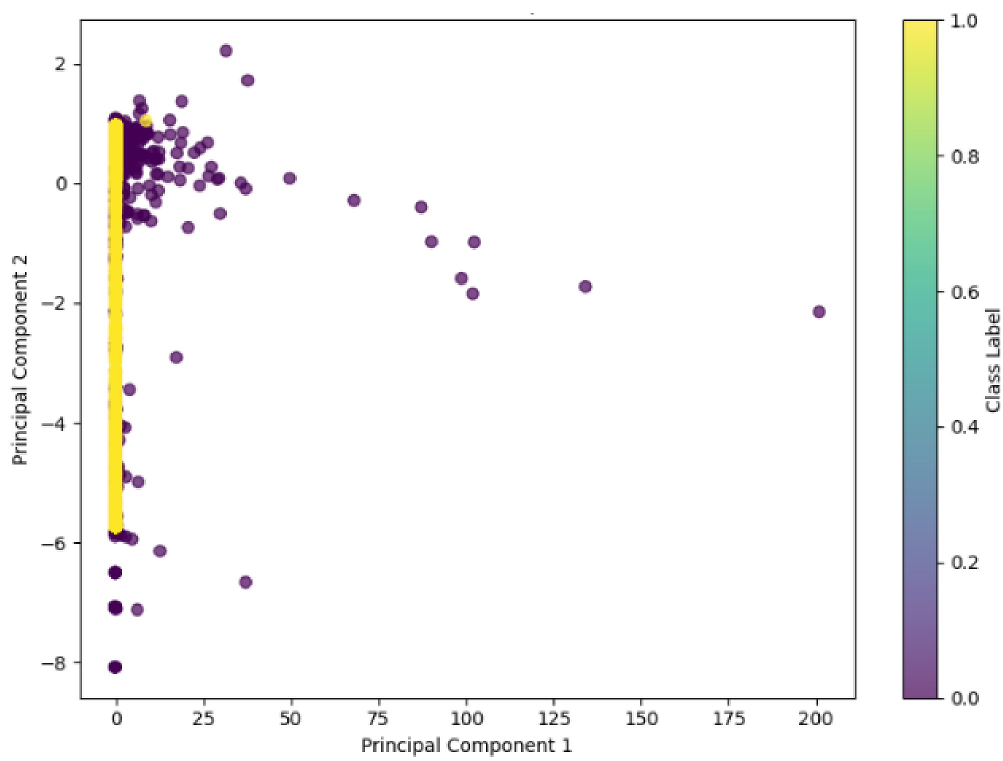


Figure 6.9: Transformed Data Scatter Plot (SMOTE-SVM)

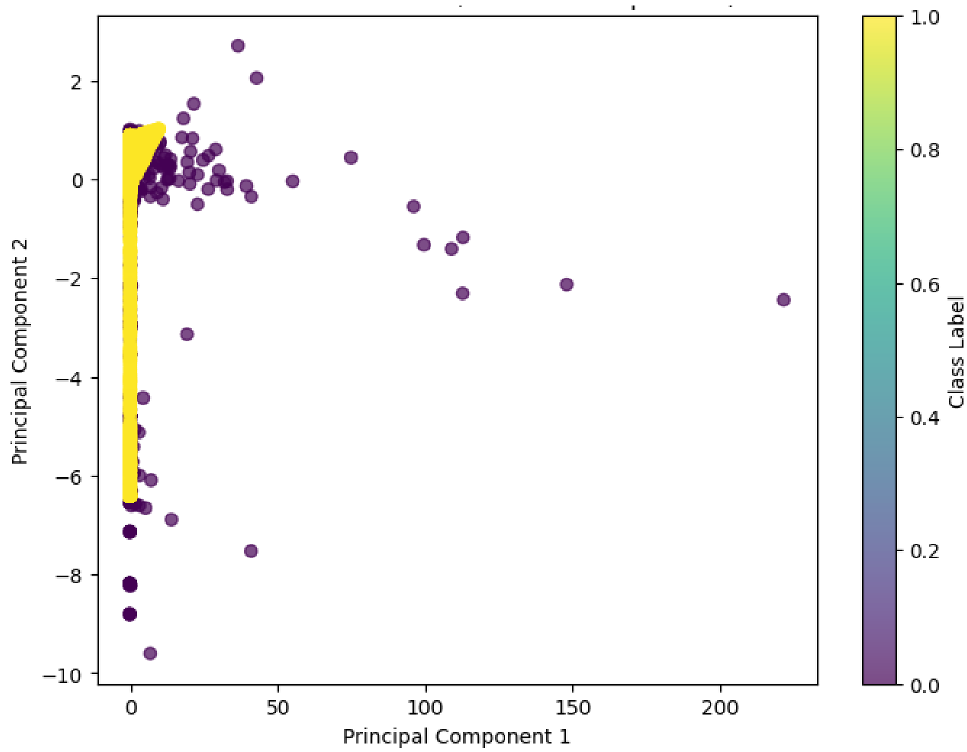


Figure 6.10: Transformed Data Scatter Plot (SMOTE-ENN)

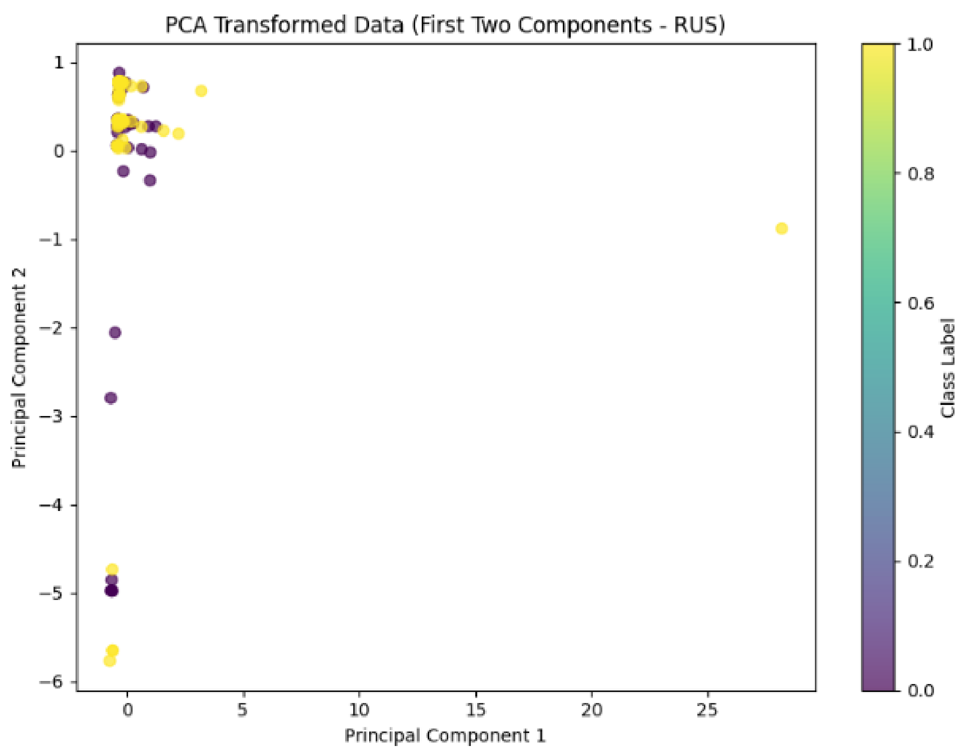


Figure 6.11: Transformed Data Scatter Plot (RUS)

6.3.Adjusted Feature Selection

In another phase of analysis, all models were retrained with an adjusted methodology to evaluate their performance under refined conditions. Specifically, SHAP was utilized exclusively for feature selection, while PCA was omitted to assess the direct impact of SHAP-based feature selection alone. To ensure robust evaluation, a 10-fold cross-validation approach was employed across all models, providing a thorough assessment of generalization performance. Additionally, the SHAP threshold for feature importance was increased to 0.2, thereby selecting only the most impactful features for the training process. This adjustment has been used to analyze the dataset further by focusing on highly influential features, reducing noise, and potentially enhancing model interpretability and performance. These modifications provide a critical comparative framework to examine the performance of models trained with SHAP-selected features versus the previous combination of SHAP and PCA.

6.4. Evaluation Metrics Analysis

To provide a comprehensive evaluation of the trained models, Tables 6.2 to 6.4 present a structured comparison of the 98 distinct models generated in this study. These models are derived by combining various resampling techniques (e.g., SMOTE-ENN, Borderline-SMOTE, ADASYN, SMOTE-SVM, and RUS), feature selection and dimensionality reduction approaches (e.g., SHAP and PCA), and K-fold cross-validation methods (5-fold and 10-fold). The performance of each model is assessed across five evaluation metrics: F1 Score, balanced accuracy, ROC-AUC, PR-AUC, and Specificity. This tabular representation allows for a detailed comparison of how different combinations influence model effectiveness in addressing the backorder prediction task.

Since classification accuracy is unable to capture model performance in class imbalance (CI) problems, especially for the minority classes, and does not account for the disparity in class sizes, we have used balanced accuracy as an evaluation metric. Classification accuracy can lead to misleading conclusions when dealing with imbalanced datasets (Ahmad Khan et al., 2024). In the current study, the XGBoost and Stacking models, combined with SMOTEENN and Random Under-Sampling (RUS) as resampling techniques, and SHAP for feature selection, using 10-fold cross-validation, achieved the highest balanced accuracy scores of 0.876 and 0.888, respectively.

To evaluate the ability of a classifier to distinguish between classes, the AUC (Area Under the Curve) was used as a summary score of the ROC curve. Better model performance in class distinction is indicated by a higher AUC. Perfect classification is represented by a model with an AUC of 1, but all negatives are misclassified as positives by a model with an AUC of 0 and vice versa (Ahmad Khan et al., 2024). The findings demonstrate that, in terms of AUC, ensemble learning models perform better than conventional machine learning models. Moreover, in the adjusted approach—where PCA was omitted and the SHAP threshold was increased—Stacking, using SMOTEENN as a resampling technique and SHAP for feature selection, achieved the highest AUC score of 0.954, surpassing all other models (see Tables 6.2 to 6.4).

An assessment metric for binary classification issues is the AUC-ROC curve. Plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR) at different threshold levels is done by the AUC-ROC curve. Figures 6.12 to 6.25 provide a visual representation of the comparative performance through ROC curves for the trained models. These curves provide information on each model's capacity to distinguish between backordered and non-backordered instances by highlighting the trade-off between the true positive rate (sensitivity) and the false positive rate. By visually comparing the ROC curves, Figures 6.12 to 6.25 emphasizes the strengths and weaknesses of various resampling and modeling combinations in achieving optimal classification outcomes. Together, tables 6.2 to 6.4 and figures 6.12 to 6.25 serve as critical tools for analyzing model performance, with tables 6 focusing on quantitative comparisons and Figure 6 providing intuitive visual insights into classifier discrimination power.

Table 6-2: Comprehensive Evaluation Metrics of Model Training: Integration of Resampling Techniques, ML and Ensemble Methods, Feature Selection (SHAP and PCA), and 5-Fold CV

Resampling Techniques	Machine Learning Models Ensemble Learnings	Evaluation Metrics				
		ROC-AUC	PR-AUC	F1 Score	B-ACC	Specificity
SMOTEENN	Neural Network	0.637	0.013	0.023	0.571	0.83
	K Nearest Neighbour	0.608	0.092	0.09	0.598	0.977
	Random Forest	0.826	0.044	0.135	0.589	0.99
	Gradient Boosting	0.797	0.047	0.108	0.586	0.985
	LightGBM	0.78	0.037	0.081	0.569	0.983
	XGBoosting	0.811	0.042	0.121	0.616	0.982
	Stacking	0.829	0.042	0.096	0.558	0.991
SMOTESVM	Neural Network	0.669	0.014	0.046	0.549	0.973
	K Nearest Neighbour	0.583	0.049	0.069	0.542	0.99
	Random Forest	0.792	0.075	0.087	0.53	0.998
	Gradient Boosting	0.796	0.048	0.071	0.529	0.996
	LightGBM	0.821	0.038	0.032	0.513	0.994
	XGBoosting	0.854	0.044	0.033	0.513	0.994
	Stacking	0.781	0.072	0.04	0.514	0.997
ADASYN	Neural Network	0.701	0.033	0.067	0.567	0.978
	K Nearest Neighbour	0.594	0.107	0.083	0.583	0.979
	Random Forest	0.839	0.049	0.083	0.543	0.993
	Gradient Boosting	0.791	0.053	0.115	0.587	0.987
	LightGBM	0.761	0.039	0.139	0.603	0.988
	XGBoosting	0.784	0.046	0.154	0.619	0.987
	Stacking	0.863	0.051	0.071	0.555	0.985
Borderline-SMOTE	Neural Network	0.711	0.024	0.063	0.591	0.963
	K Nearest Neighbour	0.584	0.051	0.07	0.542	0.99
	Random Forest	0.735	0.101	0.095	0.53	0.998
	Gradient Boosting	0.825	0.089	0.074	0.529	0.996
	LightGBM	0.831	0.039	0.036	0.513	0.995
	XGBoosting	0.664	0.019	0.062	0.603	0.956
	Stacking	0.728	0.08	0.093	0.53	0.998

Table 6-3: Comprehensive Evaluation Metrics of Model Training: Integration of Resampling Techniques, ML and Ensemble Methods, Feature Selection (SHAP and PCA), and 10-Fold CV

Resampling Techniques	Machine Learning Models Ensemble Learnings	Evaluation Metrics				
		ROC-AUC	PR-AUC	F1 Score	B-ACC	Specificity
SMOTEENN	Neural Network	0.728	0.03	0.045	0.638	0.902
	K Nearest Neighbour	0.608	0.092	0.09	0.598	0.977
	Random Forest	0.802	0.074	0.1	0.572	0.987
	Gradient Boosting	0.797	0.047	0.108	0.586	0.985
	LightGBM	0.78	0.037	0.081	0.569	0.983
	XGBoosting	0.811	0.042	0.121	0.616	0.982
	Stacking	0.839	0.045	0.119	0.573	0.991
SMOTESVM	Neural Network	0.669	0.014	0.046	0.549	0.973
	K Nearest Neighbour	0.583	0.049	0.069	0.542	0.99
	Random Forest	0.763	0.071	0.091	0.53	0.998
	Gradient Boosting	0.796	0.048	0.071	0.529	0.996
	LightGBM	0.821	0.038	0.032	0.513	0.994
	XGBoosting	0.854	0.044	0.033	0.513	0.994
	Stacking	0.83	0.072	0.078	0.53	0.997
ADASYN	Neural Network	0.594	0.029	0.051	0.562	0.968
	K Nearest Neighbour	0.594	0.107	0.083	0.583	0.979
	Random Forest	0.808	0.073	0.081	0.543	0.992
	Gradient Boosting	0.825	0.049	0.128	0.588	0.989
	LightGBM	0.761	0.039	0.139	0.603	0.988
	XGBoosting	0.78	0.038	0.138	0.617	0.985
	Stacking	0.815	0.042	0.12	0.588	0.988
Borderline-SMOTE	Neural Network	0.736	0.018	0.038	0.535	0.976
	K Nearest Neighbour	0.584	0.051	0.07	0.542	0.99
	Random Forest	0.735	0.102	0.095	0.53	0.998
	Gradient Boosting	0.804	0.106	0.113	0.545	0.996
	LightGBM	0.809	0.039	0.033	0.513	0.995
	XGBoosting	0.848	0.044	0.098	0.544	0.995
	Stacking	0.816	0.085	0.091	0.53	0.998
Random Under Sampling	Neural Network	0.839	0.023	0.056	0.782	0.847
	K Nearest Neighbour	0.873	0.362	0.033	0.784	0.663
	Random Forest	0.888	0.0795	0.043	0.801	0.760
	Gradient Boosting	0.878	0.044	0.039	0.810	0.714
	LightGBM	0.889	0.034	0.052	0.860	0.782
	XGBoosting	0.828	0.036	0.045	0.763	0.807
	Stacking	0.897	0.041	0.048	0.827	0.780

Table 6-4: Comprehensive Evaluation Metrics of Model Training: Integration of Resampling Techniques, ML and Ensemble Methods, Feature Selection (SHAP), and 10-Fold CV

Resampling Techniques	Machine Learning Models Ensemble Learnings	Evaluation Metrics				
		ROC-AUC	PR-AUC	F1 Score	B-ACC	Specificity
SMOTEENN	Neural Network	0.894	0.086	0.065	0.856	0.838
	K Nearest Neighbour	0.742	0.185	0.115	0.671	0.967
	Random Forest	0.895	0.141	0.218	0.681	0.987
	Gradient Boosting	0.915	0.156	0.22	0.651	0.99
	LightGBM	0.93	0.189	0.202	0.636	0.99
	XGBoosting	0.944	0.136	0.146	0.876	0.94
	Stacking	0.954	0.144	0.265	0.638	0.995
ADASYN	Neural Network	0.834	0.085	0.056	0.795	0.841
	K Nearest Neighbour	0.7	0.171	0.127	0.659	0.974
	Random Forest	0.908	0.131	0.268	0.668	0.992
	Gradient Boosting	0.908	0.17	0.218	0.592	0.997
	LightGBM	0.934	0.138	0.161	0.576	0.995
	XGBoosting	0.947	0.166	0.17	0.823	0.959
	Stacking	0.907	0.147	0.216	0.622	0.993
SMOTESVM	Neural Network	0.822	0.066	0.09	0.692	0.947
	K Nearest Neighbour	0.646	0.17	0.178	0.635	0.988
	Random Forest	0.917	0.165	0.054	0.515	0.999
	Gradient Boosting	0.946	0.156	0.14	0.546	0.998
	LightGBM	0.926	0.193	0.12	0.545	0.997
	XGBoosting	0.937	0.189	0.266	0.758	0.984
	Stacking	0.894	0.233	0.205	0.562	0.999
Borderline-SMOTE	Neural Network	0.806	0.112	0.101	0.641	0.969
	K Nearest Neighbour	0.677	0.178	0.165	0.619	0.989
	Random Forest	0.904	0.161	0.105	0.531	0.999
	Gradient Boosting	0.936	0.194	0.208	0.577	0.998
	LightGBM	0.935	0.21	0.2	0.577	0.997
	XGBoosting	0.929	0.187	0.262	0.787	0.981
	Stacking	0.941	0.237	0.205	0.562	0.999
Random Under Sampling	Neural Network	0.785	0.019	0.041	0.634	0.892
	K Nearest Neighbour	0.866	0.113	0.033	0.792	0.647
	Random Forest	0.92	0.168	0.062	0.828	0.844
	Gradient Boosting	0.933	0.106	0.069	0.887	0.837
	LightGBM	0.926	0.116	0.071	0.851	0.859
	XGBoosting	0.93	0.091	0.065	0.857	0.839
	Stacking	0.938	0.117	0.069	0.888	0.839

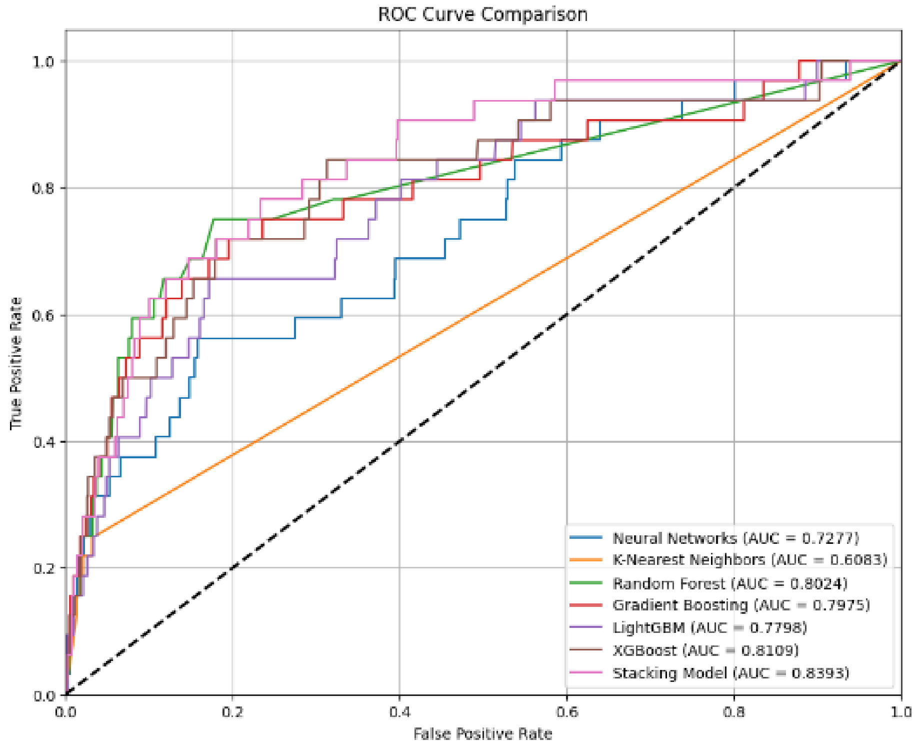


Figure 6.12: ROC Curve - SMOTEENN 10-Fold Cross Validation

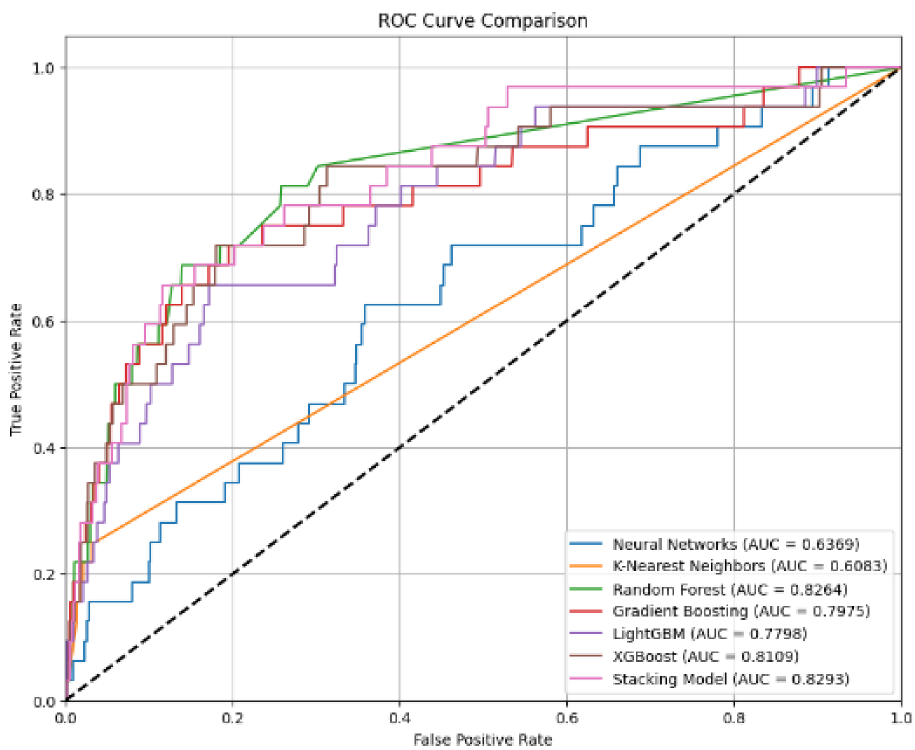


Figure 6.13: ROC Curve - SMOTEENN 5-Fold Cross Validation

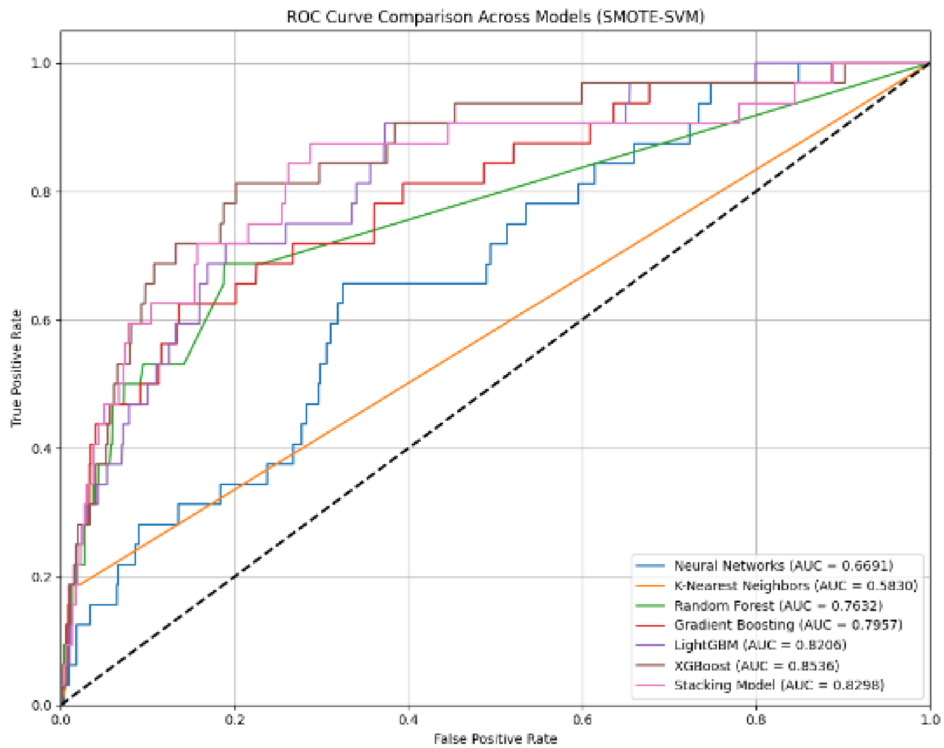


Figure 6.14: ROC Curve - SMOTE-SVM 10-Fold Cross Validation

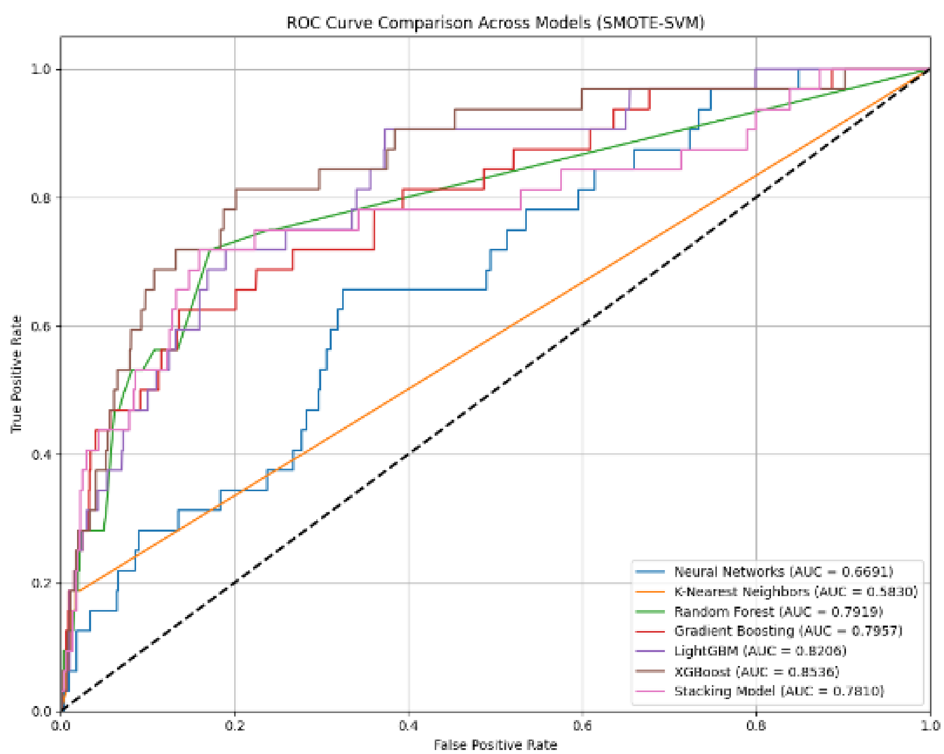


Figure 6.15: ROC Curve - SMOTE-SVM 5-Fold Cross Validation

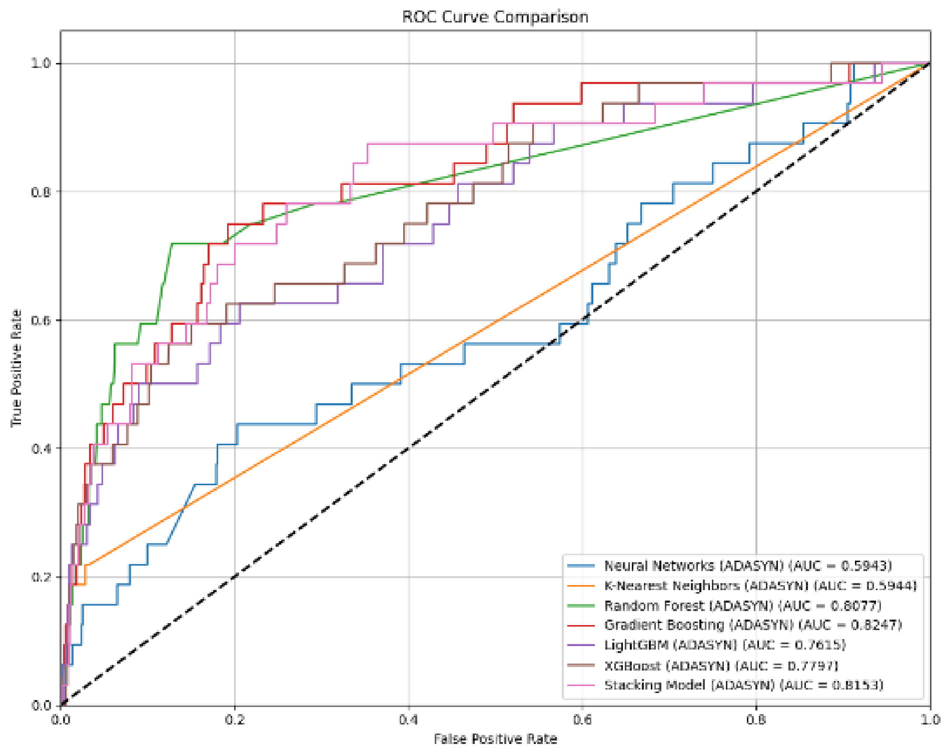


Figure 6.16: ROC Curve - ADASYN-10 Fold Cross Validation

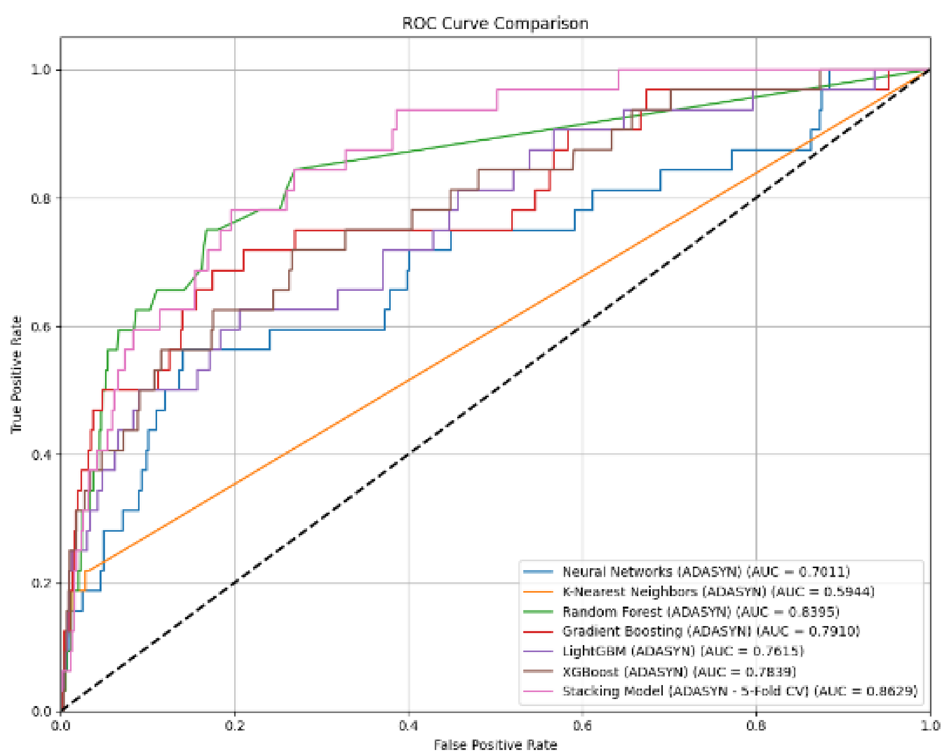


Figure 6.17: ROC Curve - ADASYN– 5-Fold Cross Validation

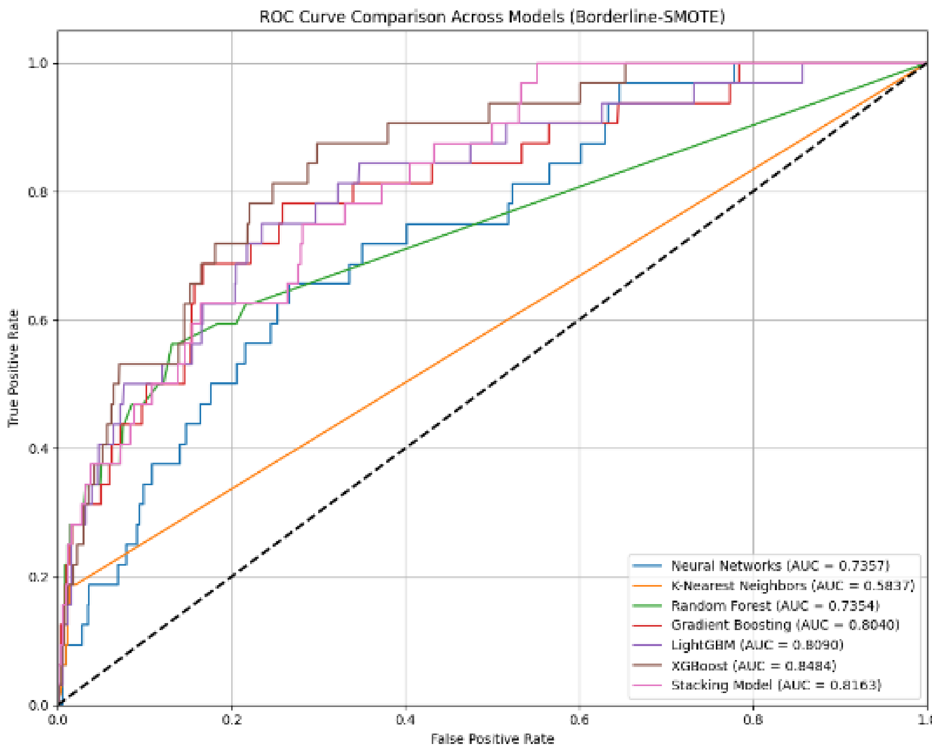


Figure 6.18: ROC Curve - Borderline-SMOTE-10 Fold Cross Validation

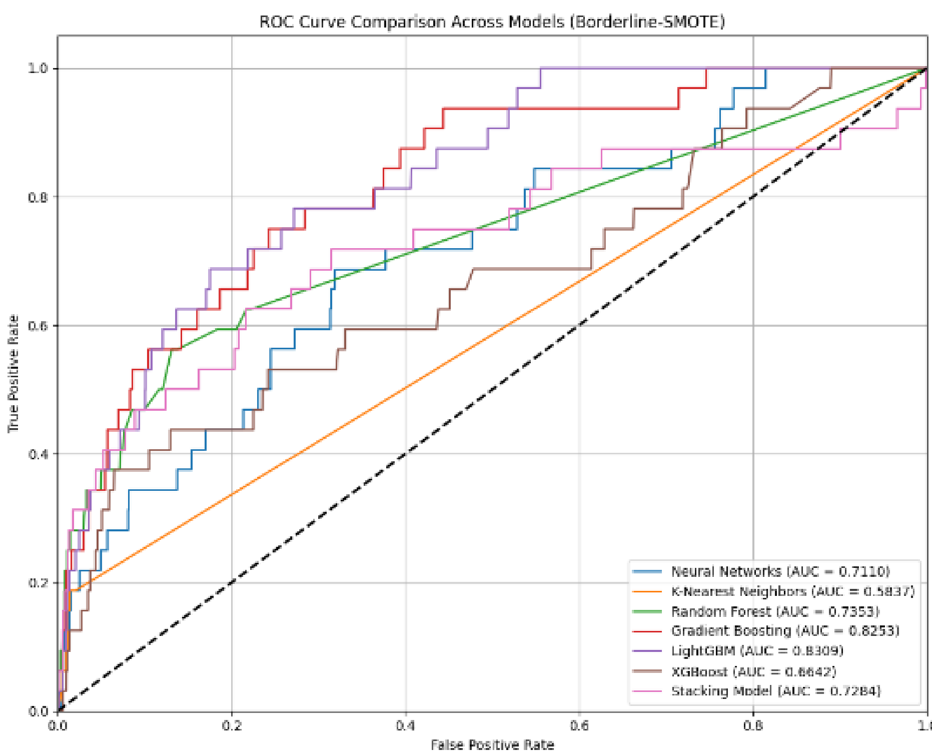


Figure 6.19: ROC Curve - Borderline-SMOTE 5-Fold Cross Validation

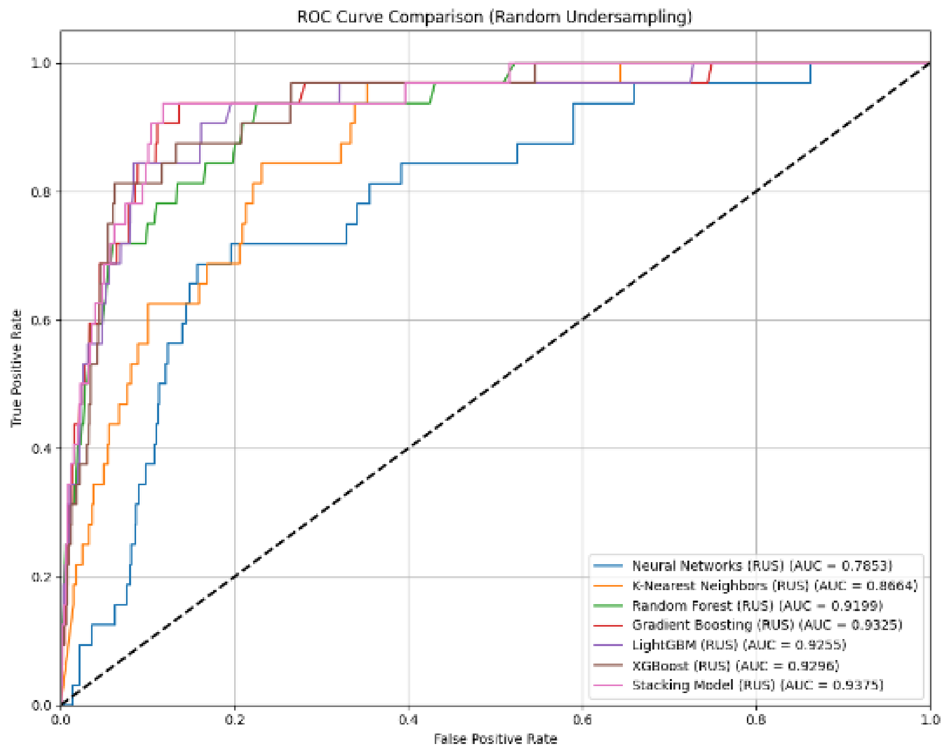


Figure 6.20: ROC Curve - Random Under Sampling-10 Fold Cross Validation (Without PCA)

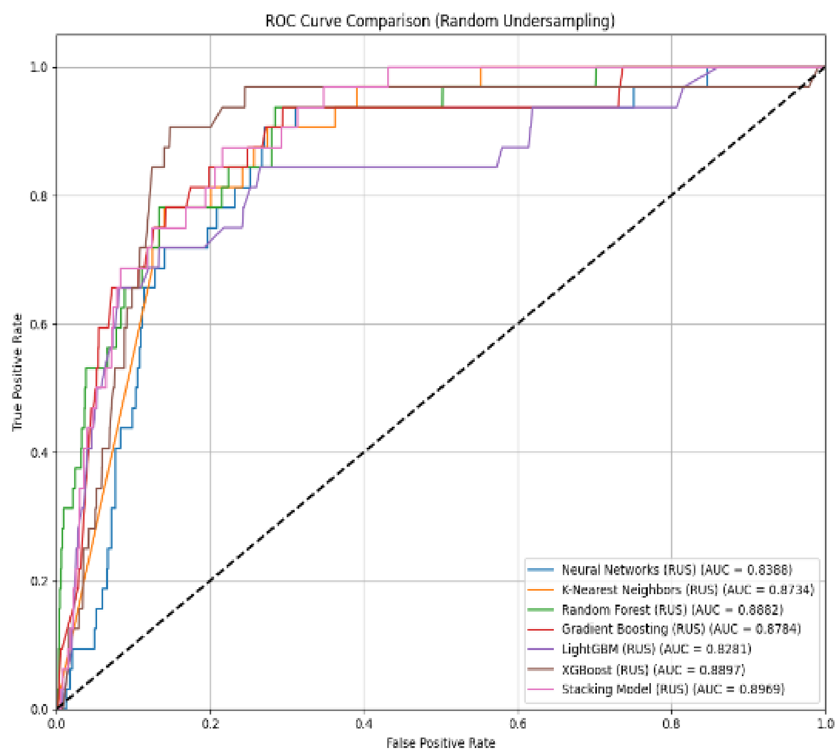


Figure 6.21: ROC Curve - Random Under Sampling-10 Fold Cross Validation

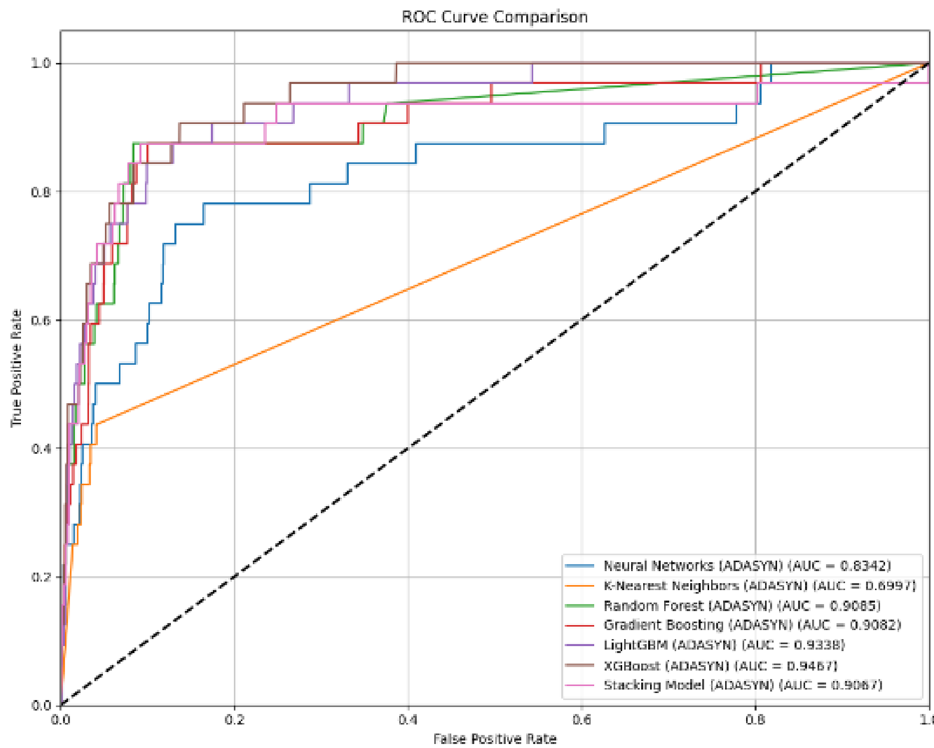


Figure 6.22: ROC Curve - ADASYN -10-Fold Cross Validation (Without PCA)

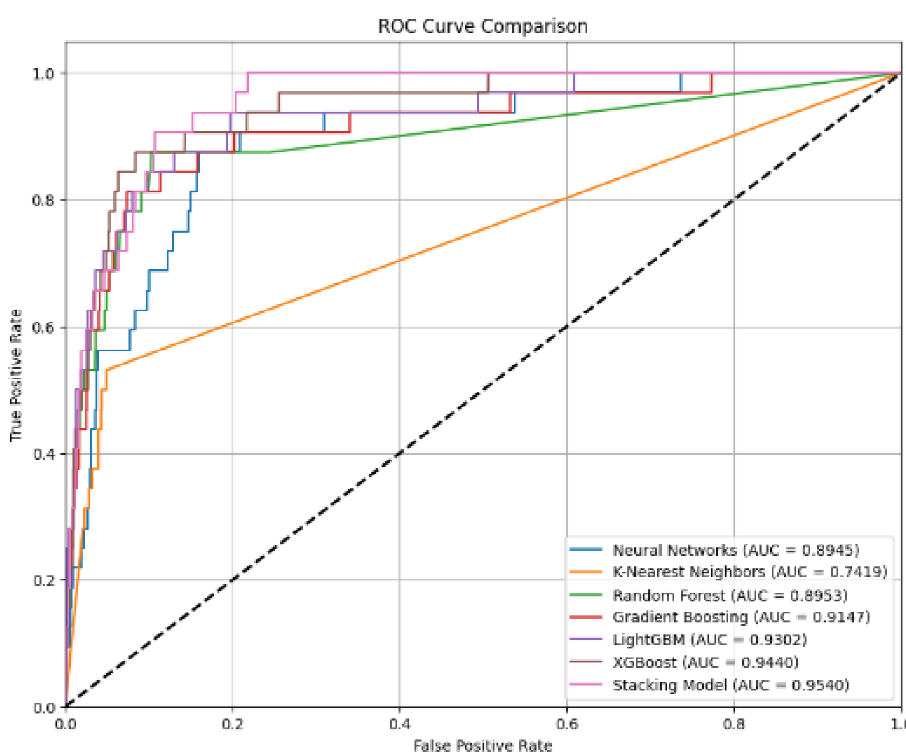


Figure 6.23: ROC Curve - SMOTEENN -10-Fold Cross Validation (Without PCA)

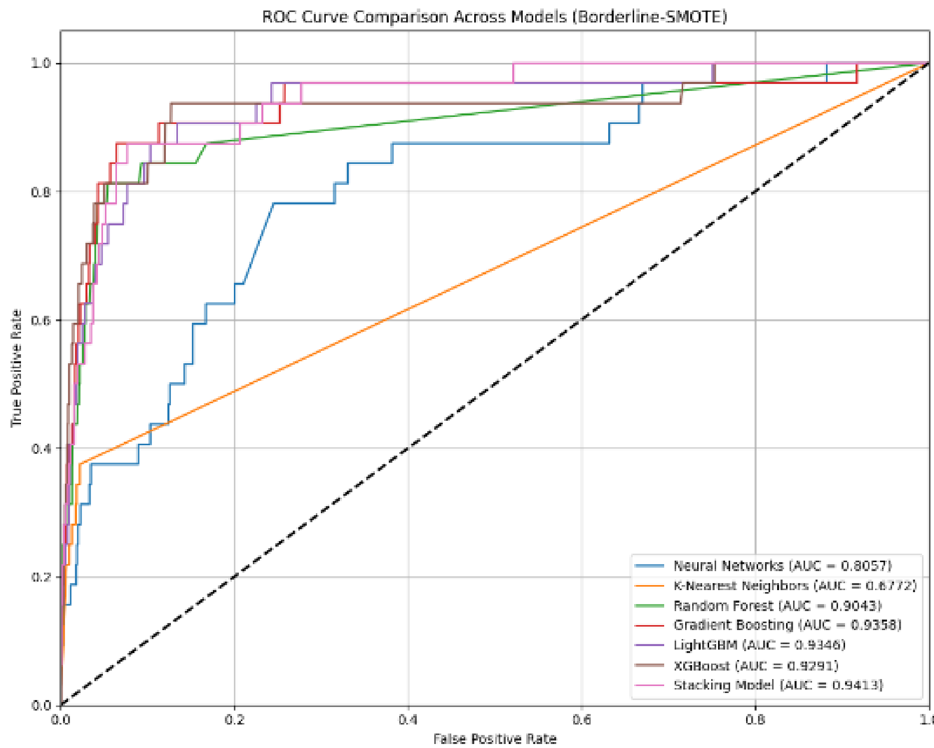


Figure 6.24: ROC Curve - Borderline-SMOTEENN -10-Fold Cross Validation (Without PCA)

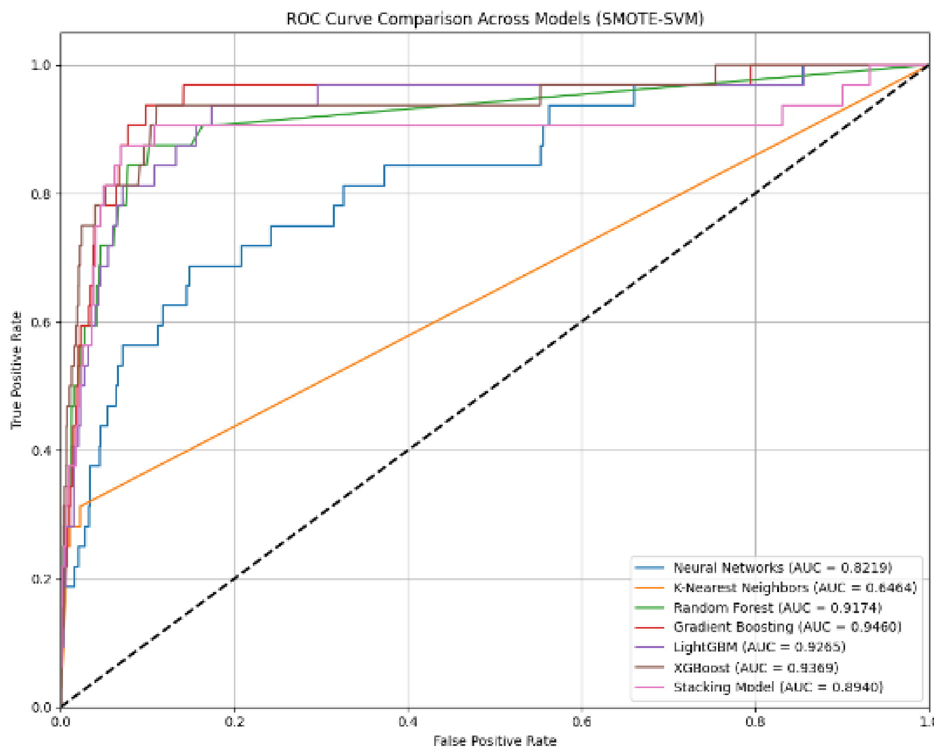


Figure 6.25: ROC Curve - SMOTESVM -10-Fold Cross Validation (Without PCA)

7. Discussion of the results

The results of our research indicate that traditional machine learning models struggle with class imbalance compared to ensemble techniques. This broader comparison provides a comprehensive understanding of model performance, identifying strengths and limitations within the supply chain context. Furthermore, this research advances the field by integrating ensemble techniques, such as XGBoost, LightGBM, and Stacking, which demonstrated superior performance in handling class imbalance and improving prediction accuracy. In addition to model selection, this study emphasizes the critical role of resampling techniques (SMOTE-ENN, ADASYN, RUS) in addressing class imbalance, ensuring that predictions are more reliable and balanced. By combining ensemble learning with effective resampling techniques, this research offers a novel and practical approach to improving predictive modeling in supply chain management.

Class imbalance is a prevalent issue in a variety of applications, such as credit card fraud detection, corporate bankruptcy prediction, and loan approval modeling (Makki et al., 2023); medical imaging for brain disease diagnosis (Shoeibi et al., 2023); image segmentation and classification (Kaur and Singh, 2023); medical diagnostics (Zhu et al., 2018); disease diagnosis (Kim and Hwang, 2022); and prediction methods in biotechnology and medicine (Osama et al., 2023). This issue is also prevalent in the manufacturing and retail industries, where imbalanced datasets are a frequent obstacle in predictive modeling. As datasets grow larger and more complex, the challenge for machine learning models to effectively handle class imbalance becomes increasingly significant, often requiring advanced resampling techniques and feature selection methods to achieve balanced and accurate predictions.

In this study, we explored 98 possible combinations of training models by applying various machine learning and ensemble learning techniques combined with resampling methods, feature selection using SHAP, and dimensionality reduction using PCA. However, our findings indicate that the combination of resampling techniques with both SHAP and PCA did not yield satisfactory results in terms of AUC and balanced accuracy across all models. In contrast, when dimensionality reduction was excluded, and the SHAP threshold was increased (resulting in 10 selected features), all models showed significant performance improvements.

For instance, combining resampling techniques with SHAP resulted in higher AUC-ROC and balanced accuracy (B-ACC) values across all models, particularly for ensemble learning methods. Among these, Stacking consistently outperformed other models. Specifically, with the combination of SMOTEENN and SHAP, Stacking achieved an AUC-ROC of 0.954 and a balanced accuracy of 0.638, while the XGBoosting Model reached an AUC-ROC of 0.944. This trend was also observed with other SMOTE-based techniques such as Borderline-SMOTE and SMOTE-SVM.

Interestingly, Random Under Sampling (RUS) combined with SHAP showed comparable performance to other resampling techniques, with all ensembles learning models achieving AUC-ROC scores above 0.92 and specifically balanced accuracy values (higher than 0.85) that was notably better than those achieved with SMOTE and ADASYN. Under this configuration, Stacking achieved an AUC-ROC of 0.938 and a balanced accuracy of 0.888. Even when dimensionality reduction (PCA) was applied alongside SHAP, RUS maintained strong performance, suggesting its robustness as a resampling method in this context.

Our results suggest that dimensionality reduction using PCA is not effective or applicable for this analysis, as it consistently underperformed compared to setups without PCA. However, Random Under Sampling emerged as a key technique for backorder prediction, achieving the highest balanced accuracy across all trained models compared to other resampling methods.

Additionally, our findings highlight the superior performance of SMOTEENN compared to other SMOTE-based and other resampling techniques. This emphasizes the role of Edited Nearest Neighbors (ENN) in reducing noise, resulting in a cleaner and more balanced dataset that significantly enhances model accuracy in detecting backorders. Notably, to the best of our knowledge and until the preparation of this research, this study is the first to explore the application of SMOTEENN on this specific dataset as a severely imbalanced one, marking an innovative contribution to the literature.

Lastly, our research demonstrates that ensemble learning models perform better than conventional machine learning models, with Stacking continuously outperforming other ensemble techniques. However, in some cases, the XGBoosting Model demonstrated slightly higher balanced accuracy values, indicating that both approaches are highly effective for imbalanced backorder prediction tasks. These findings reinforce the importance of carefully selecting resampling techniques and model architectures to address class imbalance and improve prediction accuracy.

Our results reflect the low values of the F1 score and AUC-PR indicating the model struggles with effectively identifying the minority class. The model is obviously biased towards predicting the majority class due to the imbalance in the dataset, which results in poor performance on the minority class. This has a direct impact on the F1 score and AUC-PR. These metrics emphasize minority class performance and are highly affected by a low recall or precision.

Regarding the main research question, the following points are noteworthy:

The results indicate that while advanced machine learning models such as K-nearest neighbors (KNN) and neural networks (NN) face challenges in handling class imbalance, they still outperform traditional forecasting methods in backorder prediction accuracy within supply chain management. Despite their limitations in addressing imbalanced data, these models demonstrate improved predictive capabilities, making them more effective than conventional approaches in identifying potential backorders.

For the second question, the findings of this research emphasize the critical role of ensemble learning techniques—specifically staking model, and XGBoost—in enhancing the accuracy and reliability of backorder prediction models. Among these, the Stacking model demonstrated the highest overall performance, achieving the best AUC-ROC and balanced accuracy scores across various resampling techniques. These results confirm that ensemble learning methods outperform traditional machine learning models, making them a superior choice for backorder prediction in supply chain management.

To answer question three (Q3), the findings of this study demonstrate that data preprocessing techniques, particularly resampling methods, have a significant impact on the performance of machine learning models in backorder prediction. Among the resampling methods tested, SMOTEENN consistently outperformed other SMOTE-based techniques by effectively balancing the dataset while reducing noise, leading to enhanced model accuracy. Additionally, random under sampling (RUS) emerged as a highly effective technique, achieving the highest balanced accuracy across all trained models, surpassing SMOTE and ADASYN in performance. These results highlight the importance of carefully selecting resampling strategies to mitigate class imbalance issues in backorder prediction. Furthermore, while missing value imputation was not the primary focus of this study, advanced imputation methods can further refine data quality and contribute to model stability and reliability.

Finally, feature selection using SHAP played a crucial role in improving backorder prediction accuracy by identifying the most relevant features while eliminating less impactful ones. When the SHAP threshold was increased, resulting in 10 selected features, all models exhibited significant performance improvements, particularly in terms of AUC-ROC and balanced accuracy. However, dimensionality reduction using PCA failed to yield similar improvements and, in some cases, led to decreased model performance. This is likely because PCA transforms features into uncorrelated components, potentially discarding critical domain-specific information necessary for accurate backorder prediction. In contrast, SHAP preserves interpretability and feature importance, ensuring that the most meaningful variables contribute to model predictions. Thus, while feature selection enhances model efficiency, PCA does not appear to be a suitable dimensionality reduction technique for this problem.

Results show that stacking model outperforms other models. Stacking incorporates multiple machine learning algorithms, each trained with different learning mechanisms (e.g., tree-based models like XGBoost and LightGBM, distance-based models like KNN, and deep learning models like Neural Networks). By combining models with varied strengths, stacking enhances generalization and predictive robustness.

Stacking is an ensemble learning technique that combines multiple base models to enhance predictive performance. The meta-learner in stacking is trained on the predictions of these base models, learning how to optimally combine their outputs. Instead of relying on a single model's decision, the meta-learner refines predictions by assigning appropriate weights to each base model's contribution, leading to better overall accuracy. Traditional ensemble methods like bagging (e.g., random forest) primarily reduce variance, while boosting methods (e.g., Gradient Boosting, XGBoost, LightGBM) reduce bias by sequentially improving weak learners. Stacking benefits from both approaches by integrating models that minimize both bias and variance, leading to a well-balanced predictive model (Soni, 2021 and Bajaj, 2024)

In this study, resampling techniques (such as SMOTEENN and RUS) played a crucial role in balancing the dataset. Stacking model further enhances performance by allowing models trained on resampled data to collectively refine predictions, reducing misclassification in the minority class. The study demonstrated that feature selection with SHAP significantly improved model performance. The stacked model, with its multiple base learners, was able to effectively utilize selected features and optimize feature interactions that single models might overlook. The results indicated that the stacked model achieved the highest AUC-ROC (0.954) and performed well in balanced accuracy, outperforming other ensemble methods like XGBoost and LightGBM. This highlights its ability to consistently distinguish between backordered and non-backordered products.

This study advances the field of backorder prediction within supply chain management by implementing a more comprehensive methodology than previous research. Unlike prior studies, which typically employed only one or two resampling techniques, this research systematically evaluated five different resampling techniques (SMOTE-ENN, Borderline-SMOTE, SMOTE-SVM, ADASYN, and RUS), allowing for a more in-depth analysis of class imbalance handling. Furthermore, feature selection and feature engineering techniques, particularly SHAP and PCA, were incorporated to assess their impact on model performance, which was absent in most previous studies. The implementation of diverse evaluation metrics, including F1 Score, balanced accuracy, ROC-AUC, PR-AUC, and Specificity, ensures that this research provides a more holistic assessment of model performance, in contrast to prior studies that relied solely on accuracy—an unreliable measure for imbalanced datasets.

Comparing these findings with previous studies, Hajek and Zoynul Abedin (2020) utilized SMOTE and CBUS as resampling techniques, achieving an ROC-AUC of 91.57% with a Random Forest model. However, they did not incorporate feature selection, potentially limiting the interpretability of their results. Similarly, Islam and Amin (2020) reported higher AUC values (0.959 for Distributed Random Forest and 0.946 for GBM) using SMOTE, yet their study did not explore feature selection or consider a broader range of resampling methods. De Santis et al. (2017) achieved comparable results, with AUC values of 0.9482, 0.9441, and 0.9478 for Gradient Boosting, Random Forest, and BLAG, respectively, using SMOTE and RUS. However, they did not incorporate feature selection or feature engineering techniques, which may have affected their models' robustness. Ntakolia et al. (2022) employed Recursive Feature Elimination (RFE) with Random Forest for feature selection and used RUS as the resampling technique. They achieved AUC scores of 0.95 for RF, 0.80 for Logistic Regression, and 0.84 for SVM, demonstrating that feature selection can enhance model performance. In contrast, this study employed SHAP, which provides more interpretability than RFE by quantifying individual feature contributions. Additionally, the present research tested multiple machine learning models and ensemble learning approaches, providing a more comprehensive analysis. Shajalal et al. (2022) focused on CNN models, incorporating ADASYN and SMOTE as resampling techniques and SHAP for feature selection, achieving an AUC of 0.9489. However, their reliance on accuracy as the primary evaluation metric undermines the reliability of their conclusions due to the severe class imbalance in the dataset. Similarly, Ali et al. (2024) used random down sampling and feature importance for feature selection but reported only an accuracy of 0.88 for Random Forest, neglecting critical metrics like ROC-AUC and balanced accuracy.

One of the most significant distinctions of this study is the superior performance of the Stacking model, which was not identified as the best-performing model in any prior research. While previous studies have primarily highlighted Random Forest, GBM, or CNN as top-performing models, this research demonstrates that stacking model, particularly when combined with SMOTEENN or RUS as resampling techniques and SHAP for feature selection, achieves the highest AUC-ROC (0.954) and balanced accuracy (0.888). This finding emphasizes the advantage of integrating multiple base models into a meta-learner to optimize predictive accuracy and robustness. Overall, this study contributes to the literature by demonstrating that a more comprehensive approach—including diverse resampling methods, rigorous feature selection, and appropriate evaluation metrics—can significantly enhance backorder prediction models. The results confirm the effectiveness of ensemble learning, particularly Stacking and XGBoost, in handling imbalanced datasets and improving predictive performance beyond what previous studies have achieved.

8. Conclusion

In this chapter, we will discuss the limitations that influenced our findings and provide suggestions for future research. This study contributes to the literature by demonstrating that the integration of advanced machine learning techniques, such as ensemble learning models and innovative data preprocessing methods, can significantly enhance backorder prediction in supply chain management.

Additionally, this work emphasizes how crucial feature selection, resampling, and hyperparameter tuning strategies are for enhancing machine learning algorithms' performance on imbalanced datasets. By addressing class imbalance with methods like SMOTEENN and ADASYN, leveraging SHAP for feature importance analysis, and employing optimized cross-validation strategies, this research provides a robust framework for handling the complexities of real-world supply chain datasets. In this study the combination of different resampling techniques, feature selection methods, hyperparameter tuning, machine learning and ensemble learning techniques led to training of totally 98 different models. The specific findings of this research emphasize on the critical role of methodology in developing reliable predictive models for inventory management and backorder forecasting.

8.1. Remarks

Backorder management plays an essential role in inventory and supply chain management. Incorrect predictions about backorders can disrupt inventory control and production processes. It is the reason that accurate predictions are crucial to reducing the risk of backorders, even though they are relatively rare. This study focuses on using various resampling techniques and machine learning methods, including ensemble learning, to tackle the imbalanced dataset and make precise backorder predictions.

The findings show that ensemble learning outperformed classic machine learning models (e.g., Neural Networks and K-Nearest Neighbors) across all combinations of resampling methods and feature selection techniques. Our research contributes to the existing literature on ensemble learning models such as Gradient Boosting and Random Forest, highlighting their effectiveness in backorder prediction. Additionally, by employing various resampling methods—including SMOTE, RUS, and ADASYN—we demonstrated the critical role these techniques play in addressing the challenge of severely imbalanced classes.

In this study, the SHAP technique has been employed to evaluate the contribution of features to the performance of the predictive model for backorder classification. SHAP also provides interpretability for model predictions by identifying the impact of individual features, thereby enhancing the transparency and explainability of the machine learning models. The original dataset includes 22 variables; however, SHAP identified 10 key features as the most influential.

This study has made several significant contributions within the supply chain context.

First, the research delivers a robust and reliable forecasting model that can be effectively used for backorder prediction and inventory management. The proposed model is applicable in the production and retail sectors for both raw materials and finished goods, helping to minimize the risks of backorders, client loss, stockouts, or excess inventory—factors that can significantly impact a company's profitability.

Additionally, by emphasizing scalable and dependable solutions, this research seeks to improve decision-making procedures, save operating expenses, boost supply chain effectiveness, and lessen the bullwhip impact, which can upset the supply-demand balance.

Finally, this thesis provides a practical and actionable solution that can be implemented to improve supply chain and inventory management practices across various industries.

8.2. Managerial Implications

Managers and decision-makers are among the beneficiaries of this research. The managerial implications of this research are significant for companies seeking to improve operational efficiency and decision-making in supply chain management, particularly in addressing backorder challenges. By demonstrating the superior performance of ensemble learning models such as XGBoost and stacking model, this study provides a clear pathway for managers to implement advanced predictive techniques to mitigate the impact of class imbalance in their datasets. The findings emphasize the importance of selecting appropriate resampling methods, with SMOTEENN and Random Under Sampling emerging as particularly effective strategies for enhancing model accuracy and balanced predictions. These insights can guide managers in refining their data preprocessing workflows to improve forecast reliability.

Moreover, the research highlights that while feature selection using SHAP adds considerable value by identifying the most impactful variables, dimensionality reduction through PCA may not be suitable for all predictive contexts, particularly for backorder forecasting. This underscores the need for targeted, problem-specific approaches when implementing machine learning solutions. Managers can leverage these findings to allocate resources more effectively, prioritize the adoption of ensemble learning methods, and focus on hybrid resampling techniques that deliver cleaner and more balanced datasets, ultimately leading to improved forecasting precision and operational resilience in dynamic supply chains.

Based on the findings of this research, several key recommendations can be made for supply chain managers seeking to enhance backorder prediction and optimize inventory management through machine learning:

8.2.1. Focus on the Most Important Variables for Backorder Prediction

Through SHAP-based feature selection, the most influential variables in predicting backorders were identified. Supply chain managers should prioritize these key variables when implementing predictive models:

- Inventory levels (National_inv): Critical for determining stock availability.
- Lead time: Helps in estimating delays in replenishment.
- Forecasting variables (forecast_3_month, forecast_9_month): Provide predictive insights into future demand.
- Sales history (1-month, 3-month, 6-month, and 9-month forecasts): Essential for tracking demand trends.
- Supplier performance (perf_6_month_avg and perf_12_month_avg): Helps assess reliability.

By focusing on these variables, supply chain managers can ensure that predictive models capture the most relevant information for backorder forecasting.

8.2.2. Choose the Right Machine Learning Model

The findings indicate that Stacking consistently outperforms other models in terms of predictive accuracy and robustness. However, if managers are looking for a single, stand-alone model, XGBoost is the most effective individual model for backorder prediction.

- Best individual model: XGBoost
  - Strong performance in handling imbalanced datasets.
  - High AUC-ROC and balanced accuracy compared to other models.
  - Well-suited for real-world supply chain applications.
- Best overall model: stacking model
  - Combines the strengths of multiple models.
  - Learns from the predictions of base models, optimizing overall performance.
  - Demonstrated superior results with SMOTEENN and RUS resampling techniques.

Managers should prioritize Stacking for the most accurate backorder predictions while considering XGBoost if computational efficiency or simplicity is a concern.

8.2.3. Implementing and using machine learning models in supply chain operations

To effectively implement machine learning in supply chain management, managers should follow these steps:

- Data collection & preparation
  - Ensure accurate and comprehensive data collection, particularly for the key variables identified.
  - Use advanced imputation techniques (such as MICE) to handle missing values instead of discarding incomplete records.
- Handling class imbalance
  - Apply resampling techniques (SMOTEENN, RUS) to balance datasets before training models.
  - Use proper evaluation metrics like balanced accuracy, AUC-ROC, and PR-AUC, instead of relying solely on Accuracy, which may be misleading in imbalanced datasets.
- Model deployment & integration
  - Integrate predictive models into supply chain management systems (e.g., ERP software, inventory tracking platforms).
  - Implement real-time backorder alerts based on model outputs to allow for proactive decision-making.
  - Use explainability tools like SHAP to interpret model predictions and improve trust among stakeholders.
- Continuous monitoring & model updating
  - Regularly retrain models with updated data to maintain accuracy over time.
  - Monitor model performance and adjust hyperparameters as needed to optimize results.

By adopting these recommendations, supply chain managers can significantly improve their ability to predict and prevent backorders, leading to better inventory planning, reduced stockouts, optimized supply chain efficiency, and improved customer satisfaction.

8.3.Limitation

A key limitation of this research is the use of a stratified subset of the original dataset, which contained over 1.6 million observations. Due to computational constraints, a reduced dataset of 16,000 samples was employed for analysis. While this approach preserved the class balance and key characteristics of the original data, it may have limited the model's ability to capture all underlying patterns in the larger dataset. Additionally, GridSearchCV, a robust hyperparameter tuning method, was not utilized because it was computationally demanding. Given the scope of the study, which involved evaluating seven different machine learning and ensemble learning models combined with five resampling techniques, allocating the necessary resources to perform GridSearchCV was not feasible.

8.4. Future research

Other hyperparameter tuning methods or the efficiency of other feature extraction approaches on classification problems including backorders or imbalanced datasets could be the subject of future research. In general, this research offers significant perspectives for scholars and professionals engaged in supply chain forecasting and backorder classification.

Future research can focus on and develop other aspects of this study. For instance, exploring other ensemble learning methods, such as bagging, or analyzing classical machine learning methods could provide further improvements. Additionally, incorporating more resampling techniques and data augmentation methods that primarily focus on the minority class would enhance model performance. Future research could explore the integration of other hyperparameter tuning methods, such as GridSearchCV or Bayesian Optimization, to address the computational challenges posed by GridSearchCV.

- Future research could introduce a cost-sensitive learning approach that prioritizes backorder predictions with an economic impact model. This could be similar to the approach by Hajek and Abedin (2020) but integrated into the training phase of the machine learning models used in this study.
- Most studies discussed in this field rely on historical data. However, as supply chains become more dynamic, real-time data could provide better predictive capabilities.
- Real-Time Data Processing: Future research could integrate streaming data, such as real-time demand fluctuations, into predictive models to improve their accuracy and responsiveness to changing conditions.
- While most models focus on a specific supply chain, learning across industries could add additional value. Transfer Learning: Applying transfer learning techniques by training models on one dataset (e.g., retail) and fine-tuning them for another industry (e.g., manufacturing) could increase robustness and adaptability across domains.
- Future work could investigate how specific interactions between features (e.g., lead time and inventory levels) can be explicitly modeled using interaction-based approaches such as Deep Feature Synthesis (DFS).
- Few studies directly address the impact of sudden supply chain disruptions, such as

geopolitical events or pandemics, on backorder predictions, researching on this field is another interesting idea.

- Developing models that incorporate external data, such as global events, to anticipate backorders could involve scenario modeling and stress-testing predictions under various disruption scenarios.

References

Abbas, K., Afaq, M., Khan, T. A., and Song, W. C. (2020). A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. *Electronics, 9*(5), 852. https://doi.org/10.3390/electronics9050852.

Abbasi, B., Babaei, T., Hosseinifard, Z., Smith-Miles, K., and Dehghani, M. (2020). Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management. *Computers & Operations Research, 119*, 104941. https://doi.org/10.1016/j.cor.2020.104941.

Abdi, H., and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433-459. https://doi.org/10.1002/wics.101.

Adana, S., Cevikparmak, S., Celik, H., and Uvet, H. (2019, November). Predicting backorders using machine learning techniques. *Proceedings of the 50th Annual Conference of the Decision Sciences Institute*.

Aguilar-Palacios, C., Muñoz-Romero, S., and Rojo-Álvarez, J. L. (2020). Cold-start promotional sales forecasting through gradient boosted-based contrastive explanations. *IEEE Access, 8*, 137574-137586. https://doi.org/10.1109/ACCESS.2020.3011805.

Ahmed, F., Hasan, M., Hossain, M. S., and Andersson, K. (2022). Comparative performance of tree-based machine learning classifiers in product backorder prediction. In P. Vasant, G.-W. Weber, J. A. Marmolejo-Saucedo, E. Munapo, and J. J. Thomas (Eds.), *Intelligent computing and optimization* (Vol. 569, pp. 572–584). Springer International Publishing. https://doi.org/10.1007/978-3-031-19958-5_54.

Ali, A., Jayaraman, R., Azar, E., and Maalouf, M. (2024). Maximizing supply chain performance leveraging machine learning to anticipate customer backorders. *Computers and Industrial Engineering, 194*, 110414. https://doi.org/10.1016/j.cie.2024.110414.

Arowolo, M. O., Aigbogun, H. E., Michael, P. E., Adebiyi, M. O., and Tyagi, A. K. (2023). A predictive model for classifying colorectal cancer using principal component analysis. In *Data Science for Genomics* (pp. 205–216). Elsevier. https://doi.org/10.1016/B978-0-323-98352-5.00004-5.

Awe, O. O., and Vance, E. A. (Eds.). (2020). *Practical statistical learning and data science methods: Case studies from LISA 2020 global network, USA*. STEAM-H: Science, Technology, Engineering, Agriculture, Mathematics & Health Series.

Azur, M. J., Stuart, E. A., Frangakis, C., and Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research, 20*(1), 40–49. https://doi.org/10.1002/mpr.329.

Bader-El-Den, M., Teitei, E., & Perry, T. (2019). Biased random forest for dealing with the class imbalance problem. *IEEE Transactions on Neural Networks and Learning Systems, 30*(7), 2163–2172. https://doi.org/10.1109/TNNLS.2018.2878400.

Bajaj, G. (2024). *Bagging Reduces Variance while Boosting Reduces Bias*. Artificial Intelligence in Plain English.

Bao, L., Liu, Z., Yu, Y., and Zhang, W. (2018). On the decomposition property for a dynamic inventory rationing problem with multiple demand classes and backorder. *European Journal of Operational Research, 265*(1), 99-106. https://doi.org/10.1016/j.ejor.2017.07.021.

Batista, G. E. A. P. A., Prati, R. C., and Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter, 6*(1), 20–29. https://doi.org/10.1145/1007730.1007735.

Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour*. Princeton University Press.

Belouch, M., & Hadaj, S. E. (2017). Comparison of ensemble learning methods applied to network intrusion detection. *Proceedings of the 2nd International Conference on Big Data, Cloud and Applications (BDCA)*. Association for Computing Machinery. https://doi.org/10.1145/3018896.3065830.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). *Algorithms for hyper-parameter optimization*. In Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS). Retrieved from https://www.researchgate.net/publication/216816964.

Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.

Bishop, C. M. (2013). *Pattern recognition and machine learning*. Springer.

Boctor, F. F., and Bolduc, M.-C. (2015). Inventory replenishment planning and staggering. *IFAC-PapersOnLine, 48*(3), 1416-1421. https://doi.org/10.1016/j.ifacol.2015.06.285.

Bouganim, J., and Olsson, K. (2019). Using deep learning to predict back orders: A study in the Volvo Group aftermarket supply chain. *Linköping University*. http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-157312.

Carbonneau, R., Vahidov, R., and Laframboise, K. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research, 184*(3), 1140–1154.

Chakraborty, I., Roy, D., Garg, I., Ankit, A., and Roy, K. (2020). Constructing energy-efficient mixed-precision neural networks through principal component analysis for edge intelligence. *Nature Machine Intelligence, 2*(1), 43-55. https://doi.org/10.1038/s42256-019-0151-7.

Chan, S. W., Tasmin, R., Aziati, A. N., Rasi, R. Z., Ismail, F. B., and Yaw, L. P. (2017). Factors influencing the effectiveness of inventory management in manufacturing SMEs. *IOP*

Conference Series: Materials Science and Engineering, 226(1), 012024. http://dx.doi.org/10.1088/1757-899X/226/1/012024.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research, 16*, 321–357. https://doi.org/10.5555/1622407.1622416.

Chen, J., Huang, S., Hassin, R., and Zhang, N. (2015). Two backorder compensation mechanisms in inventory systems with impatient customers. *Production and Operations Management, 24*(10), 1640–1656. https://doi.org/10.1111/poms.12360.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785.

Chinello, E., Herbert-Hansen, Z. N. L., and Khalid, W. (2019). Assessment of the impact of inventory optimization drivers in a multi-echelon supply chain: Case of a toy manufacturer. *Computers and Industrial Engineering, 139*, 106232. https://doi.org/10.1016/j.cie.2019.106232.

Datta, S., Sasso, A. M., Kiwit, N., Bose, S., Nadkarni, G., Miotto, R., & Böttinger, E. P. (2022). Predicting hypertension onset from longitudinal electronic health records with deep learning. *JAMIA Open, 5*(4), ooac097. https://doi.org/10.1093/jamiaopen/ooac097.

Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., Wamba, S. F., and Roubaud, D. (2019). Can big data and predictive analytics improve social and environmental sustainability? *Technological Forecasting and Social Change, 144*, 534–545. https://doi.org/10.1016/j.techfore.2017.06.020.

Dehghan-Bonari, E., Shajalal, M., Sharma, S., and Abiodun, O. I. (2021). Backorder prediction in inventory management: Classification techniques and cost considerations. *International Journal of Supply Chain Management, 10*(3), 56–67. https://doi.org/10.1016/j.scm.2021.07.012.

ElHafsi, M., Fang, J., and Hamouda, E. (2021). Optimal production and inventory control of multi-class mixed backorder and lost sales demand class models. *European Journal of Operational Research, 291*(1), 147-161.

Ezziane, H., Houassine, H., Moulahoum, S., & Chaouche, M. S. (2022). A novel method to identification type, location, and extent of transformer winding faults based on FRA and SMOTE-SVM. *Russian Journal of Nondestructive Testing, 58*(5), 391–404. https://doi.org/10.1134/S1061830922050047.

Feizabadi, J. (2020). Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, 1-24. https://doi.org/10.1080/13675567.2020.1850976.

Feurer, M., Springenberg, J., and Hutter, F. (2015). Initializing Bayesian hyperparameter optimization via meta-learning. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, *29*(1). https://doi.org/10.1609/aaai.v29i1.9354.

Fonseca, B., Nogueira, H., and Ravetti, G. (2019). A hybrid Lagrangian metaheuristic for the cross-docking flow shop scheduling problem. *European Journal of Operational Research, 275*(1), 139-154. https://doi.org/10.1016/j.ejor.2018.11.033.

Gaalman, G., Disney, S. M., and Wang, X. (2022). When bullwhip increases in the lead time: An eigenvalue analysis of ARMA demand. *International Journal of Production Economics, 250*, 108623. https://doi.org/10.1016/j.ijpe.2022.108623

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., and Herrera, F. (2012). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 42*(4), 463-484.

Gao, H., Ren, Q., Lv, C., et al. (2022). Supply chain management and backorder products prediction utilizing neural network and Naive Bayes machine learning techniques in big data area: A real-life case study. *Preprint*, Research Square. https://doi.org/10.21203/rs.3.rs-2020401/v1.

Garcia, Y., and Panduro, J. (2022, December). Use of a machine learning model for the reduction of backorders in the cross-docking sales process for the homecenter order service. Paper presented at the 1st Australian International Conference on Industrial Engineering and Operations Management. https://doi.org/10.46254/AU01.20220439.

Gaye, B., Zhang, D., & Wulamu, A. (2021). A Tweet Sentiment Classification Approach Using a Hybrid Stacked Ensemble Technique. *Information*, *12*(9), 374. https://doi.org/10.3390/info12090374.

Ge, D., Pan, Y., Shen, Z.-J. (Max), Wu, D., Yuan, R., and Zhang, C. (2019). Retail supply chain management: A review of theories and practices. *Journal of Data, Information and Management, 1*(1), 45–64. https://doi.org/10.1007/s42488-019-00004-z.

GeeksforGeeks. (2024, September 10). *Principal Component Analysis (PCA)*. https://www.geeksforgeeks.org/principal-component-analysis-pca/.

Goltsos, T. E., Syntetos, A. A., Glock, C. H., and Ioannou, G. (2022). Inventory – forecasting: Mind the gap. *European Journal of Operational Research, 299*(2), 397–419. https://doi.org/10.1016/j.ejor.2021.07.040.

Genton, M. (2001). Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research, 2*, 299–312.

Guo, J., Wu, H., Chen, X., & Lin, W. (2024). Adaptive SV-Borderline SMOTE-SVM algorithm for imbalanced data classification. *Applied Soft Computing, 150*, 110986. https://doi.org/10.1016/j.asoc.2023.110986.

Hajek, P., and Abedin, M. Z. (2020). A profit function-maximizing inventory backorder prediction system using big data analytics. *IEEE Access, 8*, 58982-58994. https://doi.org/10.1109/ACCESS.2020.2983118.

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *International Conference on Intelligent Computing* (pp. 878–887). Springer.

He, H., and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263-1284.

He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IJCNN)*, 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2014). *An efficient approach for assessing hyperparameter importance*. Proceedings of the 31st International Conference on Machine Learning, Beijing, China. Retrieved from https://www.researchgate.net/publication/287850221_An_efficient_approach_for_assessing_hyperparameter_importance.

Iqbal, G. M. D., Rosenberger, M., Ha, L., Gregory, S., and Anoruo, E. (2023). Classification and regression tree model to predict the probability of a backorder in uncertain supply chain. *International Journal of Supply Chain Management, 12*(4). https://doi.org/10.59160/ijscm.v12i4.6199.

Islam, M., and Amin, M. (2020). Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques.

Jamaleddyn, I., El Ayachi, R., and Biniz, M. (2023). An improved approach to Arabic news classification based on hyperparameter tuning of machine learning algorithms. *Journal of Economic Research*, *35*(1), Article 100061. https://doi.org/10.1016/j.jer.2023.100061.

Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202.

Kaur, R., & Singh, S. (2023). A comprehensive review of object detection with deep learning. *Digital Signal Processing, 132*, Article 103812. https://doi.org/10.1016/j.dsp.2022.103812.

Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation, and evaluation. *Expert Systems with Applications, 244,* 122778. https://doi.org/10.1016/j.eswa.2023.122778.

Khan, M., Qayoom, A., Nizami, M., Siddiqui, M., Wasi, S., and Raazi, S. (2021). Automated prediction of good dictionary examples (GDEX): A comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques. *Complexity, 2021*, 2553199. https://doi.org/10.1155/2021/2553199.

Kim, M., and Hwang, K.-B. (2022). An empirical evaluation of sampling methods for the classification of imbalanced data. *PLoS ONE, 17*(7), e0271260. https://doi.org/10.1371/journal.pone.0271260.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31*, 249–268.

Lanjewar, M. G., and Panchbhai, K. G. (2024). Enhancing fish freshness prediction using NasNet-LSTM. *Journal of Food Composition and Analysis, 127*, 105945. https://doi.org/10.1016/j.jfca.2023.105945.

Hong, L., and Ping,W. (2007). Bullwhip effect analysis in supply chain for demand forecasting technology. *Systems Engineering - Theory and Practice, 27(7)*, 26-33. https://doi.org/10.1016/S1874-8651(08)60044-7.

Lawal, S. O. and Akintola, K. G. (2021). A product backorder predictive model using recurrent neural network. *Iconic Research And Engineering Journals, 4*(8), 49–57.

Li, B., and Zhuang, X. (2020). Multiscale computation on feedforward neural network and recurrent neural network. *Frontiers of Structural and Civil Engineering, 14*(6), 1285–1298. https://doi.org/10.1007/s11709-020-0691-7.

Lopez, V., Fernandez, A., Garcia, S., Palade, V., and Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences, 250*, 113-141. https://doi.org/10.1016/j.ins.2013.07.007.

Lundberg, M., Erion, G., Chen, H., DeGrave, A., Prutkin, M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56-67. https://doi.org/10.1038/s42256-019-0138-9.

Lundberg, S., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv*. https://doi.org/10.48550/arXiv.1705.07874.

Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.-S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access, 7*, 93010–93022. https://doi.org/10.1109/ACCESS.2019.2927266.

Maitra, S., and Kundu, S. (2023). Backorder prediction in inventory management: Classification techniques and cost considerations. *Journal of Inventory Management and Logistics Research*, *32*(2), 104-117.

Mishra, U., Wu, J.-Z., and Sarkar, B. (2021). Optimum sustainable inventory management with backorder and deterioration under controllable carbon emissions. *Journal of Cleaner Production, 279*, 123699. https://doi.org/10.1016/j.jclepro.2020.123699.

Moshtagh, M., and Taleizadeh, A. (2017). Stochastic integrated manufacturing and remanufacturing model with shortage, rework, and quality-based return rate in a closed loop supply chain. *Journal of Cleaner Production, 141*, 1548–1573. https://doi.org/10.1016/j.jclepro.2016.09.173.

Namazkhan, M., Albers, C., and Steg, L. (2020). A decision tree method for explaining household gas consumption: The role of building characteristics, socio-demographic variables, psychological factors, and household behavior. *Renewable and Sustainable Energy Reviews, 119*, 109542. https://doi.org/10.1016/j.rser.2019.109542.

Ntakolia, C., Kokkotis, C., Karlsson, P., and Moustakidis, S. (2022). An explainable machine learning model for material backorder prediction in inventory management. *Sensors, 21*(23), 7926. https://doi.org/10.3390/s21237926.

Oluwaseyi, J. A., Onifade, M. K., and Odeyinka, O. F. (2017). Evaluation of the role of inventory management in logistics chain of an organisation. *LOGI – Scientific Journal on Transport and Logistics, 8*(2), 1–11.

Osama, S., Shaban, H., & Ali, A. A. (2023). Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: A comprehensive review. *Expert Systems with Applications, 213*(Part A), Article 118946. https://doi.org/10.1016/j.eswa.2022.118946.

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., and Swami, A. (2017, April). Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security* (pp. 506-519). https://doi.org/10.1145/3052973.3053009.

Pérez-Padilla, N. A., Garcia-Sanchez, R., Avalos, O., Gálvez, J., Bian, M., Yu, L., Shu, Y., Feng, M., and Yelian, F. D. (2024). Optimizing trigger timing in minimal ovarian stimulation for in vitro fertilization using machine learning models with random search hyperparameter tuning. *Computers in Biology and Medicine*. https://doi.org/10.1016/j.compbiomed.2024.108856.

Pillai, V. M., and Pamulety, T. C. (2013). Impact of backorder on supply chain performance: An experimental study. Paper presented at the 7th IFAC Conference on Manufacturing Modelling, Management, and Control, Saint Petersburg, Russia.

Profillidis, V. A., and Botzoris, G. N. (2019). Artificial Intelligence—Neural Network Methods. In V. A. Profillidis and G. N. Botzoris (Eds.), *Modeling of Transport Demand* (pp. 353–382). Elsevier. https://doi.org/10.1016/C2016-0-04307-6.

Ryan, W. T., Christopher, R. W., Geoffrey, M. W., and Robin, K. M. (2021**).** *Backorder prediction*. https://apps.dtic.mil/sti/citations/AD1141471.

Rodger, J. A. (2014). Application of a fuzzy feasibility Bayesian probabilistic estimation of supply chain backorder aging, unfilled backorders, and customer wait time using stochastic simulation with Markov blankets. *Expert Systems with Applications, 41*(16), 7005-7022. https://doi.org/10.1016/j.eswa.2014.05.012.

Sadaiyandi, J., Arumugam, P., Sangaiah, A. K., & Zhang, C. (2023). Stratified sampling-based deep learning approach to increase prediction accuracy of unbalanced dataset. *Electronics, 12*(21), 4423. https://doi.org/10.3390/electronics12214423.

Sai, M. J., Chettri, P., Panigrahi, R., Garg, A., Bhoi, A. K., & Barsocchi, P. (2023). An ensemble of Light Gradient Boosting Machine and Adaptive Boosting for prediction of type-2 diabetes. *International Journal of Computational Intelligence Systems, 16*(1), Article 14. https://doi.org/10.1007/s44196-023-00184-y.

Samad, M. D., Abrar, S., and Diawara, N. (2022). Missing value estimation using clustering and deep learning within multiple imputation framework. *Knowledge-Based Systems, 249*, 108968. https://doi.org/10.1016/j.knosys.2022.108968.

Santis, R., Pestana de Aguiar, E., & Goliatt, L. (2017). Predicting material backorders in inventory management using machine learning. In *Proceedings of the 4th IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, Arequipa, Peru. https://doi.org/10.1109/LA-CCI.2017.8285684.

Satpathy, S. (2020). SMOTE for imbalanced classification with Python. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/.

Seyedan, M., and Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. *Journal of Big Data, 7*(1). https://doi.org/10.1186/s40537-020-00329-2.

Shaik, M. E., Islam, M. M., and Hossain, Q. S. (2021). A review on neural network techniques for the prediction of road traffic accident severity. *Asian Transport Studies, 7,* 100040. https://doi.org/10.1016/j.eastsj.2021.100040.

Shajalal, M., Boden, A., and Stevens, G. (2022). Explainable product backorder prediction exploiting CNN: Introducing explainable models in businesses. *Electronic Markets, 32*(1). https://doi.org/10.1007/s12525-022-00599-z.

Shajalal, M., Hajek, P., and Abedin, M. Z. (2021). Product backorder prediction using deep neural network on imbalanced data. *International Journal of Production Research, 61*(1), 302-319. https://doi.org/10.1080/00207543.2022.2071212.

Shi, X., Bahnsen, A. C., and Munk, A. G. (2018). Leveraging domain knowledge in transfer learning: A case study of machine learning models for predicting backorders. *Expert Systems with Applications, 106*, 163-172. https://doi.org/10.1016/j.eswa.2018.04.006.

Shoeibi, A., Khodatars, M., Jafari, M., Ghassemi, N., Moridian, P., Alizadehsani, R., Ling, S. H., Khosravi, A., Alinejad-Rokny, H., Lam, H. K., Fuller-Tyszkiewicz, M., Acharya, U. R., Anderson, D., Zhang, Y., & Gorriz, J. M. (2023). Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: A review. *Information Fusion, 93,* 85–117. https://doi.org/10.1016/j.inffus.2022.12.010.

Smiti, S., & Soui, M. (2020). Bankruptcy prediction using a deep learning approach based on Borderline SMOTE. *Information Systems Frontiers, 22*(5), 1067–1083. https://doi.org/10.1007/s10796-020-10031-6.

Soni, B. (2021). *Stacking to Improve Model Performance: A Comprehensive Guide on Ensemble Learning in Python*. Medium.

Subramanian, N., Gunasekaran, A., Yu, K., Cheng, J., and Ning, K. (2014). Influence of performance measures on performance outcomes: A balanced scorecard perspective on the roles of cognition and emotional intelligence. *International Journal of Production Research, 52*(7), 2009–2025.

Sudharsan, M., and Thailambal, G. (2023). Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA). *Materials Today: Proceedings, 81*(Part 2), 182-190. https://doi.org/10.1016/j.matpr.2021.03.061.

Sundarakani, B., and Hamid, M. A. (2017). Big data driven supply chain analytics model for backorder prediction in manufacturing. Paper presented at the 8th IFAC Conference on Manufacturing Modelling, Management, and Control, Troyes, France. https://doi.org/10.1016/j.ifacol.2017.07.030.

Surendran, R. V., Arunachalam, M., and Muthulakshmi, A. (2019). Backorder prediction in inventory management systems using LSTM deep learning. Paper presented at the 2nd International Conference on Inventive Research in Computing Applications, Coimbatore, India. https://doi.org/10.1109/ICIRCA.2019.8899515.

Tan, X., & Lu, Q. (2021). Predicting supply chain backorder using data mining and machine learning techniques. In *Proceedings of the 2021 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, Beijing, China. https://doi.org/10.1145/3465924.3465945.

Tang, Z., and Ge, Y. (2021). CNN model optimization and intelligent balance model for material demand forecast. *International Journal of System Assurance Engineering and Management*, 1-9. https://doi.org/10.1007/s13198-021-01157-0.

Thompson, P., and Psaraftis, H. N. (2019). Machine learning techniques in supply chain management: A systematic literature review. *IFAC-PapersOnLine, 52*(13), 1779-1785. https://doi.org/10.1016/j.ifacol.2019.11.426.

Tibshirani, R., Hastie, T., and Friedman, J. H. (2017). *The elements of statistical learning* (2nd ed.). Springer.

Umoh, U. A., Eyoh, I. J., Murugesan, V. S., & Nyoho, E. E. (2022). Fuzzy-machine learning models for the prediction of fire outbreaks: A comparative analysis. In *Artificial Intelligence and Machine Learning for EDGE Computing* (pp. 563–589). https://doi.org/10.1016/B978-0-12-824054-0.00025-3.

Vukovic, D. B., Spitsin, V., Bragin, A., Leonova, V., and Spitsina, L. (2024). Forecasting firm growth resumption post-stagnation. *Journal of Open Innovation: Technology, Market, and Complexity*, *100406*. https://doi.org/10.1016/j.joitmc.2024.100406.

Wan, X., Britto, R., and Zhou, Z. (2020). In search of the negative relationship between product variety and inventory turnover. *International Journal of Production Economics, 222*, 107503. https://doi.org/10.1016/j.ijpe.2019.09.024.

Wang, D., and Tang, O. (2014). Dynamic inventory rationing with mixed backorders and lost sales. *International Journal of Production Economics, 149*, 56–67. https://doi.org/10.1016/j.ijpe.2013.10.004.

Wang, D., and Zhang, Y. (2020). Implications for sustainability in supply chain management and the circular economy using machine learning model. *Information Systems and e-Business Management*, 1-13. https://doi.org/10.1007/s10257-020-00492-9.

Wang, G., Gunasekaran, A., Ngai, E. W. T., and Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics, 176*, 98–110. https://doi.org/10.1016/j.ijpe.2016.03.014.

Wang, K.-J., Adriana, A. M., Chen, K.-H., & Wang, K.-M. (2015). A hybrid classifier combining Borderline-SMOTE with AIRS algorithm for estimating brain metastasis from lung cancer: A case study in Taiwan. *Computer Methods and Programs in Biomedicine, 119*(1), 63–76. https://doi.org/10.1016/j.cmpb.2015.03.003.

Widodo, A., and Yang, B. S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing, 21*(5), 2560–2574. https://doi.org/10.1016/j.ymssp.2006.12.007.

Xia, J., Jin, X., Yan, M., Cai, G., & Yan, J. (2018). Class weights random forest algorithm for processing class imbalanced medical data. *IEEE Access, 6*, 4641–4652. https://doi.org/10.1109/ACCESS.2018.2789428.

Yildiz, H., and Nezami, M. (2021). Integrated production and transportation planning in a supply chain with capacitated vehicles and multiple production modes. *Transportation Research Part E: Logistics and Transportation Review, 152*, 102389. https://doi.org/10.1016/j.tre.2021.102389.

Yu, K., Wang, X., Zhong, R., and Huang, G. Q. (2017). E-commerce logistics in supply chain management: Practice perspective. *Procedia CIRP, 63*, 851-856. https://doi.org/10.1016/j.procir.2017.03.327.

Zhang, M., Shi, L., Zhuo, X., and Liu, Y. (2021). A bullwhip effect weakening approach based on VMD-SVM algorithm under the background of intelligent manufacturing. *Processes, 9*(11), 1957. https://doi.org/10.3390/pr9111957.

Zhang, Y., and Chu, F. (2011). Backorder inventory model in stochastic demand environment. Paper presented at the 8th International Conference on Service Systems and Service Management, Tianjin, China. https://doi.org/10.1109/ICSSSM.2011.5959466.

Zhu, Y., and Rabinovich, E. (2021). Investigating the backorder behavior of e-commerce firms: Evidence from the software industry. *Journal of Business Logistics, 42*(1), 24-43. https://doi.org/10.1111/jbl.12281.