LogForum

> Scientific Journal of Logistics <

2024, 20 (1), 55-70

http://doi.org/10.17270/J.LOG.001007

http://www.logforum.net p-ISSN 1895-2038 e-ISSN 1734-459X

ORIGINAL PAPER

ENHANCING SUPPLY CHAIN RESILIENCE: RIME-CLUSTERING AND ENSEMBLE DEEP LEARNING STRATEGIES FOR LATE DELIVERY RISK PREDICTION

Kaoutar Douaioui[1], Rachid Oucheikh[2], Charif Mabrouki[1]

1) Faculty of Science and Technology, Hassan First University of Settat, **Route de Casablanca**
2) Physical Geography and Ecosystem Science, Lund University, **Sweden**

ABSTRACT. Background: Global supply chains are confronted with the challenge of ensuring on-time deliveries while simultaneously enhancing supply chain resilience. Conventional methods aim to address the complexities of modern supply chains, promoting the transition to intelligent and data-driven strategies.
Methods: This research represents an innovative methodology for predicting the risk of late deliveries in supply chains. The presented framework combines clustering and multiclassification techniques, where the clustering phase is executed through hyperparameter optimization and a novel metaheuristic called RIME. In the multiclassification phase, five distinct deep learning models are employed, namely, Generative Adversarial Network (GAN), Convolutional Neural Network Long Short-Term Memory (CNN-LSTM), within Ensemble learning via bagging, Ensemble learning stacking, and Ensemble learning within boosting. The three ensemble learning models are based in GAN and CNN-LSTM.
Result: This paper presents a systematic evaluation of diverse models in a risk of late delivery prediction framework. This evaluation demonstrates that Ensemble learning stacking provides the higher accuracy by 0.926, showcasing its prowess in precise predictions. Notably, Ensemble learning bagging and Ensemble learning boosting exhibit strong precision. Regression metrics reveal Ensemble learning stacking and Ensemble learning bagging's superior error minimization (MSE 0.11, MAE 0.09). This metric demonstrates that the proposed model can predict the risk level of late delivery in a supply chain with high precision.
Conclusion: This paper introduces an innovative clustering and multiclassification-based framework for predicting the risk of late deliveries. The ability of prediction late deliveries risk helps organizations to enhance supply chain resilience by adopting a proactive management risks strategy, optimizing operational processes, and elevating customer satisfaction.

Keywords: Artificial Neural Networks (ANNs), Late Delivery Risk, Supply Chain Management, Clustering, Multiclassification, Deep Learning Models, RIME Optimization Algorithm.

INTRODUCTION

In the midst of the global supply chain, companies face the challenge of not only expanding their production worldwide but also meeting the rising expectations of customers for quick and on-time delivery (Fri et al., 2021; Douaioui et al., 2021). Balancing this act requires smart strategies in streamlining processes within the dynamic world of supply chain management (Fri et al., 2019). A crucial challenge in this dynamic environment is the risk of late deliveries, significantly impacting the success of supply chain management across industries (Douaioui et al., 2021). Accurately predicting delivery times is now more important than ever, bringing about a need to minimize delays and handle these risks proactively (Ngniatedema et al., 2016). The approach taken to address the precision of late delivery risks directly

Citation: Douaioui K., Oucheikh R., Mabrouki C., 2024. Enhancing Supply Chain Resilience: RIME-Clustering and Ensemble Deep Learning Strategies for Late Delivery Risk Prediction. LogForum 20 (1), 55-70, http://doi.org/10.17270/J.LOG.001007
Received: 31.12.2023, Accepted: 11.02.2024, on-line: 14.03.2024

impacts the efficiency of the supply chain, prompting a shift from conventional methods to more innovative ones (Menchaca-Méndez et al., 2022).

The approach used in this work for handling these risks has evolved over time. Early strategies used statistical methods like analysis of variance (ANOVA) to build systems targeting global late delivery times, especially in fast-paced environments (Patel et al., 2023; Mahmud, 2023). The last ten years have seen a big change, thanks to the widespread use of artificial intelligence (AI), bringing in a new era of effective risk management (Aziz and Dowling, 2019).

AI is at the heart of this transformation, serving as a crucial tool for proactive risk management in supply chains. AI allows us to look at risks comprehensively, identify patterns, and make informed decisions to handle potential disruptions (Helo and Hao, 2022). Beyond just managing risks, AI improves the reliability of supply chains by giving us real-time data insights, making decision-making more effective and operations more efficient. Additionally, AI encourages collaboration and information exchange, improving communication and coordination among different players in the supply chain (Helo and Hao, 2022) .

With the same goal of predicting risk in the supply chain, but using AI, (Sarbas et al., 2023) historical supply chain data from a publicly available data repository was utilized to train machine learning algorithms. The considered models are logistic regression algorithm, random forest classifier algorithm, and Gaussian Naïve Bayes algorithm. Subsequently, the trained models underwent validation through k-fold cross-validation. The paper includes a comparative analysis using performance metrics on the test data such as receiver operator characteristics (ROC), precision, recall, and F1-score to identify the most effective predictive model for the delivery risk prediction problem. Generally, the random forest model shows the highest performance across multiple metrics. With the focus on the eCommerce sector, Lolla et al. (2023) tried to forecast late delivery risks through the examination of historical data employing machine learning methodologies. They evaluated various algorithms, namely Logistic Regression, XGBoost, Light GBM, and Random Forest and suggest that the hybrid approach that merges all of these models surpasses other ensemble and individual algorithms in terms of accuracy, specificity, precision, and F1-score. The objective of a paper written by Zaghdoudi et al. (2022) is to address the high number of supplier delays faced by an industrial furniture manufacturing company due to the Covid-19 pandemic. The evaluation involves testing three machine learning models: logistic regression, random forest, and Gaussian Naïve Bayes. The Random Forest model emerges as the most effective in avoiding false delivery advance alerts. Overfitting is addressed through a variable selection study for the "decision tree" model.

In another study, Zheng et al. (2023) proposed an approach based on federated learning for collective risk prediction in supply chains. They tried to address the challenge of inadequate datasets and privacy concerns by enabling organizations to tap into collective knowledge without exposing their data. The study focuses on buyers predicting order delays from shared suppliers before and after Covid-19, demonstrating the effectiveness of federated learning, particularly for buyers with limited datasets, while highlighting the impact of training data-imbalance, disruptions, and algorithm choice. The compared models were LogReg, ANN, MLP, and CNN1D, with the last one being the outperforming model in the scheme of federated learning.

Another study deals with the specific case of imbalanced class problems in predictive models, where the occurrence of delivery risks is infrequent, compared to non-risk orders (Thomas and Panicker, 2023). The study compared four models, namely K-Nearest Neighbour, Random Forest, Logistic Regression, and Support Vector Machine, with three oversampling methods: random oversampling, Synthetic Minority Over-sampling Technique (SMOTE), and SMOTE

Tomek. The results reveal that the Random Forest model, combined with SMOTE and Tomek link, achieves superior performance.

This paper proposes using deep learning approaches to specifically deal with the challenges of predicting and managing the risk of late delivery in supply chains. By using historical deliveries data from manufacturing execution systems, this paper aims to enhance supply chain resilience by offering helpful advice for managers in different supply chains phases, contributing to the ongoing improvement of risk management strategies. By achieving a high level of accuracy in risk prediction, we empower decision-makers and stakeholders within the chain to proactively address potential challenges and make informed decisions, thereby enhancing the overall efficiency and resilience of the system. So, the scientific purpose of the presented study is to pioneer innovative methodologies for the accurate prediction and management of late delivery risks within global supply chains. This research establishes a dynamic predictive modeling framework integrating clustering techniques and advanced classification models, specifically the Generative Neural Network and a hybrid architecture combining CNN with LSTM. The strategic integration of these models through ensemble learning techniques—bagging, boosting, and stacking—aims to synergize their unique capabilities, creating a comprehensive classification framework. The study not only focuses on the theoretical development of these methodologies but also provides a meticulous exploration of their practical implementation and evaluation.

This paper is organized as follows: Section 2 outlines a methodology tailored for handling late delivery risk in the supply chain, demonstrating its advantages. In Section 3, the results of the real-world applicability of these methods are explored through experimentation on actual data. Section 4 provides meticulous details on the experimental setup and comprehensive comparative results, emphasizing the relevance of the proposed method to procurement challenges. Finally, the paper concludes with remarks and outlines potential perspectives for future research, particularly in the context of applying deep learning techniques to overcome late delivery risks in the supply chain.

MATERIALS AND METHODS

The study tackles research problems in late delivery risk prediction by developing an innovative methodology integrating clustering and advanced classification models. It explores the practical use of ensemble learning strategies to enhance predictive performance, mitigate overfitting, and achieve generalizability, thus addressing the complexity of late delivery risks in global supply chains. Technically, the research problems include refining the precision of classification methods, evaluating the real-world applicability of the proposed framework, and demonstrating the impact of ensemble learning on effective risk mitigation strategies within the intricate network of the supply chain.

To achieve these goals, this section offers an extensive overview of the methodology used in this study. Figure 1 shows the framework for predicting the risk level of late delivery, the framework integrates the unique platform clustering technics in order to classify the order regions and classification methodologies according to four level of risks. This integration offers a dynamic approach to predictive modeling, enabling not only the identification of data clusters but also the accurate prediction of late delivery.

Data description:

The dataset utilized in this study, known as the "Smart Supply Chain for Big Data Analysis" dataset, was provided by DataCo Global in 2019 (Constante, 2019). This dataset was intentionally designed to support the application of Machine Learning Algorithms and analysis. It covers various supply chain activities, including provisioning, production, sales, and commercial distribution. What sets this dataset apart is its ability to seamlessly incorporate both structured and unstructured data, enabling the correlation of different data types and the generation of valuable insights.

Moreover, this dataset includes a wide range of product categories, such as Clothing, Sports equipment, and Electronic Supplies. This diversity offers researchers and analysts the opportunity to explore insights from different dimensions of the supply chain. On closer examination, this work identified a total of 49 features within the dataset, encompassing information related to orders, shipping, and comprehensive sales data. It is a combination of textual and numerical data, including order locations and quantitative sales information, with 24 character columns and 28 numeric columns, making it a comprehensive dataset suitable for in-depth analysis.

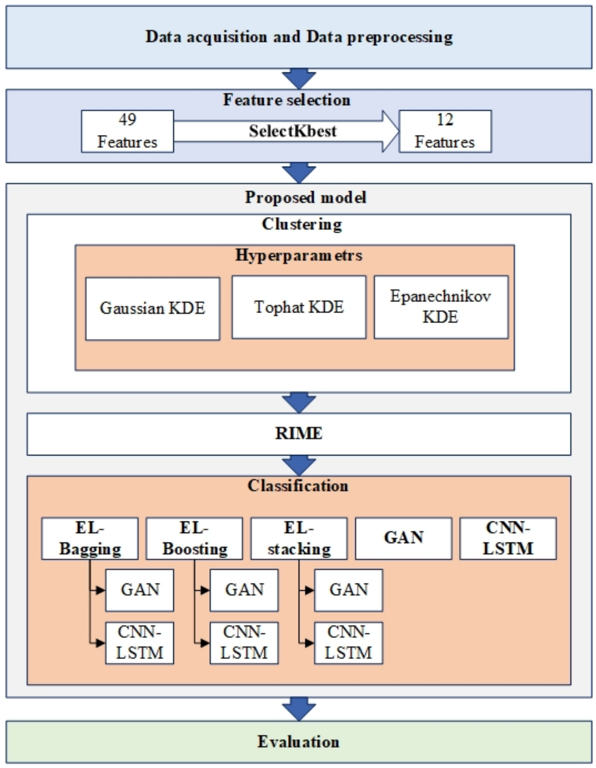

Fig. 1 : The framework used to predict the risk level of late deliveries

Data Processing:

Data pre-processing is an indispensable task in preparing raw data to ensure its suitability for the deep learning model proposed in this research (Perez et al., 2021) (Perez and Wang, 2017). Through a systematic and rigorous execution of the data preprocessing step, the dataset's full potential is unleashed. Consequently, the Min-Max scaling technique is employed to standardize the numeric attributes. By systematically and rigorously executing the step of data preprocessing, the full potential of the dataset used is unlocked. By undergoing this transformation, Min-Max normalization guarantees that all data points are proportionally adjusted to fit within the specified range (Moon et al., 2014). As a result, comparisons and analyses become more meaningful when dealing with data that originally exhibited varying scales and units. By implementing the Min-Max normalization method, data harmonization is achieved, enabling the proposed machine learning model to operate efficiently. This, in turn, enhances its predictive accuracy and analytical efficiency. Min-Max normalization is carried out using the following formula (Henderi, 2021)(Eq 1):

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

Here's a breakdown of the variables:
x_{scaled} represents the newly scaled or normalized value.
x denotes the original data point.
$x_{\min}$ stands for the minimum value observed in the dataset.
$x_{\max}$ corresponds to the maximum value present in the dataset.

Feature Selection:

Feature selection is not crucial in enhancing the performance and reducing the computational complexity of deep learning algorithms. The main goal of feature selection is to reduce the dimensionality of the data, thereby enhancing the efficiency of the model. In this particular approach, the K-best method is utilized as a simple yet effective technique for feature selection (Geng et al., 2007). This method involves the organized assessment of scores to each characteristic, and the selection of the top k characteristics based on these scores (Sánchez-Maroño et al., 2007). The scoring process consists of a variety of statistical tests, such as chi-squared for categorical features, F-scores for regression, and mutual information scores for both classification and regression. Typically, in this investigation, a carefully curated set of the top 12 features is employed. This selection not only diminishes the dimensionality of the data, but also preserves crucial attributes, thereby facilitating a more efficient and insightful analysis.

Feature Clustering:

The customer's location significantly influences delivery time, impacted by factors like supplier proximity, local infrastructure, seasonal aspects, and shipment mode. Employing clustering techniques on the 'Order Region' feature helps group data, revealing hidden patterns. Categorizing products based on 'Order Region' segments allows an insight into patterns affecting delivery times.

This detailed classification facilitates targeted analysis of potential delays within specific regions or categories, enabling proactive measures to mitigate late delivery risks. Clustering within the 'Order Region' enhances the ability to discern subtle factors impacting delivery timelines, contributing to a more accurate risk prediction strategy in supply chain management.

Selecting the right number of clusters is vital, leading to a hyperparameter tuning step to identify the optimal number. Three algorithms—Gaussian KDE, Tophat KDE, and Epanechnikov KDE—are used for this purpose.

Gaussian Kernel Density Estimation (Gaussian KDE) (Węglarczyk, 2018) is the first technique that is employed for clustering processus in this work. It is a technique used in nonparametric kernel density estimation to estimate probability density functions. It is widely applicable in various fields such as real-life settings, scientific computing, graph algorithms, machine learning, and statistics. In this work the Gaussian Kernel Density Estimation method is used in clustering processes using the Gaussian kernel function $\hat{f}(x)$ to estimate the probability density function, represented as (Eq1). This approach involves assessing changes in center point scores and minimum distances, establishing a scoring mechanism for identifying peaks that indicate potential cluster centers. The mathematical representation of the scoring mechanism (Eq2) facilitates the determination of optimal cluster numbers. This data-driven method combines density estimation and distance metrics to inform the configuration of clusters in subsequent clustering algorithms (Niu et al., 2022).

$$\frac{1}{n\sqrt{2\pi}\sigma}\sum_{i=1}^{n}\exp\left(-\frac{(x-x_i)^2}{2\sigma^2}\right) \tag{2}$$

$$\text{Score }(k) = \text{Change in Center Point Score } + \text{ Change in Minimum Distance,}$$

Here,
$\hat{f}(x)$ is the estimated probability density function at point x,
n is the number of data points,
x_i represents individual data points,
σ is the bandwidth parameter, determining the width of the Gaussian kernel.

The Tophat Kernel Density Estimation (Tophat KDE; Węglarczyk, 2018) is the second technique that is employed for clustering process in this work. This technique involves estimating the probability density function of a dataset using a Tophat kernel function. In contrast to the standard Gaussian kernel, the Tophat kernel assigns equal weight to all data points within a specified range, making it less sensitive to outliers (Niu et al., 2022). The Tophat KDE equation is given by (Eq3):

$$\hat{f}(x) = \frac{1}{n \cdot h} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (3)$$

Where
$\hat{f}(x)$ is the estimated density at point x, n is the number of data points,
x_i represents individual data points, h is the bandwidth parameter, and
K is the Tophat kernel function.

The Tophat KDE provides a robust approach for density estimation in clustering scenarios. It is used, particularly in this work, to deal with datasets containing outliers or irregularities.

The Epanechnikov Kernel Density Estimation (Epanechnikov KDE; Węglarczyk, 2018) is the third technique that is employed for clustering process in this work. This technique utilizes the Epanechnikov kernel function, a symmetric, bell-shaped kernel known for its robustness and sensitivity to local features (Niu et al., 2022). The Epanechnikov KDE equation is defined as (Eq 4):

$$\hat{f}(x) = \frac{3}{4h}\left(1 - \left(\frac{x - x_i}{h}\right)^2\right) \quad (4)$$

In this equation,
$\hat{f}(x)$ represents the estimated density at point x, h is the bandwidth parameter,
and x_i denotes individual data points.

The distinctive feature of the Epanechnikov kernel is its ability to assign zero weight to data points outside a specified range, making it particularly robust and less sensitive to outliers. This characteristic makes Epanechnikov KDE well-suited for the clustering process in this work, where outlier resilience is crucial for accurate density estimation and cluster identification.

A Graphical Exploration of Optimal Number of Clusters

Determining the most advantageous value of 'k' is of the utmost importance in order to attain effective and efficient outcomes in the subsequent stage of predicting late delivery. Therefore, the graphical analysis of the Kernel Density Estimation methods proposes the four clusters, as depicted in Figure 2.

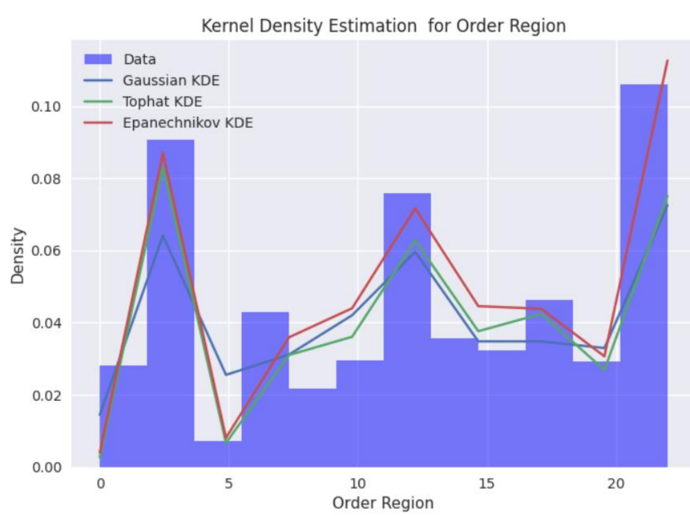

Fig. 2 :A graphical exploration of optimal number of clusters.

Clustering using RIME Optimization Algorithm

The clustering phase in this study employs the Rime Optimization Algorithm (Su et al., 2023), inspired by the intricate growth patterns of natural rime-ice. This algorithm integrates unique exploration, exploitation, and selection strategies, mimicking the formation of rime-ice influenced by temperature and environmental factors (Su et al., 2023).

Drawing from the subtle movement of soft-rime particles, a step-by-step exploration and exploitation technique is devised, establishing a soft-rime search strategy as the core method for optimization (Su et al., 2023). Inspired by the interaction of hard-rime agents, a puncture mechanism for hard-rime is proposed, fostering dimensional crossover interchange between ordinary and optimal agents to enhance solution precision (Su et al., 2023).

Building on the concept of greedy selection, an improved positive greedy selection approach is introduced to augment population diversity and minimize the risk of the algorithm getting trapped in local optima. This modification involves adjusting the selection of optimal solutions. The overall algorithm structure is elucidated in Algorithm 1, presenting the pseudo-code and flow chart for clarity and implementation coherence (Su et al., 2023).

Algorithm 1: Pseudo-code of RIME (Su et al., 2023)

Initialize the rime population R
Get the current optimal agent and optimal fitness
While $t \leq T$
  Coefficient of adherence $E = (t/T)^0.5$
If $r_2 < E$
  Update rime agent location by ***the soft-rime search strategy***
End If
If $r_3 < Normalizefitnessof(S_i)$
  Cross updating between agents **by *the hard-rime puncture mechanism***
End If
If $F(R_i^{new}) < F(R_i)$
  Select the optimal solution and replace the suboptimal solution using ***the positive greedy selection mechanism***
End If
  t=t+1
End While
`

Where
R The rime-population
S_i The rime-agents
x_{ij} The rime-particles
d The dimension of the population
i The ordinal number of rime-agent
j The ordinal number of rime-particle
$F(S_i)$: The fitness value of the agent
$F^{normr}(S_i)$: The normalized value of the current agent fitness
Rnew: The new position of the updated particle
$R_{best;j}$ The best rime agent in the rime-population
r_1; r_2; r_3 A random number

COSθ : The direction of particle
t: The current number of iterations
T : The maximum number of iterations
b : The environmental factor
*w :*The number of segments of the step function
h : The degree of adhesion
Ub_{ij}; Lb_{ij} *:* The upper and lower bounds of the escape space
E : The coefficient of being
FEs: The current number of evaluations

Cluster Validity Index

The Cluster Validity Index assesses Compactness in clustering, quantifying how tightly data points are packed within each cluster (Rendon et al., 2023). Mathematically expressed as Compactness, this metric evaluates the cohesion of clusters, crucial for gauging the clustering algorithm's effectiveness. It computes squared distances between data points and cluster centroids, providing a quantitative measure of intra-cluster tightness. This concise approach ensures a clear understanding of spatial distribution and cohesion, guiding the evaluation of clustering performance (Rammal et al., 2015). Mathematically, it is expressed as Compactness (Eq5). This function considers both intra-cluster similarity and inter-cluster dissimilarity (Rendon et al., 2023).

$$\text{Compactness} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\sum_{i=1}^{n_k} d(x_i, c_k)^2 \quad (5)$$

Where
K represents the number of clusters,
n_k is the number of data points in cluster k, x_i denotes individual data points,
c_k is the centroid of cluster k, and $d(x_i, c_k)$ is the distance between data point x_i and centroid c_k.

The goal of using this function in this work is to minimize this value, creating clusters that are internally cohesive and well-separated from each other, ensuring meaningful and distinct cluster formations in the clustering process.

The proposed classification model:

In the continuously evolving field of data science, the search for precision in classification methods has encouraged the examination of cutting-edge architectures and ensemble techniques. This research seamlessly integrates two distinct models: Model I, which is the Generative Neural Network, and Model II, which is a hybrid architecture that combines Convolutional Neural Networks (CNN) with Long Short-Term Memory networks (LSTM). The objective is to identify clusters within the intricate "order region." The combination of these models is arranged through the strong ensemble learning framework, which includes advanced techniques such as bagging, boosting, and stacking.

In the specific context of this research, the ensemble approach intricately combines the generative capabilities of Model I with the spatial-temporal expertise demonstrated by Model II. Ensemble learning has proven to be a powerful technique in machine learning, enabling the combination of diverse models to enhance predictive performance. In this study, we employed ensemble learning to synergistically integrate two distinct models: a Generative Neural Network (Model I) and a Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) hybrid (Model II). The integration was executed through three ensemble strategies: Bagging (utilizing Random Forest), Boosting (employing Train AdaBoost), and Stacking (leveraging Logistic Regression). This calculated fusion is achieved through the precise implementation of collaborative learning strategies, namely bagging, boosting, and stacking.

The ultimate goal is to develop a comprehensive classification framework that is adept at deciphering the complex clusters within the discerning domain of the "order region." This collaborative framework is dedicated to predicting the risks associated with late deliveries within the intricate network of the supply chain, along with an exhaustive exploration of tailored strategies for efficient risk management and mitigation.

Model I, the Generative Neural Network, serves as the initial foundation of our ensemble approach. This model possesses the unique ability to generate data that closely resembles the training set, thus providing a novel perspective on the classification of the "order region." By leveraging generative adversarial networks or variational autoencoders, Model I captures the intricate patterns within the order region, thereby facilitating a nuanced understanding of the underlying data distribution.

Complementing the generative capabilities of Model I, Model II adopts a hybrid architecture that combines the strengths of Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM). This design is particularly adept at handling both spatial and sequential dependencies, making it highly suitable for the multifaceted nature of order region data. In other words, this proposed architecture excels in learning hierarchical features and capturing temporal dependencies, making it a valuable component in our ensemble framework.

The CNN-LSTM hybrid not only extracts spatial features but also captures temporal patterns, ensuring a comprehensive representation of the clusters within the "order region."

The synergy between Model I and Model II is harnessed through ensemble learning, a strategic amalgamation of diverse models aimed at enhancing classification performance. This ensemble methodology encompasses three key techniques: bagging, boosting, and stacking (see figure 03, figure 04, and figure 05, respectively).

Bagging, short for Bootstrap Aggregating, involves training multiple instances of each model on distinct subsets of the training data independently, and then aggregating their predictions. The combination is performed through averaging or voting, which harnesses the collective wisdom of the ensemble to enhance overall model robustness and generalization performance. Therefore, this approach has the ability to mitigate the issue of overfitting and increase the stability of the classification. To implement the Bagging strategy, we employed the Random Forest algorithm. In our implementation, the combined predictions from Model I and Model II are used to train a RandomForest classifier, offering an ensemble approach that leverages both the predictive power of deep neural networks and the robustness of a RandomForest model for improved overall performance.

Boosting, on the other hand, takes advantage of sequential training of Model I and Model II. Each model focuses on correcting the misclassifications made by its predecessor, progressively refining the overall accuracy of the ensemble. The idea is to combine the predictions of weak learners, which are our base models, in an iterative process. It focuses on correcting errors made by previous learners by assigning higher weights to misclassified instances. The final model is a weighted combination of all weak learners, resulting in a strong, more accurate model. AdaBoost is a popular boosting algorithm that adapts over time and is effective in improving performance, especially in challenging situations, while reducing overfitting.

Finally, stacking is employed to combine the predictions of both models. A meta-model is trained to weigh the contributions of Model I and Model II, leveraging their respective strengths to produce a cohesive and robust classification outcome. In the implementation level, the stacking ensemble technique is designed by training a meta-model (Logistic Regression) on the predictions of the two base models. The base model predictions for both training and testing datasets are combined, and a logistic regression meta-model is trained on this combined prediction matrix. Finally, the performance of the stacked model is evaluated using accuracy on the testing dataset.

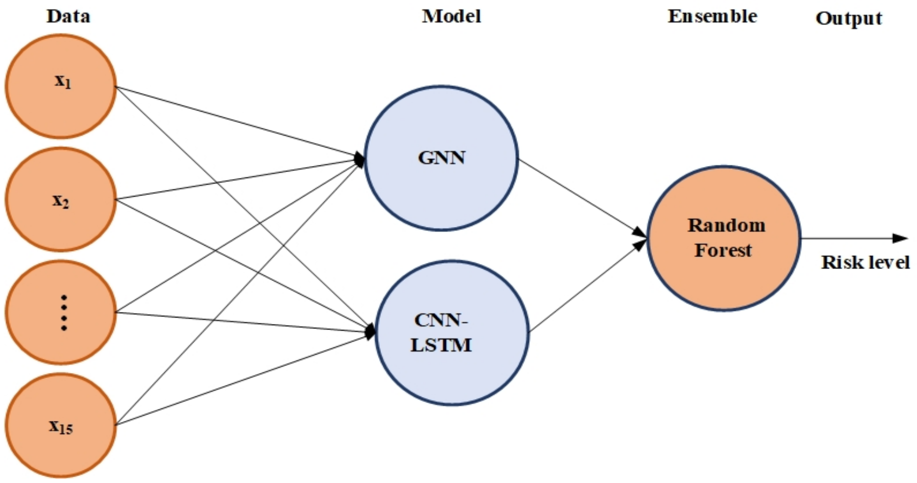

Fig. 3 : The general framework of the ensemble learning based on bagging.

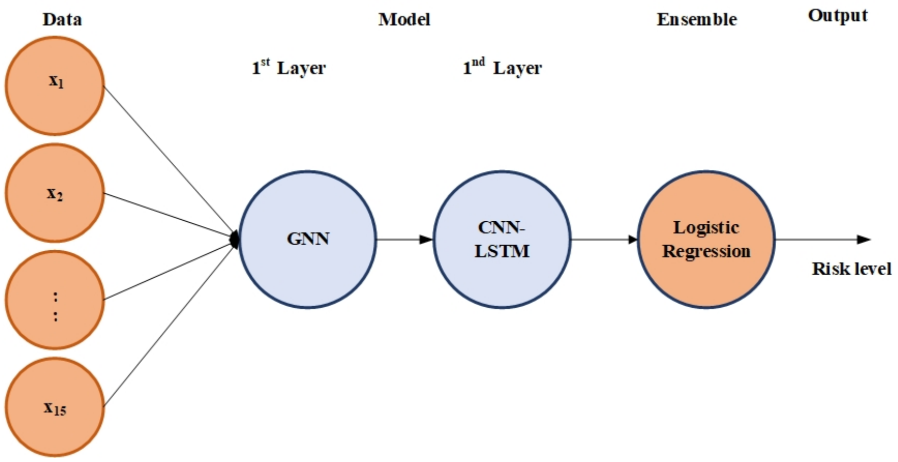

Fig. 4 : The general framework of the ensemble learning based on stacking.

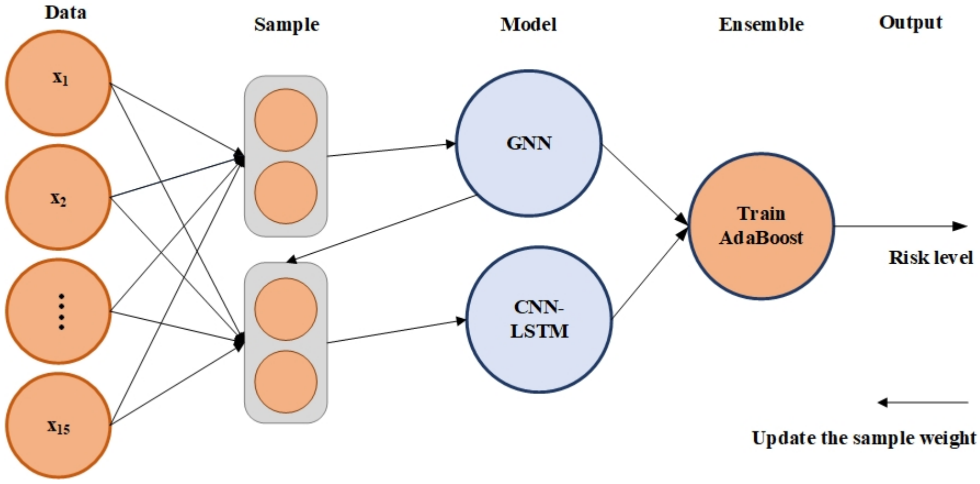

Fig. 5 : The general framework of the ensemble learning based on boosting.

RESULTS

The results included in Table 1 offer a detailed analysis of the performance of different models, each assessed across various metrics. Among the ensemble learning (EL) approaches, EL-stacking emerges as the top performer in terms of accuracy, achieving a notable 0.926. This result suggests that the stacking ensemble method excels in making accurate predictions of the target variable.

The precision score is another significant performance metric that indicates the proportion of actual positive observations among all instances predicted as positive. In terms of this metric, both EL-bagging and EL-boosting exhibit high values at 0.946, indicating a strong ability to correctly identify positive instances, i.e. existence of risk and precisely identifying its level. EL-stacking closely follows, with a precision of 0.9186. This precision metric underscores the ensemble methods' proficiency in accurately predicting positive outcomes. In our specific multi-classification scheme, this means making accurate predictions for each class, ensuring that the model correctly assigns the input data to the appropriate class labels, i.e., risk level. Furthermore, the F1-score, derived from both precision and recall scores, serves as an effective measure for assessing the overall accuracy of a classification model. Examining the F1-score, EL-stacking leads the ensemble models with a score of 0.9175, highlighting a balanced performance between precision and recall. EL-boosting and EL-bagging also demonstrate strong F1-scores, reinforcing their effectiveness in classification tasks.

For regression-oriented metrics, EL-stacking and EL-bagging showcase the lowest Mean Squared Error (MSE) and Mean Absolute Error (MAE) at 0.11 and 0.09, respectively. This signifies their superior performance in minimizing prediction errors for regression tasks. While MSE and MAE are typically associated with regression tasks, we have repurposed them for this multi-classification context to assess the degree of deviation between predicted risk levels and actual class labels, providing insight into the magnitude of misclassifications. MSE penalizes larger errors more heavily due to the squaring of differences, which means that high values of MSE indicate that the model misclassifies very high-level risk with very low-level risk. This helps identify cases where the predicted risk significantly deviates from the actual risk, regardless of the specific classes. In the other hand, MAE measures the average absolute difference between predicted and actual values, and it is less sensitive to outliers than MSE.

The Receiver Operating Characteristic Area Under the Curve (ROC-AUC) score provides an interpretation as the probability of a classifier correctly ranking a randomly chosen positive observation higher than a randomly chosen negative one. This metric is calculated by determining the area under the receiver operator characteristic curve. The AUC scores of all models fall in a good range, indicating an excellent accuracy. Notably, EL-Stacking demonstrates a perfect discriminatory power with a value equal to 1. These results indicate excellent performance in distinguishing between positive and negative instances in binary classification mode. Analyzing predictive capability, EL-stacking, and EL-boosting exhibit high R-squared values (0.8728 and 0.857), indicating strong explanatory power. Similarly, both models demonstrate high explained variance, emphasizing their ability to account for a substantial portion of the target variable's variance.

In terms of computational efficiency, EL-boosting stands out with the lowest runtime at 7.24, suggesting efficient model training. EL-stacking follows closely with a runtime of 9.01, while EL-bagging and the CNN-LSTM model have longer runtimes.

Table 1 : Performance Metrics of Late Delivery Prediction Framework

	EL-bagging	EL-stacking	EL-boosting	CNN-LSTM	GAN
Accuracy	0.906	0.926	0.906	0.915	0.908
Precision	0.946	0.9186	0.946	0.916	0.892
F1-Score	0.915	0.9175	0.9154	0.916	0.886
MAE	0.1	0.09	0.104	0.09	0.1
MSE	0.125	0.11	0.125	0.09	0.12
AUC	0.98	1	0.98	0.98	0.98
R-squared	0.857	08728	0857	0.87	0.866
Explained Variance	0.86	0.8729	0.8609	0.87	0.864
Run Time	11.27	9.01	7.24	5.18	2.71

DISCUSSION

Our research introduces a novel framework methodology for predicting the risk of late delivery in supply chains, incorporating advanced clustering techniques and deep learning models. In the domain of supply chain risk classification, the evaluation of model performance is critical for effective risk management and resilience. Among the ensemble learning models, EL-stacking stands out as a comprehensive performer, excelling in accuracy (0.926), precision (0.9186), F1-score (0.9175), and discriminatory power (1). The accurate classification of supply chain risks is fundamental to building resilience. So, for a broad comparison and considering the fact of using different datasets and models, we analyse our finding in relation to previously published papers.

(Sarbas et al., 2013) trained three models, namely logistic regression algorithm, random forest classifier algorithm, and Gaussian Naïve Bayes. The random forest model exhibited the highest AUC score (0.929). Assessing sensitivity, the recall score was highest for the random forest model (0.833), followed by logistic regression (0.762) and Naïve Bayes (0.577). The random forest model outperformed others with an F1-score of 0.859, compared to logistic regression (0.773) and Naïve Bayes (0.685). Their results collectively imply the strong performance of the random forest model across various evaluation metrics. However, generally for big datasets, deep learning that we have used outperforms classical machine learning used in this paper. Another paper, dealing particularly with the challenge of data imbalance, employs the Area Under the Curve (AUC) score as the performance metric for risk prediction (Thomas and Panicker, 2023). The comparative analysis is conducted on K-Nearest Neighbour, Random Forest, Logistic Regression, and Support Vector Machine, while employing three oversampling methods: random oversampling, SMOTE, and SMOTE Tomek. The results showed that the Random Forest model, combined with Synthetic Minority Over-sampling Technique (SMOTE) and Tomek link demonstrates superior performance with an AUC score of 0.80. Random forest also demonstrates superior performance in predicting delivery delays in another paper when compared to decision tree and naïve bayes (Zaghdoudi et al., 2022). It outperforms other machine learning models with an accuracy of 76.02%, precision of 76.43%, and F1 score of 77.96%.

EL-stacking's ability to balance precision, recall, and overall accuracy positions it is a promising model for supporting resilient supply chain decision-making. Resilience is not only about preventing risks but also effectively responding to and recovering from disruptions. The ensemble learning approach, by integrating diverse models, showcases adaptability and robustness in capturing the complexity of supply chain risk dynamics.

EL-boosting demonstrates efficiency in runtime and strong discriminatory power as measured by the AUC. While EL-bagging and

CNN-LSTM deliver respectable performance, their accuracy and precision slightly trail behind EL-stacking.

The choice of the optimal model depends on specific task requirements, necessitating a balance between accuracy, interpretability, and computational efficiency. Our findings underscore the potential of the proposed framework in enhancing risk prediction accuracy through feature engineering, exploring new variables, and considering additional data dimensions. Addressing the interpretability challenge inherent in deep learning models is crucial, and future research should focus on developing methods to clarify decisions for practitioners and stakeholders. Furthermore, integrating uncertainty analysis in predicting supply chain risk late delivery represents a promising avenue for enhancing the robustness of our methodology.

In conclusion, our research not only introduces an innovative approach to predicting supply chain risk but also provides valuable insights into the performance trade-offs among different ensemble models. The proposed framework lays the foundation for future advancements in risk prediction accuracy, interpretability, and uncertainty analysis within the landscape of supply chain management.

CONCLUSION

In this research, the objective was to mitigate the resilience challenge in the global supply chain by introducing a novel framework designed to address the complexity of predicting late delivery risks in supply chain networks. This was achieved through the application of an innovative deep learning model. Notably, the model operates as a black box, making it challenging to discern the key features influencing delivery risk. To overcome these limitations, further research should prioritize the interpretation of deep learning models using explainable artificial intelligence. It is essential to evaluate the applicability and effectiveness of the proposed ensemble learning techniques (bagging, boosting, stacking) across diverse supply chain environments. The model's performance should be tested in various environments and prediction problems within the realm of supply chain management.

The proposed framework integrates clustering and multiclassification methodologies, incorporating hyperparameter tuning and a novel metaheuristic, RIME, in the clustering phase. The multiclassification phase harnesses the power of five deep learning models—GAN, CNN-LSTM, ensemble learning through bagging, stacking, and boosting—to enhance the precision of late delivery predictions.

In conclusion, this research not only addresses the challenge of late delivery risks in global supply chains but also lays the foundation for a transformative approach to proactive risk management. The integration of advanced clustering techniques, metaheuristic optimization, and a suite of deep learning models signifies a pioneering step towards building resilience in supply chain networks. As the landscape of supply chain management continues to evolve, our framework stands as a testament to the potential of innovative methodologies in navigating the complexities of modern supply chains and fortifying them against disruptions.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Aziz, S., Dowling, M., 2019. Machine Learning and AI for Risk Management, in: Lynn, T., Mooney, J.G., Rosati, P., Cummins, M. (Eds.), Disrupting Finance, Palgrave Studies in Digital Business & Enabling Technologies. Springer International Publishing, Cham, pp. 33–50. https://doi.org/10.1007/978-3-030-02330-0_3

Constante, F., 2019. DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS. https://doi.org/10.17632/8GX2FVG2K6.5

Douaioui, K., Fri, M., Mabrouki, C., Semma, E.A., 2021. A multiobjective integrated procurement, production, and distribution problem of supply chain network under fuzziness. IFAC-PapersOnLine 54, 1104–1111. https://doi.org/10.1016/j.ifacol.2021.08.130

Fri, M., Douaioui, K., Mabrouki, C., El Alami, S., 2021. Reducing Inconsistency in Performance Analysis for Container Terminals. International Journal of Supply and Operations Management 8, 328–346. https://doi.org/10.22034/ijsom.2021.3.6

Fri, M., Fedouaki, F., Douaioui, K., Mabrouki, C., Semma, E.A., 2019. Supply chain performance evaluation models, state-of-the-art and future directions. Int. J. Eng. Adv. Technol 9, 6336–6347.

Geng, X., Liu, T.-Y., Qin, T., Li, H., 2007. Feature selection for ranking, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Presented at the SIGIR07: The 30th Annual International SIGIR Conference, ACM, Amsterdam The Netherlands, pp. 407–414. https://doi.org/10.1145/1277741.1277811

Helo, P., Hao, Y., 2022. Artificial intelligence in operations management and supply chain management: an exploratory case study. Production Planning & Control 33, 1573–1590. https://doi.org/10.1080/09537287.2021.1882690

Henderi, H., 2021. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. IJIIS : Int. J. Inform. Inform. Systems. 4, 13–20. https://doi.org/10.47738/ijiis.v4i1.73

Lolla, R., Harper, M., Lunn, J., Mustafina, J., Assi, J., Loy, C. K., & Al-Jumeily OBE, D. (2023). Machine Learning Techniques for Predicting Risks of Late Delivery. Lecture Notes on Data Engineering and Communications Technologies, 343–356. https://doi.org/10.1007/978-981-99-0741-0_25

Mahmud, S., 2023. Modelling and Optimizing Supply Chain Integrated Production Scheduling Problems (PhD Thesis). UNSW Sydney.

Menchaca-Méndez, A., Montero, E., Flores-Garrido, M., Miguel-Antonio, L., 2022. An algorithm to compute time-balanced clusters for the delivery logistics problem. Engineering Applications of Artificial Intelligence 111, 104795. https://doi.org/10.1016/j.engappai.2022.104795

Moon, I., Shin, E., Sarkar, B., 2014. Min-max distribution free continuous-review model with a service level constraint and variable lead time. Appl. Math. Comput. 229, 310–315. https://doi.org/10.1016/j.amc.2013.12.034

Ngniatedema, T., Chen, L., Guiffrida, A.L., 2016. A modelling framework for improving supply chain delivery performance. IJBPSCM 8, 79. https://doi.org/10.1504/IJBPSCM.2016.077163

Niu, D., Sun, L., Yu, M., Wang, K., 2022. Point and interval forecasting of ultra-short-term wind power based on a data-driven method and hybrid deep learning model. Energy 254, 124384. https://doi.org/10.1016/j.energy.2022.124384

Patel, P.S., Pandya, D.M., Shah, M., 2023. A holistic review on the assessment of groundwater quality using multivariate statistical techniques. Environ Sci Pollut Res 30, 85046–85070. https://doi.org/10.1007/s11356-023-27605-x

Perez, H.D., Hubbs, C.D., Li, C., Grossmann, I.E., 2021. Algorithmic approaches to inventory management optimization. Process. 9, 1–17. https://doi.org/10.3390/pr9010102

Perez, L., Wang, J., 2017. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. https://doi.org/10.48550/ARXIV.1712.04621

Rammal, A., Perrin, E., Vrabie, V., Bertrand, I., Chabbert, B., 2015. Weighted-covariance factor fuzzy c-means clustering, in: 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE). Presented at the 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), IEEE, Beirut, Lebanon, pp. 144–149. https://doi.org/10.1109/TAEECE.2015.7113616

Rendon, N., Giraldo, J.H., Bouwmans, T., Rodríguez-Buritica, S., Ramirez, E., Isaza, C., 2023. Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning. Engineering Applications of Artificial Intelligence 124, 106635. https://doi.org/10.1016/j.engappai.2023.106635

Sánchez-Maroño, N., Alonso-Betanzos, A., Tombilla-Sanromán, M., 2007. Filter Methods for Feature Selection – A Comparative Study, in: Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (Eds.), Intelligent Data Engineering and Automated Learning - IDEAL 2007, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 178–187. https://doi.org/10.1007/978-3-540-77226-2_19

Sarbas, P. et al. 2023. Development of Predictive Models for Order Delivery Risk in a Supply Chain: A Machine Learning Approach. In: Li, X., Rashidi, M.M., Lather, R.S., Raman, R. (eds) Emerging Trends in Mechanical and Industrial Engineering. Lecture Notes in Mechanical Engineering. Springer, Singapore. https://doi.org/10.1007/978-981-19-6945-4_43

Su, H., Zhao, D., Heidari, A.A., Liu, L., Zhang, X., Mafarja, M., Chen, H., 2023. RIME: A physics-based optimization. Neurocomputing 532, 183–214. https://doi.org/10.1016/j.neucom.2023.02.010

Thomas, A., & Panicker, V. V. 2023. Supply Chain Data Analytics for Predicting Delivery Risks Using Machine Learning. Asset Analytics, 159–168. https://doi.org/10.1007/978-981-99-1019-9_16

Węglarczyk, S., 2018. Kernel density estimation and its application. ITM Web Conf. 23, 00037. https://doi.org/10.1051/itmconf/20182300037

Zaghdoudi, M. A., Hajri-Gabouj, S., Varnier, C., Zerhouni, N., & Ghezail, F. 2022. Predictive Analysis Methodology for Industrial Systems: Application in Supplier Delays Prediction. 2022 IEEE Information Technologies & Smart Industrial Systems (ITSIS). https://doi.org/10.1109/itsis56166.2022.10118391

Zheng, G., Kong, L., & Brintrup, A. 2023. Federated machine learning for privacy preserving, collective supply chain risk prediction. International Journal of Production Research, 61(23), 8115–8132. https://doi.org/10.1080/00207543.2022.2164628

Kaoutar Douaioui
Faculty of Science and Technology,
Hassan First University of Settat, Route de Casablanca
e-mail: kaoutar.douaioui@gmail.com

Rachid Oucheikh
Physical Geography and Ecosystem Science,
Lund University, Sweden
e-mail: rachid.oucheikh@ju.se

Charif Mabrouki
Faculty of Science and Technology,
Hassan First University of Settat, Route de Casablanca
e-mail: charif.mabrouki@uhp.ac.ma