DOI: https://doi.org/10.15276/aait.09.2026.28
UDC 004.9: 005.93

Assessment of material supply risks in make-to-order manufacturing based on the TabPFN neural network model

Andrew L. Mrykhin[1)]
ORCID: https://orcid.org/0009-0009-1545-632X; amrykhin@gmail.com
Svitlana G. Antoshchuk[1)]
ORCID: https://orcid.org/0000-0002-9346-145X; asg@op.edu.ua. Scopus Author ID: 8393582500
[1)] Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

Relevance: Growing global instability and the vulnerability of supply chains lead to increasing risks in material procurement for industrial enterprises. The problem of forecasting and assessing material supply risks is particularly relevant in make-to-order manufacturing and high-mix, low-volume production. Such manufacturing systems, characterized by product diversity and low buffer inventories, are highly sensitive to delays in material delivery. Traditional risk management systems, designed for stable mass production, are often unable to react quickly enough to the dynamics of order-driven, high-mix environments. **The purpose** of this study is to improve risk assessment of production materials procurement through the use of machine learning for prediction of material delivery delays and development of an risk evaluation method on the base of such forecasts. To achieve this goal, **the following tasks** were solved: a model for forecasting delivery delays was developed and integrated supply risk index, based on probability of a delays and their impact introduced. The interpretability of the classifier model is also investigated, and an example of its integration into the resource planning system of the customer enterprise is provided. **Methods**: To obtain delay predictions, a multi-class classifier based on the state-of-the-art TabPFN model is used. The TabPFN model is based on a transformer neural network architecture adapted for tabular data analysis; a set of historical procurement data obtained from the customer's information systems was used for its training. **Results**: In the study, the TabPFN-based classifier model demonstrated a high degree of result reliability even when using default parameters, ordinal consistency of results, and a low tendency to underestimate supply delays, which indicates the obtained model's orientation toward risk minimization. The proposed risk assessment methodology allows for considering the impact of the forecasted supply delay on production depending on the current production program and warehouse inventories. **Novelty**: Combining multi-class delay classification based on business-defined thresholds with the state-of-the-art TabPFN model enables accurate delay forecasting even on limited datasets. Another novel aspect is the proposed approach to calculating a comprehensive risk index, which combines predicted delay probabilities with an evaluation of their impact on the production schedule. The combination of machine learning-enabled delay prediction and the proposed risk index allows for automated risk assessments on a near-real-time scale. **Conclusions**: The research demonstrates that the application of machine learning methods can enhance the accuracy and timeliness of supply risk assessment, enabling near-real-time monitoring and more informed decision-making regarding the provision of materials for production.

Keywords: risk assessment, forecasting, make-to-order manufacturing, high-mix manufacturing, machine learning, tabular data, multiclass classification, TabPFN

***For citation*: Mrykhin A. L., Antoshchuk S. G. "Assessment of material supply risks in make-to-order manufacturing based on the TabPFN neural network model". *Applied Aspects of Information Technology*. 2026; Vol.9 No.3: 415–430. DOI: https://doi.org/10.15276/aait.09.2026.28**

1. INTRODUCTION

Over the last decades, manufacturing supply chains have become increasingly complex and interdependent. Globalization, outsourcing, and product diversification have led to increased manufacturing efficiency, but at the same time have created new forms of vulnerability. Disruptions in logistics, material supply, or supplier performance now spread rapidly through deeply connected manufacturing networks, exposing enterprises to operational and financial risks [1].

Classic supply chain risk management systems, designed for stable high-volume production and focused on long-term strategic risks, rely on multi-stage processes and expert assessments [2]. Such approaches often fail to respond in time to the variability and high data intensity characteristic of Make-to-Order (MTO) and High-Mix, Low-Volume (HMLV) manufacturing [3]. In these environments, production schedules are strictly dependent on the timely delivery of materials, leaving minimal time buffers for delays and requiring faster, more adaptive risk monitoring mechanisms.

Achievements in digitalization at all stages of industrial production and the real-time availability of data from modern ERP systems pave the way for more dynamic risk assessment using automated and intelligent methods and technologies, and as a result–improved responsiveness and quality of management decisions.

In previous publications [4], the authors proposed an information model for the automated assessment of transient and limited-impact (tactical) material supply risks, and also preliminarily investigated the possibilities of its enhancement using machine learning methods [5].

In this article, the authors analyze the results of production deployment of the model proposed in [4] and propose an new model based on the advanced TabPFN [6] transformer neural network. The interpretation of the model's predictions from a risk assessment perspective and its integration with a manufacturing Enterprise's Resource Planning (ERP) system are considered. An analysis of the interpretability of the model results was performed using the SHAP [7] technique, and the relationship between the obtained results and the risk structure was analyzed.

2. PRIOR RESEARCH AND PUBLICATIONS REVIEW

The development of supply risk assessment and management issues originates in the foundational works of George Zsidisin [8], [10] and Christine Harland [9] from the 1990s and 2000s.

By 2009, Supply Chain Risk Management (SCRM) had emerged as a distinct discipline, formally recognized through its inclusion in the ISO 31000 standard [11]. In the following decade, interest in the topic of supply risks continued to grow, in a 2021 review, Amulya Gurtu and Jestin Johny identified 455 publications on this topic over the ten-year period from 2010 to 2019 [12]. Numerous approaches to risk assessment were proposed—both qualitative and quantitative, based on scores, ratings, weight coefficients, or probabilities [13], [14]. Various methodological approaches to supply chain risk management were also developed.

Despite the variety of proposed systems, certain common features can be identified:

– a multi-stage risk management process: identification, analysis, assessment, and response, often represented as a cycle similar to the Deming Cycle. This process can be quite labor-intensive and time-consuming, requiring the regular involvement of specialists and experts;

– adoption of conception of risk realization as a consequence of relatively rare, non-standard events (trigger events) that disrupt the normal course of business processes. Significant effort is directed toward identifying, classifying, and tracking these events, as well as determining their probabilities and potential consequences;

– the primary focus of SCRM is directed toward the impact of risks on the overall activities of the enterprise, financial results, or operational stability. Risk analysis procedures are often conducted only for strategic materials or the most critical suppliers.

While SCRM provides a well-developed framework for supply risk management, the models proposed within its scope are largely associated with managing strategic material supply risks and are aimed at the early identification of potential risks and the preparation of mitigation measures. Examples of such risks include the catastrophic loss of unique production capacities (e.g., due to fire), the bankruptcy of key suppliers, international sanctions, pandemics, and military actions.

At the same time, the limited flexibility of SCRM and its focus on multi-stage procedures create constraints in the case of Make-to-Order (MTO) and, especially, High-Mix, Low-Volume (HMLV) production. Such manufacturing environments are characterized by inherent instability and dynamism. Enterprises operating under the MTO + HMLV model function in conditions of high demand variability and, accordingly, maintain limited safety stocks of specialized components [3].

This variability is compounded by a complex supplier ecosystem, which involves reliance on a broad partner base that often includes niche suppliers of specific components. In a custom-order manufacturing environment, the enterprise's Master Production Schedule (MPS) is tightly linked to the supply schedule. Any delay in receiving materials due to an unforeseen supply risk can quickly lead to a production bottleneck.

The limited timeframe for effective risk response and the constant changes in the status of supply channels make rapid (ideally real-time) updates of supply risk assessments critical for maintaining the production schedule.

To address the aforementioned need for timely risk assessment, Mrykhin and Antoshchuk in [4] proposed a model for the automated assessment of small and moderate deviations in supply channels (tactical risks), which was implemented as a service within an ERP system and relied on data available in the customer's information systems.

The developed model automatically assesses supply risks for each production order using data from the ERP system. It calculates the risk of order fulfillment delay as the cumulative standard deviation of delivery times for all required materials. The risk for each individual material is determined based on deviations in segments of its supply chain, with data dynamically updated from logistics information. The model generates quantitative risk indicators, providing decision-making support in production planning and procurement in real time.

The model was put into production at an electrical engineering production facility and demonstrated the ability to generate useful risk forecasts. However, during the process of industrial use, a number of limitations also emerged, primarily centered around the following aspects:

– the model is based on the assumption that deviations follow a normal distribution;

– the model's exclusive focus on time deviations during transportation may lead to the loss of non-linear, categorical, or external factors;

– complex interactions between features cannot be modeled;

– a lack of data for individual segments undermines the stability of risk assessments.

Analysis showed that the actual distribution of delivery delays (Fig. 1) differs significantly from the normal distribution (Fig. 2). It features a sharp, high peak near zero and heavy tails relative to the center, forming a leptokurtic (peaked) rather than a normal curve. More importantly, the distribution has high positive skewness with a long, heavy right tail, reflecting probable lengthy delays in the supply of certain materials. Such prolonged procurement delays pose a critical risk to the production schedule and timely order fulfillment. Accurate assessment of these risks is vital for the effectiveness of our model.

Another significant challenge is data sparsity: for certain segments of the supply chain, only a very limited number of samples are available. Even after three years of data collection, over 35% of segments contain fewer than 10 logistics records. Such a high level of sparsity can introduce biases and undermines the stability of risk assessments for any route that includes these segments. While the problem can be partially mitigated through the application of regularization, the overall accuracy of the supply risk assessment still suffers.

Finally, we observed that delivery delays for certain supply routes exhibit significant variability. In over 16 % of routes, the range between the minimum and maximum delay exceeds 15 days, simultaneously forming a platykurtic distribution, which indicates that the variation is not caused by isolated outliers but reflects a statistically significant structural trend. Consequently, the standard deviation ceases to be an adequate measure of risk for such routes. We hypothesize that this variability is driven by factors not accounted for by the current model, and the differences in assessment stability between routes suggest that these factors have a complex, non-linear nature. In our study, we aim to identify such factors and adapt the model to incorporate them.

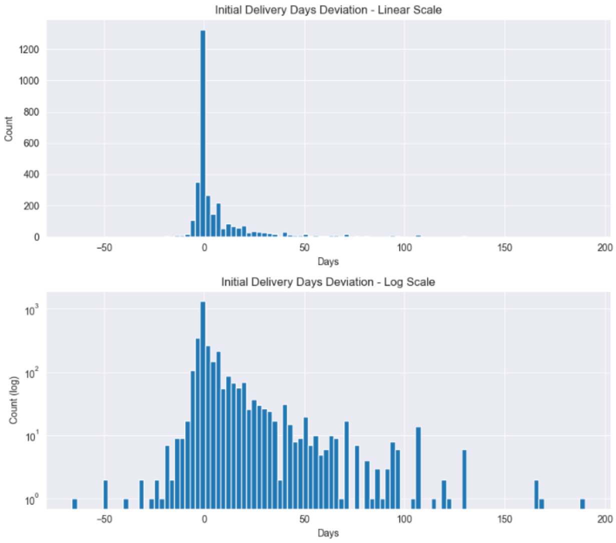

Fig 1. **Histogram of supply delay distribution, linear (top) and logarithmic (bottom) scales on the frequency axis**
Source: **compiled by the authors**

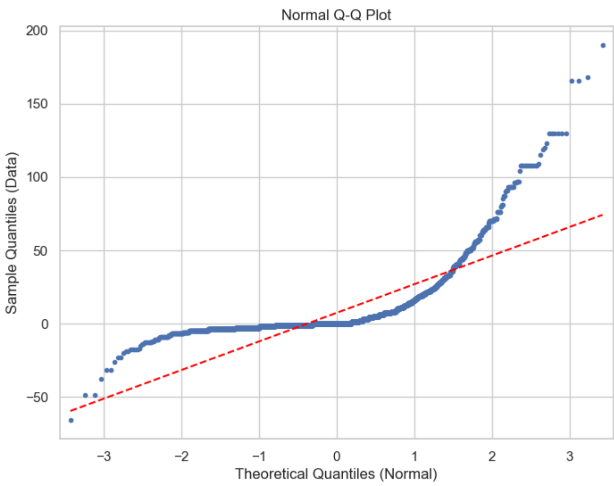

Fig 2. **Q-Q plot comparing the distribution of delays with a normal distribution**
Source: **compiled by the authors**

The aforementioned limitations indicate the need to search for more advanced methods of automated supply risk assessment.

In the review [16], Muller et al. demonstrated the effectiveness of applying machine learning models to predict material delivery delays. At the same time, it was pointed out that certain issues in existing research are insufficiently developed. In particular, most models use binary classification, without considering the varying impact of delays of different durations. Interpretability analysis is often absent, and insufficient attention is paid to the integration of models with existing ERP systems.

In addition, the predictive models described in [15], [16], [17], [18] utilize large datasets (from tens of thousands of samples) and prefer classical machine learning models.

In study [5], the use of modern machine learning methods for predicting supply delays based on historical records of production material procurement was examined. The research demonstrated that modern machine learning models are capable of providing good forecasting accuracy despite the limited number of samples and the dominance of categorical features and weak predictors in the dataset obtained from a real-world manufacturing enterprise ERP system. The authors compared the use of common gradient boosting models and a state-of-the-art transformer-based neural network model, TabPFN, for tabular data analysis, using the task of multiclass classification of supply delays as an example. Empirical confirmation was obtained showing that on small datasets, TabPFN is capable of outperforming existing models without requiring complex feature engineering or resource-intensive hyperparameter optimization.

In general [5], as well as the work of other researchers [15], [16], [17] demonstrate that the use of machine learning methods is a promising direction for improving the quality of tactical supply risk assessment for production materials. At the same time, a number of issues raised in previous works require further development:

– studies focus on predicting supply delays, while in business practice, risk is primarily defined as the product of the probability of an event and the magnitude of its impact; no mechanism for transforming forecasts into such assessments has been proposed;

– no analysis of the model's interpretability was performed, which could be useful for understanding the nature of the risks studied;

– there was no discussion of the integration of the assessment model with the ERP system and the results of industrial use;

– due to the limited size of the datasets used, enriching them with new samples and features may lead to improved forecasting quality.

In light of the above, there is significant value in developing a model capable of:

– predicting potential material delivery delays based on heterogeneous features and a historical dataset of limited size, separating delays by duration;

– automatically assessing the impact of predicted delays on the production schedule and the ability to fulfill customer orders on time;

– providing support for interpretability analysis and integration of the model into the ERP system, which will ensure transparency of the risk structure and ease of use by the users.

3. PURPOSE AND OBJECTIVES OF THE STUDY

Thus, the purpose of the study is to develop a model for assessing material supply risks in Make-to-Order and High-Mix, Low-Volume manufacturing using the state-of-the-art TabPFN machine learning model.

To achieve this goal, the following main objectives were formulated:

– development of supply delays forecasting model, that would be capable of taking into account the high dynamics and specifics of Make-to-Order and High-Mix, Low-Volume manufacturing;

– development of an effective mechanism for transforming classified supply delay forecasts into integral risk assessments based on the probability of a delay and its impact on the final result;

– performing an interpretability analysis of the machine learning model to ensure transparency and validity of the forecasting results;

– development of practical recommendations for integrating the studied model into the applied information systems of the enterprise to ensure continuous risk monitoring.

4. GENERAL DESCRIPTION OF THE MODEL

The proposed risk assessment model includes supply delay prediction and integrated supply risk calculation modules.

The material supply delay prediction module is based on a multiclass classifier using a machine learning model. Driven by data from the material procurement management subsystem of the ERP system, this module calculates the probability of a material supply falling into one of the delay classes.

A comparison of modern multi-class classification models conducted by the authors in [5] demonstrates that on datasets similar to the one studied, TabPFN transformer neural network outperforms currently dominant models based on gradient boosting in tabular data analysis. Similar conclusions were reached by the authors of studies

[19], [20], in which TabPFN prevails over other models on datasets with a small number of samples and features as well as a significant proportion of categorical features. Taking into account the results of the above comparisons, in this study, the authors focused on investigating the application of specifically the TabPFN model for forecasting supply delays based on historical data. Compared to [5], an updated and expanded dataset was used to train the model.

The integrated supply risk calculation module uses delay probabilities vector, produced by the TabPFN classifier to calculate the delay criticality score *Ds* representing probability and length of the delay. In addition, material requirements and inventory data are used to calculate material deficit impact score *Is* that reflects consequences of particular material deficit on manufacturing. Finally the integrated supply risk index *Ri*, that combines delay probability, length and material importance data is computed as a product of *Ds* and *Is*.

The outputs of the model are both the material supply delay forecast (the most probable delay class) and an integrated risk assessment index that enables a comprehensive understanding of risk by business stakeholders.

The model is implemented as a service within the enterprise's ERP system, and the display of the forecasting and risk assessment results is integrated into the system's user interface.

5. ANALYSIS AND PREPROCESSING OF MATERIAL SUPPLY DATA

For this study, the set of historical procurement data used in [5] was updated and expanded. Features related to contract terms and payment conditions for the supply were added, as well as samples accumulated in the customer's systems since previous publications. The set now includes records of 4071 material procurements by an industrial enterprise for the years 2023-2025. Each sample contains features obtained from the material and technical supply module, logistics, and finance modules.

In total, 42 features were extracted, including data on the material supplied, quantity, currency, price, amount, supplier, country, dates of order, shipment, delivery, etc., information on the method of delivery, carrier, route, and freighting, as well as information on payments, contract terms, documentation, and involved personnel. A notable characteristic of the resulting dataset is the predominance of categorical features.

Some samples contained missing values in various features, including such key fields as shipment or delivery dates. The authors strived to fill them in as much as possible at the extraction stage, using indirect data such as ERP transaction logs, information from related modules, and so on.

Basic data preprocessing was performed after extraction, which included type conversions and correction of identified data errors. We also calculated delivery deviations and the duration of transportation stages and time until material payment based on the transition dates available in the dataset.

During the exploratory data analysis, we used both visual and numerical tools to identify patterns and interconnections in our data, as well as to identify the features most useful for solving our forecasting problem. In addition to using filtering methods (such as correlation analysis and mutual information analysis), we also constructed a proxy model to as sess the predictive value of the features. The proxy model was created based on the LightGBM [21] framework, the use of which allowed for rapid computation of SHAP [6] values for different feature sets and a comprehensive assessment of their predictive significance.

Finally, 11 features were selected as the most promising predictors. The list of selected features is provided in Table 1.

Table 1. **Features Selected for Modeling**

Feature	Type	Cardinality
SUMM	Float	-
ID_TYPE_DELIVERY	Category	6
ID_CURRENCY	Category	3
ID_COUNTRY	Category	39
ID_SELLER	Category	178
ID_ELEM	Category	764
ID_CONTRACT	Category	203
PAY_WHEN	Category	2
DAYS_TO_PAY	Integer	-
DAYS_TO_SHIP	Integer	-

***Source*: compiled by the authors**

Among the features selected for modeling, there are 7 categorical ones with cardinality ranging from 2 to 764. For features with high cardinality, we tested both frequency grouping and grouping by functional similarity. However, these transformations did not lead to a significant improvement in model quality, confirming the effectiveness of TabPFN's built-in support for categorical features. Consequently, we did not perform additional encoding, apart from setting the data type for them.

For the target variable, we use a distribution into four target classes based on business-significant threshold values, which we proposed earlier in [5]. The division is based on expert assessments provided by the customer's production managers and reflects the fact that different delay durations require different management actions:

– IN_TIME (Timely Delivery) – receipt no later than the day following the planned delivery date;

– SHORT_DELAY (Short Delay): 2–5 days – usually a controllable risk that can be compensated for by existing inventory or minor adjustments to the production schedule without affecting the final order fulfillment date;MODERATE_DELAY (Moderate Delay): 5-15 days — constitutes a risk that may or may not affect order fulfillment depending on the material criticality and the volume of current buffer stock, requiring management intervention and additional analysis;

– LONG_DELAY (Long Delay): > 15 days – delays beyond this threshold generally lead to urgent transportation costs, the search for alternative procurement routes, or the failure to meet obligations regarding the order readiness date.

The resulting target classes are highly imbalanced, with most samples falling into the IN_TIME category, see Table 2. At the same time, the most problematic classes in terms of negative impact on production are MODERATE_DELAY and LONG_DELAY, and their reliable forecasting is important for preventing delays in order fulfillment and operational losses.

6. APPLICATION AND SETUP OF TABPFN

TabPFN (Tabular Prior-Fitted Network) – is a state-of-the-art transformer-based model developed specifically for tabular data classification [6]. The model implements a new paradigm that involves applying the principles of large pre-trained foundation models, common in natural language processing, to structured datasets. TabPFN implements the idea of in-context learning (ICL), similar to that used in large language models. Instead of learning a fixed mapping of features to results for a specific dataset, the model learns how to learn based on examples.

In [5], we used the second version of the model, developed by researchers at the Technical University of Munich and presented in 2025. At the time of preparing this article, the next version, 2.5, had also been released, which, according to the developers, has increased productivity and scalability [22]. However, based on the results of testing on our dataset, the new version did not show an advantage in classification quality or significantly better performance, so we continued to use TabPFN v2 while preparing this study. The advantages of the new version are likely revealed on datasets containing more than 10 thousand samples.

TabPFN provides both a native API and an interface compatible with scikit-learn Estimator, as well as an HTTP REST API via tabpfn-client. In this study, we use the scikit-learn interface (tabpfn.classifier.TabPFNClassifier). We trained the model using a random sample of approximately 80% of the samples in our dataset, leaving the rest for the test set, and activated the option balance_probabilities, explicitly instructing the model to adjust class weight coefficients during inference.

After training the model, we applied the classifier's predict_proba() function to the test set to obtain an array of probability distributions for each test sample falling into the delay classes. The probabilistic forecast obtained in this way will subsequently be used to calculate the supply risk indices. Since we also need to obtain a discrete forecast (specific supply delay classes), we applied the argmax function to the probabilistic forecast, obtaining a vector of predicted delay classes.

6.1. MODELING RESULTS

The overall prediction accuracy when applying the TabPFN v2 model to the updated dataset reached 83%, slightly exceeding the previously obtained figure of 79%. However, due to class imbalance, this metric can be misleading, as it mainly reflects success on the dominant IN_TIME class. Therefore, for a complete and objective assessment of the modeling results, we will analyze the precision, recall, and F1-score for each class separately (Table 3), and also consider the confusion matrix (Fig. 3).

Table 2. **Target Classes**

Class	Delay, days	Number of samples	Proportion of samples
IN_TIME	<= 1	2799	73%
SHORT_DELAY	2 – 5	366	9.6%
MODERATE_DELAY	6 – 15	366	9.6%
LONG_DELAY	> 15	299	7.8%

***Source*: compiled by the authors**

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

Table 3. **Classification Metrics of the Updated Dataset**

Class	Precision	Recall	F1-score	Support
IN_TIME	0.98	0.85	0.91	451
LONG_DELAY	0.83	0.84	0.84	135
MODERATE_DELAY	0.71	0.72	0.71	118
SHORT_DELAY	0.41	0.77	0.54	62

***Source*: compiled by the authors**

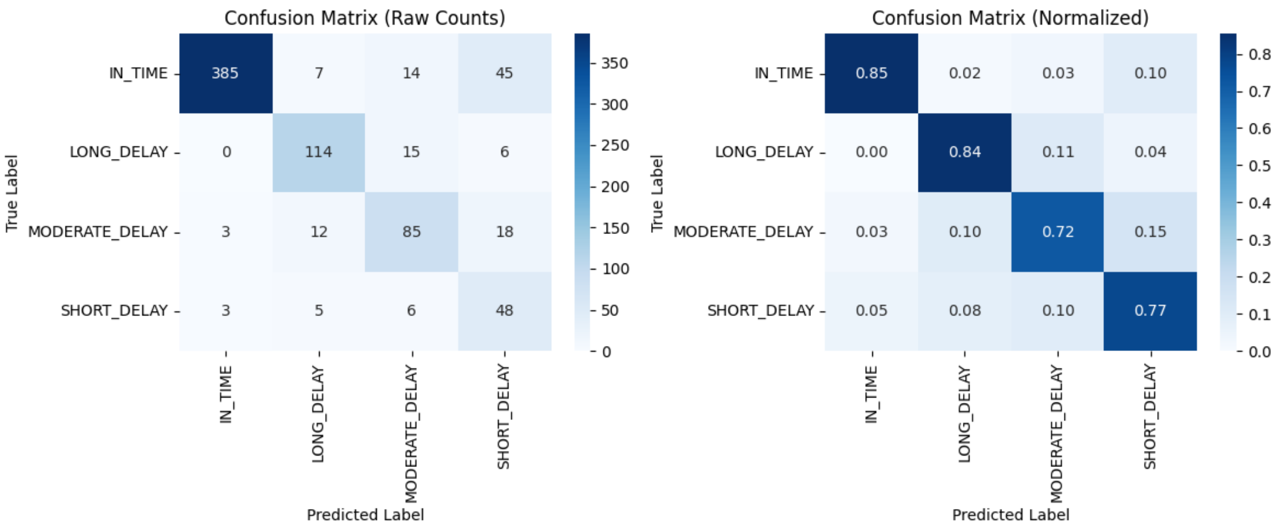

Fig. 3. **Confusion matrix for the classification results of the updated dataset**
***Source:* compiled by the authors**

The best classification performance is expectedly achieved in the dominant IN_TIME class, but it is important that samples associated with the highest risk to production, the LONG_DELAY class, are also classified with high precision and recall. The low precision in the SHORT_DELAY class is explained by the contamination of the forecast for this class by 45 samples from the neighboring IN_TIME class, meaning the model demonstrates slightly excessive conservatism in this case, which is acceptable in risk assessment.

The analysis of the model's confusion matrix demonstrates not only a generally high degree of reliability but also the ordinal consistency of the results (the confusion matrix clearly shows the concentration of data around the main diagonal). This means the model was able to learn the natural sequence of target classes IN_TIME – SHORT_DELAY – MODERATE_DELAY – LONG_DELAY, and in case of an error, the model predominantly chooses a class adjacent to the correct one, rather than a random one. For example, for the true LONG_DELAY class, the model does not predict IN_TIME in any case; instead, most errors fall into the neighboring MODERATE_DELAY class (15 instances). For forecasting supply delays, this characteristic is critical, as the model avoids gross errors, thereby minimizing their consequences. A moderate delay forecast instead of a long one will still prompt the production manager to pay attention to the supply, unlike an erroneous forecast of timely delivery.

Overall, the model demonstrates minimal leakage of classes critical to production stability (LONG_DELAY, MODERATE_DELAY) into the safe category (IN_TIME), meaning the model rarely underestimates the seriousness of a supply delay. Of all forecasts for the IN_TIME category, only 3 cases (zero significant delays and 3 moderate delays) were serious misses. Thus, the model is risk-averse.

For comparison, let's present the classification results of the previous version of the dataset (Table 4). As can be seen, increasing the number of samples and adding features led to some improvement in both overall classification accuracy and performance in individual classes, except for the precision in the SHORT_DELAY class, where 8 new samples of the IN_TIME class were mistakenly classified as SHORT_DELAY.

Table 4. **Classification Metrics of the Previous Dataset**

Class	Precision	Recall	F1-score	Support
IN_TIME	0.98	0.82	0.89	393
LONG_DELAY	0.75	0.81	0.78	103
MODERATE_DELAY	0.61	0.63	0.62	89
SHORT_DELAY	0.45	0.81	0.58	67

***Source*: compiled by the authors**

7. MODEL INTERPRETABILITY ANALYSIS

The application of complex machine learning models in a production environment requires ensuring not only high forecasting quality but also transparency of the mechanisms on which inference is based. When using models for risk management, the ability to trace the influence of features on the results is of particular importance, as it allows for the analysis of factors causing the assessed risk and the implementation of measures to prevent or minimize their impact.

During the work on [5], the authors were unable to perform interpretability analysis due to technical difficulties and time constraints, therefore, performing such an analysis was one of the main objectives when preparing this study.

Some models, such as decision trees, are inherently interpretable (the so-called “white box” architecture). In such models, the internal structure itself reflects a clear path of decision-making rules based on feature values, which directly ensures understanding of the model's predictions. At the same time, TabPFN (like other neural network-based models) is a “black box” model that does not support direct interpretability. Such models do not provide an automatic understanding of how features influence their decisions, but we can gain such understanding using proxy methods like SHAP [7] or LIME [23].

This study uses the SHAP method, which is capable of providing both global and local explanations for model predictions. SHAP (SHapley Additive exPlanations), proposed in 2017 by Scott Lundberg and Su-In Lee, is based on cooperative game theory and allows calculating the contribution of each feature using Shapley values. These values measure how much each feature contributes to the prediction, taking into account all possible combinations of features and assigning a fair weight to each of them. The method allows not only determining the overall significance of a feature but also understanding the direction of its influence (positive or negative) on the probability of each target delay class.

TabPFN supports the calculation of Shapley values through the tabpfn_extensions.interpretability module, which acts as a specialized interface between the model kernel and the SHAP library. To reconcile the TabPFN transformer neural network architecture with standard feature importance methods, this module provides a wrapper that manages the learning context during inference for correct feature processing. The get_shap_values() function of the interpretability module returns a structured Explanation object, which contains the complete matrix of Shapley values. This interpretability support allows for both detailed local analysis of individual predictions and comprehensive assessment of the global importance of features, ensuring compatibility with any visualization and statistical analysis methods within the SHAP ecosystem.

Let us consider the results of the interpretability analysis. Fig. 4 presents a global summary of feature importance for the four target classes. Each horizontal bar reflects the mean absolute SHAP value for one of the features. As can be seen from the diagram, the main features influencing the probability and duration of delay are the country of shipment, time from order to shipment, and supplier. This bar chart well illustrates the degree of influence of features on the assignment of samples to different target classes, but it does not reflect the direction or structure of this influence.

More detailed information on the distribution of Shapley values can be obtained by constructing a SHAP summary plot (Fig. 5), which uses a “beeswarm plot” visualization type. Each point on the feature axis corresponds to the feature's impact on an individual sample. The horizontal position of the point indicates the magnitude and direction of the feature's influence on the prediction. Points located to the right of the central vertical line (SHAP value > 0) indicate a positive contribution of the

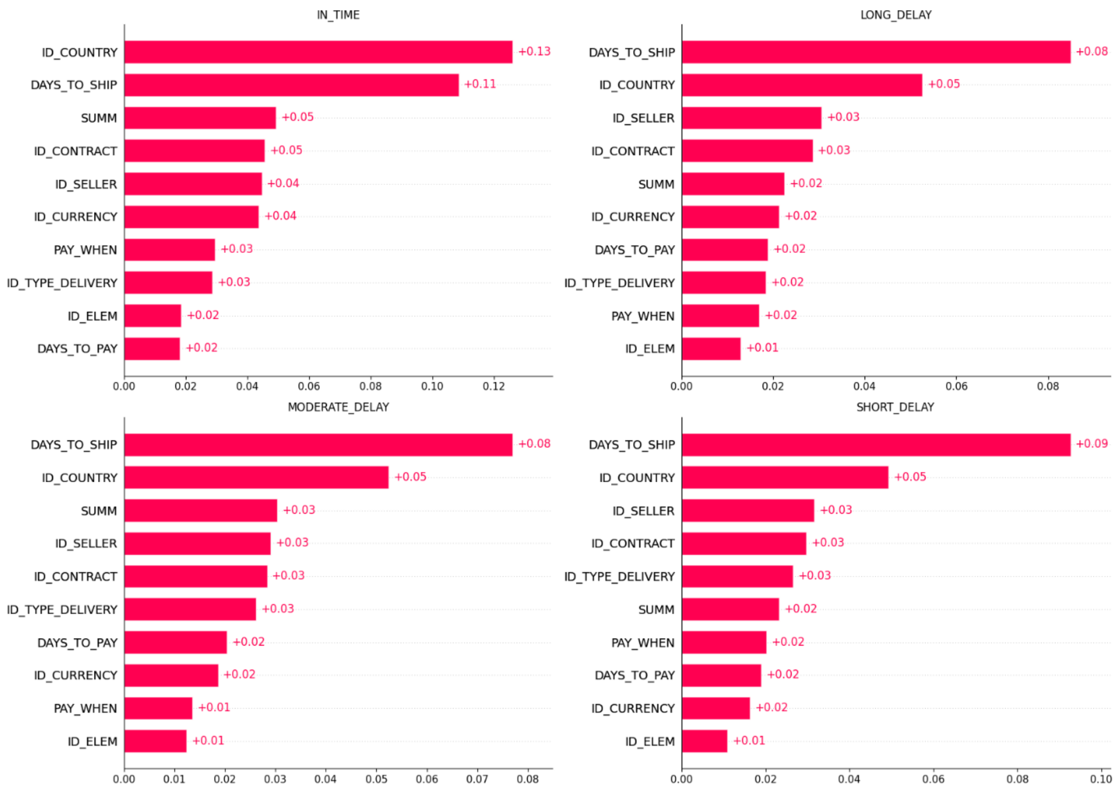

Fig. 4. **Diagrams of the overall importance of features for the target classes**

Source: compiled by the authors

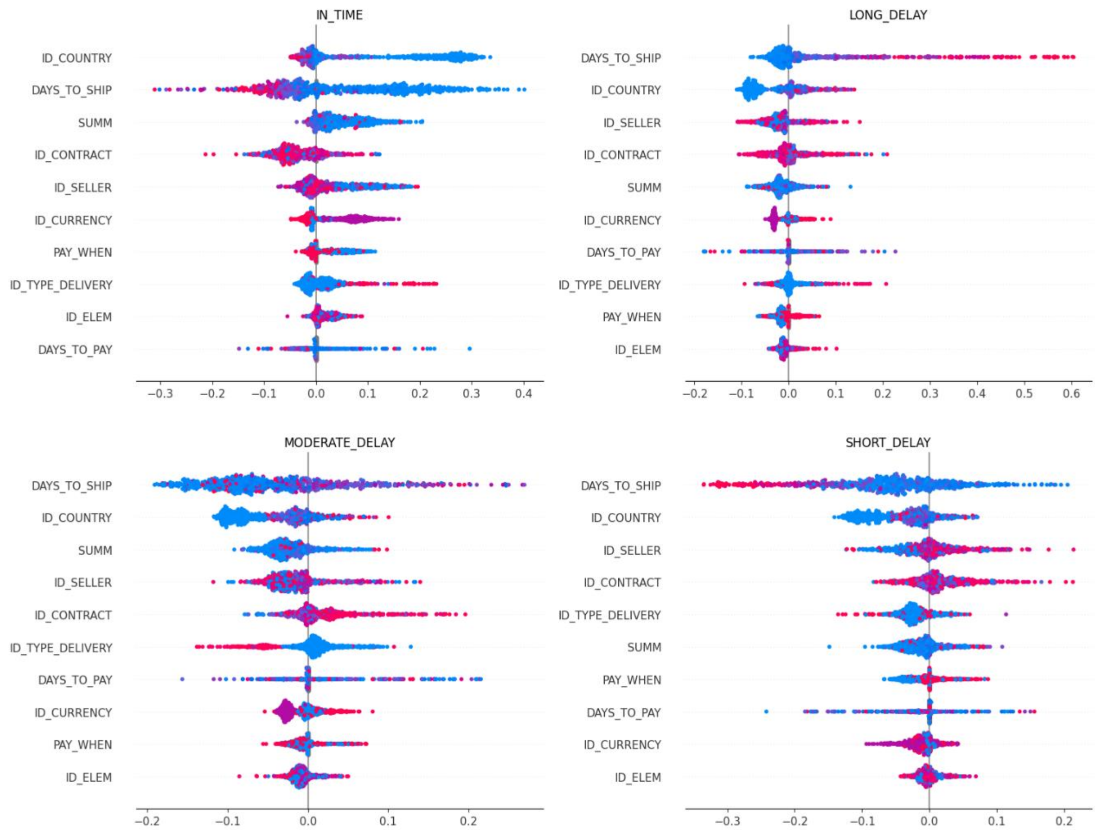

Fig. 5. **SHAP summary plots for the target classes**

Source: compiled by the authors

feature, meaning it increases the probability of the sample being assigned to a specific target delay class. Points located to the left (SHAP value < 0) indicate a negative contribution, meaning the feature reduces the probability of that class. The color coding of the points reflects the feature values and allows for exploring the relationship between the impact and the actual feature values (for example, whether a long time from order to shipment, DAYS_TO_SHIP, indicates an increased probability of untimely supply).The summary plot very clearly shows the structure of the influence of quantitative features but is significantly less informative regarding categorical ones. For instance, from the distribution of the ID_COUNTRY feature, it is almost impossible to understand the impact on the prediction of specific countries from which the supply is carried out. Detailed analysis of the influence of categorical features requires constructing separate contribution plots for the values of each such feature on the classification results. It is impossible to present such plots here for all studied features due to the publication volume constraints and the preservation of customer data confidentiality requirements, but let us consider examples of visualizing the influence of the country of shipment on the IN_TIME and LONG_DELAY classes (Fig. 6 and Fig. 7). As expected, the probability of delays for materials supplied from Ukraine is lower than for those procured abroad, but it is interesting that overall the probability of delays does not show a direct dependence on geographical distance. For example, the probability of delay for Romanian materials is significantly higher than for supplies from Poland.

In general, the results of the SHAP analysis of the interpretability of the supply delay forecasting model provide comprehensive material for studying and searching for ways to optimize the material procurement process.

8. COMPREHENSIVE RISK ASSESSMENT

In previous sections, we showed that the model presented by us is capable of satisfactorily forecasting material supply delays. However, a comprehensive supply risk assessment must consider both the probability of delays and their impact on production [11], [13], [24].

Our proposed TabPFN-based classifier directly produces a probability vector of the target classes for each classified sample (in our case, a vector of probabilities for different delay classes to occur). We need to transform this vector into a single integral assessment, taking into account that a long delay duration is associated with increased risk. We will call this assessment the delay criticality index. To solve this problem, we propose using a weighted approach by multiplying the probability of each class by the median delay for that class in days (1).

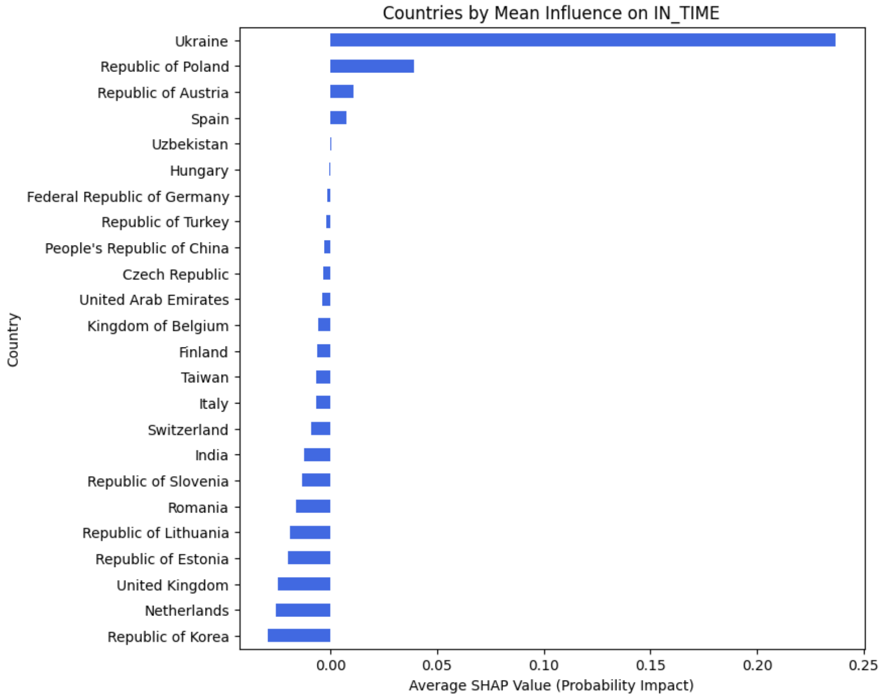

Fig. 6. **Contribution plot of COUNTRY feature on IN_TIME class**
***Source*: compiled by the authors**

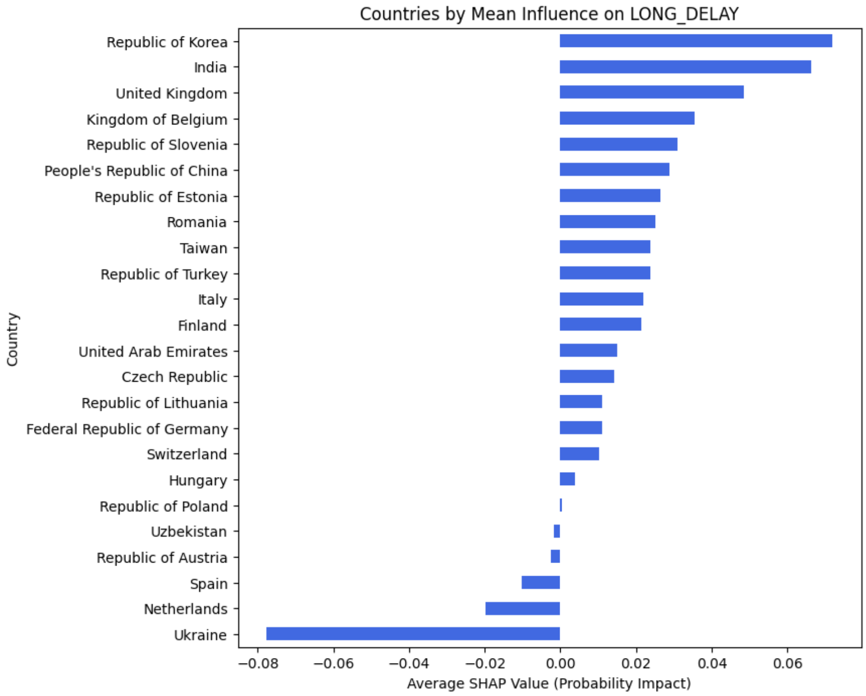

Fig. 7. **Contribution plot of COUNTRY feature on LONG_DELAY class**

***Source*: compiled by the authors**

$$Ds = \sum_{i=1}^{4} P_i \widetilde{D}_i \qquad (1)$$

where *Ds* (Delay Score) is delay criticality index; *Pi* is predicted probability of the supply falling into the i-th delay class; $\widetilde{D}_i$ is median delay of the class, calculated using the test set data. As a result of this calculation, the median for the IN_TIME class is negative due to the influence of deliveries that arrived ahead of schedule; however, we will set the weight of this class equal to zero.

An example of calculating the delay criticality score for five samples is given in Table 5.

The second component necessary for calculating a comprehensive risk assessment is the numerical expression of the impact of risk realization on production–the ability of the enterprise to fulfill customer orders on time and the stability of the production plan. A detailed assessment of the impact of material supply delays on make-to-order manufacturing with a large product variety requires considering a large number of factors: the structure of the production program, warehouse stocks and existing work in progress, the availability of similar materials and alternative production processes, analysis of critical dates for order fulfillment, and contract terms.

Table 5. **Examples of Delay Criticality Score Calculation**

Sample	**Predicted Class**								**Criticality Score *Ds***	**True Class**
	IN_TIME $\widetilde{D}_i$=0		SHORT_DELAY $\widetilde{D}_i$=3		MODERATE_DELAY $\widetilde{D}_i$=9		LONG_DELAY $\widetilde{D}_i$=33			
	Probability	Weight	Probability	Weight	Probability	Weight	Probability	Weight		
3973	0.3612	0	0.6026	1.8078	0.0253	0.2275	0.0109	0.3595	**2.3948**	IN_TIME
455	0.0061	0	0.9034	2.7103	0.0811	0.7301	0.0094	0.3088	**3.7491**	SHORT_ DELAY
336	0.0005	0	0.0397	0.1191	0.9256	8.3305	0.0342	1.1295	**9.5791**	MODERATE_ DELAY
3361	0.0004	0	0.0015	0.0045	0.2634	2.3703	0.7348	24.2478	**26.6226**	LONG_ DELAY
1394	0.0581	0	0.1154	0.3461	0.3165	2.8485	0.5101	16.8327	**20.0272**	LONG_ DELAY

Source: compiled by the authors

Developing a mechanism for such a comprehensive assessment requires a separate study, which we plan to carry out in the future; however, within the scope of this work, we propose a simplified index that takes into account the need for the material and its overall significance for the production process (2).

$$Is = (Vd/Dm) \cdot Mi \qquad (2)$$

where *Is* (Material deficit impact score) is metric of the impact of material shortage resulting from supply delay; *Vd* (Delayed volume) is the quantity of material in the supply for which the risk of delay is assessed; *Dm* (Demand for material) is the need for the material according to the Master Production Schedule, starting from the planned supply date and for the period up to the median value of the forecasted delay class; *Mi* (Material importance) is material criticality index, by default calculated as the share of orders in the production program that require the presence of this material for completion, but can be adjusted by expert decision.

An example of calculating the delay impact index for four samples is given in Table 6.

Consequently, the comprehensive risk assessment index can be calculated as:

$$Rs = Ds \cdot Is \qquad (3)$$

where *Ri* (Risk index) is combined delay and impact risk index; *Ds* (Delay Score) is delay criticality score; *Is* (Impact score) is score of the impact of material supply delay.

An example of calculating the risk assessment indices for five supplies is given in Table 7.

9. DEPLOYMENT OF MODELS AND INTEGRATION INTO THE ERP SYSTEM

Since TabPFN is supplied pre-trained, its configuration and obtaining predictions do not require specialized computing equipment. At the same time, calculating SHAP values requires a workstation with a CUDA-compatible accelerator. On an NVIDIA GeForce RTX 3060 accelerator, the SHAP calculation for our dataset took about 8 hours.

The model can be integrated into the ERP system in the form of a handler service or a batch task executed on the system's application servers. For our customer, such a handler was deployed as a Docker container with a REST interface, the call of which is performed when the status of orders changes. Obtaining an updated forecast for all active orders takes approximately 2 minutes on an Intel

Table 6. **Calculation of the Delay Impact Index**

Sample	**Material**	**Unit of measure**	**Supply volume**	**Demand for the delay period**	**Material criticality index**	**Material deficit impact score *Is***
3973	Black Polystyrene	kg	2084	3980	0.2	0.104724
455	White PVC Plasticate	kg	1216	2820	0.4	0.172482
336	Aluminum Wire Rod	kg	22050	60250	0.3	0.109793
3361	Inorganic Blue Dye	kg	100	218	0.2	0.330275
1394	Copper Drawing Stock	kg	21920	380000	0.55	0.031726

***Source*: compiled by the authors**

Table 7. **Calculation of the Comprehensive Risk Index**

Sample	**Delay Class**	**Delay Criticality Score *Ds***	**Delay Impact Score *Is***	**Risk Index *Ri***
3973	IN_TIME	2.3948	0.104724	**0.250795**
455	SHORT_DELAY	3.7491	0.172482	**0.646658**
336	MODERATE_DELAY	9.5791	0.109793	**1.051714**
3361	LONG_DELAY	26.6226	0.330275	**8.79277**
1394	LONG_DELAY	20.0272	0.031726	**0.63539**

***Source*: compiled by the author**

Xeon 6244 processor without using a graphics accelerator and with the number of processor cores available to the container limited to four. The obtained delay forecasts and risk indices are stored in the ERP system database and are available to users through the standard system interface in forms, reports, and dashboards. Additionally, notification of relevant users about forecasts of long supply delay (LONG_DELAY) has been implemented.

10. LIMITATIONS AND DIRECTIONS FOR FURTHER RESEARCH

The limitations of this study are primarily related to the fact that it is based on data available in the customer's IT systems, where the volume of available samples and features is determined by the scale of the business activities of a single enterprise, as well as the duration of data collection and the achieved level of business process automation.

As shown in this study, the available data is sufficient for the practical testing of the proposed models, but the question of their effectiveness in other production environments and with different practices for organizing material procurement has not been sufficiently explored.

The issue of assessing the impact of risks predicted by the developed models on the operational activities of the enterprise and integrating the obtained forecasts into the enterprise's risk management system is also currently developed in a simplified manner and leaves a significant area for further research.

In the future, the authors primarily plan to focus on the issue of improving the assessment of the impact of material supply delays on the enterprise's production activities and considering numerous factors in such an assessment, such as the volume of warehouse stocks of materials, available material analogs and the availability of alternative technological processes, available limits for production replanning, order priority, and the framework of contractual obligations.

Another promising area for future research is the integration of the model's outputs into the enterprise's decision-making framework, alongside the implementation of a corresponding decision support system [25].

The authors also plan to continue working on enriching the used datasets with new samples and features and optimizing the utilized models.

11. CONCLUSION

The supply risk-assessment model, proposed in this study enables automated, near-real-time evaluation of short-to-moderate (tactical) supply risks for manufacturing enterprises. The model takes into account both the supply delays predicted by the TabPFN-based multiclass classifier and their subsequent impact on the production schedule.

The authors' main contribution lies in developing a delay forecasting approach that combines multi-class delay classification based on business-defined thresholds with the state-of-the-art TabPFN model that enables accurate delay prediction even on limited datasets. Another novel aspect is the proposed approach to calculating a comprehensive risk index, which combines predicted delay probabilities with an evaluation of their impact on the production schedule. The combination of machine learning-enabled delay prediction and the proposed risk index allows for automated risk assessments on a near-real-time scale."On the updated and expanded dataset, TabPFN confirmed its ability to provide high-quality forecasts based on complex tabular data characterized by a limited number of samples and numerous categorical features. The obtained forecasts demonstrate ordinal consistency of results and a low tendency to underestimate supply delays, which allows for the integration of the model into decision-making processes as a reliable tool for warning about material supply risks.

Compared to statistical method, described in [4], the use of a pre-trained transformer neural network ensures effective consideration of heterogeneous features (including categorical ones), the possibility of universal approximation of dependencies, in particular complex non-linear functions, as well as the ability to automatically detect and analyze complex feature interactions.

A method for calculating a comprehensive material supply risk assessment is proposed, which takes into account both delivery delays predicted by the model and their impact on production.

An interpretability analysis of the model was performed using the SHAP method, showing the possibility of using its results to identify risk factors.

The model was deployed as a containerized service that queries the ERP system database in near-real-time, outputting delay predictions and risk evaluations through the standard ERP user interface.

In the future, the authors plan both to continue work on expanding the datasets and optimizing the model, as well as to improve the comprehensive risk assessment, considering additional factors of the impact of supply risks on production and the integration of modeling results into the management decision-making process.

REFERENCES

1. Choudhary, N. A., Singh, S., Schoenherr, T. & Ramkumar, M. "Risk assessment in supply chains: a state-of-the-art review of methodologies and their applications". *Ann Oper Res*. 2023; 322 (2): 565–607. DOI: https://doi.org/10.1007/s10479-022-04700-9.

2. Zsidisin, G., Gaudenzi, B. & Pellegrino, R. "Strategic sourcing: Approaches for managing supply chain risk". *Strategic Sourcing*. 2024. DOI: https://doi.org/10.1007/978-3-031-52592-6.

3. Gan, Z. L., Musa, S. N. & Yap, H. J. "A review of the high-mix, low-volume manufacturing industry". *Applied Sciences*. 2023; 13 (3): 1687. DOI: https://doi.org/10.3390/app13031687.

4. Mrykhin, A. L. & Antoshchuk, S. G. "Information model for assessing the impact of tactical material procurement risks on order fulfillment in Make-to-Order manufacturing". *Herald of Advanced Information Technology*. 2024; 7 (2): 243–252. DOI: https://doi.org/10.15276/hait.07.2024.16.

5. Mrykhin, A. L. & Antoshchuk, S. G. "Application of a novel TabPFN model for tabular data classification". *Informatics. Culture. Technique*. 2025; (2): 54–60. DOI: https://doi.org/10.15276/ict.02.2025.07.

6. Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., Schirrmeister, R. T. & Hutter, F. "Accurate predictions on small data with a tabular foundation model". *Nature*. 2025; 637: 319–326. DOI: https://doi.org/10.1038/s41586-024-08328-6.

7. Lundberg, S. M. & Lee, S. I. "A unified approach to interpreting model predictions". *Advances in Neural Information Processing Systems*. 2017. DOI: https://doi.org/10.48550/arXiv.1705.07874.

8. Zsidisin, G. A., Panelli, A. & Upton, R. "Purchasing organization involvement in risk assessments, contingency plans, and risk management: An exploratory study". *Supply Chain Management: An International Journal*. 2000; 5 (4): 187–198. DOI: https://doi.org/10.1108/13598540010347307.

9. Harland, C., Brenchley, R. & Walker, H. "Risk in supply networks". *Journal of Purchasing and Supply Management*. 2003; 9 (2): 51–62. DOI: https://doi.org/10.1016/S1478-4092(03)00004-9.

10. Zsidisin, G. "A grounded definition of supply risk". *Journal of Purchasing and Supply Management.* 2003; 9 (5): 217–224. DOI: https://doi.org/10.1016/j.pursup.2003.07.002.

11. Schlegel, G. L. & Trent, R. J. "Supply chain risk management: An emerging discipline". *CRC Press*. 2014. DOI: https://doi.org/10.1201/b17531.

12. Gurtu, A. & Johny, J. "Supply chain risk management: Literature review". *Risks*. 2021; 9 (1). DOI: https://doi.org/10.3390/risks9010016.

13. Heckmann, I., Comes, T. & Nickel, S. "A critical review on supply chain risk – definition, measure and modeling". *Omega*. 2014; 52: 119–132. DOI: https://doi.org/10.1016/j.omega.2014.10.004.

14. Ho, W., Zheng, T., Yildiz, H. & Talluri, S. "Supply chain risk management: A literature review". *International Journal of Production Research*. 2015; 53 (16): 5031–5069. DOI: https://doi.org/10.1080/00207543.2015.1030467.

15. Steinberg, F., Burggräf, P., Wagner, J., Heinbach, B., Saßmannshausen, T. & Brintrup, A. "A novel machine learning model for predicting late supplier deliveries of Low-Volume-High-Variety products with application in a german machinery industry". *Supply Chain Analytics*. 2023; 1: 100003. DOI: https://doi.org/10.1016/j.sca.2023.100003.

16. Müller, N., Burggräf, P., Steinberg, F., Sauer, C. R. & Schütz, M. "An analytical review of predictive methods for delivery delays in supply chains". *Supply Chain Analytics*. 2025; 11: 100130. DOI: https://doi.org/10.1016/j.sca.2025.100130.

17. Rezki, N. & Mansouri, M. "Machine learning for proactive supply chain risk management: predicting delays and enhancing operational efficiency". *Management Systems in Production Engineering*. 2024; 32: 345-356. DOI: https://doi.org/10.2478/mspe-2024-0033.

18. Baryannis, G., Dani, S. & Antoniou, G. "Predicting supply chain risks using machine learning: The trade-off between performance and interpretability". *Future Generation Computer Systems*. 2019; 101: 993–1004. DOI: https://doi.org/10.1016/j.future.2019.07.059.

19. Zhan, Q., Tan, Y. S., Tian, Q. & Li, P. "TabPFN: One Model to Rule Them All?". *arXiv*. 2025. DOI: https://doi.org/10.48550/arXiv.2505.20003.

20. Ye, H. J., Liu, S. Y. & Chao, W. L. "A closer look at tabpfn v2: Strength, limitation, and extension". *arXiv*. 2025. DOI: https://doi.org/10.48550/arXiv.2502.17361.

21.Ke, G., Meng, Q., Finley, T., et al. “LightGBM: A highly efficient gradient boosting decision tree”. *Advances in Neural Information Processing Systems* 2017. DOI: https://doi.org/10.5555/3294996.3295074.

22.Grinsztajn, L., Flöge, K., Key, O. et al. “TabPFN-2.5: Advancing the state of the art in tabular foundation models”. *arXiv*. 2026. DOI: https://doi.org/10.48550/arXiv.2511.08667.

23.Ribeiro, M. T., Singh, S. & Guestrin, C. “‘Why should i trust you?’: Explaining the predictions of any classifier”. *arXiv*. 2016. DOI: https://doi.org/10.48550/arXiv.1602.04938.

24.Govindan, K., Fattahi, M. & Keyvanshokooh, E. “Supply chain network design under uncertainty: A comprehensive review and future research directions”. *European Journal of Operational Research*. 2017; 263 (1): 108–41. DOI: https://doi.org/10.1016/j.ejor.2017.04.009.

25.Komleva, N. O. “Model and method for evolutionary control of the decision-making structure in intelligent systems”. *Herald of Advanced Information Technology*. 2026; 9 (2): 181–198, https://www.scopus.com/authid/detail.uri?authorId=57191858904.
DOI: https://doi.org/10.15276/hait.09.2026.13.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article
Author Svitlana G. Antoshchuk is member of the Editorial Board of this journal. This role had no influence on the peer review process or editorial decision regarding this manuscript

Received 07.04.2026
Received after revision 12.06.2026
Accepted 18.06.2026

DOI: https://doi.org/10.15276/aait.09.2026.28
УДК 004.9: 005.93

Оцінювання ризиків постачання матеріалів у виробництві на замовлення на основі нейромережевої моделі TabPFN

Мрихін Андрій Львович[1)]
ORCID: https://orcid.org/0009-0009-1545-632X; amrykhin@gmail.com
Антощук Світлана Григорівна[1)]
ORCID: https://orcid.org/0000-0002-9346-145X; asg@op.edu.ua. Scopus Author ID: 8393582500
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Актуальність. Зростаюча глобальна нестабільність та вразливість ланцюгів постачання призводять до підвищення ризиків забезпечення матеріалами промислових підприємств. Проблема прогнозування та оцінювання ризиків постачання матеріалів є особливо актуальною у виробництві на замовлення та багатономенклатурному дрібносерійному виробництві. Такі виробничі системи, що характеризуються різноманітністю продукції та низькими буферними запасами, є високочутливими до затримок у доставці матеріалів. Традиційні системи управління ризиками, розроблені для стабільного масового виробництва, часто не здатні достатньо швидко реагувати на динаміку позамовних багатономенклатурних середовищ. **Метою** цього дослідження є покращення аналізу ризиків закупівлі виробничих матеріалів завдяки використанню машинного навчання для прогнозування затримок доставки матеріалів та розробці методу оцінювання ризиків на основі таких прогнозів. Для досягнення цієї мети було вирішено такі **завдання** розроблено модель для прогнозування затримок доставки та впроваджено інтегрований індекс ризику постачання, який базується на ймовірності затримок та оцінці їхнього впливу. Також досліджується інтерпретованість моделі-класифікатора та наводиться приклад її інтеграції до системи планування ресурсів підприємства-замовника. **Методи.** Для отримання передбачень затримок використовується багатокласовий класифікатор на основі новітньої моделі TabPFN. Модель TabPFN заснована на архітектурі нейромережі-трансформера адаптованої для аналіза табулярних даних, для її навчання використан набір історичних даних щодо закупівель, отриманий з інформаційних систем замовника. **Результати.** У дослідженні модель-класифікатор на основі TabPFN продемонструвала високу ступінь достовірності результатів навіть при застосуванні параметрів за замовчуванням, ординальну узгодженість результатів та низьку схильність до недооцінки затримок постачання, що свідчить про орієнтованість отриманої моделі на мінімізацію ризиків. Запропонована методика обчислення оцінки ризику дозволяє врахувати вплив прогнозованої затримки постачання на виробництво залежно від чинної виробничої програми та складських залишків. **Наукова новизна.** Поєднання багатокласової класифікації затримок на основі визначених вимогами бізнесу меж із новітньою моделлю TabPFN забезпечує точне прогнозування затримок навіть на обмежених наборах даних. Іншим аспектом новизни є запропонований підхід до розрахунку комплексного індексу ризику, який поєднує прогнозовані ймовірності затримок з оцінкою їхнього впливу на виробничий графік. Поєднання прогнозування затримок за допомогою машинного навчання та запропонованого індексу ризику дозволяє автоматизовано отримувати оцінки ризиків у масштабі часу, наближеному до реального. **Висновки.** Дослідження демонструє, що застосування методів машинного навчання може підвищити точність та своєчасність оцінювання ризиків постачання, уможливлюючи моніторинг у режимі, наближеному до реального часу, та прийняття більш обґрунтованих рішень щодо забезпечення виробництва матеріалами.

Ключові слова: оцінювання ризиків; прогнозування; позамовне виробництво; багатономенклатурне виробництво; машинне навчання; табулярні дані; багатокласова класифікація; TabPFN

ABOUT THE AUTHORS

Andrew L. Mrykhin - Postgraduate student, Department of Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0009-0009-1545-632X; asg@op.edu.ua
***Research field*:** Information systems for manufacturing planning and control

Мрихін Андрій Львович – аспірант кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Svitlana G. Antoshchuk - Doctor of Engineering Sciences, Professor, Department Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-9346-145X; asg@op.edu.ua. Scopus Author ID: 8393582500
***Research field*:** Automated management systems and progressive information technologies; information technology; theoretical and applied aspects of multimedia data processing

Антощук Світлана Григорівна – доктор технічних наук, професор кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)