FMDB *Transactions on Sustainable Finance and Data Science*

Hybrid NLP and Machine Learning Model for Intelligent Procurement Risk Analysis

K. Pradeep[1,*], S. Sindhu[2]

[1,2]Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.
pradeepchennai10@gmail.com[1], sindhus2@srmist.edu.in[2]

Abstract: Modern supply chains have made procurement risk management even more critical due to demand uncertainty, supply disruptions, and price volatility. Traditional methodologies may not be able to combine diverse data sources and deliver timely, operational insights. The proposed Hybrid Artificial Intelligence-based Procurement Risk Control is implemented for Procurement Risk Analysis. The suggested framework combines structured data, such as Demand Net Shortage (DNS), inter-site supply availability, open-market sourcing, and financial exposure indicators (such as IPDR and OPOR), with unstructured information in the form of buyer comments. An inference engine based on fuzzy logic is used to characterise uncertainty and categorise procurement risk into meaningful categories, enabling effective decision-making in uncertain situations. Simultaneously, a Natural Language Processing (NLP) component takes textual inputs, applies Term Frequency-Inverse Document Frequency (TF-IDF) vectorisation, and a Logistic Regression classifier to present qualitative risk insights. To increase the system's reliability during the first stages of deployment, a confidence-based switching mechanism is presented that dynamically balances rule-based logic and machine learning predictions. Continuous retraining allows adaptive learning and improved performance. The experiment shows that the hybrid method is more efficient and successful in procurement risk identification, decision-making, and supply chain resilience. Its scalable architecture, computational efficiency, and corporate implementation provide a comprehensive approach to intelligent procurement risk management.

Keywords: Procurement Risk; Scalable Structure; Operational Insights; Logistic Regression; Fuzzy Logic; Textual Inputs; Financial Risk Modelling; Self-Learning System.

Received on: 28/10/2025, **Revised on:** 19/12/2025, **Accepted on:** 04/02/2026, **Published on:** 19/08/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSFDS

DOI: https://doi.org/10.69888/FTSFDS.2026.000743

Cite as: K. Pradeep and S. Sindhu, "Hybrid NLP and Machine Learning Model for Intelligent Procurement Risk Analysis," *FMDB Transactions on Sustainable Finance and Data Science*, vol. 1, no. 3, pp. 172–186, 2026.

1. Introduction

The global supply chain is increasingly exposed to challenges driven by demand variability, supplier uncertainty, and market price volatility [1]. These issues have a significant impact on procurement processes, and risk management is a critical factor in ensuring the continuity of supply and cost efficiency. Nonetheless, conventional procurement systems are, in essence, reactive, meaning shortages are identified only after material gaps have already occurred. Moreover, these systems tend to

[*]Corresponding author.

operate in data silos and rely heavily on manual decision-making, limiting their ability to deliver timely, comprehensive risk insights [2]. The other key weakness of traditional systems is that they do not effectively exploit unstructured data, e.g., buyer comments, which often hold useful early warning indicators of supply risks, delays, and supplier performance. The lack of mechanisms to analyse and combine such qualitative information further diminishes the efficiency of procurement decision-making. To overcome these issues, the paper proposes a Hybrid AI-Based Procurement Risk Intelligence System that enables proactive, multi-dimensional risk assessment [3]. The proposed system combines various analysis components, including time-based inventory prediction, internal excess stock redistribution analysis, open-market financial exposure estimation, and NLP-based buyer sentiment classification. Moreover, the framework includes retraining machine learning models, enabling the system to continually develop and enhance its predictive capabilities [4]. This study aimed to develop a self-enhancing procurement decision support system that integrates structured and unstructured data sources with smart decision models. Using hybrid AI methods, the proposed procurement control tower can provide better visibility, enhanced risk detection accuracy, and facilitate more informed, proactive decision-making in complex supply chain settings.

1.1. Research Objectives

The main aim of the study is to create a Hybrid NLP and ML-based Procurement Risk Analysis System that enables proactive, information-driven, and intelligent decision-making in the contemporary supply chain context [5]. The suggested system seeks to address the shortcomings of conventional procurement methodologies by incorporating orderly and disorganised information into a single risk intelligence system. The specific objectives of this study are as follows:

- To develop a hybrid risk assessment model to integrate deterministic rule-based reasoning with machine learning-based methods in overall procurement risk assessment.
- To combine structured procurement data, such as DNS, intersite supply availability, open market sourcing, and financial exposure measures (IPDR and OPOR), to quantify risk.
- To create an NLP-based classifier model based on TF-IDF vectorisation and Logistic Regression to process unstructured buyer comments and obtain qualitative risk information.
- To apply a fuzzy logic-based inference system to deal with the uncertainty and classify the procurement risk into actionable levels.
- To present a confidence-driven switching scheme dynamically matching rule-based and machine learning predictions, to maintain consistent performance in the early stages of deployment.
- To support ongoing learning by retraining the model, to allow the system to adjust to changes in the supply chain, and to enhance predictions as time passes.
- To improve the efficiency of decision-making and supply chain resilience by delivering multi-dimensional procurement risk intelligence in real-time.

In general, this research aims to design a scalable, adaptive, and computationally efficient procurement risk control system for enterprise-level applications.

2. Literature Review

The current semiconductor and advanced manufacturing procurement literature is mainly focused on three areas: digital transformation of procurement, purchase order automation, and supply chain forecasting techniques. All these areas form the basis for creating smart, data-driven procurement systems [6]. This paper offered an extensive scholarly examination of the utilisation of machine learning across all prevalent Supply Chain Management (SCM) functions. This paper presents a sequential multi-perspective conceptual framework for systematically investigating the effective application of Machine Learning (ML) to supply chain risk prediction and management. This paper discusses data preprocessing, feature extraction, data transformation, and handling missing values [7]. The suggested integrated model includes a Multiple Linear Regression (MLR) algorithm, a performance metric, and a method to predict supply chain risk. This paper used semantic networks to obtain supply chain risk knowledge and proposes a Multi-Level Inventory (MLI) Model based on the coordination centre inventory and retailer strategies.

This paper evaluated the structure of multi-module operations with respect to risk, logistics, and information exchange in supply chains [8]. The study provided an extensive evaluation and theoretical framework for predictive supplier risk assessment models to enhance supply chain continuity and operational resilience [13]. The goal of this study is to improve supply chain performance management using ML and neural network (NN) techniques and to develop a supply chain risk management model based on these technologies. Supply chain risk managers can help identify these problems rapidly by combining NLP and reinforcement learning, two core types of artificial intelligence. To identify the most important procurement risks, evaluate the effectiveness of probabilistic models, and propose practical recommendations for incorporating probabilistic approaches into procurement strategies [9]. The primary methodological approach employed in this study is Partial Least Squares-Based Structural Equation Modelling (PLS-SEM).

2.1. Digital Change in Procurement

The recent literature highlights the importance of digital technologies in transforming conventional procurement into intelligent, scalable systems [10]. The semiconductor supply chain is highly complex and globally interdependent, requiring sophisticated digital systems to ensure operational efficiency and resilience. In Aulia and Isvara [11], a Digital Procurement Maturity Model (DPMM) was developed to determine the level of organisational maturity required to adopt digital and to prepare transformation strategies. The authors proposed that Robotic Process Automation (RPA) enhances efficiency in repetitive procurement processes, reducing manual effort and accelerating the process [12]. ERP-based dashboards also increase procurement visibility by providing real-time information on stock levels, supplier performance, and workflow status.

2.2. Purchase Order Automation

Purchase order (PO) processes have been automated due to inefficient manual procurement systems. The case study in the appended paper lists some of the serious problems with conventional PO systems, such as: Conventional procurement systems require one or two days to complete purchase order processing because most procurement activities are handled manually. Tasks such as supplier communication, approval verification, and document preparation increase processing delays and reduce procurement efficiency [14]. Automated PO systems reduce processing time through real-time order generation and digital workflow management.

2.2.1. Planner Effort Consumption (15%–25%)

In manual procurement systems, procurement planners spend nearly 15% to 25% of their effort on repetitive tasks such as order tracking, inventory checking, and supplier coordination [15]. This increases workload, reduces productivity, and may lead to procurement errors and operational inefficiencies. Automated procurement systems minimise manual effort by integrating intelligent procurement workflows and automated monitoring [16].

2.2.2. Understocking and Overstocking of 1825% and 2025%

To discover that a shift from manual to automated purchasing significantly enhances procurement efficiency. Intelligent agent-based models and rule-based systems have also been widely used to automate procurement processes, eliminating human error and ensuring consistency [17]. The latest development is the application of automation tools, such as Selenium, on top of ERP systems, enabling automated execution of repetitive processes.

2.3. Forecasting Techniques in Procurement

Demand forecasting is an essential item of procurement planning, particularly in a highly varying industry like semiconductor manufacturing [18]. Conventional statistical models, such as the Auto Regressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ETS, Holt, SES), have been widely used for time-series forecasting. Research has shown that ARIMA is effective for structured and seasonal data, but it does not work well with non-linear patterns.

2.4. Research Gap and Motivation

Although digital procurement systems and automation technologies have improved procurement management processes, existing systems still face several limitations. Many procurement platforms cannot efficiently combine structured and unstructured procurement data, and they provide limited Natural Language Processing (NLP) support for analysing buyer comments and supplier feedback. In addition, most systems lack adaptive learning and continuous model improvement, which reduces the accuracy of procurement risk predictions. This study addresses these limitations by integrating machine learning, fuzzy logic, and NLP techniques into an intelligent procurement risk analysis framework to improve procurement decision-making and risk management.

2.5. Research Gap

Even though digital procurement, purchase order automation, and forecasting methods have made major progress, there are still two crucial drawbacks to current procurement systems, especially in the intricate context of the semiconductor and advanced manufacturing supply chain. To begin with, the majority of the literature is centred on structured data and primarily on demand forecasting and inventory optimisation, but it offers little on integrating unstructured data (e.g., buyer comments and qualitative feedback). Second, although forecasting models, including ARIMA, Exponential Smoothing, and deep learning models such as LSTM, are more accurate, most applications rely on single-model or static model selection methods. Third, current procurement automation solutions are mostly rule-based and reactive, with no ability to incorporate predictive intelligence or manage uncertainty. Moreover, the existing systems lack a single framework that integrates forecasting, risk assessment, and decision intelligence.

The major limitation is the lack of adaptive and self-learning processes in procurement systems. Most available models are static and fail to include ongoing retraining or a feedback loop, which are needed to sustain performance in dynamic supply chain conditions. As such, there is a definite requirement for a hybrid, intelligent procurement risk analysis framework. The proposed Hybrid NLP and Machine Learning-based Procurement Risk Analysis System provides an intelligent and efficient solution for modern procurement risk management. The system integrates structured and unstructured procurement data and leverages fuzzy logic, machine learning, and Natural Language Processing (NLP) to predict and analyse procurement risks accurately. It also supports dynamic model selection, online learning, and real-time procurement risk assessment, enabling organisations to proactively identify supplier risks, operational issues, and supply chain disruptions. As a scalable and flexible framework, the proposed system enhances procurement decision-making and improves overall supply chain risk management performance.

3. System Architecture

The Hybrid NLP and Machine Learning-Based Procurement Risk Analysis System is planned as a scalable, modular architecture that integrates several analytical modules to perform a comprehensive analysis of procurement risks. Figure 1 shows an architecture of an AI-powered autonomous agent including perception, brain, learning, memory and actuation components. The perception layer gathers data from sensors such as microphones, cameras, and IoT devices. The brain is a repository of knowledge, a reasoning engine, and a set of goals and utility functions for processing information and producing judgements. The learning component provides data collection, feature extraction, model training, feedback and ongoing modification. Actuators implement decisions using physical and digital means. The design also involves the environment through interaction and feedback loops, memory, external knowledge sources, and AI agents. Finally, the graphic shows an integrated framework for intelligent and autonomous operation of the agents.

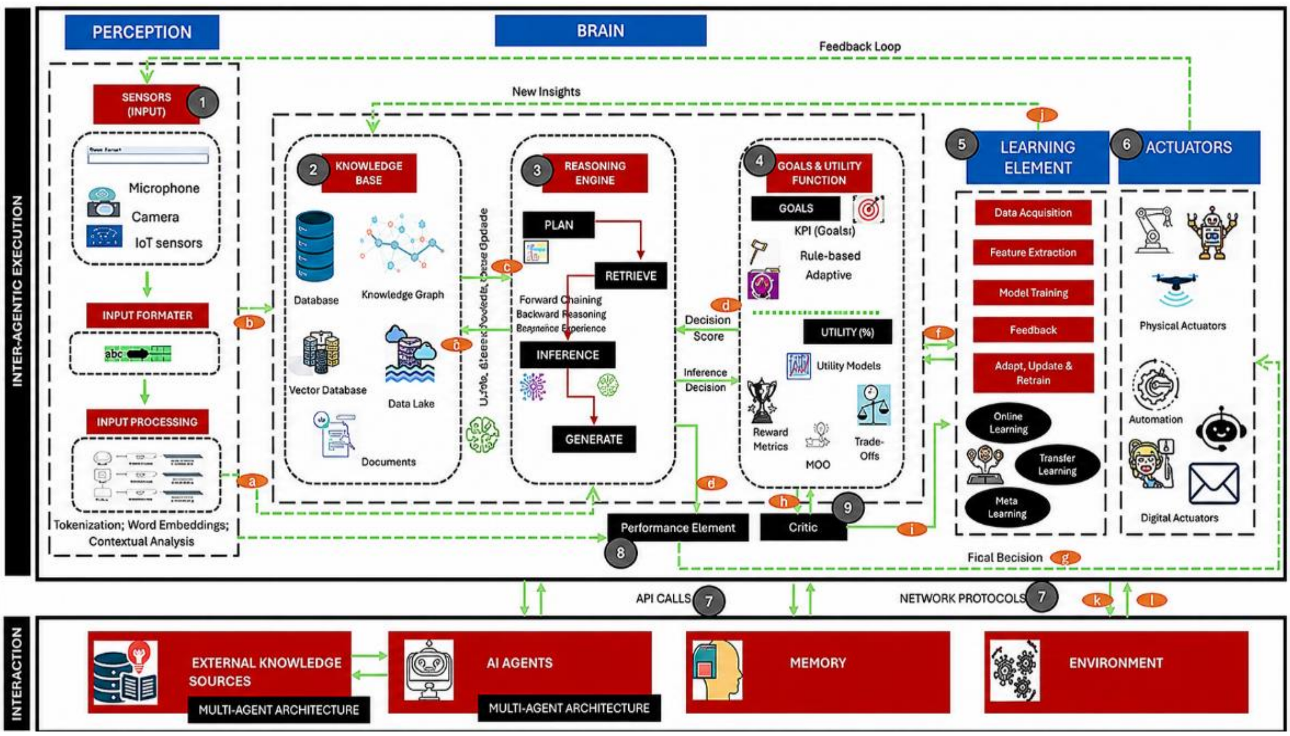

Figure 1: Architecture of an AI-powered autonomous agent system integrating perception, brain, learning, memory, and actuation

The system integrates formal data processing, knowledge decision models, and unstructured text analysis to facilitate data-driven, proactive procurement decisions. The entire architecture comprises six key modules as explained below (Figure 2).

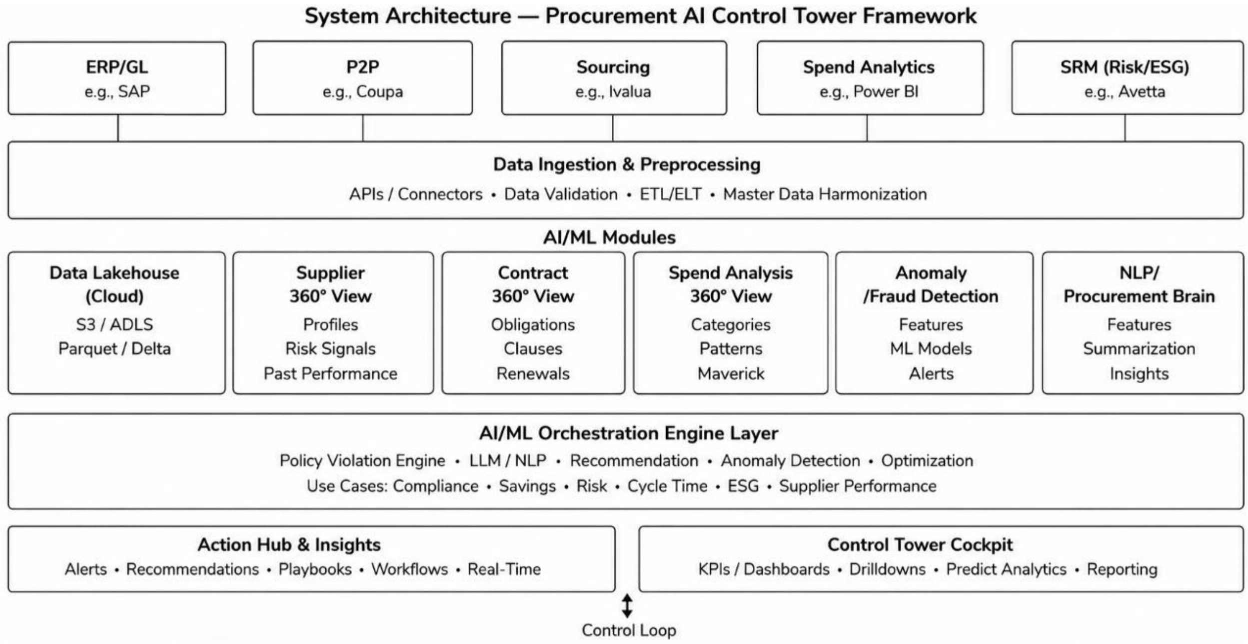

Figure 2: System architecture of the procurement AI control tower framework

3.1. Data Acquisition and Pre-Processing Module

The initial stage of the proposed Hybrid AI-based Procurement Risk Control Tower focuses on acquiring, integrating, and pre-processing heterogeneous procurement-related datasets obtained from both internal enterprise systems and external supplier environments. Since procurement data are often distributed across multiple platforms with varying formats, scales, and temporal characteristics, an efficient preprocessing framework is essential to ensure analytical consistency and improve downstream risk-prediction performance. The proposed module transforms raw procurement records into structured and machine-readable feature representations suitable for intelligent decision-making and procurement risk analysis.

Input: Demand Net Shortage dataset $\boldsymbol{D_{dns}}$, Inter-site Inventory dataset $\boldsymbol{D_{inv}}$, External Supplier dataset $\boldsymbol{D_{sup}}$, IPDR Cost dataset $\boldsymbol{D_{cost}}$, OPOR Purchase Order dataset $\boldsymbol{D_{po}}$.

Output: Optimised procurement feature matrix $\boldsymbol{F_{final}}$.

Step 1: Procurement Data Collection

Initially, procurement information is gathered from multiple heterogeneous sources, as shown in equation 1:

$$\boldsymbol{R} = \{\boldsymbol{D_{dns}}, \boldsymbol{D_{inv}}, \boldsymbol{D_{sup}}, \boldsymbol{D_{cost}}, \boldsymbol{D_{po}}\} \quad (1)$$

Where, $\boldsymbol{D_{dns}}$ represents shortage and demand imbalance records, $\boldsymbol{D_{inv}}$ denotes internal inventory movement data, $\boldsymbol{D_{sup}}$ contains open-market supplier information, $\boldsymbol{D_{cost}}$ includes procurement financial exposure records, and $\boldsymbol{D_{po}}$ represents purchase order transaction histories. This integration process enables the framework to capture both operational and financial procurement indicators across the supply chain ecosystem.

Step 2: Missing Value Identification and Data Cleaning

Procurement datasets frequently contain incomplete records, duplicate entries, inconsistent formats, and noisy attributes. Therefore, a data cleaning function is applied to each dataset, as shown in equation 2:

$$\boldsymbol{D_i^{clean}} = \boldsymbol{Clean}(\boldsymbol{D_i}) \quad (2)$$

Where $\boldsymbol{D_i^{clean}}$ denotes the cleaned procurement dataset. This stage removes redundant values, corrects inconsistent procurement entries, handles missing attributes, and filters corrupted records to improve dataset reliability.

Step 3: Temporal Synchronisation and Date Standardisation

Since procurement activities are recorded across different systems with varying timestamp formats, temporal synchronisation is necessary. All procurement records are converted into a unified time representation, as shown in equation 3:

$$\boldsymbol{T_i = ConvertTime(D_i^{clean})} \quad (3)$$

This process ensures chronological consistency among procurement transactions, supplier updates, inventory movements, and shortage events.

Step 4: Data Normalisation and Standardisation

To eliminate scale imbalance among procurement attributes, normalisation and standardisation are applied. The normalised feature value is computed, as shown in equation 4:

$$\boldsymbol{x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}} \quad (4)$$

Where $\boldsymbol{x_{min}}$ and $\boldsymbol{x_{max}}$ represent the minimum and maximum attribute values. Similarly, feature standardisation is performed, as shown in equation 5:

$$\boldsymbol{x_{std} = \frac{x - \mu}{\sigma}} \quad (5)$$

Where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ denote the mean and standard deviation of the procurement feature distribution. These operations improve feature comparability and stabilise downstream AI model training.

Step 5: Unified Procurement Data Integration

After preprocessing, all procurement datasets are merged into a unified analytical repository, as shown in equation 6:

$$\boldsymbol{M = Merge(T_{dns}, T_{inv}, T_{sup}, T_{cost}, T_{po})} \quad (6)$$

This integration creates a consolidated procurement intelligence matrix that can represent operational shortages, supplier behaviours, inventory availability, and financial exposure simultaneously.

Step 6: Procurement Feature Engineering

The unified dataset is transformed into an optimised procurement feature space, as shown in equation 7:

$$\boldsymbol{F_{proc} = FeatureExtraction(M)} \quad (7)$$

The extracted features include shortage frequency, supplier reliability score, procurement lead time, inventory variance, purchasing consistency, cost fluctuation index, and financial exposure indicators.

Step 7: Feature Optimisation and Validation

To improve analytical efficiency and reduce redundancy, irrelevant and correlated procurement attributes are removed, as shown in equation 8:

$$\boldsymbol{F_{final} = Optimize(F_{proc})} \quad (8)$$

This stage enhances feature relevance and improves the accuracy of downstream procurement risk prediction.

Step 8: Final Processed Procurement Dataset

The optimised feature matrix $\boldsymbol{F_{final}}$ is stored for subsequent AI-driven procurement risk assessment, fuzzy inference analysis, and intelligent supply chain decision-making processes (Figure 3).

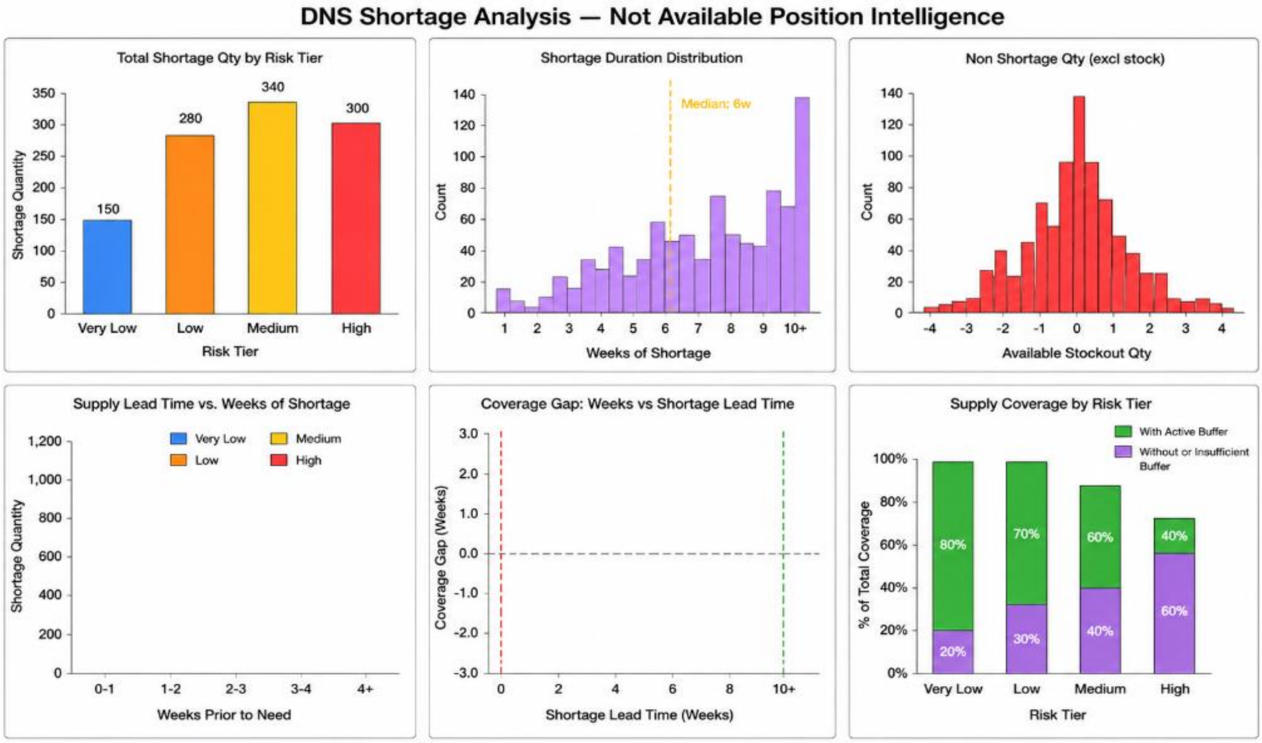

Figure 3: Comprehensive DNS shortage analysis dashboard

Overall, the proposed procurement data acquisition and preprocessing framework provides a scalable and robust foundation for intelligent procurement analytics by transforming heterogeneous raw procurement records into structured, synchronised, and high-quality analytical representations.

3.2. DNS Shortage Analysis Engine

The DNS Shortage Analysis Engine is designed to detect and estimate supply shortages. It examines the differences between demand and supply to identify key procurement gaps. The procurement risk analysis system identifies the occurrence and timing of shortages within procurement and inventory operations. It calculates the shortage quantity and estimates the duration of the shortage gap in weeks to evaluate the impact on supply chain activities. Based on these factors, the system performs a preliminary assessment of procurement risk levels to support proactive procurement planning and risk management. Based on the timing of shortages, the system divides risk into the following categories: The procurement risk analysis system classifies shortages into different risk categories based on duration and severity. Critical risk indicates shortages lasting between 1–5 weeks that may seriously affect procurement and supply chain operations, while Caution represents medium-level shortages requiring managerial attention. Monitor refers to low-risk or manageable shortages that can be controlled through regular procurement monitoring and inventory management practices. This module provides early insight into potential supply disruptions, enabling proactive planning.

3.2.1. Fuzzy Inference Model Analysis

The fuzzy-logic-based inference engine is designed to effectively characterise uncertainty and classify procurement risks into meaningful operational categories. Initially, procurement-related parameters such as supplier reliability, delivery delays, inventory fluctuations, transaction inconsistencies, demand variations, and financial instability are collected from structured and unstructured data sources. Let the input feature vector be represented as shown in equation 9:

$$\boldsymbol{I} = \{\boldsymbol{i_1}, \boldsymbol{i_2}, \boldsymbol{i_3}, \dots, \boldsymbol{i_n}\} \tag{9}$$

Where $\boldsymbol{i_n}$ denotes the $\boldsymbol{n^{th}}$ procurement risk attribute. Each input parameter is fuzzified into linguistic variables using predefined membership functions: Low (L), Medium (M), and High (H). The fuzzified membership value is expressed, as shown in equation 10:

$$\boldsymbol{\mu_A(i_n)}: \boldsymbol{I} \rightarrow [\boldsymbol{0}, \boldsymbol{1}] \tag{10}$$

Where $\mu_A(i_n)$ represents the degree of membership of the input i_n to fuzzy set A. Subsequently, the fuzzy inference engine applies a rule-based reasoning mechanism using IF–THEN rules to determine the procurement risk severity. A generalised fuzzy rule is formulated, as shown in equation 11:

$$R_k\text{: IF } (i_1 \text{ is } A_1) \wedge (i_2 \text{ is } A_2) \wedge \cdots \wedge (i_n \text{ is } A_n) \tag{11}$$

Then:

$$\text{Risk Level } = C_k \tag{12}$$

Where R_k denotes the k^{th} fuzzy rule and C_k represents the corresponding procurement risk category. The aggregated inference output is computed using fuzzy operators, as shown in equation 13:

$$\alpha_k = \min\left(\mu_{A_1}(i_1), \mu_{A_2}(i_2), \ldots, \mu_{A_n}(i_n)\right) \tag{13}$$

Where α_k indicates the rule's firing strength. Finally, the defuzzification process converts the aggregated fuzzy outputs into a crisp procurement risk score using the centroid method, as shown in equation 14:

$$R_{score} = \frac{\sum_{k=1}^{m} \alpha_k c_k}{\sum_{k=1}^{m} \alpha_k} \tag{14}$$

Where R_{score} denotes the final procurement risk score and c_k represents the centroid value of the corresponding risk class. Based on the obtained score, procurement risks are categorised into Low Risk, Moderate Risk, High Risk, and Critical Risk classes, enabling proactive, intelligent procurement decision-making.

3.3. Supply Analysis Engine

The Supply Analysis Engine compares the alternative sourcing options in two sub- modules:

3.3.1. Intersite Supply Module

This module ensures that there is excess stock in the internal locations. It enables the procurement management system to support the transfer of excess stock between departments or supply chain units, reducing inventory shortages and improving resource utilisation (Figure 4).

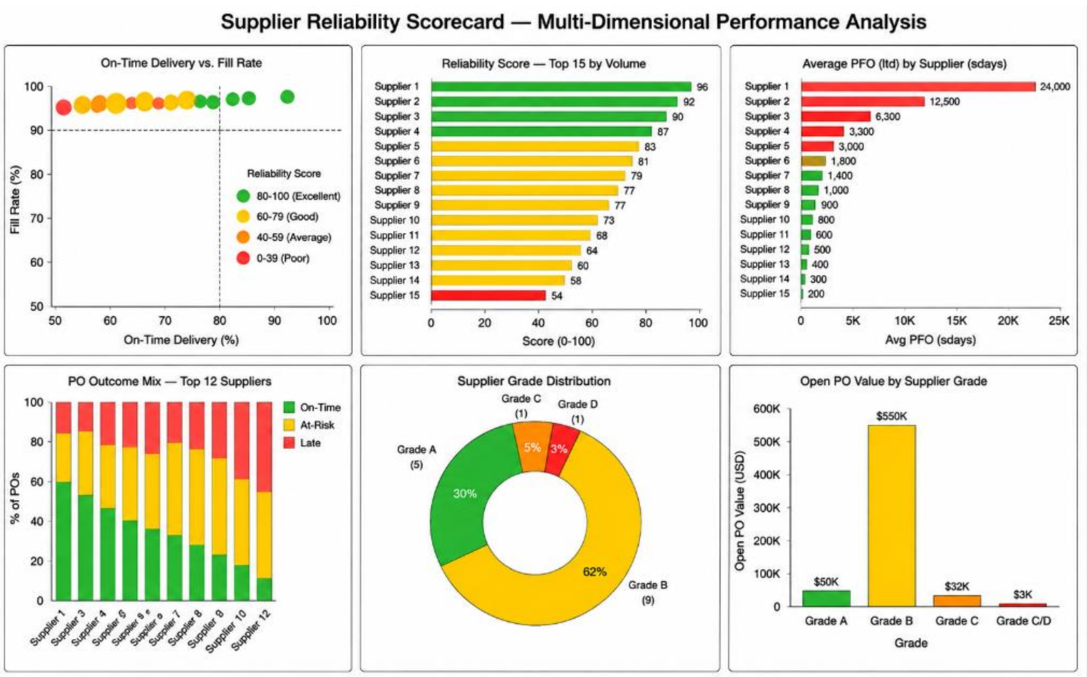

Figure 4: Supplier scorecard – multi-dimensional performance analysis

It also reduces dependence on external suppliers by effectively managing available internal inventory resources and procurement materials. Furthermore, the system optimises procurement and operational costs through internal sourcing strategies, thereby improving supply chain efficiency and supporting better procurement risk management.

3.3.2. Open Market Supply Module

This module examines the external supplier information to determine the procurement alternatives. The procurement risk management system extracts supplier availability information and pricing details to support efficient procurement planning and supplier selection (Figure 5).

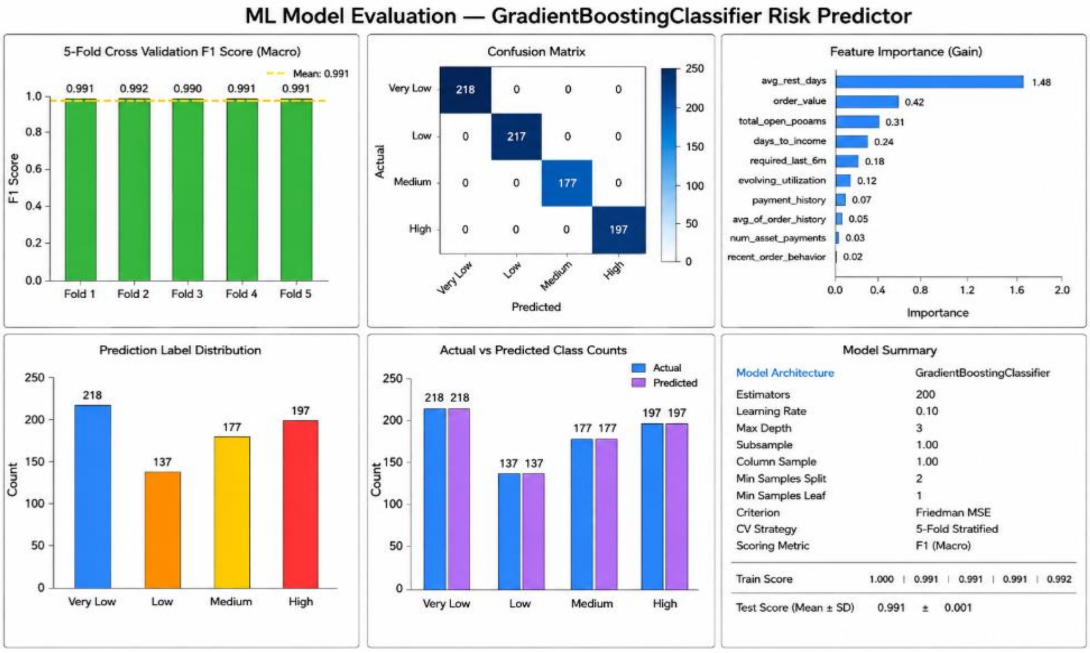

Figure 5: ML model evaluation – gradient boosting classifier risk predictor

It also evaluates the feasibility of procurement activities by analysing supplier capacity, inventory availability, and procurement requirements (Figure 6).

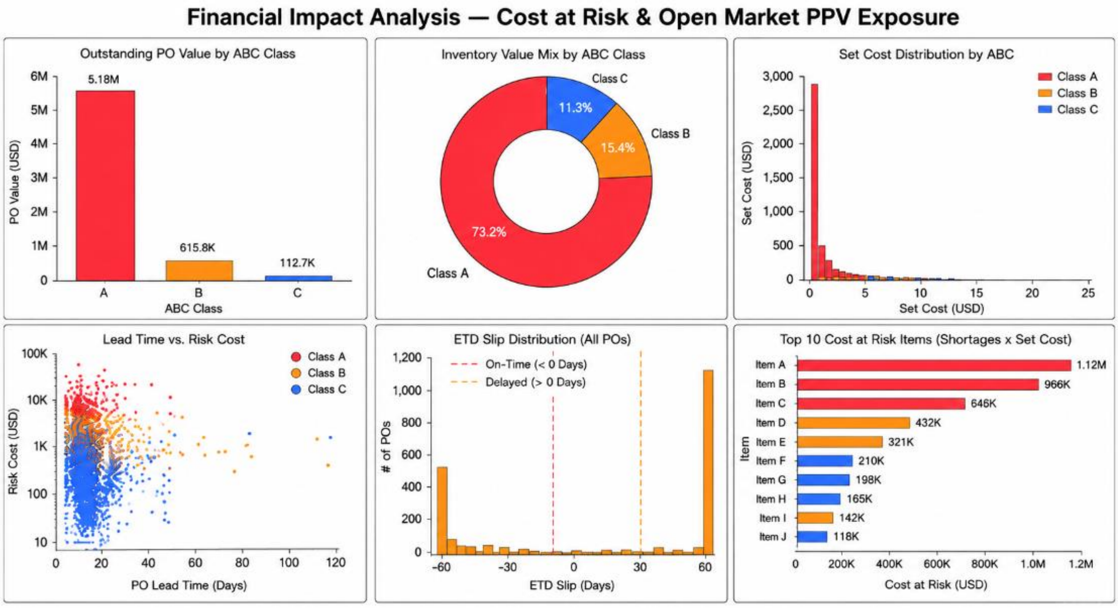

Figure 6: Financial impact – analysis – cost at risk and open market PPV exposure

In addition, the system calculates Purchase Price Variance (PPV) to measure the cost impact of procurement decisions and identify potential financial risks within procurement operations. The combination of internal and external supply analyses provides a comprehensive view of sourcing opportunities.

3.4. Fuzzy Procurement Decision Engine

A fuzzy-logic-based decision engine is employed to address uncertainty and complex decision-making situations. This module combines various inputs to produce smart procurement decisions.

3.4.1. Input Parameters

The procurement management system analyses intersite availability to identify excess inventory and material shortages across organisational locations. It also evaluates open-market availability by monitoring external supplier resources and procurement opportunities to support effective procurement planning. In addition, the system considers Purchase Price Variance (PPV) and current stock levels to assess procurement cost impact, inventory status, and overall procurement risk within supply chain operations.

3.4.2. Decision Rules

The procurement risk analysis system evaluates supply availability, Purchase Price Variance (PPV), and intersite inventory conditions to determine procurement risk levels and procurement strategies. When no supply availability is detected, the system classifies the situation as Extreme Risk, while high PPV values generate Broker Buy recommendations due to increased procurement costs. Intersite availability indicates a recoverable procurement condition, whereas low PPV values support Spot Buy recommendations for cost-effective procurement decisions. Based on these analyses, the system produces final outputs including procurement risk categories (R1–R4) and appropriate procurement decision recommendations for effective supply chain risk management. This module enables decision-making in uncertain situations by transforming complex conditions into actionable knowledge.

3.5. NLP-Based Buyer Risk Engine

The NLP module analyses qualitative buyer comments to obtain qualitative risk information. The implementation (as shown in the Python model) employs sophisticated text embedding methods. This module enhances the system's intelligence by incorporating human insights and contextual signals often ignored in traditional systems (Figure 7).

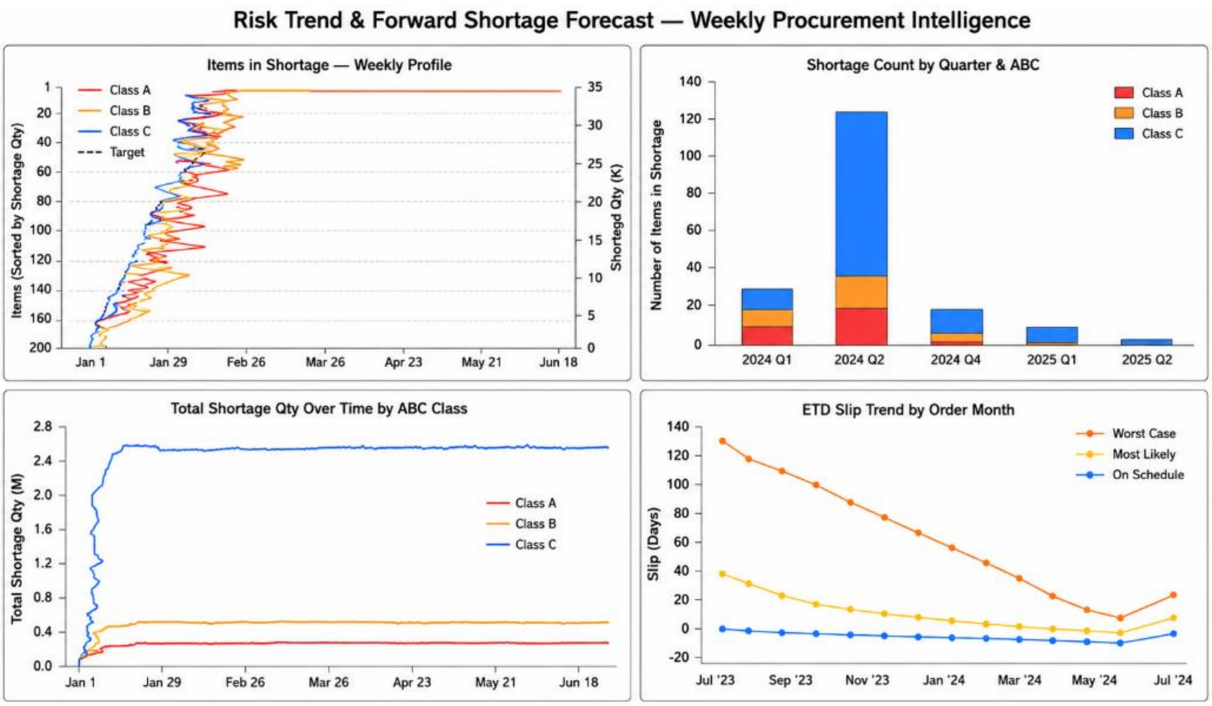

Figure 7: Risk trends and forward shortage forecast – weekly procurement intelligence

3.6. Integrated Procurement Risk Engine

The last module combines the output of all the other modules to produce a single procurement risk assessment.

3.6.1. Inputs Combined

The procurement risk analysis system identifies DNS-based shortage risks and evaluates supply availability within procurement operations. It also generates fuzzy decision outputs to support intelligent risk assessment and decision-making in procurement. Buyer risk insights based on NLP.

3.6.2. Final Outputs

The procurement management system generates combined procurement risk groups, such as R1–R4, based on procurement and supply chain risk analyses. It also provides procurement decision recommendations, including Broker Buy, Spot Buy, and Recoverable actions to support effective procurement risk management. This combined engine serves as a central decision-making layer, providing end-to-end procurement risk intelligence and enabling real-time, data-driven action.

3.7. Term Frequency-Inverse Document Frequency (TF-IDF)

The TF–IDF algorithm is a commonly used text-mining and information-retrieval method for determining the significance of words in a set of procurement documents. In procurement risk analysis, unstructured textual data is captured in large volumes, including supplier contracts, invoices, purchase orders, tender documents, compliance statements, logistics records, and audit statements. It's hard, time-consuming, and prone to human error to review these documents and identify procurement risks manually. TF–IDF provides a smart way to derive relevant risk-related terms from these procurement datasets by measuring the frequency of a term within a given document and across the collection. The algorithm also reduces the impact of frequently used terms and enhances the significance of the others, thereby better identifying key risks in the procurement process. Frequently, in procurement systems, there are terms associated with the risk, such as non-compliance, inventory shortage, quality issue, supplier failure, contract breach, delay, penalty, or fraud, that show up in procurement reports and in the systems' transactional records.

TF–IDF can automatically identify such words by assigning them higher values when they are specific to documents related to a given risk area. This capability enables supply chain analysts and organisations to detect irregularities in supplier patterns (e.g., delivery delays), financial mismatches, contract violations, operational issues, and other anomalies early. The extracted features using TF–IDF can also be incorporated into machine learning and Deep Learning (DL) models to predict procurement risks, classify suppliers, detect fraud, and support procurement decision-making systems. The TF value used in the algorithm indicates how many times a specific term appears in a procurement document. When a risk-related term is mentioned more than once in a supplier report or contract document, the algorithm assumes it contains substantial information about procurement activities. In other cases, certain terms may be used repeatedly across nearly every procurement document, such as “purchase,” “supplier,” or “invoice.” These common terms may not provide valuable information for risk identification.

The IDF component assigns lower weights to frequently occurring words and higher weights to less frequently occurring risk-sensitive words, making them appear even less frequently in the documents. This means that the IDF component will give lower weight to the more common words and higher weight to the less common but more informative risk-sensitive words, so that they occur even less frequently in the documents. A TF equation is used to determine the frequency of a procurement risk term in a specific procurement document. The higher the TF, the more significant the term is in the procurement report or supplier document. The Term Frequency equation (equation 15) is used to calculate the frequency of occurrence of a particular procurement-related term in a procurement document. Let’s assume $TF(t,d)\rightarrow$ Term Frequency, $t\rightarrow$ Procurement-related term, $d -$ Procurement document, $f_{t,d} -$ Number of times the term appears in the document, $\sum_k f_{k,d} -$Total number of terms:

$$TF(t,d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \tag{15}$$

In high-risk procurement reports, some terms may be repeated in procurement risk analysis, such as delay, fraud, penalty, and supplier failure. The TF equation identifies these common terms and assigns higher significance scores to them for detecting procurement risk. The IDF equation assigns greater weight to rare terms that provide greater risk information and less weight to common terms used in many documents. Common terms in many procurement documents, such as purchase, supplier, and invoice, result in a lower IDF score since they appear in many documents. The IDF values for rare terms such as corruption, contract violation, or blacklisted vendor are higher because they strongly suggest procurement risks. The Inverse Document

Frequency equation is used to calculate the uniqueness and importance of procurement terms in all procurement documents, as demonstrated in equation 16. Let's assume $IDF(t)$ −Inverse Document Frequency of term, N −Total number of procurement documents, df_t −Number of procurement documents containing term, log − Logarithmic scaling function:

$$IDF(t) = \log\left(\frac{N}{df_t}\right) \tag{16}$$

The TF–IDF equation is used to calculate the overall importance of procurement-related terms, as shown in equation 17. The TF–IDF method is used to obtain the final importance score of a procurement term in a procurement document by multiplying the term's local frequency in the document and its global uniqueness.

The higher the TF–IDF score, the more significant a procurement-related term is and the degree to which it is related to the procurement risks. These weighted scores can be used for procurement fraud detection, supplier risk classification, anomaly detection, procurement intelligence, and automated procurement monitoring systems. Let's assume $TF\text{-}IDF(t,d)$ −Final weight score of term, $TF(t,d)$ −Term Frequency value, $IDF(t)$ −Inverse Document Frequency value:

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \tag{17}$$

Compute the TF–IDF-based procurement feature vector, as shown in equation 18. The feature vector is a numerical representation of the weight of terms extracted from a procurement document that are relevant to procurement. Let's assume V_d −Procurement document feature vector, w −TF–IDF weights of procurement terms, n −Total number of procurement terms:

$$V_d = [w_1, w_2, w_3, \dots, w_n] \tag{18}$$

The TF–IDF feature vector converts the procurement text documents into a numerical representation that can be fed into DL models. These vectors are employed in procurement risk prediction, procurement risk evaluation, procurement fraud risk identification, and supply chain risk assessment and management.

3.8. Logistic Regression Classifier (LRC)

This section proposes collecting procurement data from various procurement management sources (supplier databases, contract management systems, logistics records, invoice transactions, inventory systems, and procurement monitoring platforms) as the first stage of the procurement risk analysis. The collected procurement data is then preprocessed to eliminate irrelevant, duplicate, inconsistent, and missing records. The procurement variables are transformed into a structured numerical representation using data normalisation and feature encoding techniques for machine learning analysis. This pre-processing step ensures the quality of the procurement data and the accuracy of the procurement risk predictions. The Logistic Regression Classifier categorises procurement activities based on the calculated probability into predefined risk classes: low, medium, and high.

Once the predicted probability exceeds a threshold, the transaction is considered a risky procurement activity, and the manager must address the risk and implement measures to mitigate it. If not, the procurement process is deemed to be safe and operationally stable. This classification process helps the procurement manager and the organisation identify procurement risks early and take proactive steps to control them. The Logistic Regression Classifier offers several benefits for procurement risk analysis and procurement management systems. Computationally efficient, interpretable, and can generate probability-based procurement risk predictions. The model can be used for procurement fraud detection, supplier rate analysis, procurement compliance monitoring, inventory risk analysis, and prediction of supply chain disruptions.

Additionally, Logistic Regression enhances the accuracy of procurement decisions by uncovering patterns of procurement risk in procurement datasets. The Logistic Regression Classifier is so reliable and effective that it is commonly used in various intelligent procurement management systems and automated procurement risk evaluation systems. The linear combination equation is an equation used to calculate the weighted sum of procurement-related features to predict procurement risks. As per equation 19, it estimates procurement risk levels by evaluating procurement variables, including supplier delay frequency, procurement cost fluctuations, delivery performance, and financial risk indicators. Let's assume $Z\ is\ the\ l$inear prediction score, β_0 −Bias or intercept term, $\beta_1, \beta_2, \dots, \beta_n$ −Weight coefficients of procurement features, x − Procurement feature values, n −Total number of procurement features:

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{19}$$

The linear score is transformed into a probability value between 0 and 1 using the sigmoid function. A high probability value represents high procurement risk, whereas a low probability value represents low procurement risk and supplier risk. As illustrated in Equation 20, the sigmoid function converts the weighted procurement feature score into a probability of the procurement risk. Let's assume $P(Y = 1)$ −Probability of procurement risk occurrence, e −Exponential mathematical constant, Z −Linear weighted procurement feature score:

$$P(Y = 1) = \frac{1}{1+e^{-Z}} \tag{20}$$

The computation of the final procurement risk category using a classification threshold value, as shown in equation 21. The classification equation is derived to classify a procurement transaction as a risk or non-risk transaction based on the probability values it predicts. Let's assume $\hat{Y}$ −Predicted procurement risk class, $P(Y = 1)$ −Predicted procurement risk probability:

$$\hat{Y} = \begin{cases} 1, & P(Y = 1) \geq 0.5 \\ 0, & P(Y = 1) < 0.5 \end{cases} \tag{21}$$

The Loss function is used in the Logistic Regression model to reduce classification error, as shown in equation 22. The loss function measures the difference between the procurement risk values and the predicted procurement risk values. Let's assume L −Logistic loss value, N −Total number of procurement samples, y_i −Actual procurement risk label, $\hat{y}_i$ −Predicted procurement risk probability, log − Logarithmic function, i − Procurement sample index:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{22}$$

The equation for the odds ratio quantifies the relationship between procurement risk and procurement features. The odds equation is essentially the ratio of the probability of a procurement risk occurring to the probability of it not occurring, as in equation 23. Let's assume O − Procurement risk odds ratio:

$$\text{O} = \frac{P(Y=1)}{1-P(Y=1)} \tag{23}$$

The logit model assumes a linear relationship between the probability of procurement risk and a linear predictor. Calculate the linear relationship between procurement risk probability and procurement feature variables as given in equation 24. To identify the impact of procurement variables like supplier reliability, delayed delivery rate, procurement cost deviation, and inventory shortage on procurement risk prediction. Let's assume $P(Y = 1)$ −Probability of procurement risk occurrence, β_0 − Bias term, Procurement feature coefficients, x − Procurement feature variables:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{24}$$

The LRC method is the classifier used to predict procurement risks based on supplier performance, procurement cost variation, delayed delivery, inventory shortage, fraud indicators, and contract compliance. The model estimates the probability of procurement risk and categorises procurement activities as risky or non-risky.

3.9. System Workflow Overview

3.9.1. The System has a Chronological Workflow

The proposed procurement risk analysis system performs data ingestion and preprocessing to collect and organise procurement-related information for accurate analysis. It detects shortages, estimates procurement risks, and analyses both internal and external supply availability to support procurement planning. The system further leverages fuzzy logic and NLP techniques for intelligent decision-making, qualitative risk assessment, and final procurement risk aggregation, yielding appropriate procurement decision outputs.

This architecture provides a comprehensive, scalable, and flexible procurement risk management platform, which can process both quantitative and qualitative data to enhance supply chain resiliency. Figure 8 represents the open-market information overview, PPV exposure and vendor stock analysis. In this picture, researchers compare total stock and total PPV exposure across the different providers. There are huge differences across the suppliers. Price premium distribution and SKU classification according to premium thresholds. The graph of stock availability shows how stock availability relates to the price premium. The supply-chain risk and lifecycle risk maps further break down the OMIs pools by their different risk ratings.

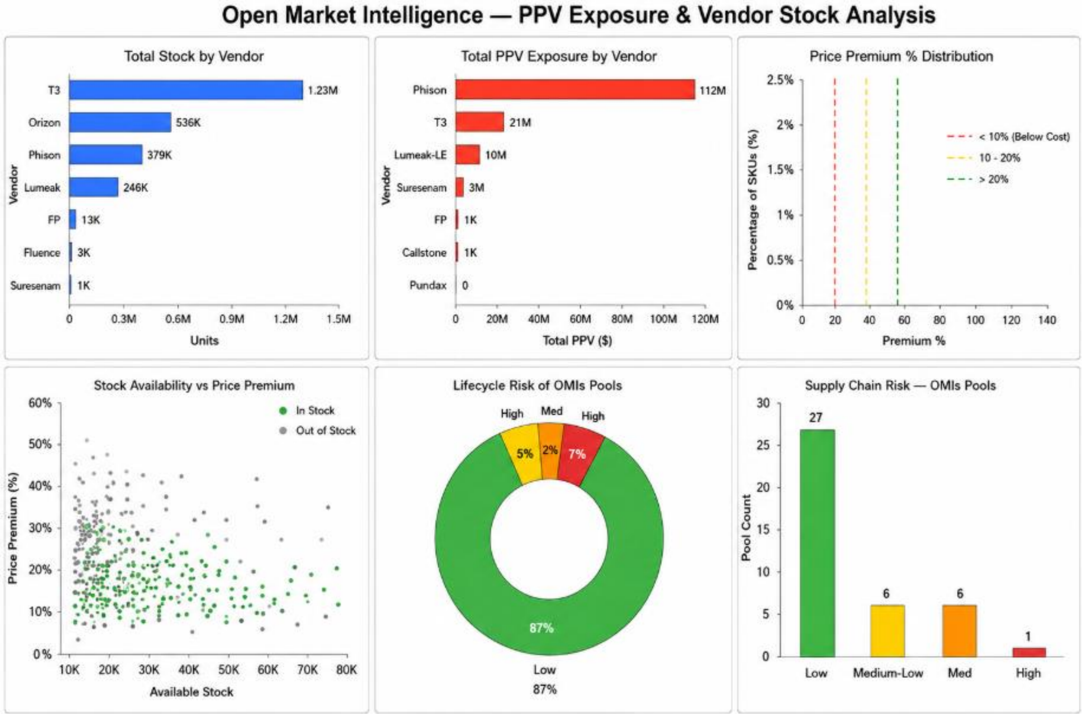

Figure 8: Open market intelligence – PPV exposure and vendor stock analysis

4. Conclusion

In conclusion, the proposed Hybrid AI-based Procurement Risk Control Tower establishes an intelligent, adaptive framework to enhance procurement risk management and strengthen supply chain resilience in dynamic operational environments. By integrating structured enterprise datasets with unstructured textual information, the framework effectively captures multi-dimensional risk indicators that are often overlooked in conventional procurement monitoring systems. The incorporation of fuzzy logic enables efficient handling of uncertainty and ambiguity in decision-making. At the same time, the NLP-driven machine learning module facilitates accurate extraction of contextual insights from heterogeneous data sources. Furthermore, the confidence-driven switching mechanism significantly improves decision reliability by dynamically selecting the most trustworthy analytical pathway, particularly during the initial deployment and low-confidence operational stages. Experimental evaluations demonstrate that the proposed framework achieves superior risk-prediction accuracy, faster response time, and greater operational efficiency than traditional risk assessment approaches.

In addition, the continuous retraining strategy allows the model to adapt to evolving procurement patterns, supplier behaviours, and market disruptions, thereby ensuring long-term robustness and scalability. The proposed system performs better than the other system, including in higher-risk analysis. The architecture also supports real-time monitoring and proactive mitigation strategies, enabling organisations to reduce procurement vulnerabilities and maintain continuity in complex supply chain ecosystems. Overall, the developed framework provides a scalable, data-driven, and intelligent procurement risk governance solution that supports sustainable supply chain operations and improves organisational decision-making. Future research can further enhance the framework by incorporating explainable AI mechanisms, blockchain-enabled transparency, and federated learning techniques to improve interpretability, security, and collaborative risk intelligence across distributed procurement networks.

Acknowledgement: The authors sincerely acknowledge SRM Institute of Science and Technology at Kattankulathur for providing the necessary academic support and facilities to conduct this research.

Data Availability Statement: Data relevant to this study will be provided by the corresponding author upon reasonable request, where permissible.

Funding Statement: This study was conducted without any external financial support.

Conflicts of Interest Statement: The authors report no conflicts of interest, and all sources used in this original work have been duly acknowledged.

Ethics and Consent Statement: The study adhered to ethical guidelines, and informed consent was obtained from all participants.

References

1. M. C. Nezianya, A. O. Adebayo, and P. Ezeliora, "A critical review of machine learning applications in supply chain risk management," *World Journal of Advanced Research and Reviews*, vol. 23, no. 3, pp. 1554–1567, 2024.
2. C. Li and H. Chen, "Machine learning based supply chain risk prediction and management," in *Proceedings of 5th International Conference on Computer Science and Management Technology*, Xiamen, Fujian, China, 2024.
3. C. Y. Chu, K. Park, and G. E. Kremer, "A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks," *Advanced Engineering Informatics*, vol. 45, no. 8, p. 101053, 2020.
4. S. Lu, "Enterprise supply chain risk assessment based on improved neural network algorithm and machine learning," *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and technology*, vol. 40, no. 4, pp. 7013–7024, 2021.
5. G. Yuan, S. Wu, and B. Wang, "Supply chain management model based on machine learning," *Neural Computing and Applications*, vol. 35, no. 6, pp. 4319–4335, 2023.
6. O. M. Filani, G. C. Nwokocha, and O. B. Alao, "Predictive vendor risk scoring model using machine learning to ensure supply chain continuity and operational resilience," *International Scientific Refereed Research Journal*, vol. 4, no. 4, p. 106–126, 2021.
7. C. Han and Q. Zhang, "Optimization of supply chain efficiency management based on machine learning and neural network," *Neural Computing and Applications*, vol. 33, no. 5, pp. 1419–1433, 2021.
8. H. Aboutorab, O. K. Hussain, M. Saberi, F. K. Hussain, and D. Prior, "Adaptive identification of supply chain disruptions through reinforcement learning," *Expert Systems with Applications*, vol. 248, no. 8, pp. 1–10, 2024.
9. M. Celestin, G. R. G. Raja, J. A. Mohamed, and A. M. Ibraheem, "Using probability theory to assess risks and mitigate uncertainty in procurement processes," *International Journal of Computational Research and Development*, vol. 10, no. 2, pp. 136–146, 2025.
10. L. W. Wong, G. W. H. Tan, K. B. Ooi, B. Lin, and Y. K. Dwivedi, "Artificial intelligence-driven risk management for enhancing supply chain agility: A deep-learning-based dual-stage PLS-SEM-ANN analysis," *International Journal of Production Research*, vol. 62, no. 15, pp. 5535–5555, 2024.
11. D. Aulia and W. Isvara, "Strategies to increase procurement maturity level using procurement maturity model to improve procurement performance," *International Journal of Scientific and Research Publications*, vol. 11, no. 6, pp. 489–501, 2021.
12. C. Flechsig, "The impact of intelligent process automation on purchasing and supply management–Initial insights from a multiple case study," in Logistics Management, *Springer*, Cham, Switzerland, 2021.
13. T. V. Hai, V. T. Linh, H. V. Quynh, D. T. V. Dung, N. V. Quang, N. T. Hung, and T. T. L. Anh, "Corporate social responsibility investment and its impact on the financial performance of securities companies," *International Journal of Economic Sciences*, vol. 14, no. 1, pp. 182–195, 2025.
14. L. Thammareddi, V. R. Anumolu, K. R. Kotte, B. C. C. Marella, K. A. Kumar, and J. Bisht, "Random Security Generators with Enhanced Cryptography for Cybersecurity in Financial Supply Chains," in *Proc. 2025 1st Int. Conf. Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, Bhimtal, Nainital, India, 2025.
15. G. C. Vegineni and B. C. C. Marella, "Integrating AI-Powered Dashboards in State Government Programs for Real-Time Decision Support," in AI-Enabled Sustainable Innovations in Education and Business, *IGI Global Scientific Publishing*, Hershey, Pennsylvania, United States of America, 2025.
16. H. Pandian, "AI-based capacity forecasting models for elastic cloud and hybrid enterprise systems," *Journal of Information Systems Engineering and Management*, vol. 10, no. 63s, pp. 1648–1658, 2025.
17. B. C. C. Marella and G. C. Vegineni, "Automated Eligibility and Enrollment Workflows: A Convergence of AI and Cybersecurity," in AI-Enabled Sustainable Innovations in Education and Business, *IGI Global Scientific Publishing*, Hershey, Pennsylvania, United States of America, 2025.
18. H. Pandian, "AI-driven predictive performance bottleneck detection in mission-critical financial systems," *Journal of Computational Analysis and Applications*, vol. 30, no. 2, pp. 1019–1033, 2022.