AI-Based Supply Chain Risk Prediction System Using Machine Learning

Dr. T. Amalraj Victoire[1], K. Deveeswar[2]
[1] Associate prof, Dept of MCA
[2] Dept of MCA
[1, 2] Sri ManakulaVinayagar Engineering College, Madagadipet, Puducherry – 605 107, India.

Abstract- *Supply chain systems are highly vulnerable to disruptions caused by factors such as transportation delays, traffic congestion, weather conditions, and inventory fluctuations. Traditional risk assessment methods are largely manual, time-consuming, and often fail to provide accurate and timely insights. This project presents an intelligent web-based system for predicting supply chain risk using machine learning techniques.*

The proposed system utilizes a Random Forest classifier to analyze key operational parameters, including delay, traffic, weather, inventory, order value, and port delay. Based on these inputs, the system classifies supply chain risk into three categories: Low, Medium, and High. The application is developed using the Flask framework and integrates a user-friendly interface for manual data entry, CSV-based batch prediction, and real-time analytics through a dashboard.

The system also includes data storage using SQLite and visualization features such as charts and export functionalities for Excel and image formats. Experimental results indicate moderate model performance, with an accuracy of approximately 61.9% and balanced accuracy of around 67%, highlighting the need for further optimization. Despite these limitations, the system demonstrates the practical application of machine learning in supply chain risk prediction and provides a foundation for future enhancements in predictive logistics systems.

Keywords: Supply Chain Risk, Machine Learning, Random Forest, Predictive Analytics, Flask, Data Visualization.

I. INTRODUCTION

Supply chain management is a critical component of modern business operations, ensuring the efficient movement of goods from suppliers to end customers. However, supply chains are inherently complex and are frequently affected by uncertainties such as transportation delays, traffic congestion, adverse weather conditions, fluctuating demand, and inventory shortages. These uncertainties introduce risks that can disrupt operations, increase costs, and reduce customer satisfaction.

Traditionally, supply chain risk assessment has been performed using manual analysis and rule-based decision-making. Such approaches are often subjective, time-consuming, and incapable of handling large volumes of data effectively. With the increasing availability of data and advancements in computational techniques, machine learning has emerged as a powerful tool for analyzing complex patterns and making accurate predictions.

This project focuses on the development of an AI-based supply chain risk prediction system that leverages machine learning algorithms to classify risk levels. The system uses key operational parameters such as delay, traffic, weather, inventory levels, order value, and port delays to predict whether a particular scenario falls under Low, Medium, or High risk. By automating the risk prediction process, the system aims to support better decision-making and proactive risk management.

The application is implemented as a web-based platform using the Flask framework, enabling users to input data, perform predictions, and visualize results through an interactive dashboard. In addition to real-time predictions, the system also supports batch processing through CSV uploads and provides export features for further analysis.

Although the current implementation demonstrates moderate predictive accuracy, it highlights the potential of integrating machine learning into supply chain management systems. The proposed solution serves as a foundation for developing more advanced, scalable, and accurate predictive models in the logistics domain.

II. LITERATURE SURVEY

Recent advancements in machine learning have significantly influenced supply chain risk prediction and decision-making processes. Traditional methods relied on statistical analysis and manual evaluation, which often failed to capture complex patterns. Researchers have increasingly adopted data-driven approaches to improve prediction accuracy and efficiency in supply chain operations.

Several studies have explored classification algorithms such as Logistic Regression and Decision Trees for predicting operational risks. These models provided basic insights but struggled with non-linear relationships and high-dimensional data. As a result, their performance was limited in dynamic and uncertain supply chain environments.

Ensemble learning methods, particularly Random Forest, have gained attention due to their ability to improve prediction accuracy and reduce overfitting. Research indicates that Random Forest performs well in handling multiple input variables and noisy datasets, making it suitable for supply chain applications. However, it may still face challenges in handling imbalanced data.

Support Vector Machines and K-Nearest Neighbors have also been used for classification tasks in logistics. While SVM offers strong performance in high-dimensional spaces, it requires careful parameter tuning. KNN, on the other hand, is simple but computationally expensive and sensitive to irrelevant features.

More recent studies have focused on advanced models such as XGBoost and Gradient Boosting techniques. These models have demonstrated superior performance in many predictive tasks due to their ability to handle complex relationships and optimize errors iteratively. However, they require more computational resources and expertise.

Overall, existing research highlights that machine learning can significantly enhance supply chain risk prediction. However, challenges such as data quality, feature selection, and model interpretability remain critical. Therefore, there is a need for systems that balance accuracy, efficiency, and practical implementation in real-world scenarios.

III. THEORETICAL FRAMEWORK

The proposed system is based on a data-driven predictive framework that utilizes machine learning techniques to analyze supply chain risk. Instead of relying on manual evaluation, the system processes historical and real-time data to identify patterns and relationships among critical operational factors. The core objective is to classify risk levels into Low, Medium, and High categories using structured input data.

The framework applies classification algorithms, particularly the Random Forest model, to handle multiple input features such as delay, traffic, weather, inventory, order value, and port delay. These features are selected based on their influence on supply chain disruptions. Data preprocessing techniques, including cleaning, encoding, and normalization, are used to improve model performance and ensure consistency in predictions.

The system follows a predictive analytics approach where trained models generate outputs based on learned patterns. The framework emphasizes accuracy, reliability, and scalability, enabling the system to support decision-making processes. It also provides a foundation for integrating more advanced models and improving prediction efficiency in future implementations.

IV. METHODOLOGY

The proposed system follows a structured methodology to predict supply chain risk using machine learning techniques. Initially, relevant data such as delay, traffic, weather conditions, inventory levels, order value, and port delay are collected in a structured format. Since raw data may contain inconsistencies, preprocessing steps such as handling missing values, encoding categorical variables, and normalization are performed to improve data quality and ensure reliable analysis.

After preprocessing, the dataset is divided into training and testing sets. Multiple machine learning models can be considered, but the Random Forest algorithm is selected due to its robustness and ability to handle multiple features effectively. The model is trained using the training dataset and evaluated on the testing dataset using performance metrics such as accuracy and balanced accuracy.

Finally, the trained model is integrated into a Flask-based web application. The system allows users to input data manually or upload CSV files for batch prediction. The predicted results are stored in a database and visualized through an interactive dashboard, enabling better analysis and decision-making.

Table: Comparison of Existing System vs Proposed System

Feature	Existing System	Proposed System
Core Function	Manual risk assessment	ML-based risk prediction
Evaluation Method	Subjective analysis	Data-driven prediction
Accuracy	Inconsistent and low	More consistent and improved
Data Usage	Limited	Multiple features

	parameters	considered
Processing Speed	Slow and manual	Fast and automated
Prediction Capability	No future prediction	Predicts risk levels
Model Usage	No ML models	Uses Random Forest
Visualization	Minimal or none	Dashboard with charts
Decision Support	Limited	Strong analytical support
Scalability	Difficult to scale	Scalable with data

V. EXISTING SYSTEM

In many organizations, supply chain risk assessment is primarily conducted using traditional methods that rely on manual observation, historical reports, and basic statistical analysis. Decision-making is often based on the experience and judgment of managers rather than data-driven insights. These methods typically involve reviewing past delays, supplier performance, and logistics issues without the support of automated predictive tools. As a result, the process is slow, reactive, and unable to respond effectively to dynamic changes in the supply chain environment.

One of the major limitations of the existing system is its inability to handle large volumes of data and identify complex relationships between multiple factors. Important variables such as traffic conditions, weather disruptions, and inventory fluctuations are often analyzed independently rather than collectively. This fragmented approach reduces the accuracy of risk assessment and fails to capture hidden patterns that could indicate potential disruptions. Additionally, the lack of real-time analysis makes it difficult to take proactive measures.

Another critical drawback is the absence of predictive capability. Traditional systems focus only on past and present data, without forecasting future risks. This leads to unexpected delays, increased operational costs, and reduced efficiency. Furthermore, the lack of visualization tools and integrated data platforms limits effective decision-making, highlighting the need for an advanced, automated system.

VI. PROPOSED SYSTEM

The proposed system introduces a machine learning-based approach to predict supply chain risk using a data-driven methodology. Unlike traditional methods, this system analyzes multiple operational factors such as delay, traffic, weather, inventory, order value, and port delay simultaneously. By leveraging a Random Forest classifier, the system identifies patterns in historical data and classifies risk levels into Low, Medium, and High categories. This enables organizations to move from reactive decision-making to proactive risk management.

The system is implemented as a web-based application using the Flask framework, providing an interactive interface for users to input data and obtain predictions in real time. It also supports batch processing through CSV uploads, allowing large datasets to be analyzed efficiently. All prediction results are stored in a database, enabling historical tracking and further analysis. The integration of visualization tools helps users understand trends and patterns clearly.

Additionally, the proposed system includes dashboard analytics and export features, making it practical for real-world usage. Although the current model shows moderate accuracy, it provides a scalable foundation for future improvements. The system enhances decision-making, reduces manual effort, and demonstrates the effective application of machine learning in supply chain risk prediction.

6.1 Proposed System Components:

A. Data Input Module:

This module allows users to enter supply chain parameters manually or upload CSV files. It ensures flexibility in handling both single and bulk data inputs.

B. Data Preprocessing Module:

The system cleans and prepares the input data by handling missing values, encoding categorical data, and standardizing inputs for accurate predictions.

C. Machine Learning Model Module:

This module uses the Random Forest algorithm to analyze input features and classify risk levels. It is responsible for training, testing, and generating predictions.

D. Database Management Module:

All prediction results and user data are stored in a structured SQLite database, enabling easy retrieval and management of historical records.

E. Visualization and Dashboard Module:

This component displays prediction results using charts and graphs, helping users analyze trends and patterns effectively.

F. Export Module:

The system provides options to export results in Excel format and download dashboard visuals as images for reporting purposes.

G. User Interface Module:

A simple and responsive web interface built with HTML, CSS, and JavaScript ensures smooth user interaction and accessibility.

VII. SYSTEM ARCHITECTURE

The system architecture is designed as a layered structure that integrates user interaction, data processing, machine learning prediction, and result visualization. At the input layer, users provide supply chain data either through manual entry forms or CSV file uploads. This data is passed to the backend, where preprocessing operations such as cleaning, encoding, and normalization are performed. The processed data is then forwarded to the machine learning model, which analyzes the input features and predicts the corresponding risk level.

In the processing layer, the Random Forest model acts as the core prediction engine, generating outputs based on learned patterns from historical data. The results are stored in a SQLite database for future reference and analysis. The output layer presents the predictions through an interactive dashboard, which includes charts and summary statistics. Additionally, export functionalities allow users to download reports in Excel and image formats. This architecture ensures smooth data flow, modular design, and efficient integration of all system components.

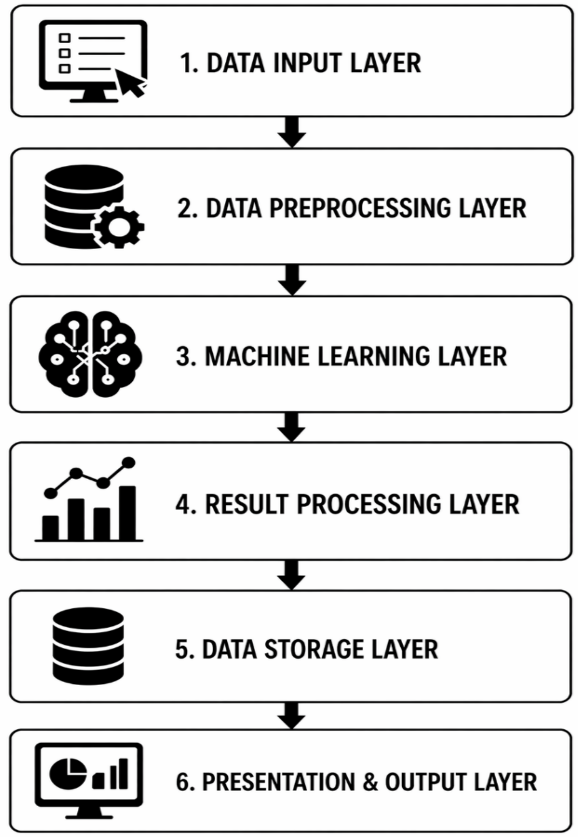

Fig 7.1: Layered Architecture of Supply Chain Risk Prediction System

VIII. IMPLEMENTATION

The implementation of the proposed system is carried out using a combination of web development and machine learning technologies. The frontend is developed using HTML, CSS, and JavaScript to create an interactive and user-friendly interface. The backend is built using the Flask framework in Python, which handles user requests, processes input data, and communicates with the machine learning model. The system supports both manual data entry and CSV file uploads, allowing flexibility in handling different types of input.

The machine learning component is implemented using Scikit-learn, along with Pandas and NumPy for data manipulation and preprocessing. A Random Forest classifier is trained on the dataset using selected features such as delay, traffic, weather, inventory, order value, and port delay. The trained model is then integrated into the Flask application to generate real-time predictions. The model outputs risk levels categorized as Low, Medium, or High based on input parameters.

For data storage, SQLite is used to maintain prediction records and user-related data. The system also includes a dashboard built using Chart.js to visualize predictions through graphs such as bar charts, pie charts, and line charts. Additional features like Excel export using OpenPyXL and dashboard image export enhance the usability of the system.

IX. EVALUATION METRICS

The performance of the proposed system is evaluated using standard classification metrics to ensure reliable prediction of supply chain risk levels. Accuracy is used as a primary metric to measure the overall correctness of the model by comparing predicted values with actual outcomes. However, since the dataset may contain class imbalance, balanced accuracy is also considered to provide a more realistic evaluation across all risk categories. These metrics help in understanding how well the model performs in classifying Low, Medium, and High risk levels.

In addition to accuracy, other important metrics such as precision, recall, and F1-score are used to evaluate the model's effectiveness for each class. Precision measures the correctness of positive predictions, while recall indicates the model's ability to identify actual positive cases. The F1-score provides a balance between precision and recall. Together, these metrics offer a comprehensive assessment of model performance and highlight areas where improvement is required.

X. RESULTS AND ANALYSIS

The results of the proposed system indicate moderate performance in predicting supply chain risk levels using the Random Forest model. The overall accuracy of the model is approximately 61.9%, with a balanced accuracy of around 67%, showing that the model performs reasonably across different classes but still lacks strong predictive power.

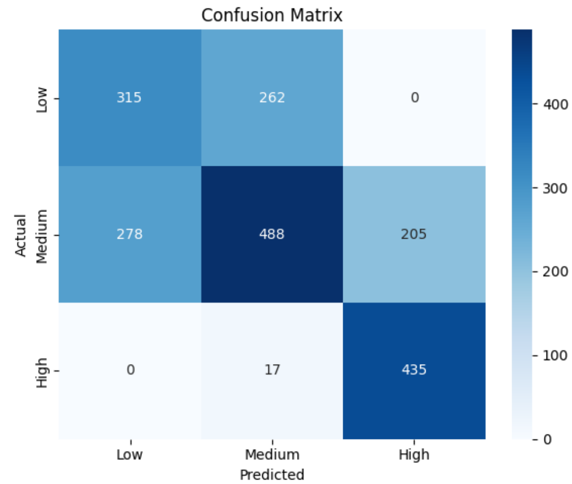

Fig 10.1: Confusion Matrix of Risk Prediction Model

The confusion matrix reveals that the model is more accurate in identifying High Risk scenarios, while it struggles to correctly classify Medium Risk cases, indicating class imbalance and overlapping feature patterns.

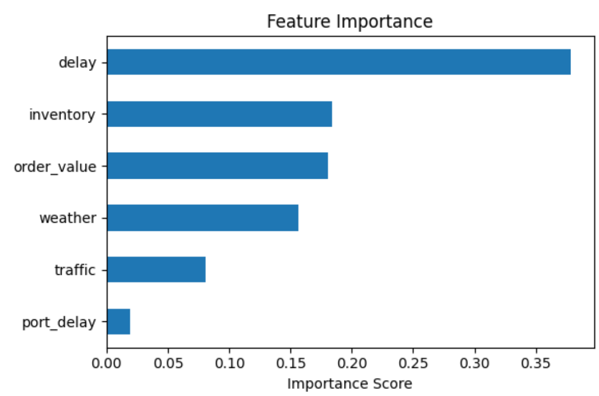

Fig 10.2: Feature Importance of Input Variables

Further analysis of feature importance shows that factors such as delay, traffic, and weather have a significant impact on prediction outcomes, whereas features like port delay contribute less to the model's decision-making. Precision and recall values vary across classes, with higher scores for extreme categories (Low and High) and lower performance for the Medium class. These results highlight the need for improved data quality, better feature engineering, and advanced models such as boosting techniques to enhance prediction accuracy and reliability.

XI. APPLICATIONS

The proposed supply chain risk prediction system can be applied in various logistics and business operations to improve decision-making and efficiency. Organizations can use the system to identify potential risks in transportation, inventory management, and supplier operations before they occur. By predicting risk levels in advance, companies can take preventive measures such as adjusting delivery schedules, optimizing routes, or maintaining safety stock levels. This helps in reducing delays, minimizing operational costs, and improving overall supply chain performance.

In addition, the system can be integrated into enterprise resource planning (ERP) and logistics management platforms to provide real-time analytics and automated insights. It is also useful for e-commerce companies, manufacturing industries, and distribution networks that rely heavily on timely deliveries. The visualization dashboard further supports managers in analyzing trends and making strategic decisions. Overall, the system enhances operational transparency and supports proactive risk management.

XII. ADVANTAGES AND LIMITATIONS

The proposed system offers several advantages by introducing automation and data-driven decision-making into supply chain risk management. It reduces dependency on manual analysis and improves prediction efficiency by using machine learning techniques. The system can process multiple factors simultaneously, enabling more accurate and consistent risk assessment compared to traditional methods. The integration of a user-friendly dashboard with visualization tools helps in better understanding of trends and supports quick decision-making. Additionally, features such as CSV upload, real-time prediction, and data export enhance usability and make the system practical for real-world applications.

However, the system also has certain limitations that affect its overall performance. The current model achieves only moderate accuracy, which limits its reliability in critical scenarios. It struggles particularly with medium-risk classification due to overlapping data patterns and possible class imbalance. The system also depends heavily on the quality and size of the dataset, and limited features may restrict prediction capability. Furthermore, the use of basic infrastructure such as SQLite and Flask may not be suitable for large-scale deployment without further optimization and scalability improvements.

XIII. ETHICAL AND PRACTICAL CONSIDERATIONS

The implementation of an AI-based supply chain risk prediction system raises several ethical and practical concerns that must be addressed. Data privacy is a primary issue, as the system may process sensitive operational or business information that requires secure storage and controlled access. There is also a risk of bias in predictions if the training data is incomplete or unrepresentative, which can lead to inaccurate or unfair decision-making. Over-reliance on automated predictions may reduce human oversight, potentially causing critical errors in high-risk situations. From a practical perspective, the system requires consistent data quality, proper maintenance, and periodic model updates to remain effective. Additionally, organizations must ensure transparency in how predictions are generated so that users can trust and interpret the system's outputs responsibly.

XIV. FUTURE ENHANCEMENTS

The system can be improved by increasing model accuracy through advanced algorithms such as XGBoost and Gradient Boosting. Incorporating techniques like hyperparameter tuning and handling class imbalance can enhance prediction performance. Expanding the dataset and including additional relevant features will also improve reliability.

Further enhancements include deploying the system on cloud platforms for scalability and integrating real-time data sources for dynamic predictions. Implementing secure authentication mechanisms and role-based access control can strengthen system security. Adding advanced analytics such as ROC curves and detailed reports will provide deeper insights for decision-making.

XV. CONCLUSION

The proposed system demonstrates the application of machine learning techniques in predicting supply chain risk using a data-driven approach. By integrating a Random Forest model with a web-based platform, the system provides an efficient solution for analyzing multiple operational factors and classifying risk levels. It reduces manual effort and supports proactive decision-making through real-time predictions and visualization tools.

However, the current system achieves only moderate accuracy, indicating the need for further improvement in data quality and model optimization. Despite this limitation, the project establishes a strong foundation for developing more advanced and scalable predictive systems. With future enhancements, the system can become a reliable tool for improving supply chain efficiency and risk management.

REFERENCES

[1] Leo Breiman (2001). Random Forests. Machine Learning Journal, 45(1), 5–32.

[2] Jerome H. Friedman (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics.
[3] IBM (2023). What is Supply Chain Risk Management? IBM Documentation.
[4] McKinsey & Company (2022). Risk, Resilience, and Rebalancing in Global Value Chains.
[5] Scikit-learn Documentation (2024). Machine Learning in Python.
[6] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016). Deep Learning. MIT Press.
[7] World Economic Forum (2021). Building Resilient Supply Chains.
[8] Christopher M. Bishop (2006). Pattern Recognition and Machine Learning. Springer.
[9] Amazon Web Services (2023). Machine Learning for Supply Chain Optimization.
[10] Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009). The Elements of Statistical Learning. Springer.