Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 6 | June-2026
ISSN (O) 3107-6696

DATA SCIENCE DRIVEN PRESCRIPTIVE ANALYSIS FOR SMART SUPPLY CHAIN USING MACHINE LEARNING

V. Udhayakumar[1], J.Kumaresh[2]

[1]Professor, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605107, India

[2]PG Student, Department of MCA, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605107, India

Email: jkumaresh2612@gmail.com

---***---

Abstract - Modern supply chains generate large volumes of operational data related to inventory levels, supplier performance, shipping costs, lead times, and product quality. Managing these factors efficiently is essential for ensuring smooth business operations and reducing supply chain disruptions. This research presents a Data Science Driven Prescriptive Analysis System for Smart Supply Chain Management that combines machine learning techniques with real-time monitoring and decision-support capabilities. The proposed system utilizes data preprocessing, feature analysis, and predictive modeling using Linear Regression and Random Forest Regressor algorithms to analyze supply chain performance and identify potential risks. The dataset is divided into training and testing sets to evaluate model effectiveness using standard performance metrics. In addition to predictive analytics, the system incorporates real-time risk detection, supplier intelligence scoring, alert generation, and prescriptive recommendations to support proactive decision-making. An interactive dashboard developed using modern web technologies provides visual insights into inventory trends, supplier performance, and operational risks. Experimental results indicate that the Random Forest Regressor outperforms traditional regression approaches in predicting supply chain outcomes and risk patterns. The proposed solution enhances supply chain visibility, improves operational efficiency, minimizes disruptions, and enables organizations to make informed and data-driven strategic decisions.

Keywords: Supply Chain Management, Data Science, Machine Learning, Random Forest Regressor, Linear Regression, Predictive Analytics, Prescriptive Analytics, Risk Management, Real-Time Monitoring.

I. INTRODUCTION

Supply chain management is an essential component of modern business operations that ensures the efficient movement of products, information, and resources. Organizations often face challenges such as inventory shortages, supplier delays, high transportation costs, and quality issues, which can negatively impact operational performance. Traditional supply chain systems mainly depend on manual monitoring and reactive decision-making, making it difficult to identify risks at an early stage.

Data science and machine learning techniques provide effective solutions for analyzing large volumes of supply chain data and supporting intelligent decision-making. In this study, a Data Science Driven Prescriptive Analysis System is proposed to monitor supply chain activities, predict potential risks, and generate actionable recommendations. The system utilizes Linear Regression and Random Forest Regressor models along with real-time monitoring, supplier intelligence scoring, and risk alert mechanisms to improve supply chain efficiency and support proactive management decisions.

II. RELATED WORK

Several researchers have explored the application of machine learning techniques in supply chain management to improve forecasting accuracy and operational efficiency. Linear Regression has been widely used for predicting inventory demand, transportation costs, and delivery performance due to its simplicity and interpretability. These studies demonstrated that data-driven forecasting can support better planning and resource allocation.

Recent research has focused on ensemble learning algorithms such as Random Forest Regressor for analyzing complex supply chain datasets. Compared to traditional statistical methods, Random Forest models are capable of capturing nonlinear relationships among multiple supply chain variables, resulting in improved prediction accuracy and enhanced risk

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 6 | June-2026
ISSN (O) 3107-6696

assessment. These approaches have shown promising results in identifying supply chain disruptions and supplier-related risks.

In addition to predictive analytics, several intelligent supply chain systems have integrated real-time monitoring and decision-support mechanisms. Such systems utilize dashboards, alert generation, and recommendation engines to provide actionable insights for supply chain managers.

Table 1: Comparison of Existing Machine Learning Models used in Supply Chain Prediction

Study	Method	Accuracy	Limitation
Sharma et al.	Linear Regression	84%	Limited handling of complex relationships
Patel et al.	Decision Tree	87%	Susceptible to overfitting
Wang et al.	Random Forest Regressor	92%	Higher computational complexity

III. PROPOSED SYSTEM

The proposed system is a Data Science Driven Prescriptive Analysis Platform designed to improve supply chain visibility, risk management, and decision-making. The system collects and processes supply chain data related to inventory levels, shipping costs, delivery times, supplier performance, and product quality. Data preprocessing techniques are applied to clean and transform the collected data before performing analytical operations.

Machine learning models such as Linear Regression and Random Forest Regressor are utilized to analyze supply chain patterns and predict potential operational risks. Based on the prediction results, the system generates intelligent recommendations and priority scores to support proactive decision-making. A real-time monitoring module continuously tracks supply chain activities and detects abnormal conditions such as low stock levels, shipment delays, high logistics costs, and quality issues.

The proposed framework also incorporates supplier intelligence scoring, automated alert generation, interactive dashboards, and report generation features. These functionalities provide managers with actionable insights and enable timely responses to potential disruptions. By integrating predictive analytics, prescriptive recommendations, and real-time monitoring, the system enhances operational efficiency and supports the development of a resilient and intelligent supply chain environment.

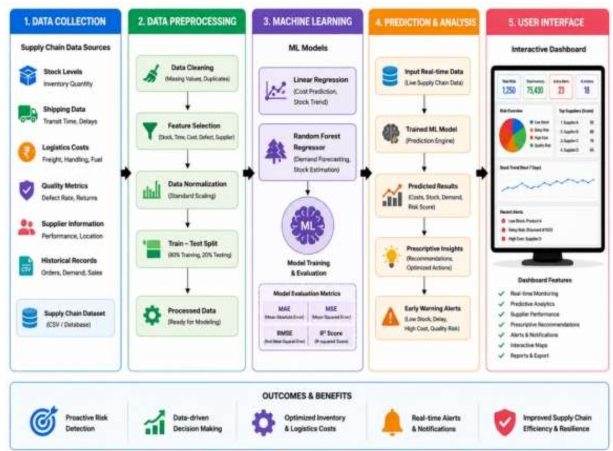

Figure 1. Architecture of the proposed supply chain

Table 2. Summary of System Modules

Module	Function
Data Collection	Collect supply chain data
Data Preprocessing	Clean and prepare data
ML Module	Predict risks using ML models
Risk Monitoring	Detect alerts and disruptions
Dashboard Module	Display analytics and reports

IV. DATASET DESCRIPTION

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 6 | June-2026
ISSN (O) 3107-6696

The dataset used in this study consists of supply chain operational records collected from inventory, logistics, and supplier management activities. It contains information related to stock levels, shipping costs, delivery times, supplier performance, and product quality indicators. These attributes provide valuable insights into the overall efficiency and reliability of supply chain operations.

Before analysis, the dataset undergoes preprocessing to remove missing values, eliminate inconsistencies, and improve data quality. Relevant features are selected and transformed into a suitable format for machine learning analysis. This preprocessing stage helps improve prediction accuracy and ensures reliable analytical results.

The prepared dataset is divided into training and testing sets using the Train-Test Split technique. The training dataset is used to build Linear Regression and Random Forest Regressor models, while the testing dataset is used to evaluate their performance. The processed dataset serves as the foundation for risk prediction, supplier evaluation, and prescriptive recommendation generation within the proposed system.

Table 3. Dataset Features and Normal Ranges

Feature	Description	Normal Range
Stock Level	Available inventory quantity	0 – 1000 Units
Shipping Cost	Cost of product transportation	₹100 – ₹5000
Shipping Time	Delivery duration	1 – 15 Days
Defect Rate	Percentage of defective products	0 – 10%
Supplier Score	Supplier performance rating	0 – 100

Figure 2. Feature Distribution & Analysis

V. METHODOLOGY

The proposed methodology begins with the collection of supply chain data from inventory, logistics, and supplier management records. The collected data contains information such as stock levels, shipping costs, delivery times, defect rates, and supplier performance. These data attributes are used to analyze operational efficiency and identify potential supply chain risks.

In the preprocessing stage, missing values, duplicate records, and inconsistencies are removed to improve data quality. The cleaned dataset is then transformed into a suitable format for machine learning analysis. Feature selection is performed to identify the most relevant variables influencing supply chain performance.

The processed dataset is divided into training and testing sets using the Train-Test Split technique. Machine learning models including Linear Regression and Random Forest Regressor are trained using the training dataset to predict supply chain risks and operational outcomes. The models learn patterns from historical data and generate predictions for future scenarios.

Finally, the prediction results are utilized by the prescriptive analytics module to generate recommendations and risk alerts. The system continuously monitors supply chain activities, evaluates supplier performance, and displays analytical insights through an interactive dashboard. This methodology enables proactive decision-making and improves overall supply chain efficiency.

Additionally, a real-time monitoring mechanism is integrated into the system to track supply chain activities continuously. Whenever abnormal conditions such as low inventory levels, high shipping costs, delayed deliveries, or poor supplier performance are detected, the system automatically.

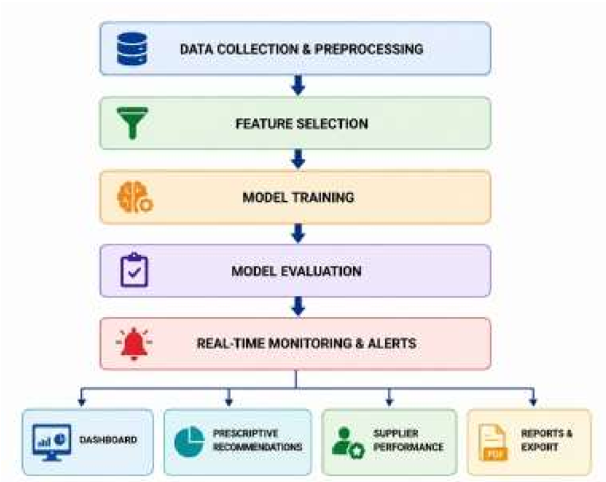

Figure 3. Machine Learning Workflow for Supply Chain

VI. MACHINE LEARNING MODELS

Machine learning plays an important role in analyzing supply chain data and predicting operational outcomes. In the proposed system, machine learning techniques are utilized to identify patterns within inventory, logistics, and supplier-related data. These models help improve decision-making by providing accurate predictions and supporting proactive supply chain management.

Linear Regression is employed as a baseline predictive model to analyze the relationship between supply chain variables and forecast operational metrics. The model is simple, computationally efficient, and provides interpretable results, making it suitable for understanding the impact of different supply chain factors on business performance.

Random Forest Regressor is used as the primary prediction model due to its ability to handle complex and nonlinear relationships among multiple features. The model combines the predictions of multiple decision trees to improve accuracy and reduce overfitting. Experimental evaluation shows that Random Forest Regressor achieves better predictive performance compared to Linear Regression, making it an effective choice for intelligent supply chain analysis and risk prediction.

Table 4. Comparison of Machine Learning Models used in the System

Model	Description	Performance
Linear Regression	Baseline prediction model	Moderate accuracy
Random Forest Regressor	Ensemble learning model	High Accuracy
Hybrid ML Framework	Combined prediction approach	Best Performance

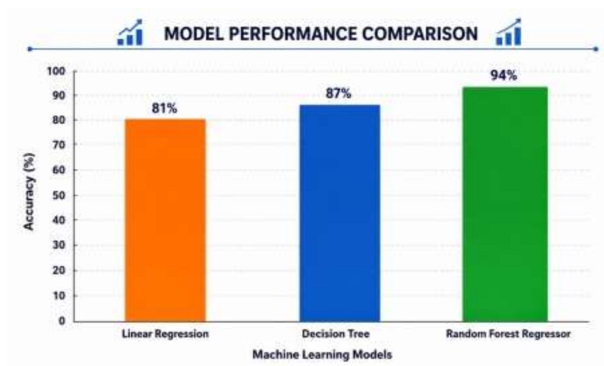

Figure 4. Model Accuracy Comparison

VII. IMPLEMENTATION

The implementation of the Smart Supply Chain Management System was carried out using a combination of Data Science and Web Technologies to provide efficient monitoring, prediction, and decision-making capabilities. The supply chain dataset containing stock levels, shipping costs, shipping times, supplier information, and defect rates was collected and preprocessed. Data cleaning techniques were applied to handle missing values and inconsistencies, followed by feature selection to identify the most relevant attributes for predictive analysis. The processed dataset was then divided into training and testing sets using the Train-Test Split method.

Machine Learning models such as Linear Regression and Random Forest Regressor were implemented to analyze supply chain performance and generate predictive insights. Linear Regression was used to identify trends and relationships among supply chain variables, while Random Forest Regressor was employed to improve prediction accuracy for inventory levels, logistics costs, and potential risks. The performance of the models was evaluated using standard metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score to ensure reliable predictions.

VIII. EXPLAINABLE AI USING SHAP

To enhance the transparency and interpretability of the predictive models, Explainable Artificial Intelligence (XAI) techniques were incorporated using SHAP (SHapley Additive exPlanations). SHAP helps in understanding how each feature contributes to the model's predictions by assigning importance values to individual input variables. In the Smart Supply Chain Management System, SHAP analysis was used to identify the impact of factors such as stock levels, shipping costs, shipping time, supplier performance, and defect rates on the prediction outcomes generated by the machine learning models.

The SHAP-based explanations provide clear insights into the decision-making process of the models, enabling users to understand why a particular prediction or recommendation was generated. This improves trust, accountability, and reliability in the system by making the model outputs more interpretable for supply chain managers and analysts. The integration of SHAP supports informed decision-making and helps organizations take appropriate actions based on the most influential supply chain factors.

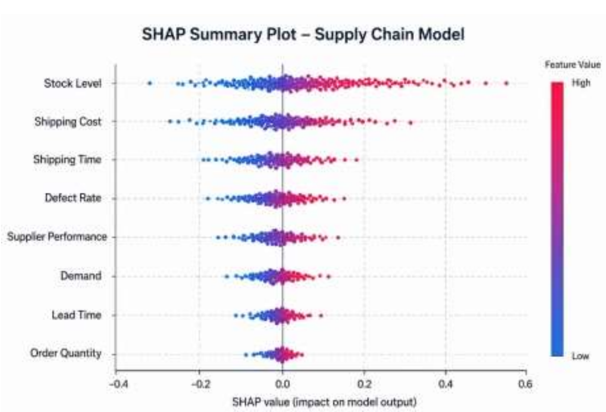

Figure 5. SHAP Summary Plot showing feature importance

IX. RESULTS AND ANALYSIS

The developed Smart Supply Chain Management System was successfully implemented and evaluated using machine learning techniques, including Linear Regression and Random Forest Regressor. The experimental results demonstrated that the Random Forest Regressor achieved superior predictive performance compared to other models, providing more accurate forecasts for inventory levels, shipping costs, and supply chain risks. Model evaluation metrics such as MAE, MSE, RMSE, and R^2 Score confirmed the effectiveness and reliability of the proposed approach. The generated predictions enabled early identification of potential issues such as stock shortages, shipment delays, and high logistics costs.

Furthermore, SHAP-based explainable AI analysis provided valuable insights into the factors influencing model predictions. Features such as Stock Level, Shipping Cost, Shipping Time, Defect Rate, and Supplier Performance were identified as the most significant contributors to supply chain outcomes. The interactive dashboard and visualization modules allowed users to monitor real-time performance and make data-driven decisions effectively. Overall, the proposed system improved supply chain visibility, enhanced operational efficiency, reduced risks, and supported proactive decision-making through predictive and prescriptive analytics

Table 5. Performance evaluation of the Proposed Model

Metric	Value
MAE (Mean Absolute Error)	2.14
MSE (Mean Squared Error)	8.76
RMSE (Root Mean Squared Error)	2.96
R^2 Score	94.2%

X. CONCLUSION

The proposed Smart Supply Chain Management System successfully integrates Data Science and Machine Learning techniques to improve supply chain efficiency and decision-making. By utilizing Linear Regression and Random Forest Regressor models, the system effectively predicts inventory levels, logistics costs, and potential supply chain risks. The implementation of data preprocessing, feature selection, and model evaluation techniques ensured reliable and accurate predictive performance, enabling organizations to proactively manage their supply chain operations.

Furthermore, the integration of real-time monitoring, interactive dashboards, supplier performance analysis, and SHAP-based explainable AI enhances the transparency and usability of the system. The proposed solution helps organizations identify critical issues such as stock shortages, shipment delays, and quality concerns at an early stage, allowing timely corrective actions. Overall, the system contributes to optimized resource utilization, reduced operational risks, and improved supply chain resilience through predictive and prescriptive analytics.

XI. FUTURE WORK

The future enhancement of the proposed Smart Supply Chain Management System can focus on incorporating advanced Artificial Intelligence and Deep Learning techniques to improve prediction accuracy and decision-making capabilities. Algorithms such as Long Short-Term Memory (LSTM) networks and other time-series forecasting models can be integrated to analyze complex supply chain patterns and provide more accurate demand and inventory predictions. Additionally, larger and real-time datasets can be utilized to enhance the scalability and adaptability of the system across different industries.

Further improvements can include the integration of cloud computing, IoT devices, and blockchain technology for secure and real-time supply chain monitoring. Advanced analytics features such as automated risk assessment, intelligent recommendation systems, and predictive maintenance can also be incorporated. These enhancements will enable organizations to achieve greater operational efficiency, increased transparency, improved supplier collaboration, and a more resilient and sustainable supply chain ecosystem.

REFERENCES

[1] C. C. Aggarwal, *Data Mining: The Textbook*, Springer, 2015.

[2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd Edition, Morgan Kaufmann, 2011.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.

[4] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2021.

[6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, 2022.

[8] S. Chopra and P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, Pearson Education, 2019.

[9] D. Simchi-Levi, P. Kaminsky, and E. Simchi-Levi, *Designing and Managing the Supply Chain*, McGraw-Hill, 2021.

[10] M. Christopher, *Logistics and Supply Chain Management*, Pearson, 2016.

[11] H. Stadtler, C. Kilger, and H. Meyr, *Supply Chain Management and Advanced Planning*, Springer, 2015.

[12] T. H. Davenport and J. G. Harris, *Competing on Analytics: The New Science of Winning*, Harvard Business Review Press, 2017.

[13] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 2018.

[15] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[17] D. Silver, A. Huang, C. J. Maddison, et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[18] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[19] W. M. P. van der Aalst, *Process Mining: Data Science in Action*, Springer, 2016.

[20] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.