International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

# Multi-Target Decision Analytics for Logistics Delay Prediction and Operational Efficiency Enhancement

Lakshmana Rao Battarusetty[1*], Sk. Khaja Rasool[2], Tangisetti Mani Venkata Sai Sandeep[3], Shaik Saifuddin[3], Vadlapudi Abhilash[3], Vanam Charan[3]
[1]Professor, [2]Assistant Professor, [3]UG Student, [1,2,3]Department of Computer Science and Engineering
[1,2,3]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India
*Correspondence: Lakshmana Rao (lakshman@gist.edu.in)

**Abstract**
The rapid growth of logistics operations has created massive, high-velocity datasets involving routing behaviour, shipment delays, operational constraints, and resource allocation signals. Traditional logistics systems, built on manual decision-making and rigid rule-based logic, can't keep up with this fast-changing environment. They struggle with scalability, lack predictive power, and often react too slowly to disruptions, leading to inefficiencies, inaccurate delay handling, and limited operational visibility. To address these limitations, the proposed system introduces an end-to-end Classification and Regression Trees (CART) machine learning pipeline that automates data preparation, model training, evaluation, and prediction across key logistics tasks. The system leverages K-Nearest Neighbors Classifier ad (KNN), Decision Tree (DT) Classifier and Huber-based models implemented using Stochastic Gradient Descent **(**SGD) estimators to predict shipment delay durations, classify rerouting requirements, determine resource allocation codes, and evaluate shipment feasibility. It supports both regression and classification workflows, integrates model persistence for reuse, and offers single-input and batch prediction capabilities through an intuitive Flask-based web application. By learning patterns from historical logistics data and producing real-time predictive insights, the proposed system significantly boosts accuracy, adaptability, and responsiveness in logistics operations. This creates a smarter, more proactive decision-making environment that reduces delays, enhances operational efficiency, and strengthens overall supply chain performance.

**Keywords:** Shipment Delay Prediction, Classification and Regression Trees, Rerouting Optimization, Machine Learning, KNN.

## 1. INTRODUCTION

In current logistics, supply chain management (SCM) is very important because it ensures that products are moved efficiently and smoothly from their manufacturers to consumers across international boundaries. It is therefore essential that these worldwide supply chains are optimized to guarantee both recovery and efficiency in terms of cost and time. In the current condition of Industry 4.0, the dynamic and uncertain supply chain makes the old school of deterministic models limited in handling volatile conditions [1]. These models seldom account for uncertainties like demand changes, disruptions, or variations in lead times, yet all these are critical to the overall performance of the supply chain. To overcome the constraints above, recent developments have centred on amalgamating probabilistic and fuzzy decision-making methodologies, as shown in Fig. 1. Probabilistic models enable deeper insight into uncertainties and variabilities by embracing the chances of different events taking place, while fuzzy logic plays a role in managing imperfect and uncertain input, reflecting real-life situations more appropriately [2]. These approaches represent a significant departure from deterministic models and offer greater robustness and flexibility for supply chain operations.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Fig. 1: Activity Prediction Pipeline using Smart Logistic

To further enhance decision-making in supply chain management (SCM), it is advisable that uncertainty theory be employed. This principle states that the risks within supply chains should be incorporated into stochastic, fuzzy logic processes to build stronger and more flexible systems [3]. Superior overall performance of the supply chain necessitates this improvement in risk management as well as informed decision-making. As supply chains continue to grow and change, it becomes increasingly important to use optimization models that can manage these complexities [4]. In this changing environment, goal programming (GP) stands out as a powerful multi-objective optimization technique. Charnes and Cooper first introduced it more than fifty years ago. Its flexibility and effectiveness make it especially valuable within the complex and fluctuating landscape of supply chains. GP enables decision-makers to address multiple, often conflicting objectives simultaneously [5]. To improve optimization models, dependent chance constraints (DCC) are used to capture how different uncertain parameters depend on each other. By considering the joint distribution of uncertain factors, DCC-based models offer holistic risk management, significantly enhancing optimization in supply chains. This integration allows for more accurate and reliable optimization results, reflecting real-world conditions.

## 2. LITERATURE SURVEY

Fatorachian, H. al [6] presented a predictive analytics framework integrating digital twin technology, IoT-enabled logistics data, and cybernetic feedback loops to improve last-mile delivery accuracy, congestion management, and sustainability in smart cities. Grounded in Systems Theory and Cybernetic Theory, the framework models urban logistics as an interconnected network, where real-time IoT data enable dynamic routing, demand forecasting, and self-regulating logistics operations. By incorporating machine learning-driven predictive analytics, the study demonstrates how AI-powered logistics optimization can enhance urban freight mobility. The cybernetic feedback mechanism further improves adaptive decision-making and operational resilience, allowing logistics networks to respond dynamically to changing urban conditions.

Aljohani, A al. [7] proposed an innovative strategy that makes use of machine learning as well as predictive analytics approaches. Traditional supply chain risk management frequently uses post-event analysis as well as historical data, which restricts its ability to address real-time interruptions, on the other hand, promotes a futuristic methodology that uses predictive analytics to foresee possible disruptions. Based on contextual and historical data, machine learning models can be trained to find patterns and correlations as well as anomalies that point to imminent dangers. Organizations can identify risks as they arise and take preventative measures by incorporating these models into a real-time monitoring system.

Pasupuleti, V al. [8] depicted the advanced machine learning (ML) techniques to enhance logistics and inventory man-agement. Using historical data from a multinational retail corporation, including sales, inventory levels, order fulfillment rates, and operational costs, we applied a variety of ML algorithms, in-cluding regression, classification, clustering, and time series analysis. result the application of these ML models resulted in significant improvements across key operational areas. They achieved a 15% increase in demand forecasting accuracy, a 10% reduction in overstock and stockouts, and a 95% accuracy in predicting order fulfillment timelines.

Chen, W al. [9] identified and analysed prominent challenges within sustainable logistics, such as reducing carbon emissions, minimizing waste generation, and optimizing transportation routes while considering ecological factors. They also explored emerging trends in AI-driven logistics optimization, such as the integration of real-time data analytics, blockchain technology, and autonomous systems, which hold immense potential for enhancing efficiency and sustainability.

Khatib, E.J al. [10] proposed such scenarios and challenges and proposed 5G technology as a global unified connectivity solution, and also proposed a system for exploiting the application-specific optimization capabilities of 5G networks to better cater for the needs of Smart Logistics. An application traffic modelling process is proposed, along with a proactive approach to network optimization that can improve the Quality of Service and reduce connectivity costs.

Kashem, M.A al. [11 aimed to model AI's and blockchain's role in supply chain optimization by conducting a systematic literature review based on the idealized framework of Rejeb et al. (2022) and the SALSA mechanism. In addition, this paradigm-shifting approach will provide fairer views and options for managing forecasting, planning, monitoring, and reporting across the entire supply chain. The emphasis remains on real-time accuracy, easy access, and optimization of operational indicators such as sales, visibility, and end-to-end supply chain operations always and from any location. It will be an eye-opening experience to enable stakeholders and partners to communicate information collaboratively, consistently, and efficiently.

Woschank, M al. [12] analysed the scientific literature on artificial intelligence, machine learning, and deep learning in the context of Smart Logistics management in industrial enterprises. Furthermore, based on the results of the systematic literature review, the authors present a conceptual framework, which provides fruitful implications based on recent research findings and insights to be used for directing and starting future research initiatives in the field of artificial intelligence (AI), machine learning (ML), and deep learning (DL) in Smart Logistics.

Riad, M al. [13] proposed a conceptual framework for strengthening supply chain resilience through AI integration. The framework leverages AI technologies to improve key aspects of supply chain resilience, including risk management, operational efficiency, and real-time visibility. Result/Conclusions: Additionally, it underscored the importance of collaborative relationships with supply chain partners, enabled by AI-powered data-sharing and communication tools that foster trust and coordination within the network. Originality/Value: This comprehensive framework offered a strategic approach to integrating AI into supply chain management, highlighting its potential to significantly enhance resilience, operational efficiency, and sustainability, thereby empowering organizations to navigate the complexities of modern supply chains more effectively.

Mohsen, B.M al. [14] depicted an innovative framework that integrates artificial intelligence (AI), autonomous vehicles (AVs), and Internet of Things (IoT) technologies to address these challenges. The framework leveraged real-time data from IoT-enabled infrastructure to optimize route planning, enhance traffic signal control, and enable predictive demand management for delivery services. By incorporating AI-driven analytics, the proposed approach aims to improve traffic flow, reduce congestion, and minimize the carbon footprint of urban logistics, contributing to the development of more sustainable and efficient smart cities. This work highlights the potential for combining these technologies to transform urban logistics, offering a novel approach to enhancing delivery operations in densely populated areas.

Rajabzadeh, M al. [15] employed rigorous grounded theory coding procedures, supported by the analytical capabilities of maxqda software. The culmination of the meta-synthesis endeavour culminates in the conceptual representation of IoT adoption within the agricultural logistics domain. This representation is underpinned by the identification of three overarching macro categories/constructs, namely IoT Technology Adoption, encompassing facets such as IoT implementation requisites, ancillary technologies essential for IoT integration, impediments encountered in IoT implementation, and the multifaceted factors that influence IoT adoption; IoT-Driven Logistics Management, encompassing IoT-based warehousing practices, governance-related considerations, and the environmental parameters entailed in IoT-enabled logistics; and the Prospective Gains Encompassing IoT Deployment, incorporating the financial, economic, operational, and sociocultural ramifications ensuing from IoT integration.

## 3. PROPOSED METHODOLOGY

The proposed system is a comprehensive predictive analytics platform designed to enhance supply chain efficiency by proactively identifying potential delays and disruptions in logistics operations. It operates as an end-to-end framework that integrates data acquisition, preprocessing, feature engineering, model training, evaluation, and prediction into a seamless workflow, as shown in Fig 2. The system is capable of handling both regression and classification tasks for critical logistics outcomes such as shipment delays, rerouting requirements, resource allocation, and shipment feasibility. Users interact with a web-based interface that allows exploratory data analysis, single and batch predictions, and visualizations, while the backend ensures model persistence, scalability, and real-time insights. By learning from historical logistics data, the system enables actionable, data-driven decision-making that reduces operational inefficiencies and strengthens overall supply chain performance.

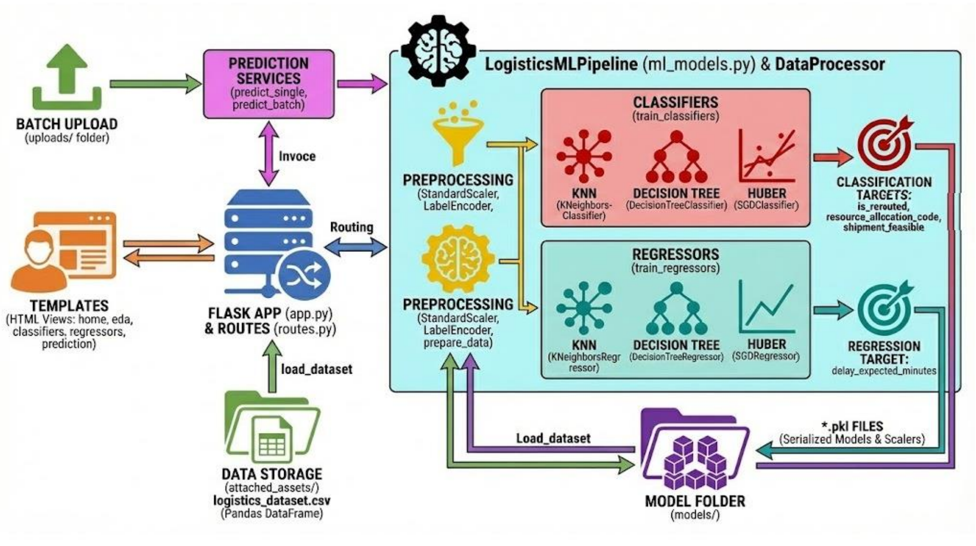



Fig. 2: Proposed system architecture for smart logistics.

**Data Acquisition and Integration:** The first step involves collecting logistics data from multiple sources, including historical shipment records, routing behavior logs, operational constraints, and resource allocation information. This data is consolidated and stored in a structured format suitable for analytics and model training. Data validation and quality checks are applied to ensure accuracy, completeness, and consistency, addressing issues such as missing or inconsistent entries. By integrating real-time and historical data streams, the system can maintain an up-to-date representation of logistics operations, providing a reliable foundation for predictive analysis and decision-making.

**Data Preprocessing and Feature Engineering:** Once the raw data is collected, it undergoes preprocessing to ensure it is clean, consistent, and ready for machine learning. Missing or erroneous values are imputed using statistical methods, while categorical variables are converted into numerical formats to allow algorithmic processing. Numerical features are scaled and normalized to maintain uniformity across the dataset. Feature engineering is then applied to extract meaningful attributes from the raw data, capturing patterns that reflect routing behavior, shipment performance, and operational constraints. This step is crucial for improving model performance and ensuring that predictions are accurate and interpretable.

**Exploratory Data Analysis (EDA):** Exploratory Data Analysis is performed to understand the characteristics of the dataset and inform model selection. This includes analyzing feature distributions, checking for correlations between variables, and identifying missing values or inconsistencies. Target variables for both regression and classification tasks are examined to understand their distribution and potential imbalances. Outliers and anomalies are detected and analyzed for their impact on predictive accuracy. EDA results are visualized through charts, histograms, and summary statistics, providing actionable insights into the data and guiding subsequent preprocessing and model training steps.

**Model Training and Evaluation:** During model training, the pre-processed dataset is split into training and testing subsets to ensure unbiased evaluation. Machine learning models are trained to handle regression tasks, such as predicting shipment delay durations, as well as classification tasks, such as determining rerouting requirements, resource allocation codes, and shipment feasibility. The performance of each model is evaluated using metrics appropriate to the task, including accuracy, precision, recall, F1-score for classification, and MSE, MAE, RMSE, and $R^2$ for regression. The best-performing models are selected and saved for deployment, ensuring consistent and reliable predictions in subsequent steps.

**Prediction and Decision Support:** The prediction module allows the system to provide actionable insights for both individual shipments and batches of data. Users can input a single shipment's features to receive real-time predictions for expected delay, rerouting requirements, and other operational decisions. Batch predictions are also supported, allowing multiple shipments to be processed simultaneously, which aids in operational planning and resource allocation. The prediction results are structured and visualized to help logistics managers interpret and act upon the insights, thereby enabling proactive interventions that minimize delays and optimize overall efficiency.

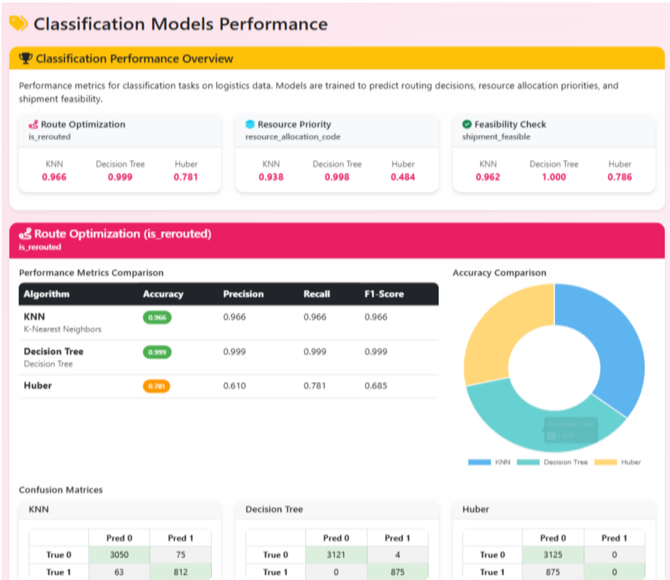
**Model Persistence and Reusability:** To ensure efficiency and scalability, trained models and preprocessing components are saved for future use. This model persistence allows the system to make predictions without retraining from scratch each time new input data is processed. The framework also supports incremental updates and retraining when new data becomes available, maintaining the system's accuracy over time. Logging and version control are maintained to monitor model performance and track changes, ensuring reliability, auditability, and reproducibility of predictive insights in the logistics environment.

International Journal of
# DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

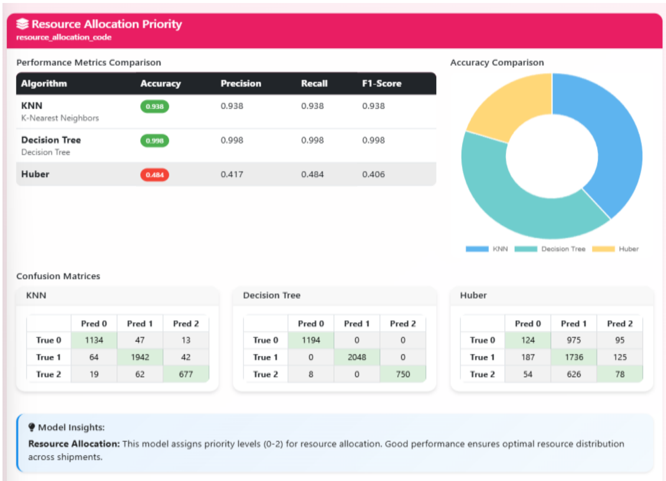
**Web-Based Interaction:** Finally, the system provides an interactive web interface that allows users to access all functionalities, including EDA, model evaluation, and predictions. Users can upload datasets, perform single or batch predictions, and visualize results through an intuitive dashboard. The interface supports real-time feedback, displaying predictions alongside key metrics and visualizations to aid decision-making. Secure session handling and responsive design ensure that the platform is user-friendly and scalable, allowing logistics stakeholders to interact with predictive insights effectively and make timely, informed operational decisions.

## 5. RESULT DESCRIPTION

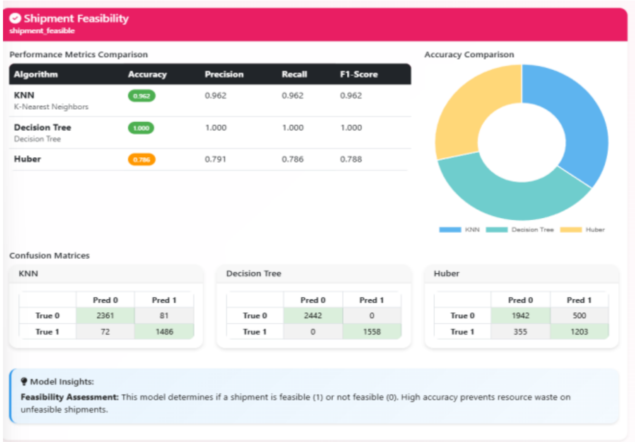
The results and discussion section presents the performance evaluation of the developed logistics prediction system using multiple machine learning models. It focuses on analyzing how effectively the models predict shipment delays, rerouting decisions, resource allocation, and feasibility. The results include various evaluation metrics such as accuracy, precision, recall, F1-score for classification tasks, and MSE, MAE, RMSE, and $R^2$ score for regression tasks. Comparative analysis is carried out to identify the best-performing model among KNN, DT, and Huber. The section also highlights patterns observed from predictions and discusses their impact on operational efficiency.



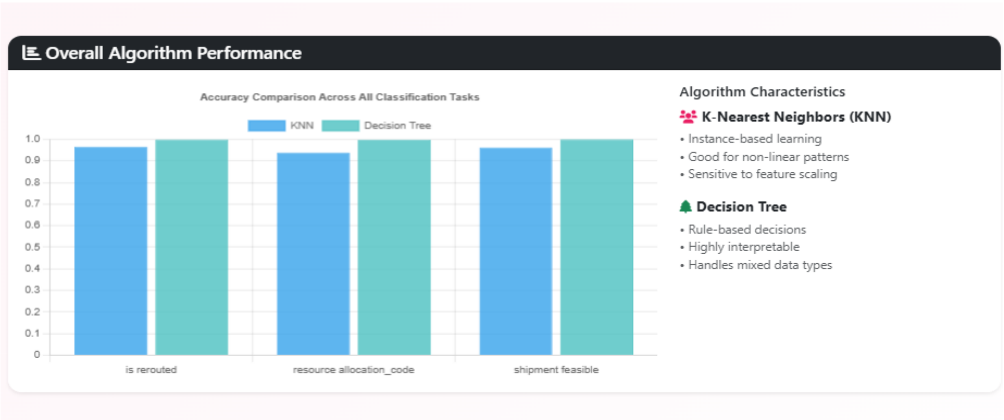

(a)




(b)



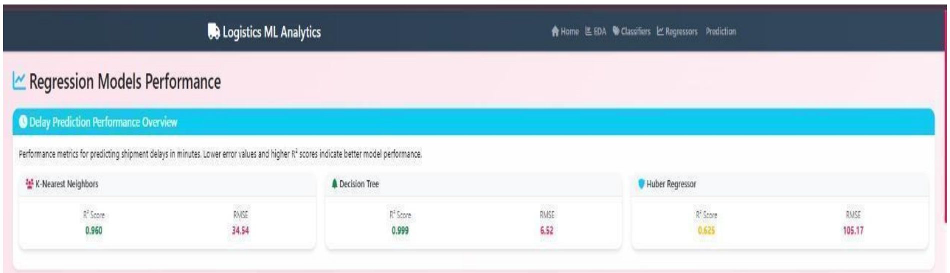

(c)



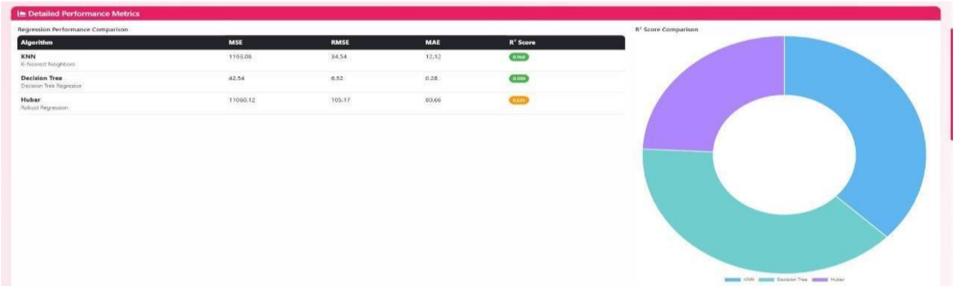

(d)

Fig. 3 (a), (b), (c), (d) Classification performance of the KNN, DT and Huber ML models in Logistics Analytics Platform

The Fig. 3 (a) overall classification performance overview of KNN, DT and Huber Model and shows Route Optimization (is_rerouted) performance with accuracies of 0.966 (KNN), 0.999 (DT), and 0.781 (Huber), alongside confusion matrices. (b) displays Resource Allocation Priority (resource_allocation_code) with accuracies of 0.938 (KNN), 0.998 (DT), and 0.484 (Huber), with corresponding confusion matrices. (c) presents Shipment Feasibility (shipment_feasible) with accuracies of 0.962 (KNN), 1.000 (DT), and 0.786 (Huber), including confusion matrices. (d) provides an Overall Algorithm Performance comparison across tasks, showing accuracies and algorithm characteristics for KNN and DT.

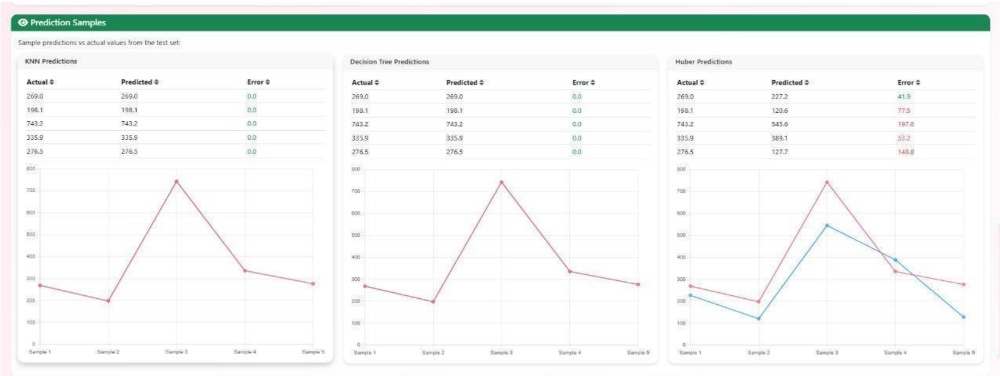



(a)

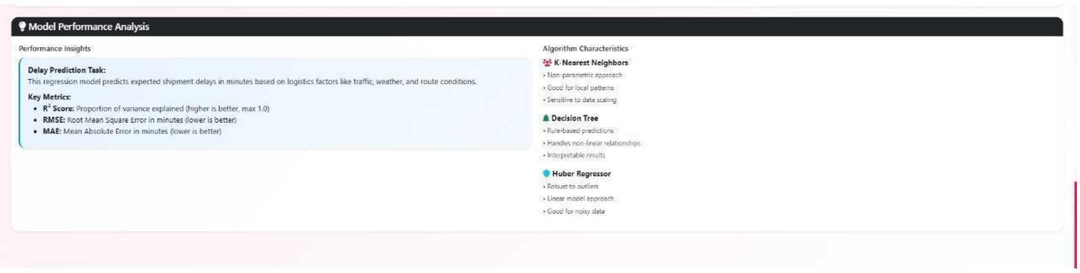



(b)


(c)



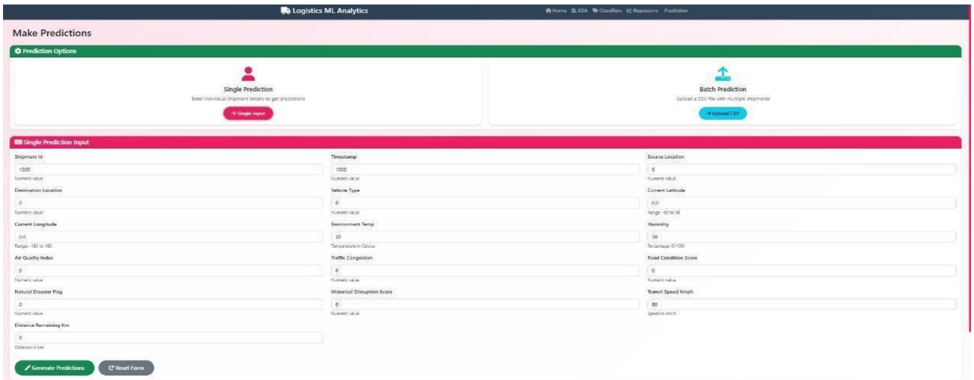

(d)



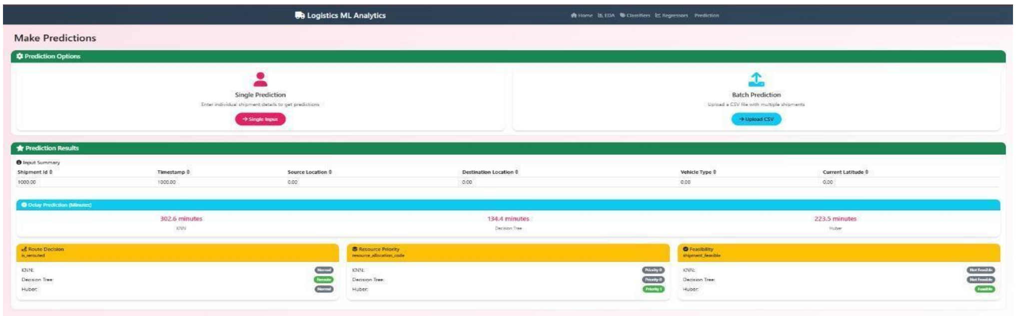

(e)

Fig. 4 (a), (b), (c), (d), (e) Regression performance of the KNN, DT and Huber ML models in Logistics Analytics Platform

The Fig. 4 (a) shows the overall Regression Models Performance i.e, Delay Prediction Performance Overview with $R^2$ scores of 0.960 (KNN), 0.999 (DT), and 0.425 (Huber), and RMSE values of 34.54, 6.32, and 108.17 respectively. (b) provides Detailed Performance Metrics, comparing MSE, RMSE, MAE, and $R^2$ scores, with DT leading at 0.999 $R^2$. (c) illustrates Error Metrics Comparison with a bar chart highlighting higher errors for Huber. (d) presents Prediction Samples, showing actual vs. predicted values and errors for each model across samples. (e) includes Model Performance Analysis, detailing performance insights and algorithm characteristics for the regression task.




(a)




(b)

Fig. 5: Prediction on test data. (a) classification. (b) regression

The Fig. 5(a) shows the Make Predictions interface with options for Single Prediction, where users can input details like shipment_id and environmental conditions to get a delay prediction, and Batch Prediction, allowing CSV file uploads for multiple predictions. (b) displays Prediction Results, showing a sample output with a predicted delay of 202.1 minutes using KNN, 154.4 minutes using DT, and 223.1 minutes using Huber, along with classification results for is_rerouted, resource_allocation_code, and shipment_feasible.

**Comparative Analysis**

The table 1 shows the performance of multiple classification algorithms applied to route optimization. The DT model achieves near-perfect results across all metrics, showing exceptional capability in identifying both feasible and high-risk routes. KNN Classifiers also performs strongly with balanced accuracy, precision, recall, and F1-score values, indicating reliable and stable predictions. The Huber classifier, however, shows noticeably weaker performance, particularly in precision and F1-score, suggesting difficulty in modelling the complex, non-linear interactions present in logistics data.

Table. 1: Classification for Target Column route optimization.

| Algorithm | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| KNN | 0.966 | 0.966 | 0.966 | 0.966 |
| DT | 0.999 | 0.999 | 0.999 | 0.999 |
| Huber | 0.781 | 0.610 | 0.781 | 0.685 |

## 5. CONCLUSION

This research demonstrates the effective application of machine learning models for predictive analytics in logistics operations. By leveraging a clean and diverse dataset containing shipment-level operational features, the study focused on both regression and classification tasks relevant to logistics: shipment delay prediction, rerouting detection, resource allocation prioritization, and feasibility assessment. Through the development of a structured pipeline, encompassing preprocessing, exploratory data analysis, feature scaling, model training, and evaluation, the research ensured a robust and reproducible methodology. The experimental results revealed that the DT algorithm consistently outperformed other models across both classification and regression tasks. It achieved near-perfect accuracy in classification (0.998-1.000) and a remarkably high $R^2$ score of 0.999 for regression, underscoring its ability to capture nonlinear relationships and feature interactions within the dataset. KNN emerged as a reliable alternative, delivering strong but slightly less precise results. In contrast, the Huber model showed significant limitations, with relatively low accuracy in classification and weak fit in regression, indicating its unsuitability for logistics-based predictive modelling in this context. The findings establish the DT as the most suitable predictive algorithm for logistics operations, offering high interpretability, consistency, and computational efficiency. These strengths are critical in logistics, where decisions must often be transparent, reliable, and explainable for operational managers and stakeholders. The results also highlight that integrating machine learning into logistics workflows can significantly improve shipment planning, resource allocation, and disruption management.

**REFERENCES**

[1] Anon, S.Y.; Amin, S.H.; Baki, F. Third-Party Reverse Logistics Selection: A Literature Review. *Logistics* 2024, *8*, 35.

[2] Kim, J.W.; Jeong, B.Y.; Park, M.H. A Study on the Factors Influencing Overall Fatigue and Musculoskeletal Pains in Automobile Manufacturing Production Workers. *Appl. Sci.* 2022, *12*, 3528.

[3] Ganesan, V.K.; Sundararaj, D.; Srinivas, A.P. Adaptive Inventory Replenishment for Dynamic Supply Chains with Uncertain Market Demand. In *Industry 4.0 and Advanced Manufacturing*; Lecture Notes in Mechanical Engineering; Chakrabarti, A., Arora, M., Eds.; Springer: Singapore, 2021; Available online: https://link.springer.com/chapter/10.1007/978-981-15-5689-0_28

[4] McKinsey & Company. Supply Chain 4.0—The Next-Generation Digital Supply Chain. McKinsey Insights. 2021. Available online: https://www.mckinsey.com/business-functions/operations/our-insights/supply-chain-40–the-next-generation-digital-supply-chain

[5] McKinsey & Company. Future Supply Chains: Resilience, Agility, Sustainability. McKinsey Insights. 2023. Available online: https://www.mckinsey.com/capabilities/operations/our-insights/future-proofing-the-supply-chain

[6] Fatorachian, H.; Kazemi, H.; Pawar, K. Enhancing Smart City Logistics Through IoT-Enabled Predictive Analytics: A Digital Twin and Cybernetic Feedback Approach. *Smart Cities* 2025, *8*, 56. https://doi.org/10.3390/smartcities8020056

[7] Aljohani, A. Predictive Analytics and Machine Learning for Real-Time Supply Chain Risk Mitigation and Agility. *Sustainability* 2023, *15*, 15088. https://doi.org/10.3390/su152015088

[8] Pasupuleti, V.; Thuraka, B.; Kodete, C.S.; Malisetty, S. Enhancing Supply Chain Agility and Sustainability through Machine Learning: Optimization Techniques for Logistics and Inventory Management. *Logistics* 2024, *8*, 73. https://doi.org/10.3390/logistics8030073

[9] Chen, W.; Men, Y.; Fuster, N.; Osorio, C.; Juan, A.A. Artificial Intelligence in Logistics Optimization with Sustainable Criteria: A Review. *Sustainability* 2024, *16*, 9145. https://doi.org/10.3390/su16219145

[10] Khatib, E.J.; Barco, R. Optimization of 5G Networks for Smart Logistics. *Energies* 2021, *14*, 1758. https://doi.org/10.3390/en14061758

[11] Kashem, M.A.; Shamsuddoha, M.; Nasir, T.; Chowdhury, A.A. Supply Chain Disruption versus Optimization: A Review on Artificial Intelligence and Blockchain. *Knowledge* 2023, *3*, 80-96. https://doi.org/10.3390/knowledge3010007

[12] Woschank, M.; Rauch, E.; Zsifkovits, H. A Review of Further Directions for Artificial Intelligence, Machine Learning, and Deep Learning in Smart Logistics. *Sustainability* 2020, *12*, 3760. https://doi.org/10.3390/su12093760

[13] Riad, M.; Naimi, M.; Okar, C. Enhancing Supply Chain Resilience Through Artificial Intelligence: Developing a Comprehensive Conceptual Framework for AI Implementation and Supply Chain Optimization. *Logistics* 2024, *8*, 111. https://doi.org/10.3390/logistics8040111

[14] Mohsen, B.M. AI-Driven Optimization of Urban Logistics in Smart Cities: Integrating Autonomous Vehicles and IoT for Efficient Delivery Systems. *Sustainability* 2024, *16*, 11265. https://doi.org/10.3390/su162411265

[15] Rajabzadeh, M.; Fatorachian, H. Modelling Factors Influencing IoT Adoption: With a Focus on Agricultural Logistics Operations. *Smart Cities* 2023, *6*, 3266-3296. https://doi.org/10.3390/smartcities6060145.