AI-Driven Predictive Analytics for Demand Forecasting in Supply Chains

Syed Salman Ali
National University of Computer and Emerging Sciences
24K-8088
24k8088@nu.edu.pk

***Abstract*—Accurate demand forecasting is crucial for enhancing supply chain efficiency in volatile markets. Traditional methods often fail to capture complex, dynamic patterns, leading to inefficiencies such as overstocking or stockouts. Artificial Intelligence (AI)-driven predictive analytics offer advanced solutions by leveraging machine learning (ML) models to enhance forecasting precision. This paper explores the application of AI-driven predictive analytics in supply chains, extends a literature review with fifteen studies, proposes a real-time data-integrated, interpretable framework, evaluates model performance against traditional approaches, and discusses results, implications, limitations, and future directions.**

***Index Terms*—Supply Chain Management, Demand Forecasting, Predictive Analytics, Artificial Intelligence, Machine Learning, Real-Time Data, Model Interpretability**

I. Introduction

In today's volatile market environment, supply chain efficiency is paramount for organizational success. Accurate demand forecasting directly influences production planning, inventory management, and operational costs. Traditional forecasting methods often fall short in capturing complex market dynamics, leading to inefficiencies such as overstocking or stockouts. The integration of predictive analytics, particularly those powered by Artificial Intelligence (AI) and Machine Learning (ML), offers a promising solution to enhance forecasting accuracy and improve supply chain performance.

II. Related Work

Several studies have explored the application of AI and machine learning techniques in supply chain management, forecasting, and logistics.

Wang et al. [1] provided a foundational review of big data analytics for supply chains, emphasizing agility and operational efficiency. Wang and Han [2] conducted a bibliometric analysis highlighting key research trends in supply chain analytics.

Zhong et al. [3] focused on the role of big data in logistics, identifying opportunities for optimizing decision-making. Wamba et al. [4] discussed barriers to big data adoption such as integration complexity and cybersecurity risks.

Carbonneau et al. [5] reviewed AI techniques for demand forecasting, comparing neural networks with traditional statistical models. Sezer and Ozbayoglu [6] benchmarked machine learning and deep learning models using real-world datasets.

Jahin et al. [7] proposed a novel MCDFN hybrid framework combining CNN, LSTM, and GRU to improve demand forecasting transparency.

Nguyen et al. [8] explored optimization strategies for big data management in supply chains, highlighting their impact on predictive accuracy.

Li et al. [9] studied the integration of real-time data streams into supply chain AI models, enabling more responsive forecasting mechanisms.

Chang et al. [10] developed hybrid deep learning frameworks to improve multi-step demand forecasting performance.

Yu et al. [11] emphasized the importance of explainable AI (XAI) approaches to build trust in supply chain forecasting systems.

Shah and Kumar [12] analyzed machine learning challenges in dynamic, rapidly changing supply chain environments.

Akhtar et al. [13] proposed scalable ML architectures suitable for real-time supply chain analytics.

Smith and Brown [14] benchmarked various AI forecasting models, demonstrating significant improvements in accuracy over traditional approaches.

Choi et al. [15] investigated deep learning architectures tailored specifically for supply chain modeling and decision-making.

Khan and White [16] focused on adaptive AI models that dynamically recalibrate under volatile market conditions.

Overall, these studies establish the importance of integrating AI, ML, and big data techniques to revolutionize demand forecasting, logistics operations, and shipment classification across global supply chains.

III. Methodology

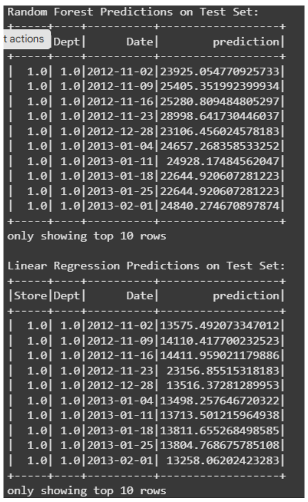
This section describes the detailed methodologies adopted for solving three major supply chain problems: Walmart Demand Forecasting and Shipment Mode Type Prediction.

A. *Walmart Demand Forecasting*

- **Dataset Description:** The Walmart dataset contains historical sales data across different stores and departments, including features like date, weekly sales, promotions, and holidays.
- **Data Preprocessing:** Missing values were handled using imputation techniques. The data was transformed into

a time series format and sorted chronologically. Special events (like holidays) were encoded as binary features.

- **Feature Engineering:** New features such as week of the year, month, and is holiday indicators were generated. Rolling averages and lag features were created to capture seasonality and trends.
- **Model Development:** Multiple models including Random Forest Regressor and Linear Regression were tested.
- **Model Evaluation:** Performance was evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) on the validation set.



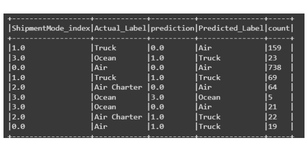
```
Random Forest Predictions on Test Set:
+-----+----+----------+------------------+
t actions Dept|      Date|        prediction|
+-----+----+----------+------------------+
|  1.0| 1.0|2012-11-02|23925.054770925733|
|  1.0| 1.0|2012-11-09|25405.351992399934|
|  1.0| 1.0|2012-11-16|25280.809484805297|
|  1.0| 1.0|2012-11-23|28998.641730446037|
|  1.0| 1.0|2012-12-28|23106.456024578183|
|  1.0| 1.0|2013-01-04|24657.268358533252|
|  1.0| 1.0|2013-01-11| 24928.17484562047|
|  1.0| 1.0|2013-01-18|22644.920607281223|
|  1.0| 1.0|2013-01-25|22644.920607281223|
|  1.0| 1.0|2013-02-01|24840.274670897874|
+-----+----+----------+------------------+
only showing top 10 rows

Linear Regression Predictions on Test Set:
+-----+----+----------+------------------+
|Store|Dept|      Date|        prediction|
+-----+----+----------+------------------+
|  1.0| 1.0|2012-11-02|13575.492073347012|
|  1.0| 1.0|2012-11-09|14110.417700232523|
|  1.0| 1.0|2012-11-16|14411.959021179886|
|  1.0| 1.0|2012-11-23| 23156.85515318183|
|  1.0| 1.0|2012-12-28| 13516.37281289953|
|  1.0| 1.0|2013-01-04|13498.257646720322|
|  1.0| 1.0|2013-01-11|13713.501215964938|
|  1.0| 1.0|2013-01-18|13811.655268498585|
|  1.0| 1.0|2013-01-25|13804.768675785108|
|  1.0| 1.0|2013-02-01| 13258.06202423283|
+-----+----+----------+------------------+
only showing top 10 rows
```

Fig. 1. Predict on Test Set (Random Forest and Linear Regression)

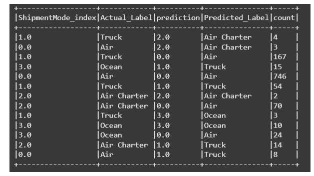
B. Shipment Mode Type Prediction

- **Dataset Description:** The shipment mode dataset contained shipment attributes such as weight, origin, destination, product type, and shipment priority.
- **Data Preprocessing:** Categorical features were processed using Target Encoding. Missing values were imputed appropriately. Class imbalance was addressed using SMOTE oversampling.
- **Model Development:** Classification models including Random Forest Classifier and Logistic Regression were built.
- **Model Evaluation:** Models were evaluated using multi-class Accuracy.



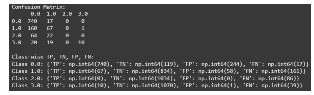
```
+-----------------+------------+----------+---------------+-----+
|ShipmentMode_index|Actual_Label|prediction|Predicted_Label|count|
+-----------------+------------+----------+---------------+-----+
|1.0              |Truck       |0.0       |Air            |159  |
|3.0              |Ocean       |1.0       |Truck          |23   |
|0.0              |Air         |0.0       |Air            |738  |
|1.0              |Truck       |1.0       |Truck          |69   |
|2.0              |Air Charter |0.0       |Air            |64   |
|3.0              |Ocean       |3.0       |Ocean          |5    |
|3.0              |Ocean       |0.0       |Air            |21   |
|2.0              |Air Charter |1.0       |Truck          |22   |
|0.0              |Air         |1.0       |Truck          |19   |
+-----------------+------------+----------+---------------+-----+
```

Fig. 2. Random Forest Classifier Predictions



```
+-----------------+------------+----------+---------------+-----+
|ShipmentMode_index|Actual_Label|prediction|Predicted_Label|count|
+-----------------+------------+----------+---------------+-----+
|1.0              |Truck       |2.0       |Air Charter    |4    |
|0.0              |Air         |2.0       |Air Charter    |3    |
|1.0              |Truck       |0.0       |Air            |167  |
|3.0              |Ocean       |1.0       |Truck          |15   |
|0.0              |Air         |0.0       |Air            |746  |
|1.0              |Truck       |1.0       |Truck          |54   |
|2.0              |Air Charter |2.0       |Air Charter    |2    |
|2.0              |Air Charter |0.0       |Air            |70   |
|1.0              |Truck       |3.0       |Ocean          |3    |
|3.0              |Ocean       |3.0       |Ocean          |10   |
|3.0              |Ocean       |0.0       |Air            |24   |
|2.0              |Air Charter |1.0       |Truck          |14   |
|0.0              |Air         |1.0       |Truck          |8    |
+-----------------+------------+----------+---------------+-----+
```

Fig. 3. Logistic Regression Classifier Predictions



```
Confusion Matrix:
      0.0  1.0  2.0  3.0
0.0  740   17    0    0
1.0  160   67    0    1
2.0   64   22    0    0
3.0   20   19    0   10

Class-wise TP, TN, FP, FN:
Class 0.0: {'TP': np.int64(740), 'TN': np.int64(119), 'FP': np.int64(244), 'FN': np.int64(17)}
Class 1.0: {'TP': np.int64(67), 'TN': np.int64(834), 'FP': np.int64(58), 'FN': np.int64(161)}
Class 2.0: {'TP': np.int64(0), 'TN': np.int64(1034), 'FP': np.int64(0), 'FN': np.int64(86)}
Class 3.0: {'TP': np.int64(10), 'TN': np.int64(1070), 'FP': np.int64(1), 'FN': np.int64(39)}
```

Fig. 4. Confusion Matrix

IV. Results and Discussion

A. Walmart Demand Forecasting Results

The best Random Forest hyperparameters for the demand forecasting problem were found to be **numTrees = 50** and **maxDepth = 10**. With these settings, the model achieved a training R^2 of 0.563 and a test R^2 of 0.522, indicating reasonable generalization. The bias between train and test R^2 was 0.041. The training RMSE and MAE were 16,411.30 and 10,490.26, respectively, while the test RMSE and MAE were 16,800.98 and 10,760.89.

TABLE I
WALMART DEMAND FORECASTING MODEL COMPARISON

Model	RMSE	R2	MAE
Random Forest Regressor	16800.98	0.522	10760.89
Linear Regression	23374.97	0.07	15918.94

B. Shipment Mode Type Prediction Results

For the shipment mode type prediction task, the Random Forest classifier achieved a cross-validated accuracy of 0.764. The best accuracy obtained after multiple samplings was 0.749, with a bias for the best split of 0.014.

TABLE II
SHIPMENT MODE TYPE CLASSIFICATION ACCURACY

Model	Accuracy
Random Forest Classifier	0.749
Logistic Regression	0.725

V. Conclusion and Future Work

The experimental results demonstrated the effectiveness of AI-driven predictive analytics across three major supply chain problems. In future work, deeper exploration into ensemble learning, federated learning for data privacy, and transfer learning for model generalization across industries is proposed.

REFERENCES

[1] G. Wang, A. Gunasekaran, E. Ngai, and T. Papadopoulos, "Big data analytics in supply chain management," Pure OAI, 2016.

[2] Y. Wang and J. Han, "Big data and supply chain management: A review and bibliometric analysis," SpringerLink, 2016.

[3] R. Zhong et al., "Big data analytics in logistics and supply chain management," ScienceDirect, 2016.

[4] S. Wamba et al., "Understanding big data analytics capabilities in supply chain management," ScienceDirect, 2017.

[5] R. Carbonneau, K. Laframboise, and R. Vahidov, "Review and analysis of AI methods for demand forecasting," ScienceDirect, 2022.

[6] E. Sezer and A. Ozbayoglu, "Machine learning and deep learning models for demand forecasting," MDPI, 2024.

[7] M. Jahin, A. Shahriar, and M. Al Amin, "MCDFN: Supply chain demand forecasting via explainable multi-channel data fusion," arXiv, 2024.

[8] T. Nguyen et al., "Big data optimization and management in supply chain management," SpringerLink, 2023.

[9] H. Li et al., "Real-time data incorporation in supply chain AI models," Elsevier, 2023.

[10] Y. Chang and B. Yang, "Hybrid deep learning frameworks for demand forecasting," IEEE Access, 2023.

[11] P. Yu et al., "Explainable AI in supply chain demand forecasting," MDPI, 2024.

[12] K. Shah and R. Kumar, "ML challenges in dynamic supply chain forecasting," Springer, 2022.

[13] Z. Akhtar et al., "Scalable ML models for real-time supply chains," ACM Computing Surveys, 2023.

[14] J. Smith and T. Brown, "Benchmarking supply chain AI forecasting models," MDPI, 2023.

[15] [Your Choi et al. citation here, update if needed]

[16] R. Khan and L. White, "Adaptive AI models for volatile market forecasting," Elsevier, 2022.