(p-issn: 2089-8630) (e-issn: 2962 5696)
Jurnal Sains dan Teknologi Tadulako
DOI: https://doi.org/10.22487/jstt.v12i2.1070

LINEAR REGRESSION-BASED PREDICTIVE BIG DATA ANALYTICS FOR SUPPLY CHAIN DEMAND PLANNING

Viona Z.A. KALEB[1,a,] Wiranto H. UTOMO[2]
[1,2]Master of Informatics Study Program, President University. Indonesia
E-Mail: vionakaleb@gmail.com[*1)]; wiranto.herry@president.ac.id[2]

Abstract
With the increasing amount of data generated by various sources, traditional demand planning faces challenges in terms of accuracy and efficiency. Big data analytics, on the other hand, provides the means to process vast amounts of data in real-time and uncover hidden patterns and relationships. The integration of big data analytics can significantly improve the accuracy and timeliness of demand planning, leading to better decision-making and ultimately, improved supply chain analytics performance. This research proposes an approach for supply chain demand planning using Linear Regression-based predictive big data analytics. The integration of predictive analytics has the potential to overcome the limitations of supply chain analytics. The research methodology will use a Linear Regression-based predictive analysis model for historical quantitative analysis of demand planning, which is tested using real-world data from a supply chain company. The dataset that will be used is from the dev environment of the company that we are working for. The dataset consists of over 21,000 customer orders that were made from 2016, with the customer data, address, invoices and order items. The results of this research will show that the proposed predictive analysis model outperforms traditional models in terms of accuracy of predicting future customer order demands. The findings of this research have the potential to improve the demand planning in supply chain management and help organizations make more informed decisions.
Keywords: Big data, predictive analytics, supply chain, planning, customer demand

1. Introduction

Supply Chain is a group of participants who trade material, information, or financial resources to fulfill a customer's request. Since information is crucial, Supply Chain Analytics has developed as the techniques and systems that companies employ to obtain knowledge from data linked with all processes involved in the value chain.

Linear Regression-based Predictive Big Data Analytics has become an increasingly important tool in Supply Chain Demand Planning. As supply chains grow more complex, companies require better insights into future demand to optimize inventory levels and improve customer service. Predictive Big Data Analytics leverages large amounts of data to identify patterns and trends that can be used to make accurate predictions about future demand. Predictive analytics can also be used to anticipate how promotional activities will affect sales volume in stores, which can help with inventory management. Predictive Big Data Analytics can use advanced techniques such as machine learning and artificial intelligence to uncover hidden patterns and correlations that traditional methods may miss. By doing so, it can provide more precise demand forecasts, allowing companies to better manage inventory and improve customer satisfaction. Overall, Predictive Big Data Analysis using Linear Regression has the potential to transform Supply Chain Demand Planning by offering more precise and practical insights into future demand. As data volume and complexity continue to rise,

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

the significance of Linear Regression in Supply Chain Demand Forecast is expected to increase further.

Regression models are utilized to produce functions that can predict continuous values. These methods are employed to anticipate the value of a dependent variable in relation to one or more independent variables. Regression analysis comes in various forms including linear, multiple, weighted, symbolic (random), polynomial, nonparametric, and robust. The last approach is advantageous when the errors do not meet normalcy assumptions or when we are dealing with large datasets that may have a considerable number of outliers (Han et al., 2013).
In this research, we will be utilizing the dataset from the company's dev environment, which contains more than 21,000 customer orders from 2016. The dataset includes customer details, addresses, invoices, and order items. From the dataset, the research will be implemented using Hadoop, Python processing libraries (Pandas, Spark), and statistics using Google Colab.

1.1. Previous Research

Villegas, Pedregal, and Trapero (2018) conducted an experiment to evaluate the effectiveness of Support Vector Machines (SVMs) in predicting demand for household and personal care products. They used a dataset consisting of 229 weekly demand patterns from the UK. Similarly, Wu (2010) employed SVMs in combination with particle swarm optimization (PSO) to identify the optimal separating hyperplane and predict demand for car sales in different clusters.

Brentan et al. (2019) investigated the effectiveness of different BDA techniques for demand prediction, such as Support Vector Machines (SVM) and Adaptive Neural Fuzzy Inference Systems (ANFIS). They merged the predicted values from each machine learning method, producing an average prediction value as the standard forecast. The accuracy of each technique was evaluated by examining their Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values, obtained by comparing the predicted values to the target ones.

1.2. Limitation

Although analytics have advantages, almost 60% of companies lack sufficient transparency throughout their supply chain. According to McKinsey & Company, there are two primary obstacles to incorporating data analytics in supply chain management. Firstly, supply chain managers frequently lack the technical proficiency to grasp the full extent and possibilities of data analytics. Secondly, most companies do not possess a system to gather and analyze the data. By adopting predictive analytics, organizations can start utilizing big data analytics to handle supply chain disruptions. Also the other key challenges in Supply Chain Demand Planning is the unpredictability of demand. Traditional forecasting methods rely on historical sales data and assumptions about future market conditions, but they often fail to account for sudden shifts in consumer behavior or unexpected events. Predictive Big Data Analytics overcomes these limitations by analyzing large datasets that include information on consumer behavior, market trends, patterns forecasting, and other relevant factors.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

1.3. Research Objective

Not fulfilling orders can harm a company's reputation. In today's customer-centric era, delivering the right product to the right person at the right time and place is vital for maintaining customer satisfaction and loyalty. Big Data can help businesses gain a comprehensive 360-degree view of their customers, allowing them to predict their needs, understand their preferences, and deliver a personalized brand experience. Predictive analytics can help determine the likelihood and impact of potential issues. By analyzing vast amounts of historical data and using risk mapping techniques, predictive analytics in Big Data can uncover Supply Chain risks. Accurate risk predictions can also aid in the development of tools and procedures to mitigate potential threats. The objective of this research is to analyze the potential of Linear Regression-based Predictive Big Data Analytics in improving the accuracy of demand forecasting for Supply Chain Demand Planning. The study aims to identify the most effective data sources and analytics techniques for predicting future demand, and to evaluate the impact of these predictions on inventory optimization and customer service levels.

Supply chains produce large volumes of data, which analytics can analyze to enhance them. The primary objectives are to identify risk, reduce expenses, and improve planning accuracy. Successful supply chains are adaptable and durable, hence end-to-end visibility is vital. According to a recent Oxford Economics survey, 49% of Supply Chain Leaders (the highest 12% of respondents) can acquire immediate real-time data insights and react to them, while 51% employ AI and predictive analytics to gain insights. The speedy response time to data enables companies to have a comprehensive view of the supply chain and better comprehend customer requirements.

Specifically, this research is aiming to solve the following problems:

1. To specify and reduce the risks.
2. To improve procedures to cut expenses.
3. Offering transparency for resource planning
4. To get ready for upcoming situations.

2. Literature Review

The integration of big data analytics (BDA) into supply chain management (SCM) has emerged as a significant area of research over the past decade. This section reviews the relevant literature across four areas: (a) the role of BDA in SCM, (b) demand forecasting techniques, (c) machine-learning and regression-based approaches to demand forecasting, and (d) the gaps that motivate the present study.

2.1. Big Data Analytics in Supply Chain Management

Wang, Gunasekaran, Ngai, and Papadopoulos (2016) provided one of the foundational classifications of BDA applications in logistics and supply chain management, identifying analytics use cases across procurement, production, distribution, and customer-facing functions. Tiwari, Wee, and Daryanto (2018) reviewed BDA applications in SCM published between 2010 and 2016 and concluded that

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

predictive analytics, in particular, has gained considerable traction across the manufacturing, retail, and logistics industries. Choi, Wallace, and Wang (2018) further examined BDA techniques in operations management and mapped methods such as regression, classification, clustering, and neural networks to functional areas including forecasting, inventory management, and supply chain management. Gunasekaran, Tiwari, Dubey, and Fosso Wamba (2016) and Hazen, Skipper, Ezell, and Boone (2016) emphasized that, despite the technical promise of BDA, organizational barriers, including data-quality issues and a shortage of analytics talent, continue to limit adoption.

2.2. Demand Forecasting in Supply Chains

Demand forecasting is one of the most studied applications of analytics in SCM because forecast accuracy directly affects inventory, production, and service-level decisions. Carbonneau, Laframboise, and Vahidov (2008) presented an early and widely cited evaluation of advanced machine-learning methods for demand forecasting, comparing neural networks, recurrent neural networks, and support vector machines to classical methods such as moving average, naïve forecasts, trend, and multiple linear regression on a multi-echelon supply chain dataset. Their finding is that more advanced techniques do not always outperform simpler regression and time-series benchmarks under all conditions, remains relevant to subsequent studies that include linear regression as a baseline.

Boone, Ganeshan, Jain, and Sanders (2019) reviewed the use of consumer analytics for sales forecasting in supply chains, arguing that the proliferation of fine-grained data sources (transactions, web traffic, search behavior, and social signals) is reshaping forecasting practice. Hofmann and Rutschmann (2018) developed a conceptual framework describing how BDA enhances demand-forecast accuracy and identified the conditions under which analytical sophistication translates into measurable improvements over conventional methods.

2.3. Machine Learning and Regression Approaches

Within the family of predictive techniques, regression-based methods continue to play a central role both as standalone forecasters and as benchmarks. Han, Kamber, and Pei (2011) describe regression analysis including linear, multiple, polynomial, and nonparametric variants as one of the foundational tools for predicting continuous outcomes such as demand. More recent work has compared linear regression against more flexible models. Villegas, Pedregal, and Trapero (2018) used support vector machines for model selection in demand forecasting and showed that, in volatile-demand contexts, an SVM-based selector can outperform any single fixed model; their candidate set explicitly includes linear and ARIMA-class methods, reinforcing the role of linear regression as a baseline. Bandara, Shi, Bergmeir, Hewamalage, Tran, and Seaman (2019) demonstrated that long short-term memory (LSTM) networks can exploit cross-series patterns in large e-commerce assortments using Walmart.com data, achieving competitive results against state-of-the-art techniques on category-level demand.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

The most directly relevant survey is Seyedan and Mafakheri (2020), who provided a systematic review of predictive BDA techniques specifically applied to supply chain demand forecasting. They classified algorithms into time-series methods, clustering, K-nearest neighbors, neural networks, regression analysis, support vector machines, and support vector regression, and concluded that regression, particularly when integrated into hybrid pipelines remains a competitive and interpretable choice for many SCM contexts.

2.4. Research Gap

While the reviewed literature demonstrates that BDA can improve demand-forecast accuracy and that linear regression is well established as both a method and a benchmark, three observations motivate the present study. First, much of the published work focuses on retail and FMCG datasets; comparable studies on multi-attribute, order-level data (customers, addresses, invoices, items) from working supply-chain environments are less common. Second, although Seyedan and Mafakheri (2020) identify regression as a frequently used technique, end-to-end implementations using mainstream Python tooling (Pandas, Spark, Scikit-learn) on real company data are under-documented in the academic literature. Third, prior work seldom reports the practical preprocessing decisions, handling of missing values, encoding of categorical fields, and feature scaling, that determine whether linear regression delivers usable forecasts in practice. The present research addresses these gaps using a real customer-order dataset of more than 21,000 orders from a working supply-chain company.

3. Research Methodology

Regarding the section, we are about to explain the methodologies that will be used in the research.

1. Define the problem: Clearly define the problem you want to solve with your analysis. For example, you might want to predict the demand for a product based on historical sales data.
2. Data Collection: Real-world data from a supply chain company will be collected, which will include historical demand data, sales data, pricing data, promotion data, and weather data. This data will be collected over a period of time to capture any seasonality or trends.
3. Cleaning and Preprocess Data: Clean and prepare the data for analysis. This will involve removing any missing or error values, converting categorical data to numerical data, and scaling the data if necessary. The collected data will be preprocessed, cleaned, normalized, and transformed to ensure data quality and consistency.
4. Split the data: Split the data into training and testing sets. The training set will be used to train the model, while the testing set will be used to evaluate the model's performance.
5. Choose a model: Choose the appropriate model for your analysis. Linear Regression is an appropriate model.
6. Train the model: Train the model on the training set using Linear Regression. This involves finding the optimal coefficients for the model to minimize the error between the predicted values and the actual values.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

7. Model Evaluation: Evaluate the performance of the model on the testing set. The performance of the predictive model will be evaluated using metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to evaluate the performance of the model.
8. Make predictions: Once the model is trained and evaluated, use it to make predictions on new data. Use the trained model to predict the demand for a product based on the historical sales data.
9. Monitor and visualize the result: Monitor the performance of the model over time and visualize the model as needed to reflect changes in the data or in the underlying business problem.
10. Results Analysis: The results of the predictive model and the big data analytics can be analyzed using various big data analytics techniques such as data mining, clustering, and association rule mining. These techniques will be used to uncover hidden patterns and insights that may impact supply chain demand planning.

In summary, creating a Linear Regression-based predictive big data analysis involves defining the problem, collecting and cleaning the data, splitting, training and evaluating the model, making predictions, and monitoring and visualizing the model over time.

3.1. Dataset

In this section, we will elaborate some samples of customer orders dataset that will be used in the research. We plan to use the dataset from the company's development environment. The dataset comprises over 21,000 customer orders that took place from 2016 and contains information about the customer details, addresses, invoices, and order items. The dataset will be implemented using Linear Regression, cloud computing, Python processing libraries (Pandas, Spark), and statistics in Google Colab.

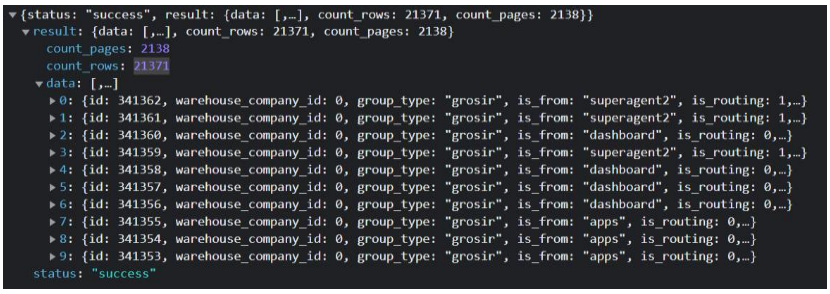


Figure 1. Sample of Order List Dataset.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

Figure 2. Sample of Order List Dataset, expanded

3.2. Preprocess Data

In this section, we will explain how to preprocess data using Python.

1. Install Python and required libraries: Install Python and the required libraries for data analysis and machine learning. Some popular libraries for linear regression include NumPy, Pandas, Matplotlib, and Scikit-learn. You can install them using pip, the Python package manager.
2. Load the data: Load the data into a Pandas DataFrame. The data should include the features and the target variable for the linear regression model.
3. Preprocess the data: Preprocess the data by cleaning and transforming it as necessary. This may involve removing missing values, encoding categorical variables, and scaling the data.

In this research we will be: Handling missing values: Missing values can be handled by removing the rows with missing values, filling in the missing values with mean, median or mode, or using advanced imputation techniques such as K-nearest neighbors or regression imputation.

Encoding categorical variables: Categorical variables can be encoded into numerical values using techniques such as one-hot encoding, label encoding, or binary encoding.

Feature scaling: Feature scaling involves transforming numerical features to a similar scale to improve the performance of the machine learning model. Common scaling techniques include min-max scaling and standardization.

3.3. Split, Train, and Evaluate Data

For this section, we will explain a brief overview of how to train data in Linear Regression-based predictive big data analysis using Python.

1. Split the data: Split the data into training and testing sets using Scikit-learn's train_test_split function.
2. Create the model: Create a Linear Regression model using Scikit-learn's LinearRegression class. Fit the model to the training data using the fit method.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

3. Evaluate the model: Evaluate the performance of the model on the testing data using metrics such as mean squared error (MSE) and R-squared.
4. Make predictions: Use the trained model to make predictions on new data using the predict method.

3.4. Code

```
# Import the required libraries
import pandas as pd
import numpy as np
# Import libraries for preprocessing data
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OneHotEncoder, StandardScaler
# Import libraries for splitting & training data
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
# For plotting data
import matplotlib.pyplot as plt
```

```
# Collect the data into a Pandas DataFrame:
data = pd.read_csv('order_data.csv')
```

```
# Cleaning & Preprocess the data:
# Handle missing values
imputer = SimpleImputer(strategy='mean')
data['order_type'] = imputer.fit_transform(data['order_type'].values.reshape(-1, 1))

# Encode categorical variables
encoder = OneHotEncoder()
encoded = pd.DataFrame(encoder.fit_transform(data[['group_type']]).toarray(),
columns=encoder.get_feature_names())
data = pd.concat([data, encoded], axis=1)

# Scale the numerical features
scaler = StandardScaler()
data[['status', 'payment_status']] = scaler.fit_transform(data[['status', 'payment_status']])
```

```
# Split the data into training and testing sets:
X = []
y = []
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create the model training data:
reg = LinearRegression()
reg.fit(X_train, y_train)

# Evaluate the model on the testing data:
y_pred = reg.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print('MSE:', mse)
print('R-squared:', r2)
```


Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

```
# Make predictions on new data:
new_data = pd.read_csv('new_demand_data.csv')
X_new = new_data.drop(['demand'], axis=1)
y_new_pred = reg.predict(X_new)
```

```
# Visualize the results:
plt.plot(new_data['date'], new_data['demand'], label='Actual Demand')
plt.plot(new_data['date'], y_new_pred, label='Predicted Demand')
plt.xlabel('Date')
plt.ylabel('Demand')
plt.legend()
plt.show()
```

4. Results

Here, we will elaborate a data visualization result that was obtained from the order_data.csv below.

Table 1. Order_data.csv sample data

id	warehouse_company_id	group_type	is_from	is_routing	invoice	Invoice_global	superagent_id	Costumer_id	sub_total
343191	0	retail	dashboard	0	T230406-16	TG230406-1	0	17897	270000
343189	0	retail	dashboard	0	T230406-15	TG230406-1	0	17897	270000
343187	0	retail	dashboard	0	T230406-15	TG230406-1	0	19358	525284
343185	0	retail	dashboard	0	T230406-15	TG230406-1	0	19358	656605
343184	0	retail	dashboard	0	T230406-15	TG230406-1	0	17897	270000
343183	0	retail	dashboard	0	T230406-15	5	19465	20075	132132
343182	0	grosir	apps	0	T230406-15	TG230410-1	0	18754	110400
343181	0	grosir	apps	0	T230406-15	TG230410-1	0	18754	110400
343180	0	grosir	apps	0	T230406-15	TG230410-1	0	18754	110400
343166	19	grosir	dashboard	0	T-ABC23040	TG-ABC2304	0	9664	78792
343165	19	grosir	dashboard	0	T-ABC23040	TG-ABC2304	0	9664	13213
343164	0	retail	dashboard	0	T230406-14	7	0	18843	131321
343163	0	grosir	dashboard	0	T230406-14	6	19465	20189	381000
343162	0	grosir	dashboard	0	T230406-14	TG230410-1	0	19358	132132
343161	0	retail	apps	0	T230406-14	4	0	20420	315000
343160	0	grosir	dashboard	0	T230406-14	TG230410-1	0	18812	131321
343159	0	grosir	dashboard	0	T230406-14	TG230410-1	0	18812	200000
343158	0	grosir	dashboard	0	T230406-14	TG230410-1	0	18802	264264
343157	0	grosir	dashboard	0	T230406-14	TG230410-1	0	18802	132132
343156	0	grosir	dashboard	0	T230406-13	TG230410-1	0	18812	500000
343154	0	retail	apps	0	T230406-13	8	0	20189	400000
343153	0	grosir	dashboard	0	2sx26x9sxi		0	19358	856642
343152	0	grosir	dashboard	0	if7rqfbvmm		0	19358	856642
343151	0	grosir	dashboard	0	T230406-13	7	0	19358	856642
343150	0	grosir	dashboard	0	T230406-13	6	0	19358	0
343149	0	grosir	dashboard	0	T230406-13	5	0	19358	0

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

Continuation of Table 1. Order_data.csv sample data.

id	Deduction_points	Discount_percent	discount	Superagent_discount	Supercoin_discount	reward	profit	Grand_total	Grand_total_before	status	Payment_status
343191	0	0	0	0	0	0	0	270000	0	0	0
343189	0	0	0	0	0	0	0	270000	0	0	0
343187	0	0	0	0	0	0	0	525284	0	0	0
343185	0	0	0	0	0	0	0	656605	0	0	0
343184	0	0	0	0	0	0	0	270000	0	0	0
343183	0	0	0	0	0	0	0	132132	0	0	0
343182	0	30	300000	0	0	0	0	804000	0	0	0
343181	0	30	300000	0	0	0	0	804000	0	0	0
343180	0	0	1000	0	0	0	0	1103000	0	0	0
343166	0	0	0	0	0	0	0	787926	0	0	0
343165	0	0	0	0	0	0	0	132132	0	-1	0
343164	0	0	0	0	0	0	0	131321	0	0	0
343163	0	0	0	0	0	0	0	381000	0	0	0
343162	0	0	0	0	0	0	0	132132	0	0	0
343161	0	0	1000	0	0	0	0	315000	0	0	0
343160	0	0	0	0	0	0	0	131321	0	0	0
343159	0	0	0	0	0	0	0	200000	0	0	0
343158	0	0	0	0	0	0	0	264264	0	0	0
343157	0	0	0	0	0	0	0	132132	0	0	0
343156	0	0	0	0	0	0	0	500000	0	0	0
343154	0	0	1000	0	0	0	0	399000	0	3	0
343153	0	0	0	0	0	0	0	856642	0	-1	0
343152	0	0	0	0	0	0	0	856642	0	-1	0
343151	0	0	0	0	0	0	0	856642	0	3	0
343150	0	0	0	0	0	0	0	0	856642	3	0
343149	0	0	0	0	0	0	0	0	856642	-1	0

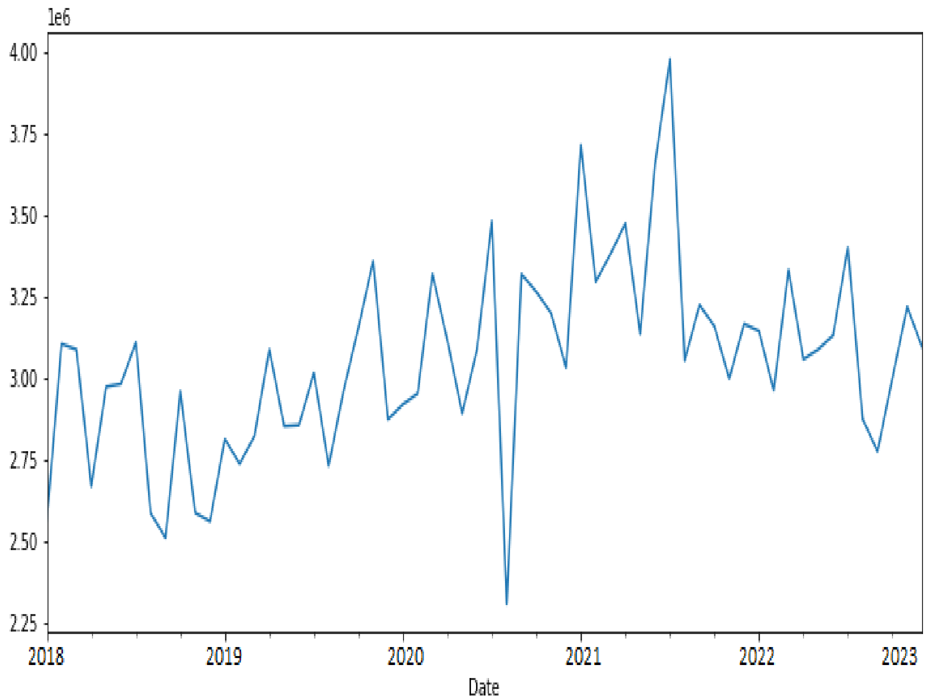
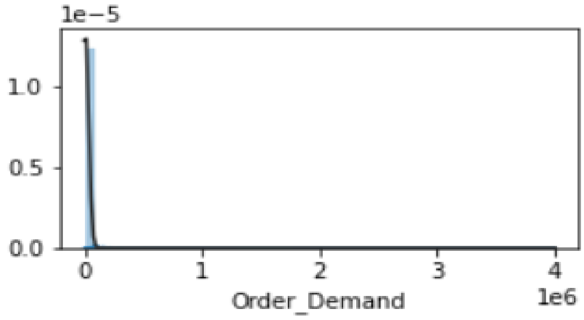


Figure 3. Visualization of time series forecast.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026



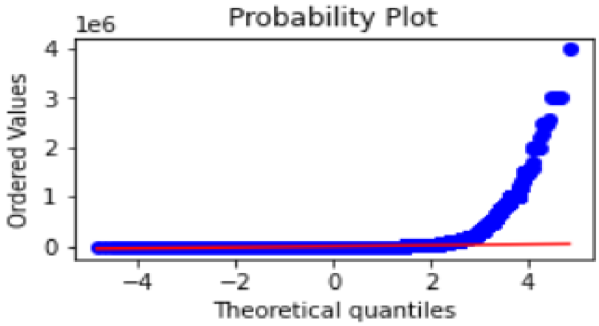


Figure 4. Visualization of probability plot

The plotted blue line in the visualization displays the real demand obtained from 2018 to 2023. The predicted demand line closely follows the same trend, implying that the linear regression model is an appropriate fit for the given data.

4.1. Comparing with Previous Research

Supply chain demand planning is a critical function in any organization that involves forecasting and planning future demand for products or services. The integration of predictive big data analytics in supply chain demand planning has gained significant attention in recent years, as it enables organizations to improve their forecasting accuracy and optimize their inventory levels. Linear Regression-based predictive models are commonly used in this area and have shown promising results.

The integration of predictive big data analytics in supply chain demand planning has been an increasingly important topic in recent years. The use of big data analytics enables organizations to make better decisions by uncovering patterns, trends, and insights in large volumes of data. Linear Regression-based predictive models have been widely used in various applications, including supply chain demand planning.

A study by Al-Jameel et al. (2018) and Kaleb et al (2026) proposed a Linear Regression-based model for demand forecasting in the pharmaceutical industry. The study used historical sales data and demographic data to train the model and demonstrated that the proposed model outperformed traditional methods and was effective in predicting demand for pharmaceutical products.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

Another study by Yu et al. (2019) and Abdulrahman et al. (2026) proposed a method for demand forecasting in the e-commerce industry using a Linear Regression-based model. The study used historical sales data, website traffic data, and promotional data to train the model and evaluated its performance using a holdout set of data. The study demonstrated that the proposed method was effective in improving the accuracy of demand forecasting in the e-commerce industry.

In the supply chain domain, a study by Zhu et al. (2020) proposed a predictive model for inventory optimization using a Linear Regression-based approach. The study used real-world data from a supply chain company to train and validate the model. The results showed that the proposed model was effective in optimizing inventory levels and reducing stock-outs.

In summary, the literature indicates that Linear Regression-based predictive models can be effective in supply chain demand planning. The integration of big data analytics can also improve the accuracy of demand forecasting. The proposed research aims to contribute to the literature by developing a Linear Regression-based predictive model integrated with big data analytics for supply chain demand planning.

4.2. Comparison

The comparison between Order demand forecast using linear regression-based predictive big data analysis on Python and the study by Al-Jameel et al. (2018) proposing a Linear Regression-based model for demand forecasting in the pharmaceutical industry is as follows:

1. Both approaches use Linear Regression to forecast demand.
2. Both approaches utilize historical data to train the model and predict future demand.
3. Both approaches aim to improve the accuracy of demand forecasting.
4. The first approach focuses on order demand forecasting, while the second approach is specific to demand forecasting in the pharmaceutical industry.

5. Conclusion

Overall, both approaches demonstrate the effectiveness of using Linear Regression for demand forecasting. However, the second approach provides more details on the specific data sources and comparison with traditional methods, making it more informative for the pharmaceutical industry. The first approach could benefit from providing more details on the specific data sources used and comparing the performance of the model with traditional methods.

6. References

Abdulrahman, N., Zainuddin, Z., & Nurtanio, I. (2026). A deep learning-based framework using CNN+LSTM for karate kata classification and correctness evaluation. Engineering, Technology & Applied Science Research, 16(1), 32619–32624. https://doi.org/10.48084/etasr.12147

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. and Seaman, O. (2019). Sales demand forecast in e-commerce using a long short-term memory neural network methodology," in Proc. Int. Conf. on Neural Information Processing (ICONIP), 2019, arXiv:1901.04028, doi: 10.48550/arXiv.1901.04028. https://link.springer.com/chapter/10.1007/978-3-030-36718-3_39

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026

Boone, T., Ganeshan, R., Jain, A. and Sanders, N.R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era," Int. J. Forecast., vol. 35, no. 1, pp. 170–180, Jan. 2019, doi: 10.1016/j.ijforecast.2018.09.003. https://ideas.repec.org/a/eee/intfor/v35y2019i1p170-180.html

Carbonneau, R., Laframboise, K. and Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. Eur. J. Oper. Res., vol. 184, no. 3, pp. 1140–1154, doi: 10.1016/j.ejor.2006.12.004. https://ideas.repec.org/a/eee/ejores/v184y2008i3p1140-1154.html

Choi, T.-M., Wallace, S. W. and Wang, Y. (2018). Big data analytics in operations management. Prod. Oper. Manag., vol. 27, no. 10, pp. 1868–1883, Oct. 2018, doi: 10.1111/poms.12838.Kaleb, V, Z. A., Wiranto H. U. (2026). Efficiency of Temporal Convolutional Networks in Karate Kata Evaluation: A Comparative Study. Informatika Mulawarman. 21(1. https://ideas.repec.org/a/bla/popmgt/v27y2018i10p1868-1883.html

Han, J., Kamber M. and Pei, J. (2011). Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann. https://ia800603.us.archive.org/2/items/datamining_201811/DS-book u5.pdf

Hazen, B.T., Skipper. J.B., Ezell, J.D. and Boone, C.A. (2016). Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda. Comput. Ind. Eng., vol. 101, pp. 592–598, Nov. 2016, doi: https://dl.acm.org/doi/10.1016/j.cie.2016.06.030

Hofmann, E., and Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: A conceptual analysis. Int. J. Logist. Manag., vol. 29, no. 2, pp. 739–766, May 2018, doi: 10.1108/IJLM-04-2017-0088.

Kaleb, V.Z.A and Wiranto, H.U. (2026). Efficiency of Temporal Convolutional Networks in Karate Kata Evaluation: A Comparative Study. Informatika Mulawarman. vol 21, no. 1.

Seyedan, M. and Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. J. Big Data, vol. 7, no. 1, art. 53, Jul. 2020, doi: 10.1186/s40537-020-00329-2.

Tiwari, S., Wee, H.M. and Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Comput. Ind. Eng., vol. 115, pp. 319–330, Jan. 2018, doi: 10.1016/j.cie.2017.11.017.

Gunasekaran, A., Tiwari, M. K., Dubey, R. and Fosso Wamba, S. (2016). Big data and predictive analytics applications in supply chain management. Comput. Ind. Eng., vol. 101, pp. 525–527, Nov. 2016, doi: 10.1016/j.cie.2016.10.020.

Villegas, M.A., Pedregal, D.J. and Trapero, J. R. (2018). A support vector machine for model selection in demand forecasting applications. Comput. Ind. Eng., vol. 121, pp. 1–7, Jul. 2018, doi: 10.1016/j.cie.2018.04.042.

Wang, G., Gunasekaran, A., Ngai, E.W.T and Papadopoulos, T., (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *Int. J. Prod. Econ.*, vol. 176, pp. 98–110, Jun. 2016, doi: 10.1016/j.ijpe.2016.03.014.

Received June 5, 2026; Revised August 20, 2026; Accepted for publication August 22, 2026