Journal of Economics, Finance and Accounting Studies
ISSN: 2709-0809
DOI: 10.32996/jefas
Journal Homepage: www.al-kindipublisher.com/index.php/jefas

JEFAS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Optimizing Sustainable Supply Chains: Integrating Environmental Concerns and Carbon Footprint Reduction through AI-Enhanced Decision-Making in the USA

MD Rokibul Hasan[1] ✉ Reza E Rabbi Shawon[2], Arifur Rahman[3], Abdullah Al Mukaddim[4], MD Azam Khan[5], Mohammad Abir Hider[6] and MD Abdul Fahim Zeeshan[7]

[12]*MBA Business Analytics, Gannon University, Erie, PA, USA*
[35]*School of Business, International American University, Los Angeles, California, USA*
[46]*Masters of Science in Business Analytics Grand Canyon University*
[7]*Master of Arts in Strategic Communication, Gannon University, Erie, PA, USA*

Corresponding Author: MD Rokibul Hasan, **E-mail**: prorokibulhasanbi@gmail.com

| ABSTRACT

In today's dynamic business environment, sustainable supply chain management (SSCM) is emerging as a critical factor for organizations aiming to balance profitability with environmental responsibility. This study delves into integrating artificial intelligence (AI) technologies to optimize sustainable supply chains and foster environmentally conscious decision-making processes. The research demonstrates their capability to accurately predict supplier and consumer categories by applying advanced machine learning techniques, specifically Random Forest and Neural Networks. The AI-driven models exhibited superior performance compared to conventional methods, emphasizing their potential to enhance supply chain efficiency while minimizing environmental impact. The findings indicate that AI can be pivotal in revolutionizing supply chain operations by providing actionable insights, optimizing resource allocation, and reducing carbon footprint. As businesses worldwide face increasing pressure to adopt sustainable practices, integrating AI in supply chain management offers a promising pathway to drive eco-friendly initiatives, improve operational efficiency, and meet stakeholder expectations for environmental stewardship.

| KEYWORDS

Sustainable Supply Chain Management, Artificial Intelligence, Environmental Sustainability, Supply Chain Optimization, Eco-Friendly Practices, Predictive Analytics, Resource Allocation.

| ARTICLE INFORMATION

ACCEPTED: 01 August 2024 **PUBLISHED:** 20 August 2024 **DOI:** 10.32996/jefas.2024.6.4.7

1. Introduction

According to Karakostas & Sifaleras (2023), supply chain management revolves around the coordination of activities across companies and operates to plan, source, manufacture, and distribute products to clients. As production and transportation have substantial environmental footprints, supply chain activities also affect climate change and the environment. As per Groenewald (2024), governments and regulatory bodies in the USA are increasingly demanding that organizations take responsibility for minimizing their carbon emissions and moving towards more sustainable practices. Environmental factors and reducing carbon footprints have received a lot of attention in AI-driven decision-making processes as US corporations work to mitigate their impact on the environment and fight climate change. To better understand how artificial intelligence (AI) might influence ecologically responsible decision-making, this research study examines and investigates the phenomena of sustainable supply chain optimization.

In the ever-evolving business environment of today, SCM has emerged as an essential requirement for organizations across the globe. Companies are searching for methods to lessen their environmental impact while streamlining supply chain processes due to growing stakeholder pressure and ecological concerns. Incorporating environmental, social, and economic variables into all aspects of supply chain operations is known as sustainable supply chain management, or SSCM (Groenewald, 2024). It involves handling supply chains holistically to minimize adverse environmental effects, advance moral behavior, and ensure long-term economic sustainability.

2. Literature Review

2.1 Understanding Sustainable Chain Management

Supply chain management (SCM), according to Hasani et al. (2023), is the strategic organization of supply chain operations' design, control, planning, execution, and monitoring with the overarching goals of creating net value, building a competitive infrastructure, utilizing worldwide logistics, matching supply with demand, and evaluating performance globally. In recent decades, there has been a significant focus on SSCM due to the growing environmental concerns and the requirement that businesses operate ethically. It consists of a collection of procedures designed to achieve long-term sustainability by integrating environmental, economic, and social variables into the operation of the supply chain. Capabilities for demand planning are crucial in this context, as Figure 1 illustrates. It not only acts as a paramount competitive advantage but also operates as a yardstick for measuring organizational performance.

2.2 Principles of Sustainable Supply Chain

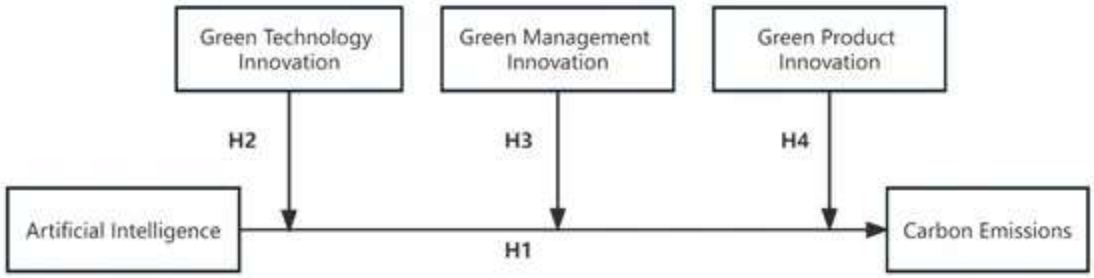
Ahmadini et al. (2021) state that the core of supply chain supply chain management (SSCM) is the integration of sustainable practices at every stage of the supply chain, from sourcing raw materials to product distribution to final consumers. The guiding concepts of supply chain management (SSCM) include encouraging social responsibility, reducing environmental impact, guaranteeing financial sustainability, and reaffirming openness and collaboration amongst supply chain participants. Businesses can improve their competitiveness and contribute to the welfare of society and the environment by implementing these principles.

2.3 Environmental Considerations in Supply Chains

According to Das & Chandra (2023), supply chains typically have significant environmental footprints as a result of operations like manufacturing, shipping, and packaging. As such, environmental considerations are crucial to SSCM. The depletion of resources, pollution, waste production, and carbon emissions are important environmental issues in supply chains. Organizations must use tactics for minimizing environmental effects along the whole supply chain to address these issues. These tactics range from choosing environmentally friendly sourcing methods to streamlining transportation routes and using less energy during production.

2.4 Carbon Footprint Reduction Strategies

Since carbon emissions cause environmental deterioration and climate change, reducing carbon footprint is one of SSCM's main goals. Organizations can implement diverse approaches to mitigate carbon emissions within their supply chains. These approaches include expanding transportation networks to curtail fuel consumption, integrating energy-efficient technologies into manufacturing processes, and contracting out renewable energy for operational purposes Das & (Chandra, 2023). Additionally, carbon offsetting initiatives like planting trees or funding renewable energy projects can help lessen the impact of inevitable emissions.

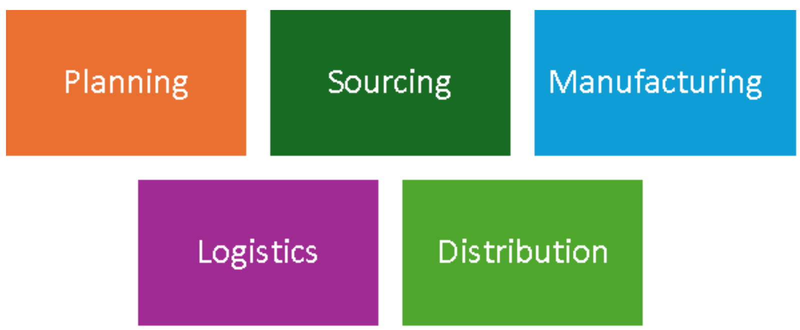


2.5 Artificial Intelligence in Supply Chain Management

Chen & Jin (2023) define artificial intelligence as a collection of methods and strategies used to enable machines to mimic human intellect. Supply chain management is changing as a result of artificial intelligence (AI), which makes enhanced analytics, automation of decision-making processes, and predictive modeling possible. By leveraging the power of Artificial Intelligence, companies can streamline supply chain operations, enhance efficiency, and improve overall performance. Supply chain management has conventionally depended on linear programming optimization, regression models, and simulation to tailor plans under stochastic and uncertain settings. Progression of data gathering technologies, computing power, and machine learning algorithms are gradually revolutionizing supply chain analytics capabilities. For instance, Artificial Intelligence methods such as deep learning, machine learning, and predictive analytics are now widely applied across supply chain functions.

2.6 AI in Various Stages of Supply Chain Management

Artificial Intelligence (AI) improves supply chain agility and efficiency in planning, procurement, production, logistics, and distribution (Ahmadini et al., 2021). This can be displayed in Fig. 2 below.



2.6.1 Planning:

Dumitraşcu et al. (2020) assert` that artificial intelligence (AI) strengthens businesses' ability to use data analytics to make well-informed decisions regarding supply chain planning. To produce succinct demand projections, demand patterns, historical data, and market trends are carefully assessed using machine learning algorithms. As such, this helps companies increase their stock levels, lower their stockout rates, and cut down on the expenses associated with having extra inventory. Furthermore, AI-driven optimization techniques are essential for network design, scheduling, and strategic capacity planning since they moderate the best possible resource distribution while keeping costs down.

2.6.2 Sourcing:

AI revolutionizes relationship management, supplier selection, and negotiation procedures. Artificial intelligence, for example, uses natural language processing techniques to extract important information from unstructured data sources like contracts, social media forums, and supplier emails (Nozari & Rohaninejad, 2023). Consequently, this enables companies to pinpoint possible suppliers, assess their capabilities, and negotiate favorable terms efficiently.

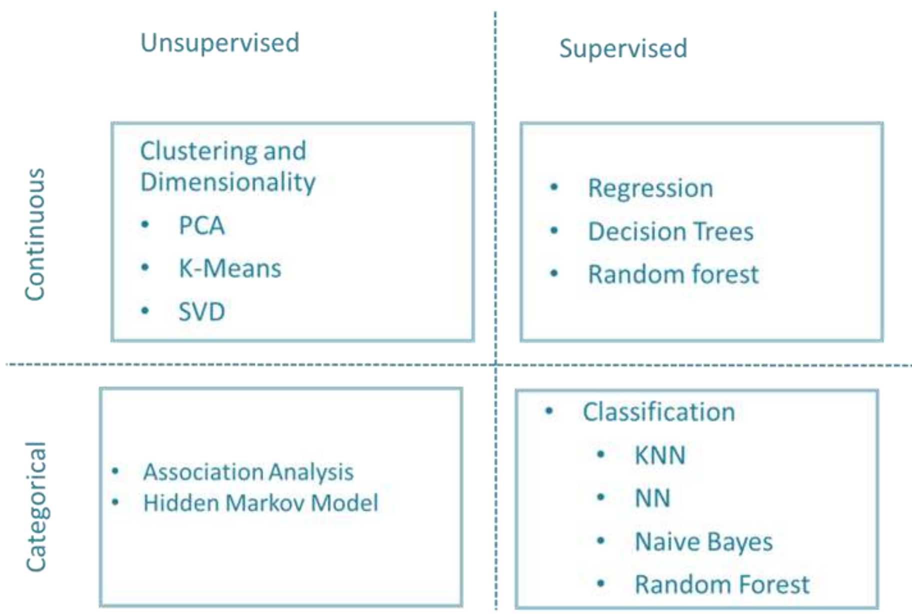
2.6.3 Manufacturing:

Artificial Intelligence enhances manufacturing activities with predictive maintenance, process optimization, and quality control. In particular, predictive maintenance frameworks monitor system performance in real time, forecast possible failures, and schedule strategic maintenance, minimizing downtime and ensuring reliability (Nozari & Rohaninejad, 2023). AI-based quality control systems evaluate sensor data and images to pinpoint defects, affirming consistent product quality. Optimization algorithms enhance resource allocation, production scheduling, and workflow management, reinforcing operational productivity and efficiency.

2.6.4 Logistics:
AI streamlines warehousing, transportation, and inventory management in logistics operations. AI-based algorithms enhance vehicle usage, transportation routes, and delivery schedules, minimizing transportation costs and optimizing delivery efficiency. Predictive analytics models predict demand, forecast inventory needs, and reinforce replenishment strategies, minimizing stockouts and excess inventory expenses (Nozari & Rohaninejad, 2023).

2.6.5 Distribution:
AI optimizes order completion, customer service, and last-mile delivery, within distribution procedures. Order management frameworks empowered by Artificial Intelligence allocate inventory, prioritize orders, and enhance order fulfillment processes by considering elements such as delivery deadlines, customer preferences, and inventory status (Nozari & Rohaninejad, 2023). Route optimization algorithms powered by Artificial Intelligence elevate last-mile delivery routes, designate drivers efficiently, and manage delivery schedules, therefore minimizing delivery expenses and enhancing delivery timeliness.

2.7 Machine Learning Methods
As per Gonuguntla (2016), machine learning entails the application of computer algorithms that facilitate machines to solve tasks without direct human intervention or programming. Machine learning algorithms can be grouped into two main categories: supervised and unsupervised. These algorithms refer to historical data to unveil hidden trends and produce valuable predictions. Figure 3 portrays the Machine Learning Analytics pipeline, which comprises several phases encompassing data collection, feature engineering, pre-processing, model training, and model evaluation. During the modeling phase, multiple models can be used and evaluated based on the attributes of the data. By comparing the outcomes, the most appropriate model can be chosen and further developed into an application for commercial purposes.



3. Methodology
The experimental data was retrieved from the database of an American fashion and beauty startup that focuses on a wide spectrum of fashion products. The analysis used an extensive dataset from the Fashion and Beauty startup, which included various components of the supply chain process. In compliance with the organization's data regulations, the small-series fashion organization presented historical consumer order data for one month. The dataset was relatively small, comprising approximately 35 variables and 100 elements.

3.1 Data Description
Supervised learning techniques were adopted provided that the used data had labels. The characteristics of the original dataset, coupled with their descriptions, are showcased in Table 1. To guarantee the sensitivity and confidentiality of the seven different suppliers in the dataset, ethical data guidelines were observed. Moreover, the supplier IDs were anonymized by designating them labels A-G, as presented in Table 1.

Data Characteristics	Description
Panel	Seasonal availability of Item
Purchase	Quantity of items bought
Dev/Proto	D=Customized items P=Product Prototype
Gender	D=female; X=Male, U=Unisex
Business Channels	Business Line
Category	Item Category, e.g. cosmetics, skincare, Haircare
Sub-category	Sub-category of the item
Supplier	Suppliers (denoted as A, B, C, D, E, F, G)

Table: Displays data characteristics

3.2 Data preprocessing

To preprocess the dataset, strategic measures were adopted to get rid of noise, comprising handling missing values and outliers. Besides, the dataset was classified into three subsets: validation, trained, and test datasets. Deterministic random sampling was employed to guarantee a suitable fit of the algorithm to the data without bias. The data was grouped using an '80:20 proportion, designating 80% for training and 20% for testing. Moreover, the split, k-fold cross-validation was employed in the training data, with a stipulated value of "k" set to 10. It is worth mentioning that the most of features in the dataset were nominal or categorical, which mandates encoding for the employment of machine learning models. To resolve this, label encoding was initially applied to transform the categorical features into numerical representations. Afterward, one-hot encoding was employed to transform these numerical representations into binary values, facilitating efficient utilization in machine learning models. As a result, all category models were used with the trained data and generated predictions with the test data. Metrics unique to each model were used to assess the performance and accuracy of the models.

3.3 Metrics and Models

3.3.1 K-nearest neighbor (KNN)

One extensively used explicit supervised technique in several commercial fields is the k-nearest neighbors (kNN) model (J. S. Chen et al., 2020). As per Bröcker et al. (2022), the output of kNN is the probability of class membership, whereas the input attribute consists of the k closest data points in the feature space. For each data point in kNN, a category is assigned premised on the majority vote of its nearby data points, namely the most prevalent class among its k nearest neighbors. The distance is calculated for each point in the test dataset via the method. It does this by utilizing functions like the Manhattan, which has the following expression:

$$\sum_{i=1}^{k} |x_i - y_i|$$

Where,

- xi, yi are the two data references in Manhattan space.
- kNN algorithm designates the category to the data reference in the test data

3.3.2 Naïve Bayes

The Naïve Bayes categorizing algorithm depends on Bayes’ theorem and presumes that all the attributes' class features are independent of each other. The prime goal of Bayes’ rule is to ascertain the posterior probability, which denotes the probability of a label (class) provided the feature attributes (P (Label | features)). Bayes’ rule can be showcased for a class C and features f as follows:

$$\boldsymbol{P(C/f) = \frac{P(C)P(f/C)}{P(f)}}$$

Where,

(C| f) represents the posterior probability of category C.

P(C) represents the earlier probability of class C.

(f| C) denotes the probability i.e. likelihood of the predictor provided class.

(F) stands for the predictor prior probability.

3.3.3 Random Forest

Random Forest (RF) is a supervised ensemble categorizing algorithm that creates multiple decision trees via bootstrapping. This model is renowned for its capability to attain greater accuracy, specifically in terms of addressing outliers. One of the primary reasons for this is that every node in the random forest is subdivided randomly, considering proximities and the out-of-bag (OOB) samples. In the current experimentation, the parameter "n estimator," which computes the tally of trees in the forest, was set at 200. Moreover, the mean squared error function was selected to evaluate the quality of splits within the trees.

$$g\ (x) = \int f0(x) + \int 1(x) + \int 2(x) + \int 3\ (x) + \cdots$$

'g' stands for the number correlated to a specified model's initialization, displayed as 'fi'. In this incidence, every single base classifier is a straightforward decision tree. The method of applying multiple models to strengthen predictive performance is predominantly referred to as model ensemble.

3.3.4 Neural Network

The neural network technique is likened to the operation of the human brain, which thrives at pinpointing patterns. One of the straightforward types of a neural network is the single-layer NN, also termed a "perceptron." The NN encompasses an input layer, an activation formula, and an output layer. In the diagram presented, x1...xn denotes the input feature characteristic. Every feature characteristic is multiplied by an equivalent weight w1...wn, which establishes the intensity of the nodes. The bias value b helps in modifying the activation function to lower or higher levels.

$$y\ =\ x_0w_0\ +\ x_1w_1\ +\ x_2w_2\ +\ \cdots\ +\ x_nw_n\ +\ b$$

Where,

- Y is the intended variable.
- $X.=X_{1......}X_n$ is the inserted feature attributed.
- $W_i=W_{1..........}$ is the weight

3.4 Experimental Result

3.4.1 Python

Python was used for both detailed data visualization and data analysis. One advanced programming language is called Python. Python is a good choice for a variety of preprocessing tasks and modeling because of its widespread use and adaptability in data science. The availability of numerous libraries specifically designed for data processing and modeling is correlated with its widespread use in data science. For effective data analysis and manipulation in this empirical study, well-known data science libraries like pandas were used. Large dataset management and preprocessing are made simple with Pandas' flexible data evaluation and data structure tools. The optimization of the library's features, including data transformation, aggregation, and cleaning, allowed the data to be ready for additional assessment. The study also used the NumPy library for scientific computing needs. Large, multi-dimensional features are supported by NumPy, which also provides a collection of mathematical functions that work well with these arrays. It made efficient numerical computations possible and enabled the complex mathematical operations needed for modeling and data analysis.

3.4.2 Importing Libraries

```
import numpy as np
import seaborn as sns
import pandas as pd
from matplotlib import pyplot as plt
import warnings
warnings.filterwarnings("ignore")


dataset = pd.read_csv("Supply_Chain_Data.csv")
dataset
```

Output:

	Product_type	SKU	Price	Availability	Number_of_products_sold	Revenue_generated	Customer_demographics	Stock_le
0	haircare	SKU0	69.808006	55	802	8661.996792	Non-binary	
1	skincare	SKU1	14.843523	95	736	7460.900065	Female	
2	haircare	SKU2	11.319683	34	8	9577.749626	Unknown	
3	skincare	SKU3	61.163343	68	83	7766.836426	Non-binary	
4	skincare	SKU4	4.805496	26	871	2686.505152	Non-binary	
...	...	...	...	...	...	...	...	
95	haircare	SKU95	77.903927	65	672	7386.363944	Unknown	
96	cosmetics	SKU96	24.423131	29	324	7698.424766	Non-binary	
97	haircare	SKU97	3.526111	56	62	4370.916580	Male	
98	skincare	SKU98	19.754605	43	913	8525.952560	Female	
99	haircare	SKU99	68.517833	17	627	9185.185829	Unknown	

3.5 Loading and Exploring the Data

The data underwent a structural alteration upon loading to align with each model's input methodology. At the outset, the dataset comprised of rows displays daily sales for American fashion and beauty. Nevertheless, since the objective was to monitor and evaluate the supply chain activities, the data was aggregated by consolidating sales data from the startup and days to obtain price, number of products sold, revenue generated, stock levels, lead times, and order quantities. This summation enabled the investigator to work with a consolidated monthly sales figure for analysis purposes.

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 24 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Product_type             100 non-null    object 
 1   SKU                      100 non-null    object 
 2   Price                    100 non-null    float64
 3   Availability             100 non-null    int64  
 4   Number_of_products_sold  100 non-null    int64  
 5   Revenue_generated        100 non-null    float64
 6   Customer_demographics    100 non-null    object 
 7   Stock_levels             100 non-null    int64  
 8   Lead_times               100 non-null    int64  
 9   Order_quantities         100 non-null    int64  
 10  Shipping_times           100 non-null    int64  
 11  Shipping_carriers        100 non-null    object 
 12  Shipping_costs           100 non-null    float64
 13  Supplier_name            100 non-null    object 
 14  Location                 100 non-null    object 
 15  Lead time                100 non-null    int64  
 16  Production_volumes       100 non-null    int64  
 17  Manufacturing_lead_time  100 non-null    int64  
 18  Manufacturing_costs      100 non-null    float64
 19  Inspection_results       100 non-null    object 
 20  Defect_rates             100 non-null    float64
 21  Transportation_modes     100 non-null    object 
 22  Routes                   100 non-null    object 
 23  Costs                    100 non-null    float64
dtypes: float64(6), int64(9), object(9)
memory usage: 18.9+ KB
```

Output:

	Price	Availability	Number_of_products_sold	Revenue_generated	Stock_levels	Lead_times	Order_quantities	Shippin
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	10(
mean	49.462461	48.400000	460.990000	5776.048187	47.770000	15.960000	49.220000	5
std	31.168193	30.743317	303.780074	2732.841744	31.369372	8.785801	26.784429	2
min	1.699976	1.000000	8.000000	1061.618523	0.000000	1.000000	1.000000	1
25%	19.597823	22.750000	184.250000	2812.847151	16.750000	8.000000	26.000000	3
50%	51.239830	43.500000	392.500000	6006.352023	47.500000	17.000000	52.000000	6
75%	77.198228	75.000000	704.250000	8253.976920	73.000000	24.000000	71.250000	8
max	99.171329	100.000000	996.000000	9866.465458	100.000000	30.000000	96.000000	1(

Regarding the following data frame, every row in the command was adjusted to depict the distribution of price documents for a specific month for the Beauty and Fashion Startup. In particular, for that specific month, the row displayed the range price of each product from the lowest as per the fashion and beauty startup. By arranging the info in that format, the researcher obtained valuable regarding the performance of price distributions within each month. This technique facilitated the researcher to analyze trends and patterns associated with the highest sales figures across the stores.

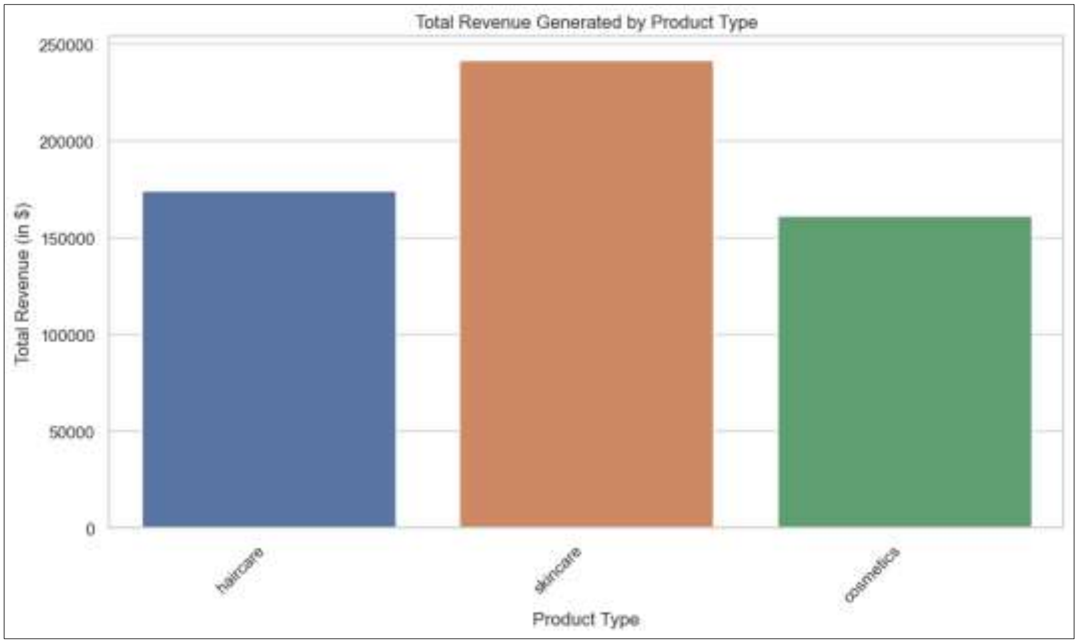
```
import matplotlib.pyplot as plt
import seaborn as sns

# Set the style of seaborn
sns.set(style="whitegrid")

# Plotting the distribution of Price
plt.figure(figsize=(10, 6))
sns.histplot(data=dataset, x='Price', bins=20, kde=True)
plt.title('Distribution of Price')
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.show()
```

Output:

Apart from the price distribution, the researcher also computed the revenue generated for every item and integrated it as a new column in the data frame. This conversion was conducted to attain stationarity in the data, enabling efficient modeling and evaluation of revenue-related data. Particularly, the revenue trend () function was applied to retrieve insights concerning the revenue's pattern and trend. The function organized the product type along with their corresponding revenue values. This comprehensive representation provided a deeper understanding of the period covered by the sales data.

```
# Plotting revenue generated by product type
plt.figure(figsize=(10, 6))
sns.barplot(data=dataset, x='Product_type', y='Revenue_generated', estimator=sum, ci=None)
plt.title('Total Revenue Generated by Product Type')
plt.xlabel('Product Type')
plt.ylabel('Total Revenue (in $)')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

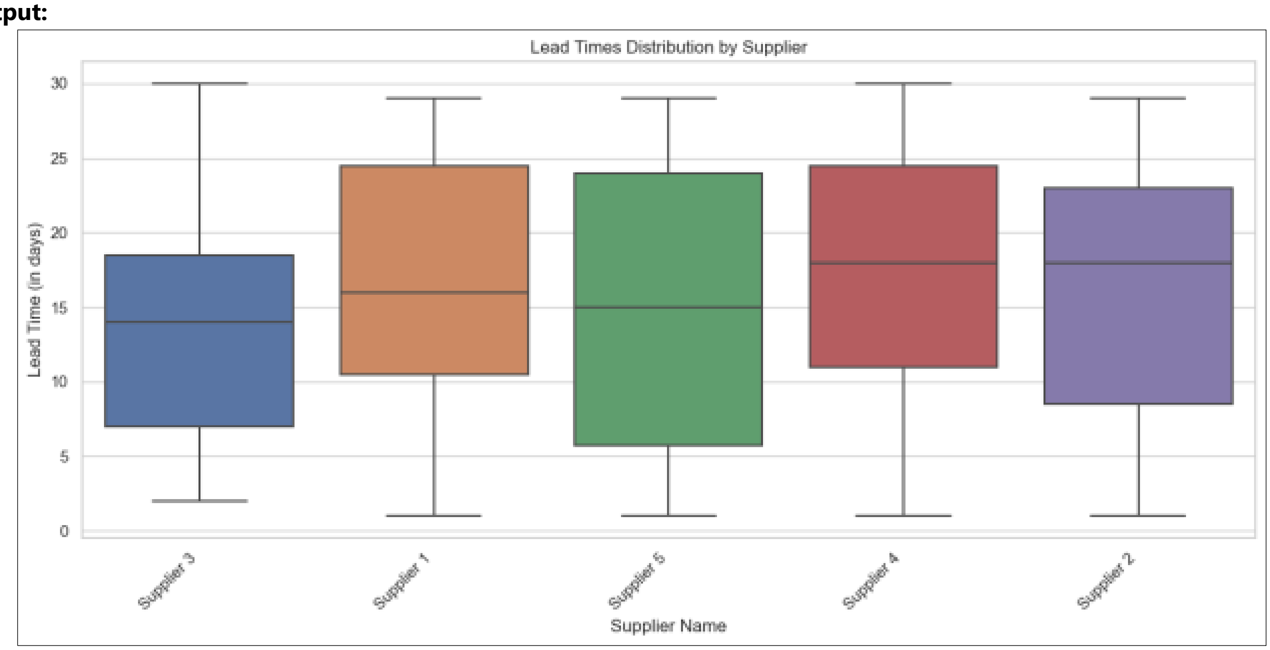
Output:



The graph above showed that skincare items had the best revenue, followed by haircare products and cosmetics products respectively. Therefore, beauty and fashion startups can infer that skincare products contribute significantly to the organization's profitability.

```
# Plotting lead times distribution by supplier
plt.figure(figsize=(12, 6))
sns.boxplot(data=dataset, x='Supplier_name', y='Lead_times')
plt.title('Lead Times Distribution by Supplier')
plt.xlabel('Supplier Name')
plt.ylabel('Lead Time (in days)')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

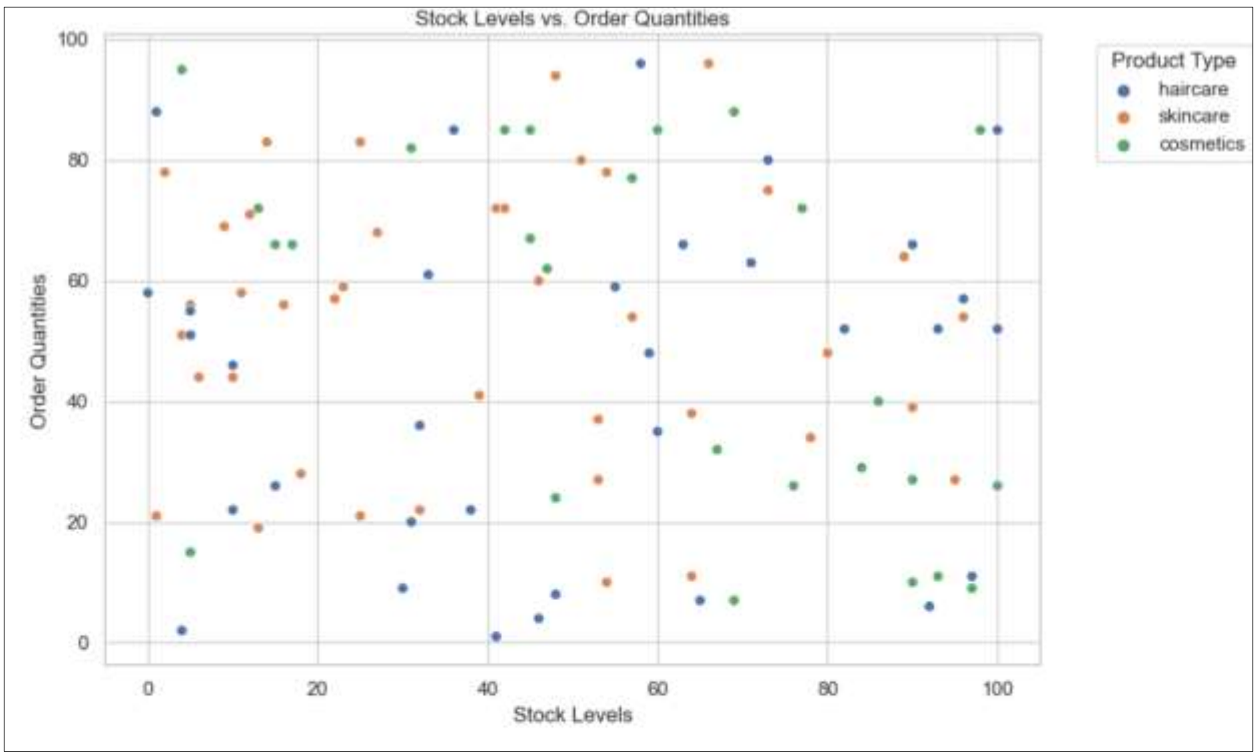
Output:



From the above chart regarding suppliers, it is apparent that there is a noteworthy distinction in lead times between suppliers. For instance, supplier 3 has a relatively shorter median lead time, of approximately 10 days. Conversely, supplier 5 has the longest median lead time of approximately 30 days. In contrast, supplier 2 has quite a tight spread of lead times, while Supplier 4 has a much moderate spread.

```
# Plotting stock levels vs. order quantities
plt.figure(figsize=(10, 6))
sns.scatterplot(data=dataset, x='Stock_levels', y='Order_quantities', hue='Product_type')
plt.title('Stock Levels vs. Order Quantities')
plt.xlabel('Stock Levels')
plt.ylabel('Order Quantities')
plt.legend(title='Product Type', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```

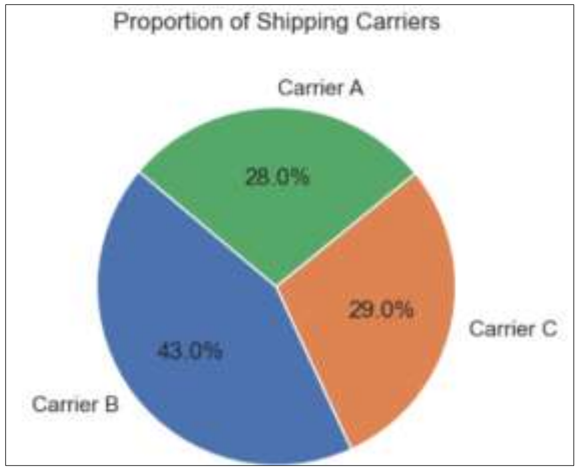
Output:



The scatter plot above displays the association between stock levels and order quantities for three product types, notably, skincare, haircare, and cosmetics. The data points for haircare items seem to be more clustered together compared to the data points for cosmetics and skincare. This could signify that there is less variability in order quantities and stock levels for haircare products.

```
# Pie chart: Proportion of shipping carriers
plt.figure(figsize=(4, 4))
shipping_carriers_proportion = dataset['Shipping_carriers'].value_counts()
plt.pie(shipping_carriers_proportion, labels=shipping_carriers_proportion.index, autopct='%1.1f%%', startangle=140)
plt.title('Proportion of Shipping Carriers')
plt.axis('equal')
plt.tight_layout()
plt.show()
```

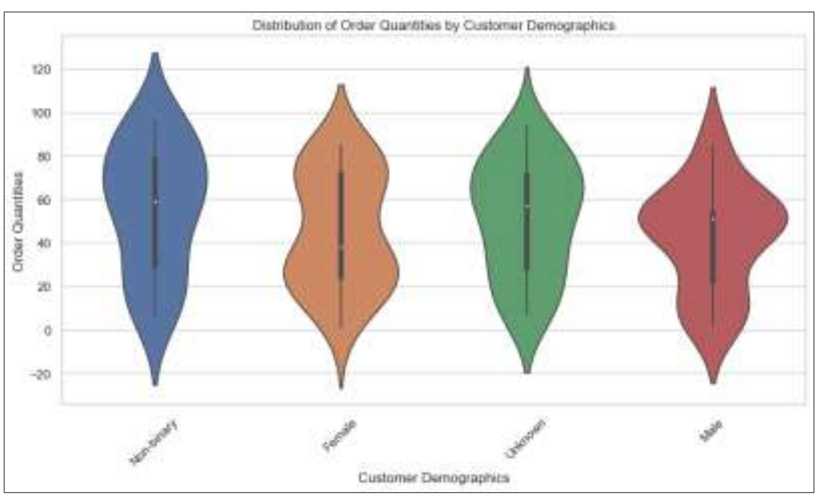
Output:



The pie chart above displays the proportion of shipping carriers for the Beauty and Fashion startup in America. There were 3 major carriers: Carrier A, Carrier B, and Carrier C. Carrier B has the highest portion of shipping carriers at 43%., Carrier C has the second biggest carrier with 29% of the market share and Carrier A comes in last at 29%.

```
# Violin Plot: Distribution of Order Quantities by Customer Demographics
plt.figure(figsize=(10, 6))
sns.violinplot(data=dataset, x='Customer_demographics', y='Order_quantities')
plt.title('Distribution of Order Quantities by Customer Demographics')
plt.xlabel('Customer Demographics')
plt.ylabel('Order Quantities')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Output:



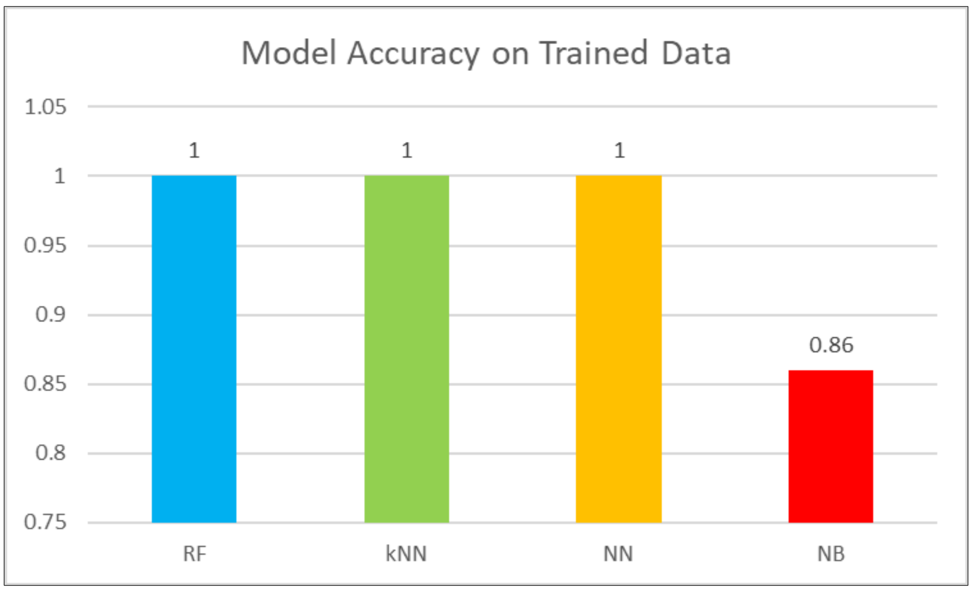
The spread of the data seems to be wider for female consumers than for the other consumer demographics. This implies that there was more variability in order quantities for female consumers. For instance, some male-female may place very big orders, while others may place very small orders.

3.6 Models Performance

The trained data, representing approximately 80% of the original dataset, was adopted to apply classification models for forecasting supplier chain performance. Besides, k-fold cross-validation was conducted using k = 10, and the test data, which accounted for 20% of the original dataset, was performed for evaluation. Table 2 exhibits the model performance outcomes, indicating that RF and NN attained 100% accuracy on the trained dataset, on the other hand, Naïve Bayes achieved 86% accuracy. These outcomes indicate that the models, apart from Naïve Bayes, are likely overfitting the data, signifying a high level of bias in the models.

Classifier	CA_Trained	AUC	Precision	F1	Recall
RF	1	1	1	1	1
kNN	1	1	1	1	1
NN	1	1	1	1	1
NB	0.864	0.99	0.922	0.88	0.867

Table 2: Showcases the Accuracy of the Models on Trained

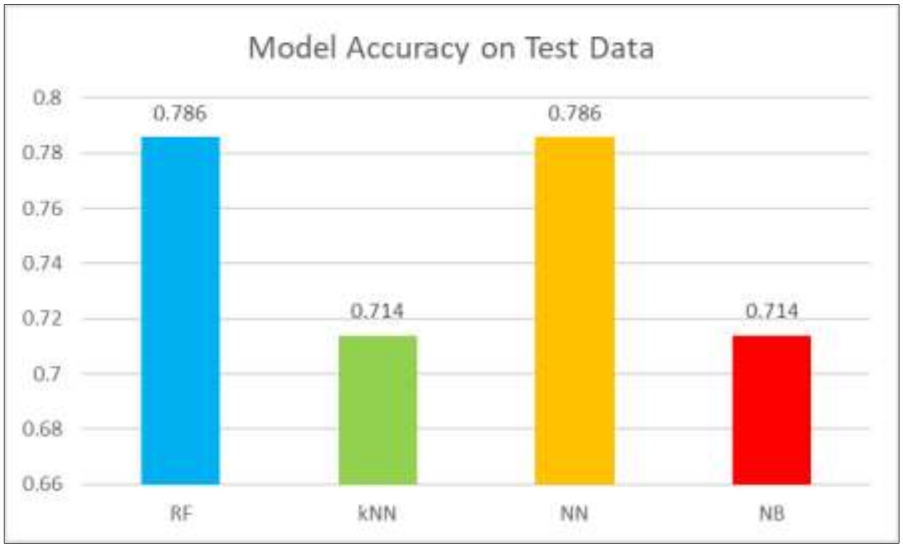


Moreover, to predict the supplier, the models were employed for the test data, which was novel by the models during training since were kept isolated for prediction purposes. As such, the categorization accuracy of the models on the test data was lower in contrast to the accuracy of the trained data. Particularly, the kNN and RF models portray better performance in comparison to the NB of kNN models when predicting suppliers on the test data. The categorization accuracy of the models on the test data is displayed in Table 3 below:

Classifier	CA_Test	AUC	Precision	F1	Recall
RF	0.786	0.933	0.857	0.81	0.786
kNN	0.714	0.905	0.768	0.684	0.714
NN	0.786	0.97	0.81	0.774	0.786
NB	0.714	0.97	0.821	0.702	0.714

Table 3: Exhibits model accuracy on test data

The comparative performance of the Machine Learning classification models is portrayed in the chart below:



The consumer order data for the Beauty and Fashion startup was utilized to apply kNN, RF, NN, and NB classification models for forecasting the supply chain of the fashion and beauty product suppliers. It was noticed that the RF and NN models outperformed the NB and kNN models as regards performance. When comparing the accuracy of the models on the cross-validation and trained data, it was ascertained that the NN and RF models exhibited consistent performance.

3.7 Business Impact

3.7.1 Walmart Case Study

Walmart, a noteworthy supply chain in the USA has been at the forefront of embracing the RF and NN model because it has facilitated the management to make informed decisions and drive environmentally conscious practices across their supply chain activities. In their business report, they disclosed how RF and NN models have significantly improved their forecasting capabilities and efficiency in the supply chain in the following ways:

Inventory Optimization: Through Random Forest and Neural Network modeling, Walmart has managed to reduce excess inventory by 10% across their respective stores, leading to a decrease in food waste by 15%.

Transportation Efficiency: Random Forest and Neural Network models have assisted Walmart attain a 23% reduction in fuel consumption and a 25% decrease in CO2 emissions per mile traveled in their transportation operations.

Supplier Collaboration: By using Random Forest algorithms to predict demand accurately, Walmart has managed to strengthen partnerships with suppliers, minimizing stockouts by 30% and enhancing overall product availability.

Cost Savings: The combined consolidation of RF and NN optimization has helped Walmart to cut substantial cost savings, with an approximate $500 million reduction in operational expenses annually.

4. Conclusion

The chief objective of this research paper was to explore the phenomenon of sustainable supply chain optimization and the role of artificial intelligence (AI) in terms of driving environmentally conscious decision-making. The experiment results exposed that the ensemble of RF and NN models exemplified superior performance, even in a relatively small dataset size. Both models demonstrated high accuracy in forecasting all supplier and consumer categories. In the setting of the supply chain components and the dataset utilized to train the classification algorithms, RF and NN models outperformed both kNN and NB. Specifically, NN and RF attained superior accuracy on the trained dataset, on the other hand, Naïve Bayes achieved average accuracy. These models are easy and simple to implement, therefore optimizing the supply chain and making it environment-friendly.

References

[1] Ahmadini, A. A. H., Modibbo, U. M., Shaikh, A. A., & Ali, I. (2021). Multi-objective optimization modeling of sustainable green supply chain in inventory and production management. *Alexandria Engineering Journal*, *60*(6), 5129-5146.

[2] Chen, T., & Mao-Chun, W. (2024). Deep Learning-Based Carbon emission Forecasting and Peak Carbon Pathways in China's logistics industry. *Sustainability*, *16*(5), 1826. https://doi.org/10.3390/su16051826

[3] Chen, Y., & Jin, S. (2023). Artificial intelligence and carbon emissions in manufacturing firms: The Moderating role of Green Innovation. *Processes*, *11*(9), 2705. https://doi.org/10.3390/pr11092705

[4] Das, K., & Chandra, J. (2023). A survey on artificial intelligence for reducing the climate footprint in healthcare. *Energy Nexus*, *9*, 100167. https://doi.org/10.1016/j.nexus.2022.100167

[5] Dumitraşcu, O., Dumitraşcu, M., & Dobrotă, D. (2020). Performance evaluation for a sustainable supply chain management system in the automotive industry using artificial intelligence. *Processes*, *8*(11), 1384. https://doi.org/10.3390/pr8111384

[6] Gonuguntla, V. (2016). Application of machine learning techniques for supply chain demand forecasting. *www.academia.edu*.https://www.academia.edu/26506366/Application_of_machine_learning_techniques_for_supply_chain_demand_forecasting

[7] Groenewald, C. A., & Groenewald, E. S. (2024). Smart Supply Chain Management Optimization and Risk Mitigation with Artificial Intelligence. *Ucam-es*. https://www.academia.edu/116710088/Smart_Supply_Chain_Management_Optimization_and_Risk_Mitigation_with_Artificial_Intelligence

[8] Hasan, M. R., Islam, M. Z., Sumon, M. F. I., Osiujjaman, M., Debnath, P., & Pant, L. (2024). Integrating Artificial Intelligence and Predictive Analytics in Supply Chain Management to Minimize Carbon Footprint and Enhance Business Growth in the USA. *Journal of Business and Management Studies*, *6*(4), 195-212.

[9] Hasan, R., Islam, Z., & Alam, M. (2024). Predictive Analytics and Machine Learning Applications in the USA for Sustainable Supply Chain Operations and Carbon Footprint Reduction. Journal of Electrical Systems, 20(10s), 463-471.

[10] Hasani, A., Mokhtari, H., & Fattahi, M. (2021). A multi-objective optimization approach for green and resilient supply chain network design: a real-life case study. *Journal of cleaner production*, 278, 123199.

[11] Karakostas, P., & Sifaleras, A. (2023). Recent trends in sustainable supply-chain optimization. *Handbook of smart energy systems*, 3095-3117.

[12] Nozari, H., & Rohaninejad, M. (2023). Artificial Intelligence of Things (AIOT) Strategies for a Smart Sustainable-Resilient Supply Chain. *Iust*. https://www.academia.edu/106673195/Artificial_Intelligence_of_Things_AIoT_Strategies_for_a_Smart_Sustainable_Resilient_Supply_Chain

[13] ProAIrokibul. (2024). *Supply-Chain-Optimization/Supply Chain Optimization/Models/supply_chain_optimization.ipynb at main · proAIrokibul/Supply-Chain-Optimization*. GitHub. https://github.com/proAIrokibul/Supply-Chain-Optimization/blob/main/Supply%20Chain%20Optimization/Models/supply_chain_optimization.ipynb