ISSN No:-2456-2165

Data Science and Machine Learning: Application of Machine Learning Models to Improve Supply Chain Management of Organization: Inyange Industries, Rwanda

Kagame Fred [1] Master's, Dr. Musoni Wilson[2]
[1, 2] Master of Science In Information Technology (MICT), University of Kigali, Kigali, Rwanda.

Abstract:- For the modern industrial sector, data created by machine learning and devices, product lifecycle management (PLM) tools, production planning tools, or quality and inventory control tools has reached a volume of more than a thousand Exabyte yearly and is anticipated to rise in the next years. To store, manage, analyze, interpret, and visualize such a large volume of data, Data technologies are now required. Supply Chains (SC) are a network of locations that connect a variety of enterprises. To reduce the overall cost of the supply chain, these organizations should cooperate. This necessitates that these entities cooperate, integrate, and share information. However, there is still a disconnect between the supply chain network's ideal and actual states. the digital transformation of the supply chain is needed today more than ever. The Digital Transformation has emerged as an important preoccupation and a key strategic matter for all kinds of organizations. One of the causes could be that producers increased output in expectation of increased demand despite not knowing the consumers' actual need. The goal of this study is to identify several business applications of machine learning (ML) in supply chain management. The study examines instances of supply chain optimization that make use ofmachine learning.

I. INTRODUCTION

Supply Chains management are the network of facilities that not only includes retailers, distributers, transporters, manufacturers but also the customers. As a result, it is critical to understand the true consumption and needs of customers, as they are the primary node of every supply chain, pushing various entities to produce and distribute. The supply chain facilities have now realized how crucial coordination and teamwork are to meeting the actual demand. The groups collaborate to reduce the overall cost of the supply chain.

II. METHODOLOGY DATA ANALYSIS

Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes.

- *Cleaning the Data*

Data cleaning is the process of repairing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data from a dataset in the context of data science and machine learning. Data cleaning is the process of filtering and modifying data to make it easier to explore, understand, and model.removing the parts you don't want or need so you don't have to look at or process them Modifying the parts that you do require but aren't in the format that you require in order to use them properly. The dataset used needed the following changes to be considered clean:

-Dropping of Rows with Null Values including NaN and null rows

-Removal of all negative numbers and replacing them with absolute values

-One Hot Encoding

- *Null Values*

The majority of data science algorithms do not tolerate nulls (missing values). As a result, somethingmust be done to eliminate them before or during data analysis. There are numerous methods for dealing with nulls. Which techniques are appropriate for a given variable can be heavily influenced by the algorithms you intend to use, as well as statistical patterns in the raw data, particularly missing values andthe randomness of their locations. Furthermore, in a given data set, different techniques may be appropriate for different variables. It is sometimes advantageous to apply multiple techniques to a single variable. Finally, corrupt values are typically treated as nulls.

The figure below shows the rows in the dataset used in this project and the number of missing values in each row.

```
Train:

Date                          0
Product_ID                    0
SUPPLIER                      0
Product_DESCRIPTION           0
Product_TYPE                  0
Total_Quantity_ordered        0
Branch_Location               0
Product_Class                 0
Branch_Size                   0
consumers                     0
stock                         0
Unit_Price                    0
Total_Quantity_ delivered     0
CustomerSatifaction           0
dtype: int64

 Date                         0.0
Product_ID                    0.0
SUPPLIER                      0.0
Product_DESCRIPTION           0.0
Product_TYPE                  0.0
Total_Quantity_ordered        0.0
Branch_Location               0.0
Product_Class                 0.0
Branch_Size                   0.0
consumers                     0.0
stock                         0.0
Unit_Price                    0.0
Total_Quantity_ delivered     0.0
CustomerSatifaction           0.0
dtype: float64
```

Fig 1 Null Rows dataset

It is clear that The value 0 (all bits at zero) is a typical value used in memory to denote null. It means that there is no absence of data or simply in these an impact on the algorithm implemented. So the best option is to remove all the rows with Null Values as this is a large dataset and removing these rows have not too much of an impact on the algorithm implemented.

➢ *Categorical Encoding*

Categorical data are the variables that consist of the label values instead of numerical values. Thistype of variables is generally called as Nominal Variables. Some of the machine learning algorithms like decision trees do not require any numerical as it is able to learn from categorical data but most of the machine learning models require input and output variables to be numeric, according to new research models Categorical data must be converted into a numerical data format. Following are the several encoding techniques for the transformation of data According to new research.

SUPPLIER	Product_DESCRIPTION	Product_TYPE	Total_Quantity_ordered	Branch_Location	Product_Class	Branch_Size	consumers	stock	Unit_Price	Total_Quantity_ delivered	CustomerSatifaction
ORANGE KINK BUSINESS LTD	Inyange Orange Juice 300mL	JUICE	43142.04884	EAST BRANCH	Less fat	Level1	11.0	22.0	3922	43142.04884	Bad
ORANGE KINK BUSINESS LTD	Inyange Orange Juice 500mL	JUICE	9093.91728	EAST BRANCH	Normal	Level1	12.0	24.0	758	9093.91728	Good
ORANGE KINK BUSINESS LTD	Inyange Pineapple juice 300mL	JUICE	28904.23380	EAST BRANCH	Less fat	Level1	13.0	26.0	2223	28904.23380	Good
ORANGE KINK BUSINESS LTD	Inyange Pineapple juice 500mL	JUICE	40024.48100	EAST BRANCH	Normal	Level2	14.0	28.0	2859	40024.48100	Good
ORANGE KINK BUSINESS LTD	Inyange mango juice 500mL	JUICE	12684.35970	EAST BRANCH	Less fat	Level3	15.0	30.0	846	12684.35970	Good

Table 1 *Categorical data*

➢ *One Hot Encoding*

A one hot encoding is a representation of categorical variables as binary vectors. This first requires that the categorical values be mapped to integer values. Then, each integer value is represented as a binary vector that is all zero values except the index of the integer, which is marked with a 1

SUPPLIER	Product_DESCRIPTION	Product_TYPE	Total_Quantity_ordered	Branch_Location	Product_Class	Branch_Size	consumers	stock	Unit_Price	Total_Quantity_ delivered	CustomerSatifaction
0	0	1	715	0	0	0	6	6	2737	715	0
5	1	1	49	0	1	0	7	7	227	49	1
5	12	1	421	0	0	0	8	8	1385	421	1
5	13	1	674	0	1	1	9	9	1980	674	1
5	14	1	91	0	0	2	10	10	307	91	1

Table 2 After applying one hot encoding

ISSN No:-2456-2165

➢ *Data Visualization*

Data visualization is the graphical representation of information and data. Data visualization tools, which use visual elements such as charts, graphs, and maps, make it easy to see and understand trends, outliers, and patterns in data. Data visualization tools and technologies are critical in the Big Data world for analyzing massive amounts of data and making data-driven decisions.

➢ *multivariate plots*

Designed to reveal the relationship among several variables simultaneously. As was the case when examining relationships among pairs of variables, there are several basic characteristics of the relationshipamong sets of variables that are of interest.

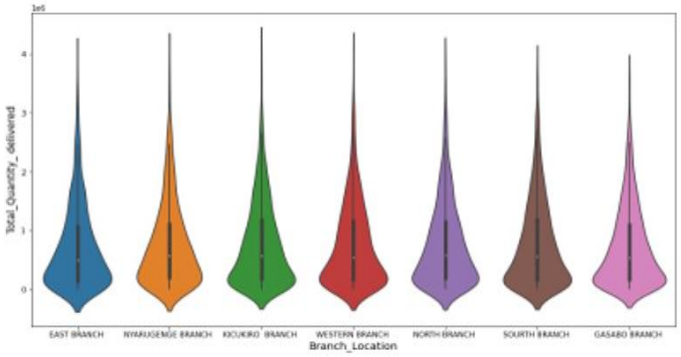

Fig 2 Branch location Vs Total Quantity delivered in The Entire 3 Years

According to Figure Among the seven branches, Kicukiro branch had the largest total quantity deliveredin 2018, 2019, and 2020, while the Gasabo branch had the lowest total quantity delivered over the periodof the three years.

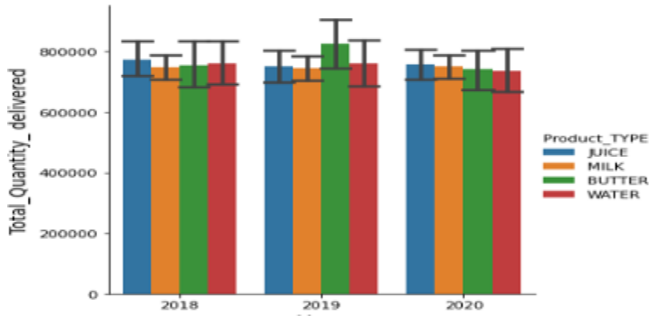

Fig 3 Graph of year in each product type Vs Total Quantity delivered

From Figure shows the total quantity delivered in each year for each product type show that In2018 the highest total quantity delivered is Juice and the lowest total quantity delivered is butter. In 2019 the highest total quantity delivered is Butter and the lowest total quantitydelivered is water. In 2020 the highest total quantity delivered milk and the lowest total quantity delivered water.

III. RESULT

➢ *Machine learning models*

First of all we divide our dataset into two variables X as the features we defined earlier and y asthe consumer the target value we want to predict.

➢ *Machine Learning Models we used:*

Linear Regression, Random Forest Regressor, Lasso Regressor, Gradient Boosting Regressor, Decision Tree Regressor, Ridge Regressor

➢ *Deep Learning Model We used*

Artificial Neurol Network

➢ *The Process of Modeling the Data:*

Importing the model, Fitting the model, Predicting demand forecasting, Regression metrics Score Metrics for Regression: Mean Absolute Error (MAE) - Mean of the absolute value oferrors (absolute distance from true value):

Mean Squared Error (MSE) - Mean of the squared value of errors (squared distance from truevalue): R^2 (coefficient of determination) - Regression score function.

➢ *Linear Regression*

Linear regression is a statistical method for modeling the relationship between a scalar response(or dependent variable) and one or more explanatory variables (or independent variables).

➢ *Future Warning.*

Mean Absolute Error: 9.07 Mean Squared Error: 127.71 R^2 Score: 0.845

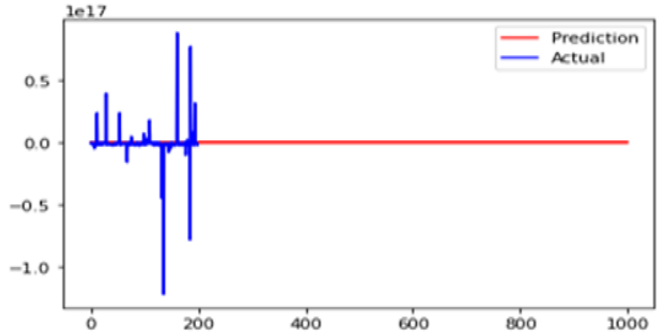

Fig 4 Linear Regression graph

Actuals Values	prediction
50.630100	41
64.282223	72
63.519379	74
62.083842	63
44.212371	34

100.725062	121
70.050853	87
Actuals Values	prediction
62.477807	67
67.682502	73

Table 3 Predicted Linear Regression

➢ *Random Forest Regressor*

Random forest is a Supervised Learning algorithm which uses ensemble learning method forclassification and regression. The following are the results obtained by the Random Forest Regressor:

Mean Absolute Error: 0.71 Mean Squared Error: 1.21 R^2 Score: 0.9985

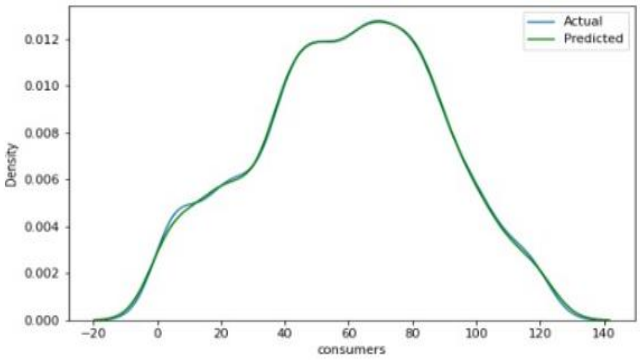

Fig 5 Graph Of Random Forest Regressor

Actual	predicted
41.699631	41
71.859016	72
73.858390	74
63.492909	63
34.386597	34
117.422593	121
87.826783	87
50.420528	50
67.062688	67
73.673656	73

Table 4 Actual And Predicted Random Forest Regressor Table

➢ *Lasso Regressor*

Lasso Regressor is trained with the dataset by using 5-fold time series cross-validation approach where80% of the data was used for training and 20% of the data was used as the test set and the performances have been measured by using the metrics MAE and MSE. The following are the results obtained by theRidge Regressor.

Mean Absolute Error: 0.18 Mean Squared Error: 0.05 R^2 Score: 0.9999

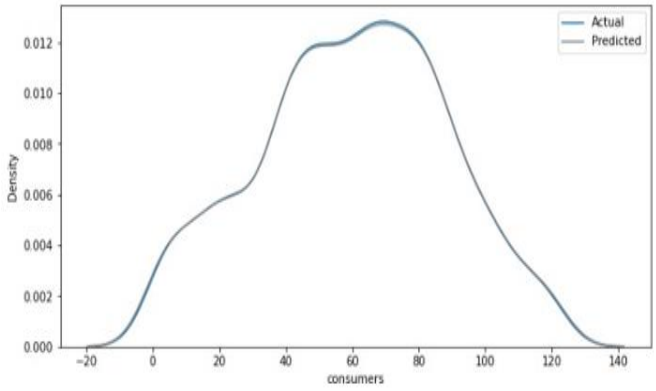

Fig 6 Graph Of Lasso Regressor

Actuals	Predicted
41.147324	41
71.856164	72
73.875833	74
62.989628	63
34.163889	34
120.514904	121
86.724431	87
50.033249	50
66.953416	67
72.852349	73

Table 5 of actual and Predicted Lasso Regressor

➢ *Decision Tree Regressor*

Decision Tree Regressor is trained with the dataset by using 5-fold time series cross- validation approachwhere 80% of the data was used for training and 20% of the data was usedas the test set and the performances have been measured by using the metrics MAE and MSE. The following are the results obtained by the Decision Tree Regressor:

Mean Absolute Error: 0.99 Mean Squared Error: 1.67 R^2 Score: 0.998

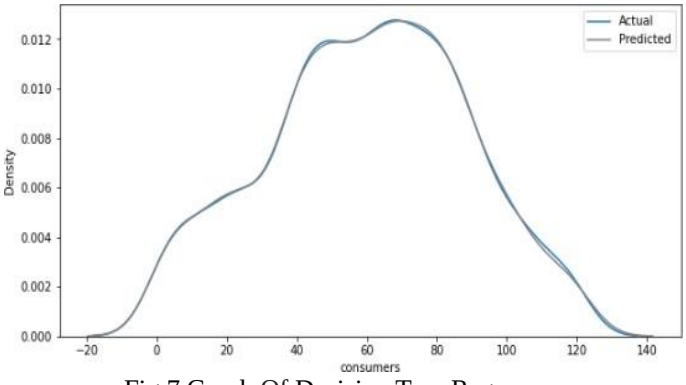

Fig 7 Graph Of Decision Tree Regressor

Actuals	Predicted
41.581250	41
73.023438	72
73.023438	74
63.008439	63
34.212670	34
117.425532	121
87.483607	87
50.975709	50
65.983333	67
73.023438	73

Table 6 of actual and Predicted Decision Tree Regressor

➢ *Ridge Regressor*

Ridge regression is a model tuning technique used to analyze data with multicollinearity. L2 regularization is performed by this method..

Mean Absolute Error: 2.56 Mean Squared Error: 13.54 R^2 Score: 0.9836

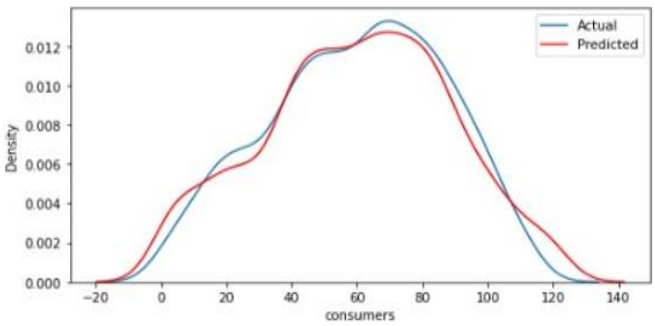

Fig 8 Graph Of Ridge Regressor

Actuals	Predicted
41.581250	41
73.023438	72
73.023438	74
63.008439	63
34.212670	34
117.425532	121
87.483607	87
50.975709	50
65.983333	67
73.023438	73

Table 7 of actual and Predicted Ridge Regressor

➢ *Gradient Boosting Regressor*

Gradient boosting is a machine learning technique used in tasks such as regression and classification, among others It returns a prediction model in the form of an ensemble of weak prediction models, usually decision trees. The following are the results obtained.

Mean Absolute Error: 0.05 Mean Squared Error: 0.01 R^2 Score: 1.0

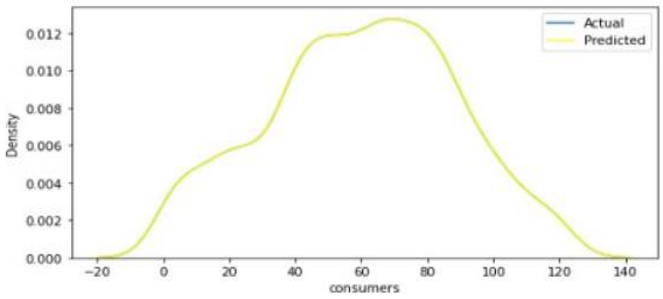

Fig 9 Graph Of Gradient Boosting Regressor

actual	predicted
41.056939	41
72.093818	72
73.910929	74
63.001873	63
33.805911	34
120.906954	121
86.992819	87
50.004601	50
66.997817	67
72.966231	73

Table 8 of actual and Predicted Gradient Boosting Regressor

➢ *A multilayer perceptron (MLP)*

The MLP model stands for Multilayer Perceptron is a type of feed-forward artificial neural network (ANN) where the information flows from the input layer towards the output layer through the hidden layer. MLP makes use of a supervised learning algorithm called backpropagation for training the network. The error is calculated by taking the difference between the network output and the actual output.

Mean Absolute Error: 3.78 Mean Squared Error: 24.36 R^2 Score: 0.9704

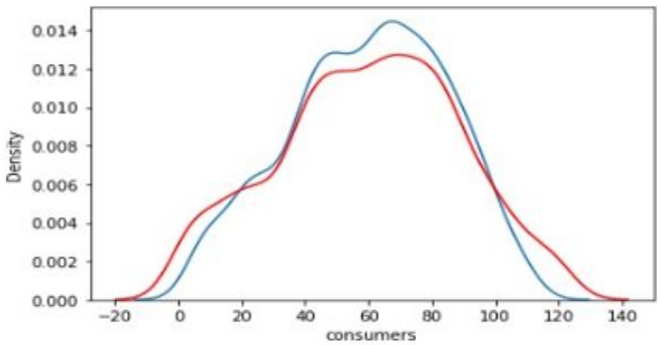

Fig 10 Graph Of MLP

➢ *Comparative study of the Fitting Models basing on Metrics*

models	MAE	MSE	R^2
Gradient Boosting Regressor	0.05	0.01	1.0000
Lasso Regressor	0.18	0.05	0.9999
Random Forest Regressor	0.71	1.21	0.9985
Decision Tree Regressor	0.99	1.67	0.9980
Ridge Regressor	2.56	13.54	0.9836
MLP	3.78	24.36	0.9704
Linear Regression	9.07	127.71	0.8450

Table 9 Metrics (MAE, MSE and R^2)

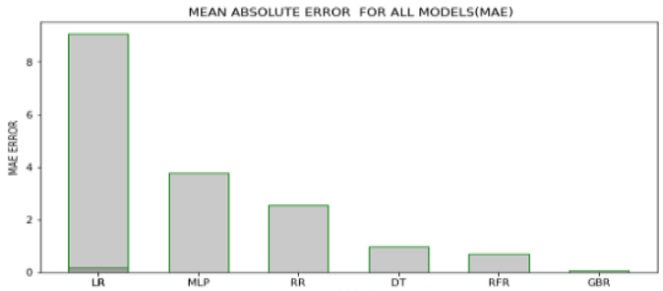

Fig 11 Mean Absolute Error obtained by regression models

Figure represents the Mean Absolute error from the results of the predictions produced by the Linear Regression , Random Forest Regressor, Regressor Lasso Regressor, Gradient Boosting,Decision Tree , Ridge Regressor. From the figure, Linear Regression has the highest, Gradient Boosting regressor has lowest MAE for the folds and thus can be said as a best-performed algorithm. Linear Regression has the highest MAE and thus it can be said as worst performer.

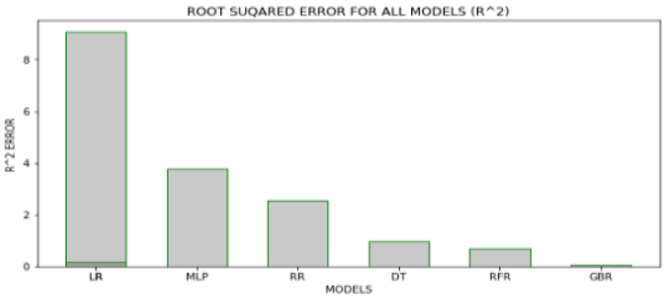

Fig 12 Comparison of R^2 obtained by regression models

➢ *Comparison of R^2 obtained by regression models*

From figure, it can be noticed that Still Gradient boosting has shown spectacular overall performance with the least MSE with the least value (0.01) and Linear Regressor has shown worstperformance with the highest MSE with the least value (127.71)

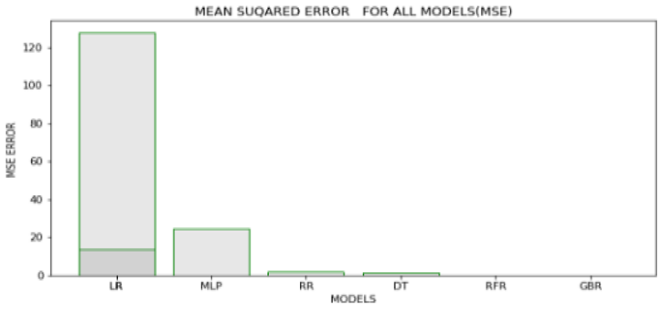

Fig 13 Comparison of MSE obtained by regression models

From figure , it can say that Gradient Boosting and Random Forest Regressor are the best performer with the lowest error and Linear Regression is the worst performer with the highest error.

➢ *Comparison of performance evaluation results*

Key concepts behind the performance of algorithms are explored as follows:

- Gradient Boosting has performed significantly well compared to the other models. This may be because of the regularization approach where the variance is reduced at the cost of some bias initiation which makes it robust to outliers and overfitting.
- Random Forest Regressor has performed surprisingly well compared to the Linear Regression, Decision Tree regressor and lasso regressor; this may be because of its generalization capability. Kernel function for the chosen parameters is set to 'Rbf', which means Radial basis function (RBF) is a function whose value depends on the distance to a center. other kernel functions like linear, radial may produce better results.
- Random Forest Regressor has not shown good performance, this may be because of overfitting problem which may be preventing from generalizing the model.
- Linear Regression produced bad results compared to others because of overfitting problem and Linear Regression is harder to tune compared to other models results
- Ridge regressor as well has not good performance on the data with the R^2 equals to 0.9704, MAE equals to 2.56 and MSE Equals to 13.54.
- Linear Regression is the worst Model for this analysis, since for both the values of R^2 which is 0.8450 and MAE: 9.07 and MSE: 127.71 all marked the unfit compared toall other models.

➢ *Comparison of performance evaluation results*

This project makes use of the GBR model stands for Gradient Boosting Regressor is a typeof machine learning boosting. It is based on the assumption that the best next model, when combined with previous models, minimizes the overall prediction error. If a small change in a case's prediction results in no change in error, the case's next target outcome is zero.

The GBR model is well suited for this project because of the following reasons:

- GBR is suitable for this project because it classified prediction problems where inputs areassigned a class or label
- GBR method is used to forecast the consumer of upcoming period. According to results there are high similarities between forecasted and actual data.
- GBR is suitable for regression prediction problems where a real-valued quantity ispredicted given set of inputs

IV. CONCLUSION

From the study results, it was found that The ability of machine learning models to analyze and learn from real-time data and historic inyange dataset helps supply chain managers to optimize the cost-saving and enhanced productivity, the study results can help a business better predict future demand. It helps Inyange industry to decide what products can be minimized when they become less profitable or understand what customer needs after the initial order 5. 0% of global companies use Artificial intelligent and machine learning , implying that supply chain efficiency must be greatly increased in the coming days. The industry must embrace AI-driven supply chain management as the solution. An efficient supply chain is critical to business success in today's competitive world. Every day, disruptive technologies such as AI and ML play an important role in making it better.

REFERENCES

[1]. Arribas, V., & Alfaro, J. A. (2018). Handbook of supply chain management. Auerbach.
[2]. Becker, T. I. (2016). neural-network computational model to improve product routing in a logisticsfacility. International Journal of Production Economics. Using an agent-based.
[3]. Berrado, A. (2022). Machine Learning in Supply Chain Management. researchgate.
[4]. Berrado, M.-I. M. (2022). Machine Learning in Supply Chain Management: A Systematic LiteratureReview. researchgate.
[5]. Boehmke, Bradley. (2019). "Gradient Boosting". . Hands-On Machine Learning with R. Chapman &Hall.
[6]. Burns, E. (2017). R programming language. Techtarget.
[7]. Dr.Yusuf Perwej. (2022, January 20). A Systematic Literature Review on the Cyber Security. Retrievedfrom HAL Open Science: https://hal.archives-ouvertes.fr/hal-03509116
[8]. Hope, C. (2021). Overview of the Python 3 programming language. Computer Hope. inyange. (2022). inyange industries.
[9]. Inyange. (2022). inyangeindustry.com. Retrieved from https://www.inyangeindustries.com/about#:~:text=Inyange%20Industries%20is%20a%20leading,%26%20selling%20pasteurized%20milk%20%26%20yoghurt.
[10]. Kanade, V. (2022). Ways Machine Learning Can Transform Supply Chain Management. spiceworks.
[11]. Kar, U. K. (2019, june 27). marutitech.com. Retrieved from marutitech: https://marutitech.com/machine-learning-in-supply-chain/
[12]. michael. (2019). Further Regression Algorithms.
[13]. Mitul. (2020). 7 Ways Machine Learning Can Solve Supply Chain Challenges. Business 2 Community.
[14]. Solanki Sandhya Makkar et Dr G.Naga Rama Devi et Vijender Kumar. (2020). Applications of Machine Learning Techniques in Supply Chain Optimization. researchgate. tech, m. (2022). machine-learning-in-supply-chain. Retrieved from https://marutitech.com/machine- learning-in-supply-chain/