ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1266

# Reverse Logistics Optimization in the Indian Dairy Industry: Forecasting Vendor Returns to Reduce Spoilage

[1]Aman Raj Singh; [2]Manish Raj; [3]Arikatla Sai Sumedha;
[4]Goalla Kartheek; [5]Bharani Kumar Depuru

Publication Date: 2025/04/03

**Abstract:- The dairy industry grapples with significant hurdles in reverse logistics, particularly due to high return rates from vendors, inefficiencies in transportation, and challenges faced by retailers. Perishable goods such as milk, yogurt, and buttermilk have limited shelf lives, making accurate demand prediction and inventory control crucial to minimizing waste and financial losses. This research focuses on predicting vendor-specific return rates to streamline the reverse logistics process.**

**We employ ARIMA and XGBoost models to forecast return rates from vendors using historical sales data, seasonal trends, and regional demand variations. By accurately predicting returns, we enable proactive redistribution of stock, ensuring products close to expiration are redirected to high-demand regions before spoilage occurs. Additionally, we propose inventory optimization strategies, including FIFO-based stock rotation and dynamic demand adjustments, to reduce waste. To enhance logistical efficiency, we implement route optimization, IoT-enabled real-time monitoring, and temperature-controlled transportation, reducing delays and maintaining product quality. Furthermore, we address retailer challenges through targeted sales training, cold chain infrastructure support, and incentive programs to improve demand planning and reduce overstocking.**

**By integrating machine learning-based forecasting, operational enhancements, and supply chain optimization, our approach improves efficiency, reduces spoilage, lowers costs, and promotes sustainability in dairy reverse logistics. The proposed framework is adaptable to other industries dealing with perishable goods and similar return-related challenges.**

***Keywords:*** *Reverse Logistics, Dairy Industry, Vendor Return Forecasting, ARIMA, LSTM, XGBoost, Supply Chain Optimization, Machine Learning, IoT.*

**How to Cite:** Aman Raj Singh; Manish Raj; Arikatla Sai Sumedha; Goalla Kartheek; Bharani Kumar Depuru (2025) Reverse Logistics Optimization in the Indian Dairy Industry: Forecasting Vendor Returns to Reduce Spoilage. *International Journal of Innovative Science and Research Technology*, 10(3), 1987-1992. https://doi.org/10.38124/ijisrt/25mar1266

## I. INTRODUCTION

The Indian dairy industry has a long and rich history, dating back thousands of years. Early records indicate that zebu cattle were domesticated, and dairy products such as milk, ghee, and curd became an integral part of Indian society. For centuries, dairy farming in India remained a small-scale, family-based activity. Most farmers relied on traditional methods, producing milk primarily for local consumption. A major transformation occurred in the 1970s with the White Revolution [1], led by Dr. Verghese Kurien and the National Dairy Development Board (NDDB) [1]. This initiative, known as Operation Flood, aimed to increase milk production, reduce dependence on imports, and make India self-sufficient in dairy production. As a result, India surpassed many developed nations and became the world's largest milk producer by the late 1997s. [1]

The growth continued, and by 2014-2015, India's milk production reached 146.3 million tonnes, contributing around 5% to the national GDP. By 2023-2024, India's milk production had further increased to 240 million tonnes, marking a 4% rise from the previous year. India now contributes approximately 25% of global milk production, reaffirming its position as the world's largest milk producer and consumer. [8]

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1266

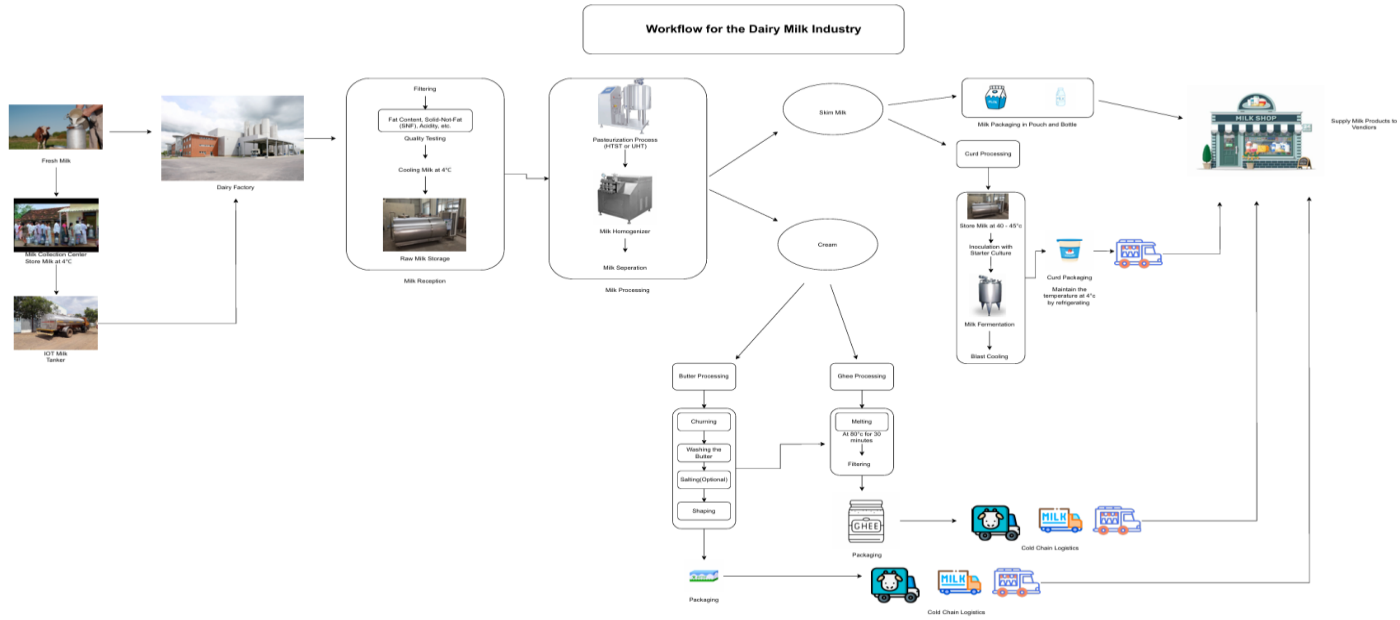


Fig 1: Workflow of the Indian dairy industry, highlighting the process from raw milk collection to processing, packaging, and distribution of dairy products.

The Indian dairy industry operates through a well-organized supply chain that ensures milk and dairy products reach consumers efficiently. The process begins with farmers, who raise cows and buffaloes, providing them with proper feed, water, and care to produce raw milk. This milk is then collected by local milkmen or collection centers, where it is tested for quality before being transported to processing plants.[2] [10]

At the processing plants, the milk undergoes various treatments such as filtering, pasteurization, homogenization, and fermentation to ensure safety and extend shelf life. The processed milk is then converted into different dairy products, including bottled milk, curd, paneer, butter, and ghee. These products are carefully packaged and stored under controlled conditions.[2] [10]

Next, the packaged dairy products are sent to distribution centers, where transportation and cold storage facilities help maintain freshness during delivery. From there, the products are supplied to retail outlets, supermarkets, and local vendors, ensuring they are available for consumers.[2] [10]

Finally, consumers purchase dairy products from retail stores, making dairy an essential part of their daily diet. This structured system ensures a steady supply of fresh dairy products while maintaining quality and hygiene throughout the process [Fig 1].[2]

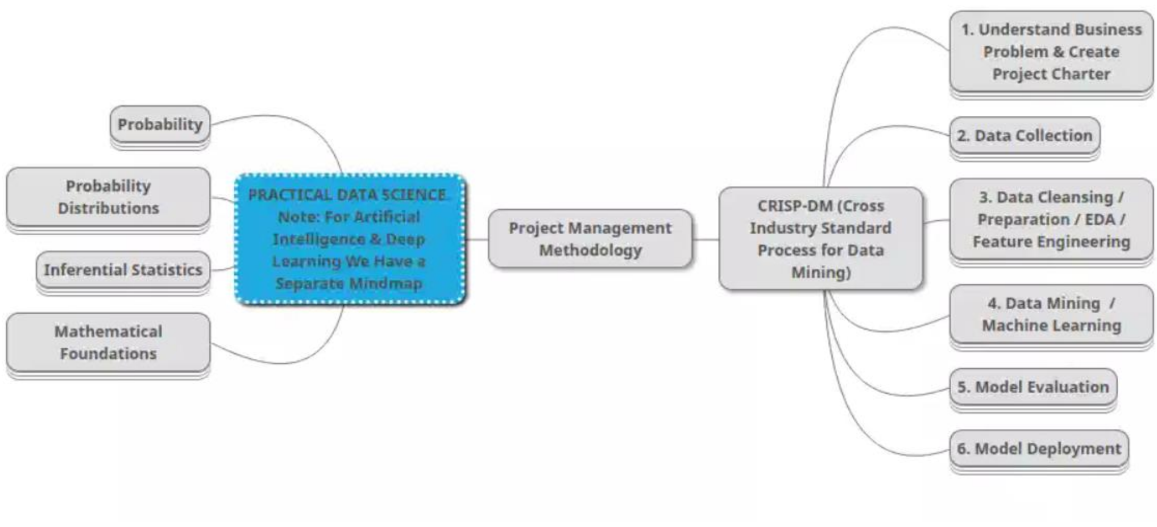


Fig. 2: CRISP-ML(Q) - approach for quality assurance across all six phases.
(Source: Mind Map - 360DigiTMG)

To develop an effective vendor return forecasting model, we follow the Cross Industry Standard Process for Machine Learning with Quality Assurance, CRISP-ML(Q) methodology, a structured approach for machine learning projects that integrates business understanding, data procewssing, model development, and quality assurance which can be easily accessed as an open-source framework on the 360DigiTMG website[Fig2].

The key objective of this study is to improve the accuracy of vendor return rate forecasting in the dairy industry through data-driven strategies and advanced machine learning techniques. This research analyses historical return data from different vendors to identify significant patterns and trends using ARIMA (AutoRegressive Integrated Moving Average) [3], LSTM (Long Short-Term Memory) [3] and XGBoost (Extreme Gradient Boosting) [5]. These models are designed to handle time-series forecasting and complex feature interactions, enhancing the precision of return rate predictions.

By evaluating the effectiveness of these machine learning algorithms, this study aims to optimize reverse logistics operations [2], reduce spoilage, and enhance supply chain efficiency. The insights gained from vendor-specific return rate forecasts will enable proactive stock redistribution, better demand-supply alignment, and minimized financial losses. The findings of this research can contribute to a more efficient and sustainable dairy supply chain, with potential applications in other perishable goods industries.

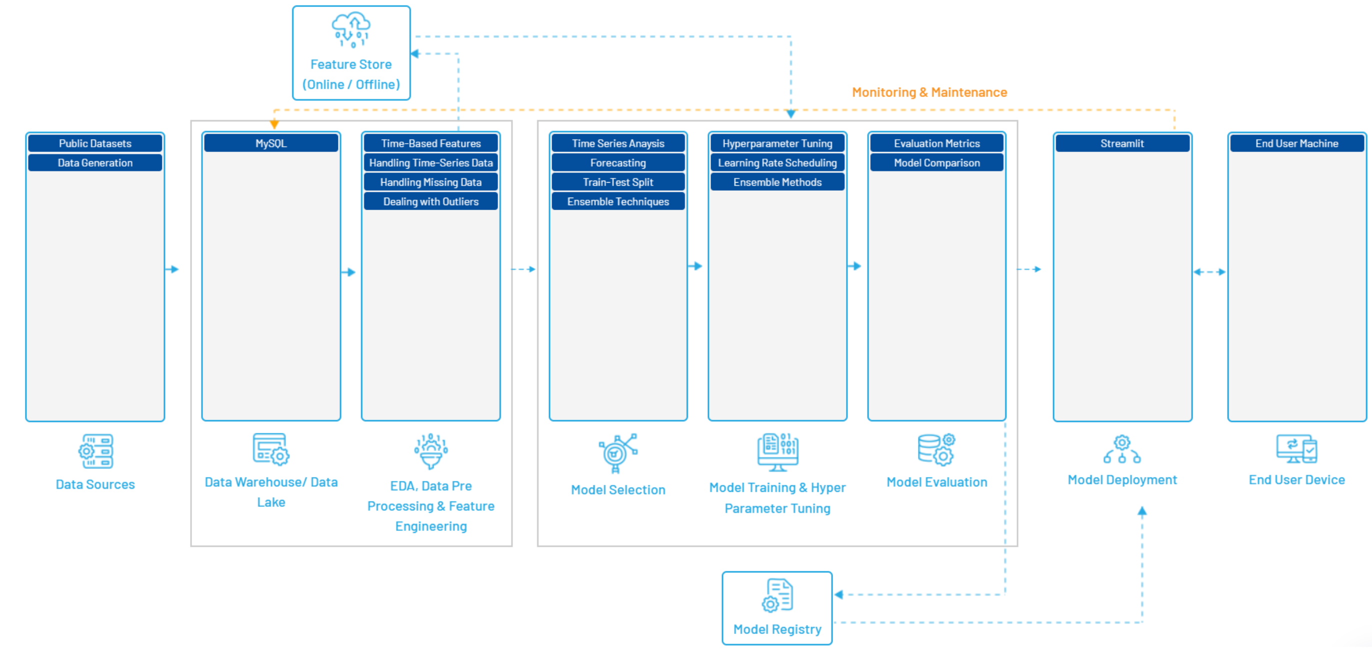


Fig. 3: Project Architecture Diagram: Project Architecture Diagram – A structured workflow from Business Understanding to Model Deployment & Monitoring (Source: Machine Learning Workflow - 360DigiTMG)

The workflow methodology (Fig. 3) follows a structured approach, beginning with Business and Data Understanding to define the problem and scope, followed by Data Collection to acquire high-quality data. EDA is performed to analyse data patterns and uncover insights for feature engineering, while Data Preprocessing cleans and transforms the data for modelling. In the Model Building phase, suitable algorithms are selected and optimized using techniques like hyperparameter tuning. Model Evaluation ensures the model's performance through metrics such as RMSE (Root Mean Squared Error) [9], MAPE (Mean Absolute Percentage Error), and MAE (Mean Absolute Error) [9]. The model is then deployed into production in the Model Deployment phase, followed by continuous Monitoring and Maintenance to ensure ongoing accuracy and address data drift.

## II. METHODS AND TECHNIQUES

➤ *Data Collection:*

Based on the observation from open source we collected open-source data also we generated, synthetic dataset for reverse logistics in the dairy industry, also we simulated product returns from retailers to warehouses. The dataset includes 5000 records with attributes such as return dates, retailer info, warehouse IDs, customer details, product name, return quantity and return reasons. Product return quantities were capped at 20% of total quantities based on product-specific constraints.

➤ *Data Preprocessing:*

Data preprocessing is essential for ensuring dataset quality before applying machine learning models. In this study, we handled missing values, outliers, and duplicates to maintain data integrity. As the dataset is time-series, we

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1266

treated missing values and outliers, which could disrupt data continuity and trend analysis using KNN Imputation (n_neighbors=5) [7] for missing values, preserving return quantity consistency across time periods. For outliers, we applied Winsorization with the IQR method (fold value = 1.5) to cap extreme values, ensuring data distribution and trends remained intact. These techniques-maintained data continuity, minimized bias, and preserved the accuracy and robustness of the analysis.

Before proceeding with model building, we tested the time-series data for stationarity using the Augmented Dickey-Fuller (ADF) test. The ADF test results confirmed stationarity, as the ADF statistic (-71.44) was significantly lower than the critical values at the 1% (-3.43), 5% (-2.86), and 10% (-2.57) significance levels. Additionally, the p-value was 0.0, providing strong evidence to reject the null hypothesis of non-stationarity. Since stationarity is a crucial assumption for time-series forecasting models, ensuring that the dataset maintains stable statistical properties over time enhances the accuracy of predictive models.[4]

➢ *Model Training:*

After confirming the stationarity of the dataset, we proceeded with model building using XGBoost [5], ARIMA [3], and LSTM [3] to forecast return quantities. The XGBoost model was trained using Outlet Info and Product Name as features, with Return Quantity as the target variable. It was configured with 100 estimators, a learning rate of 0.1, and a maximum depth of 5, and the dataset was split into 80% training and 20% testing. The ARIMA model was applied to capture time-series patterns by aggregating return quantities at the monthly level. Using an (1,1,1) order, the model was fitted to the transformed time-series data to ensure effective trend analysis. The LSTM model, a deep learning-based approach, was structured to capture sequential dependencies. The input data was reshaped, and the model was designed with 50 LSTM units, two hidden layers, and the Adam optimizer, trained over 50 epochs with a batch size of 32, using Mean Squared Error (MSE) as the loss function.[6]

➢ *Model Testing:*

Model evaluation is a critical phase in the development and implementation of any predictive model, ensuring its reliability and accuracy in real-world scenarios. In this study, we assessed the performance of our XGBoost, ARIMA, and LSTM models for return quantity forecasting using well-established evaluation metrics.

To measure predictive accuracy, we calculated Root Mean Squared Error (RMSE) for each model. The ARIMA model resulted in an RMSE of 8.05, indicating relatively higher prediction errors. The LSTM model significantly improved performance, achieving an RMSE of 0.1408, while the XGBoost model outperformed both with the lowest RMSE of 0.1345. The results demonstrate that XGBoost is the most effective model for this forecasting task, as it provides the most precise predictions.

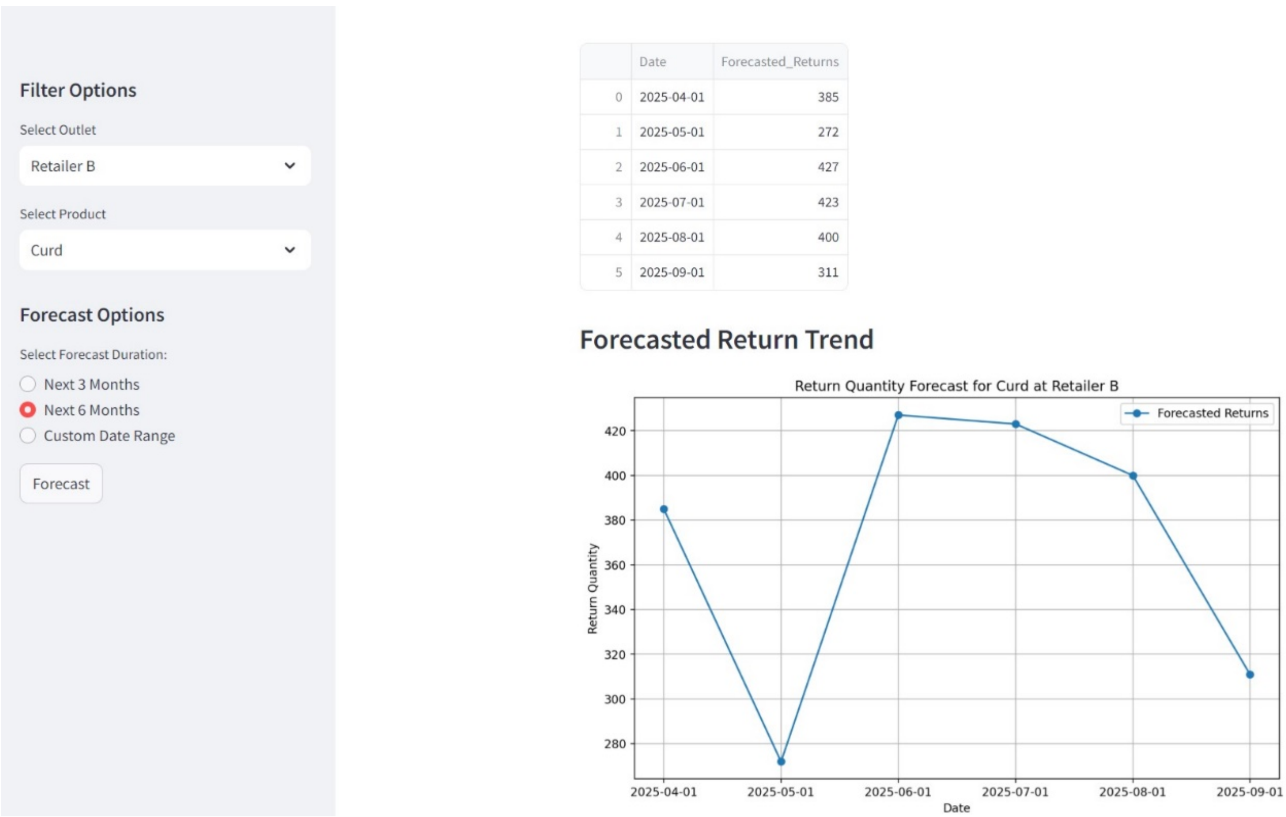


Fig. 4: Return Quantity Forecasting Model Output: Streamlit Deployment Visualization

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1266

- *Model Deployment:*

The deployment phase is a crucial step in making the predictive model accessible and usable for stakeholders. In this study, we deployed the return quantity forecasting model using Streamlit, a user-friendly and interactive web application framework. Streamlit enables seamless integration of machine learning models into a web-based interface, providing an intuitive and efficient user experience.

The deployed application allows users to select an outlet and a product category (e.g., Milk, Curd, Butter, Ghee) and choose a forecasting duration (3 months, 6 months, or a custom date range). Based on the selection, the application forecasted values in a tabular format, along with a graphical trend visualization (as shown in Fig. 4) to help users analyse return patterns over time.

Furthermore, the application provides a download option, allowing users to save the forecast results as a CSV file for further analysis. This deployment ensures that the forecasting model is not only accurate but also practical, accessible, and easy to interpret, making it a valuable tool for optimizing return quantity predictions in the dairy industry.

## III. RESULTS AND DISCUSSIONS

The implementation of machine learning models for forecasting vendor return rates in the Indian dairy industry revealed significant performance differences. Our comparative analysis of three models showed varying levels of prediction accuracy: ARIMA performed least effectively with an RMSE of 8.05, LSTM demonstrated substantial improvement with an RMSE of 0.1408, while XGBoost emerged as the superior model with the lowest RMSE of 0.1345.

XGBoost's exceptional performance can be attributed to its ability to capture complex non-linear relationships between outlet information, product categories, and return quantities. This accuracy enables dairy companies to implement strategic interventions, including proactive stock redistribution, targeted vendor support, and dynamic inventory management.

The deployed Streamlit application (Fig. 4) provides stakeholders with an accessible interface to generate forecasts based on outlet and product selection. This tool empowers decision-makers to implement timely interventions before product expiration, significantly reducing waste and associated costs throughout the supply chain.

While our approach demonstrates considerable promise, the use of synthetic data presents a limitation. Future implementations with real-world data would further validate these findings. Nevertheless, the superior accuracy of machine learning models, particularly XGBoost, offers significant potential for optimizing reverse logistics in the dairy industry, reducing spoilage, enhancing operational efficiency, and supporting sustainability goals by minimizing food waste.

## IV. CONCLUSION

This research effectively tackles the challenges of reverse logistics in the Indian dairy sector by introducing a machine learning-based framework for predicting vendor return quantities. Among the models tested, XGBoost demonstrated the highest accuracy, empowering dairy businesses to reduce spoilage, optimize inventory control, and improve supply chain performance. The framework's practicality is further enhanced by incorporating IoT-enabled monitoring, temperature-controlled transportation, and a user-friendly Streamlit-based deployment. The findings provide a scalable solution to minimize waste and promote sustainability in the dairy industry, with potential extensions to other sectors handling perishable goods. Future research should focus on validating the model using real-world datasets and exploring advanced methodologies to ensure its adaptability to changing market dynamics. This study highlights the transformative role of machine learning in creating a more efficient and resilient supply chain for perishable products.

## V. FUTURE SCOPE

The future scope of this research includes validating the framework using open-source historical data to ensure its reliability and accuracy in real-world scenarios. Advanced machine learning techniques, such as ensemble methods and hybrid models, can be explored to further enhance prediction accuracy and adaptability to complex supply chain dynamics. Integration of IoT and blockchain technology could improve traceability and transparency in reverse logistics, while dynamic demand forecasting models accounting for external factors like weather and market trends could better align supply and demand.

The framework can also be extended to other perishable goods industries to assess its broader applicability. Additionally, incorporating sustainability metrics, human-centric AI solutions, and adaptive learning mechanisms to address evolving market conditions will ensure the model remains relevant and effective. Collaborative supply chain networks and policy investigations could further support the adoption of this framework, making it a robust and scalable solution for optimizing reverse logistics in the dairy industry and beyond.

## ACKNOWLEDGMENTS

We acknowledge that with the consent from 360DigiTMG, we have used the CRISP-ML(Q) methodology (ak.1) and the ML Workflow (ak.2), which are available as open-source on the official website of 360DigiTMG.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1266

## REFERENCES


[1]. Manisha Jayprakash singh. 2024. A Study on Problem and Prospect of dairy Industry in India. IJIRT166092_PAPER.pdf

[2]. Gyanesh Kumar Sinha, Sumit Mishra. 2023. Sustainable Supply Chain Management Practices in the Dairy Industry: A Comparative Study of Leading Dairy Firms and Future Research Directives. https://arccjournals.com/journal/asian-journal-of-dairy-and-food-research/DR-2120

[3]. Sima Siami Namini, Neda Tavakoli Akbar Siami Namin. 2018. A Comparison of ARIMA and LSTM in Forecasting Time Series. https://www.researchgate.net/publication/330477082

[4]. Abhishek Kashyap, Om Ji Shukla, Bal Krishna Jha, Bharti Ramtiyal, Gunjan Soni. 2023. Enhancing Sustainable Dairy Industry Growth through Cold-Supply-Chain-Integrated Production Forecasting. https://www.mdpi.com/2071-1050/15/22/16102

[5]. Dr K. Alice, Syed Hamad ul Haq Andrab, Siddharth Anoop Srivastava. 2022. Sales Forecasting using XGBoost. https://www.techrxiv.org/doi/full/10.36227/techrxiv.21444129.v1

[6]. Jungmok Ma, Harrison M. Kim. 2016. Predictive Model Selection for Forecasting Product Returns. https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/138/5/054501/472991/Predictive-Model-Selection-for-Forecasting-Product?redirectedFrom=fulltext

[7]. Gustavo Enrique Batista, Maria Carolina Monard. 2002. A Study of K-Nearest Neighbour as an Imputation Method. https://www.researchgate.net/publication/220981745_A_Study_of_K-Nearest_Neighbour_as_an_Imputation_Method

[8]. Resmi R.S. 2024. DAIRY FARMING IN INDIA: AN OVERVIEW. https://www.jetir.org/papers/JETIR2412378.pdf

[9]. Chayuth Vithitsoontorn, Prabhas Chongstitvatana. 2022. Demand Forecasting in Production Planning for Dairy Products Using Machine Learning and Statistical Method. DOI: 10.1109/iEECON53204.2022.9741683

[10]. Abebe Tessema, Markos Tibbo. 2009. Milk processing technologies for small-scale producers https://www.researchgate.net/publication/304526128_Milk_processing_technologies_for_small-scale_producers