iDairy Forecast : Smart Product Demand Forecasting using SARIMA Model

Ridham Jasani[1], Jasmit Bajaria[2], Vedant Deshmukh[3], Dolas Keche[4], Prof. Priyanka Bhilare[5]

[1,2,3,4,] Student, Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India

[5] Assistant professor, Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India

---***---

Abstract - iDairy is an intelligent system developed to enhance the efficiency of dairy retail operations by providing precise demand predictions. It leverages sophisticated algorithms to examine past sales patterns and market dynamics, ensuring optimized inventory control while minimizing surplus. The platform integrates seamlessly with Google Sheets for live data storage, where each product is assigned a dedicated sub-sheet for systematic tracking. To account for seasonal demand fluctuations, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is implemented for forecasting. iDairy equips retailers with valuable insights to refine stock management and improve overall operational effectiveness. This project emphasizes data-driven strategies for optimizing dairy supply chain management.

Key Words: Machine Learning, Demand Forecasting, Inventory Management, SARIMA Model, Dairy Retail, Data-Driven Decision Making, Sales Analysis, Natural Language Processing

1.INTRODUCTION

The dairy sector encounters considerable challenges in effectively managing inventory due to unpredictable demand fluctuations, seasonal shifts, and the perishable nature of dairy products. Conventional inventory management techniques frequently result in either surplus stock or shortage

s, leading to financial setbacks and inefficiencies in operations. To overcome these obstacles, iDairy has been developed as an AI-powered solution that employs Machine Learning (ML) to enhance demand forecasting and optimize stock management.

iDairy incorporates the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to analyse past sales data, detect emerging patterns, and provide highly accurate demand projections. This predictive approach empowers retailers to make informed, data-driven decisions, reducing product wastage while maintaining ideal inventory levels. The system seamlessly integrates with Google Sheets, where each product is assigned an individual sub-sheet for real-time data storage and monitoring. This structured arrangement improves accessibility and enables businesses to track sales trends efficiently.

By leveraging AI-driven analytics, iDairy seeks to enhance supply chain operations, maximize profitability, and promote sustainable dairy retail practices. The project aims to revolutionize traditional dairy management by incorporating automation, allowing retailers to sustain balanced stock levels, reduce losses, and improve overall operational productivity.

1.1 PROPOSED SYSTEM

The iDairy system is designed to improve demand forecasting and inventory management in dairy retail operations. The system utilizes historical sales data to predict future demand, helping retailers optimize stock levels and reduce wastage. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used for forecasting, as it captures seasonal trends and variations in sales.

Key features of the proposed system include:

- Data Collection & Storage: Sales history is systematically recorded in Google Sheets, facilitating seamless tracking and in-depth analysis.
- Demand Forecasting Model: A SARIMA-driven model predicts upcoming sales trends based on past performance and seasonal demand variations.
- Inventory Optimization: The predicted demand helps retailers maintain balanced stock levels, mitigating risks associated with overstocking and shortages.

By integrating iDairy, retailers can streamline supply chain operations, significantly reduce product wastage, and enhance overall efficiency. This ultimately leads to improved profitability and better resource management in dairy retail.

2. LITERATURE SURVEY

Demand forecasting has been extensively researched using both traditional statistical techniques and deep learning approaches. ARIMA (Auto Regressive Integrated Moving Average) is effective for predictions in stable environments with minimal fluctuations, making it suitable for stationary time-series data. However, it struggles when dealing with datasets with strong seasonality or complex variations. In contrast, Long Short-Term Memory (LSTM) networks excel at capturing non-linear trends and seasonal dependencies, making them better suited for intricate forecasting tasks. Research indicates that training models on monthly rather than weekly data improves accuracy. Future advancements may explore the integration of attention mechanisms and global models to further refine forecasting accuracy and enhance inventory management, particularly in sectors like dairy production [1].

The application of Business Intelligence (BI) and Machine Learning (ML) in demand forecasting has gained significant traction in recent years. Various approaches, including time series analysis and rule-based models like Deep AR, have been investigated to achieve precise demand predictions. Studies highlight that DeepAR-based forecasting models offer higher accuracy, minimize losses, optimize stock levels, and enhance operational efficiency, especially as dataset size increases [2].

Recent research has also examined time series models, particularly Seasonal ARIMA (SARIMA), for forecasting commodity prices, particularly essential goods like fruits and vegetables. These models factor in seasonal fluctuations and external influences such as weather conditions, transportation expenses, and seed quality. Although these models may not always provide absolute precision, they serve as valuable tools for identifying price trends and formulating strategies to ensure affordability [3].

Demand forecasting remains a critical area of study for its role in optimizing inventory management, production planning, and market strategies. Various statistical models, especially ARIMA, have been widely utilized for time series forecasting. Studies show that both seasonal and non-seasonal ARIMA models effectively predict future demand by analyzing historical sales data. Researchers have explored model enhancements by integrating seasonality, external variables, and market trends to boost accuracy. ARIMA's flexibility across different datasets and forecasting scenarios makes it a valuable tool for various industries, offering actionable insights for data-driven decision-making [4].

Another area where forecasting is heavily researched is stock price prediction, which remains a challenging task due to market volatility and influencing factors like economic conditions, investor sentiment, and external events. Several studies have utilized ARIMA for its simplicity and efficiency in short-term forecasting. Research suggests that ARIMA delivers reasonable accuracy for individual stocks like ICICI Bank and Reliance Industries. However, its performance declines over long-term predictions due to the unpredictable nature of financial markets and the presence of multiple complexes influencing variables [5].

3. SARIMA MODEL

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an extension of the ARIMA model specifically designed to handle time-series data with seasonal patterns. While ARIMA is primarily used for non-seasonal data, SARIMA incorporates seasonal components, allowing it to effectively capture recurring patterns over fixed intervals.

The general representation of a SARIMA model is:

$$SARIMA(p,d,q)(P,D,Q)s$$

Where:

- p: The order of the non-seasonal autoregressive (AR) component, determining how many past observations influence the current value.
- d: The number of non-seasonal differences required to make the time series stationary. Stationarity is achieved by subtracting previous observations from the current one, with d specifying how many times this is performed.
- q: The order of the non-seasonal moving average (MA) component, which models the dependency between an observation and the past forecast errors.

Seasonal Components of SARIMA:

- P: The order of the seasonal autoregressive (SAR) component, capturing relationships between an observation and its past seasonal values (e.g., sales from the same month in previous years).
- D: The number of seasonal differences needed to make the seasonal pattern stationary, similar to d but applied at seasonal intervals.
- Q: The order of the seasonal moving average (SMA) component, modelling dependencies between an observation and past seasonal forecast errors.
- s: The length of the seasonal cycle, representing the number of time steps in a full seasonal period (e.g., s = 12 for monthly data with yearly seasonality, or s = 4 for quarterly data).

Key Components of SARIMA:

- Autoregressive (AR): Represented by p and P, it captures relationships between an observation and prior values. Seasonal AR terms use lags based on multiples of s (e.g., s = 12 means using data from 12 months ago).

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 05 | April 2025 www.irjet.net p-ISSN: 2395-0072

- Integrated (I): Represented by d and D, it ensures stationarity by differencing the data (removing trends and seasonal effects).

- Moving Average (MA): Represented by q and Q, it accounts for relationships between an observation and past forecast errors, either non-seasonal (q) or seasonal (Q).

By incorporating these elements, SARIMA becomes a powerful tool for time-series forecasting, especially in industries like dairy retail, where demand fluctuates based on seasonal patterns.

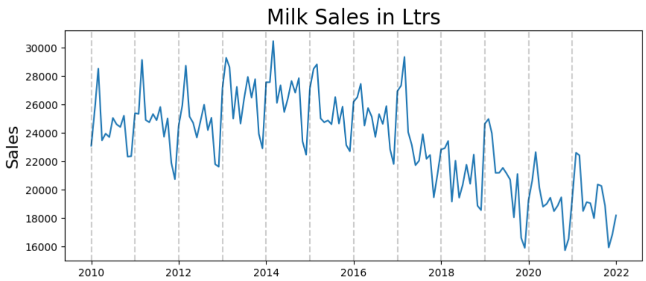


Fig.1 Milk Sales (Dataset)

The graph above represents milk sales (in Liters) over a specific period, showcasing historical sales trends. It highlights fluctuations in sales, with periodic peaks and dips that correspond to seasonal variations, shifts in demand, or external factors influencing consumption patterns. By analysing the visual data, one can observe higher sales during specific months, such as festive seasons or summer, and lower sales during other periods.

This graphical representation plays a crucial role in identifying underlying trends and seasonality in milk sales. Recognizing these patterns is essential for accurate demand forecasting, as it enables businesses to anticipate changes and optimize inventory levels. Models like SARIMA leverage such insights to enhance forecasting precision, ensuring efficient dairy retail management.

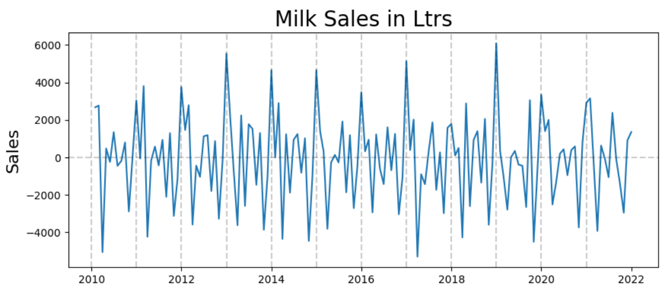


Fig.2 Milk Sales (After removing trends)

The given graph showcases milk sales (in litres) after the removal of the underlying trend component. This transformation helps in isolating seasonal fluctuations and short-term variations, making it easier to analyse sales patterns independent of long-term trends. By eliminating the trend, the graph highlights periodic changes in milk sales over time, allowing for a clearer understanding of seasonal behaviour and cyclical demand shifts.

The trend component, which represents long-term growth or decline, was removed using differencing or other de-trending techniques. As a result, the graph displays the residual or detrended data, emphasizing short-term deviations in sales. These fluctuations may be influenced by factors such as holidays, promotional activities, or weather conditions.

This detrended series is essential for further forecasting and analysis, as it offers a more precise representation of true seasonal patterns in milk sales, free from long-term influences. Studying this refined dataset is critical for developing models like SARIMA, which leverage seasonal trends and short-term variations to make accurate sales predictions.

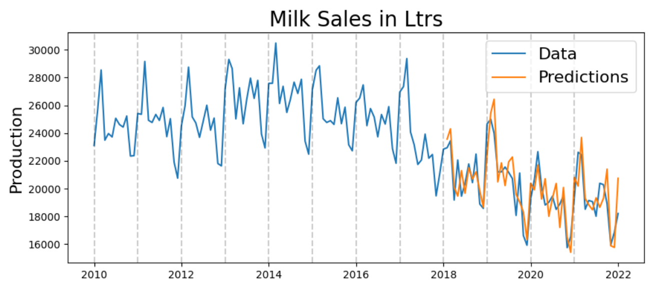


Fig.3 SARIMA Model Predictions

The graph above illustrates the forecasted milk sales (in litres) generated by the SARIMA model over a defined period. The x-axis represents the time intervals (e.g., months), while the y-axis indicates the predicted sales volume. The graph primarily consists of two elements:

- Historical Sales Data: The actual recorded sales, depicted as a continuous line connecting observed data points over time.

- Forecasted Values: The predicted sales figures, represented by a separate line derived from the SARIMA model. This forecasted line is based on patterns identified in past data, incorporating both seasonal and non-seasonal components.

The predicted line effectively mirrors the recurring trends and fluctuations seen in historical sales, demonstrating the model's ability to capture both short-term variations and long-term trends. The accuracy of these predictions is reflected in how closely the forecasted values align with actual sales figures, validating the model's effectiveness in recognizing seasonal patterns. Any discrepancies between observed and predicted values could point to forecasting errors or unforeseen factors such as shifts in consumer demand, supply chain disruptions, or external market influences. This graph serves as a crucial tool for evaluating the SARIMA model's performance, providing insights that aid in inventory optimization, demand forecasting, and strategic business planning for dairy retailers.

```
#summary of the model
print(model_fit.summary())

                                    SARIMAX Results
==========================================================================================
Dep. Variable:                              Total   No. Observations:                   97
Model:             SARIMAX(0, 1, 0)x(1, 0, [1], 12)   Log Likelihood                -826.512
Date:                            Fri, 28 Feb 2025   AIC                           1659.025
Time:                                    14:52:40   BIC                           1666.718
Sample:                                01-01-2010   HQIC                          1662.134
                                     - 01-01-2018
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.S.L12       0.9731      0.025     38.537      0.000       0.924       1.023
ma.S.L12      -0.8148      0.088     -9.274      0.000      -0.987      -0.643
sigma2        1.5e+06   8.25e-09   1.82e+14      0.000     1.5e+06     1.5e+06
===================================================================================
Ljung-Box (L1) (Q):                   9.19   Jarque-Bera (JB):                 2.06
Prob(Q):                              0.00   Prob(JB):                         0.36
Heteroskedasticity (H):               0.45   Skew:                            -0.34
Prob(H) (two-sided):                  0.03   Kurtosis:                         3.26
===================================================================================
```

Fig.5 SARIMAX Results

The results summary from the SARIMA model provides key performance metrics that assess its accuracy, complexity, and suitability for forecasting. The Akaike Information Criterion (AIC) evaluates the model's fit by balancing goodness-of-fit and complexity, with lower values indicating a better model. The Bayesian Information Criterion (BIC) functions similarly but penalizes more complex models more heavily to aid in optimal selection. The Hannan-Quinn Information Criterion (HQIC) serves as a middle ground between AIC and BIC. The Prob(H) value represents the p-value for testing whether the model's residuals exhibit white noise behavior, ensuring randomness in prediction errors. The skew metric evaluates the asymmetry of the residuals' distribution, indicating potential bias, while kurtosis measures the tailedness of the residual distribution, with higher values suggesting the presence of extreme variations. The Prob(JB) value, derived from the Jarque-Bera test, determines whether the residuals follow a normal distribution, which is crucial for reliable forecasting. These statistical indicators play a vital role in validating the SARIMA model, ensuring it effectively captures seasonal trends while maintaining a well-distributed error structure.

3.1 ACF

The Autocorrelation Function (ACF) plot shown above depicts the correlation between milk sales data and its lagged values. It serves as a crucial tool in time series analysis, helping to determine the q parameter, which represents the number of lagged forecast errors in the moving average (MA) component of the SARIMA model. By analysing the ACF plot, patterns in the data's dependencies can be identified, aiding in the selection of an appropriate model for accurate demand forecasting.

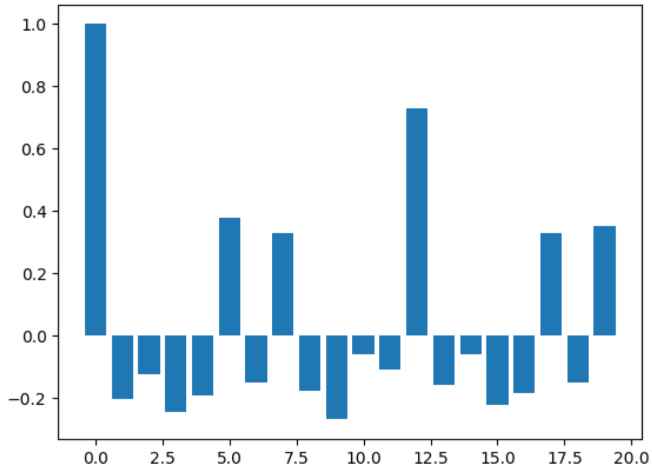


Fig.6 ACF Plot

In the ACF plot, the x-axis denotes lag values, while the y-axis represents the autocorrelation coefficient. Significant spikes that extend beyond the confidence interval, usually highlighted by blue-shaded areas, indicate a correlation between the time series and its previous values. These prominent spikes assist in identifying the optimal lag value for the moving average component, q.

From the ACF plot, we can observe the following:

- Strong autocorrelation is evident in the initial lags, indicating a pattern that relies on past values, which is a typical characteristic of time series data.
- As the lag increases, the correlation gradually weakens, signifying a reduced influence of past values on future ones, a pattern the moving average component is designed to capture.

The observations derived from the ACF plot assist in determining the appropriate q parameter for the SARIMA model. By analysing the plot, I identified the optimal q value that effectively captures the time-dependent relationships within the sales data, enhancing the model's accuracy in predicting future milk sales.

3.2 PACF

The Partial Autocorrelation Function (PACF) plot helps identify the key lags in a time series, assisting in selecting the autoregressive (AR) term for the model. In the PACF plot, the x-axis denotes the lag values, while the y-axis represents the partial autocorrelation at each lag. A notable spike at a specific lag signifies a strong correlation between the observations at that lag and the current value, after accounting for the influence of all preceding lags.

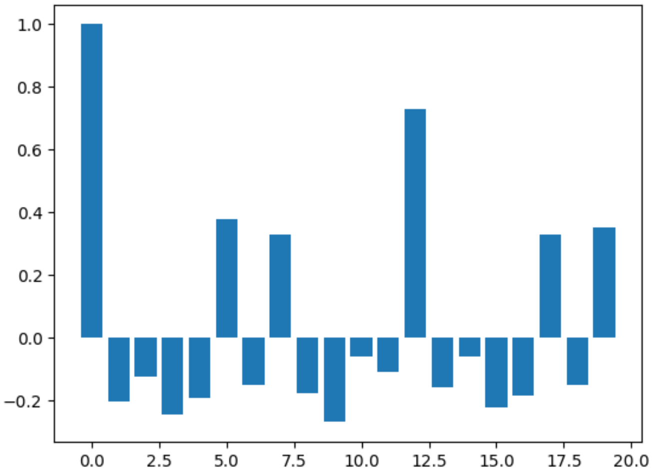


Fig.7 PACF Plot

For my SARIMA model, the PACF plot (depicted in the image) was examined to determine the suitable order of the autoregressive (AR) component. The point at which the partial autocorrelation values decline sharply served as the basis for selecting the optimal value for p (the autoregressive order). The number of prominent spikes in the PACF plot played a crucial role in identifying the appropriate p value for the SARIMA model.

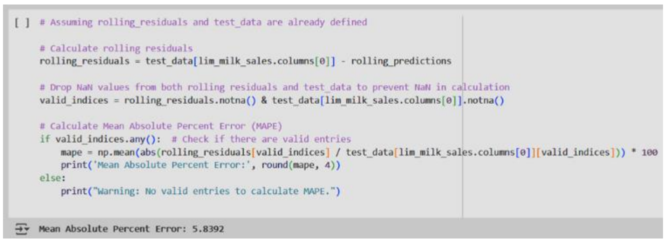
By analyzing the PACF plot, we established the number of past observations (lags) that have a direct impact on the current value, ultimately enhancing the model's precision in forecasting future milk sales.

3.3 MODEL EVALUATION

Model evaluation plays a vital role in predictive modelling, as it allows for assessing the accuracy and efficiency of a model. One of the widely used evaluation metrics is MAPE (Mean Absolute Percentage Error), which offers a straightforward and interpretable measure of forecast accuracy in percentage terms. MAPE is computed by averaging the absolute percentage errors between predicted and actual values, then multiplying by 100 to express it as a percentage. It is mathematically represented by the following formula:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \times 100$$

Fig. 1. MAPE formula.



```
[ ] # Assuming rolling_residuals and test_data are already defined

    # Calculate rolling residuals
    rolling_residuals = test_data[lim_milk_sales.columns[0]] - rolling_predictions

    # Drop NaN values from both rolling residuals and test_data to prevent NaN in calculation
    valid_indices = rolling_residuals.notna() & test_data[lim_milk_sales.columns[0]].notna()

    # Calculate Mean Absolute Percent Error (MAPE)
    if valid_indices.any():  # Check if there are valid entries
        mape = np.mean(abs(rolling_residuals[valid_indices] / test_data[lim_milk_sales.columns[0]][valid_indices])) * 100
        print('Mean Absolute Percent Error:', round(mape, 4))
    else:
        print("Warning: No valid entries to calculate MAPE.")

Mean Absolute Percent Error: 5.8392
```

Fig. 2. MAPE Results.

For our model evaluation, the Mean Absolute Percentage Error (MAPE) was computed to measure the accuracy of the predictions. The obtained MAPE value of 5.8392% signifies that, on average, the model's forecasts deviate by roughly 5.84% from the actual values. This outcome indicates strong model performance, demonstrating a high level of precision in predicting demand with minimal errors. Generally, a MAPE below 10% is regarded as an indicator of reliable forecasting, making this model well-suited for real-world applications where accurate demand estimation is essential.

4. CONCLUSIONS

In summary, the SARIMA (Seasonal Auto Regressive Integrated Moving Average) model has proven to be an effective tool for forecasting milk sales by capturing both seasonal and non-seasonal patterns in historical data. The model successfully identifies recurring trends and fluctuations, which is especially crucial in the dairy industry, where demand exhibits strong seasonality. By integrating seasonal differencing, autoregressive, and moving average components, SARIMA delivers more accurate predictions compared to simpler forecasting methods.

The forecasted values generated by the model closely align with actual sales data, as shown in the visualization, demonstrating its ability to predict both short-term and long-term trends. This predictive capability is highly beneficial for businesses, enabling better inventory management by ensuring optimal stock levels. It helps prevent issues such as overstocking or stock shortages, reducing potential financial losses and improving operational efficiency.

Despite its effectiveness, forecasting remains a continuous process. As new data becomes available, the SARIMA model should be updated and recalibrated to maintain accuracy and adapt to any evolving patterns or external factors. Additionally, the model's performance is influenced by the quality and consistency of the historical data it is trained on. Regular updates and high-quality data can help minimize errors and enhance forecasting precision.

Overall, the SARIMA model serves as a valuable forecasting tool that aids businesses in making data-driven decisions, optimizing inventory management, and improving overall efficiency. With continuous refinement and integration of additional data, it can further enhance strategic planning in the dairy sector and other industries requiring accurate demand forecasting.

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to Prof. Priyanka Bhilare for her invaluable guidance, support, and encouragement throughout this research. Her expertise, insightful feedback, and continuous motivation have played a crucial role in shaping this project. We sincerely appreciate the time and effort she dedicated to helping us refine and enhance our work. Her mentorship has been a source of inspiration, and we are truly grateful for her unwavering support.

REFERENCES

[1] A. Ganesan and A. Kannan, "Stock Price Prediction using ARIMA Model," *Dept. of Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Enathur, Kanchipuram, India.*

[2] C. Vithitsoontorn and P. Chongstitvatana, "Demand Forecasting in Production Planning for Dairy Products Using Machine Learning and Statistical Method," 2022 International Electrical Engineering Congress (iEECON), Khon Kaen, Thailand, 2022, pp. 1-4, doi: 10.1109/iEECON53204.2022.9741683

[3] M. A. Khan et al., "Effective Demand Forecasting Model Using Business Intelligence Empowered With Machine Learning," in IEEE Access, vol. 8, pp. 116013-116023, 2020, doi: 10.1109/ACCESS.2020.3003790

[4] R. Dharavath and E. Khosla, "Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices," 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), Rourkela, India, 2019, pp. 47-52, doi: 10.1109/iSES47678.2019.00023

[5] N. Tarannum and S. V. M. S, "A Brief Introduction to Demand Forecasting using ARIMA models," Dept. of Computer Science & Engineering, Rashtreeya Vidyalaya College of Engineering, Karnataka, India.