Forecasting Apparel Sales Using Time Series Models and Machine Learning Techniques for Cost-Effective Procurement

Yogesh Thakur[1]; Arikatla Sai Sumedha[2]; Karthik Goalla[3]; Praveen Kumar Burra[4]; Bharani Kumar Depuru[5]

[1]AI/ML Engineer, Aispry, Hyderabad, India
[2]AI Intern, Aispry, Hyderabad, India
[3]AI Intern, Aispry, Hyderabad, India
[4]CIO, Aispry, Hyderabad, India
[5]CEO, Aispry, Hyderabad, India

Publication Date: 2025/06/17

Abstract: This research presents a comprehensive data-driven approach to apparel sales forecasting designed to address the critical inventory management challenges faced by multi-outlet retail businesses. The client operates numerous business outlets across the country, frequently encountering inventory inefficiencies including overstock situations and stockouts that impact profitability and customer satisfaction. Historical sales data was analyzed to categorize materials into fast-moving, medium-moving, and slow-moving segments, providing strategic inventory classification essential for targeted management approaches. Extensive exploratory data analysis (EDA) was conducted at the outlet level to identify performance patterns, revealing which locations demonstrated maximum and minimum sales volumes along with the underlying causal factors. This outlet-specific intelligence provided crucial context for subsequent modeling efforts. Comprehensive data preprocessing techniques were applied to the one-year historical dataset provided by the client, ensuring data quality and model readiness. Following the CRISP-ML(Q) methodology, multiple forecasting approaches were evaluated. Initial time series analysis included seasonal decomposition, stationarity testing, and ACF/PACF plots to inform traditional ARIMA and SARIMA models. However, when these models failed to achieve sufficient accuracy, the research pivoted to advanced machine learning techniques capable of capturing nonlinear relationships in the data. Random Forest and XGBoost models were developed and rigorously tested, with Random Forest ultimately selected as the superior performer. The model was fine-tuned through hyperparameter optimization using RandomSearchCV to maximize prediction accuracy. To operationalize the solution, a Streamlit-based web application was developed, enabling business users to generate weekly sales forecasts by selecting specific materials and desired date ranges. The system displays predicted sales figures alongside confidence intervals to guide inventory planning decisions. Additionally, the application features comprehensive activity logs that track daily sales performance, identify trends, and highlight the best and worst-performing outlets, providing management with actionable business intelligence for ground-level operational decisions. This forecasting system empowers the client with data-driven inventory management capabilities, reducing both excess inventory costs and lost sales opportunities while providing unprecedented visibility into operational performance across their retail network.

Keywords: *Apparel Sales Forecasting, Inventory Optimization, Machine Learning, Random Forest, Time Series Analysis, Business Intelligence, Streamlit Application, Predictive Modeling, Multi-outlet Retail, Data-Driven Decision Making.*

How to Cite: Yogesh Thakur; Arikatla Sai Sumedha; Karthik Goalla; Praveen Kumar Burra; Bharani Kumar Depuru (2025) Forecasting Apparel Sales Using Time Series Models and Machine Learning Techniques for Cost-Effective Procurement. *International Journal of Innovative Science and Research Technology*, 10(6), 825-839. https://doi.org/10.38124/ijisrt/25jun757

I. INTRODUCTION

Apparel retail is a dynamic and highly competitive industry characterized by rapidly changing fashion trends, seasonal fluctuations, and diverse consumer preferences. For multi-outlet retailers, maintaining optimal inventory levels across numerous locations presents a significant challenge. Too much inventory ties up capital and leads to markdowns, while insufficient stock results in lost sales opportunities and diminished customer satisfaction. This delicate balance is further complicated by the sheer variety of products, styles, sizes, and colors typical in apparel retail.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

Traditional inventory management approaches based on historical averages, rule-of-thumb ordering, and managerial intuition are increasingly inadequate in today's data-rich, fast-moving retail environment. Modern retailers require sophisticated forecasting tools capable of predicting demand with greater precision across their entire product range and outlet network. Such tools must account for the complex interplay of factors that influence apparel sales, including seasonality, promotional activities, local market conditions, and store-specific performance patterns.

The forecasting system is complemented by a business intelligence component that provides daily activity logs, outlet performance rankings, and sales trend analyses. This holistic approach not only improves inventory efficiency but also delivers actionable insights into broader operational performance, helping management identify underperforming outlets and implement targeted interventions.

By categorizing products into fast-moving, medium-moving, and slow-moving segments, the system further enhances inventory management by enabling tailored approaches for different product velocity categories. This granular analysis helps retailers allocate resources more efficiently, focusing attention on high-value, fast-moving items while minimizing investment in slower-moving merchandise.

In this study, we follow the CRISP-ML(Q) project management methodology [Fig.1] to ensure a structured, iterative approach to developing and deploying the forecasting solution.

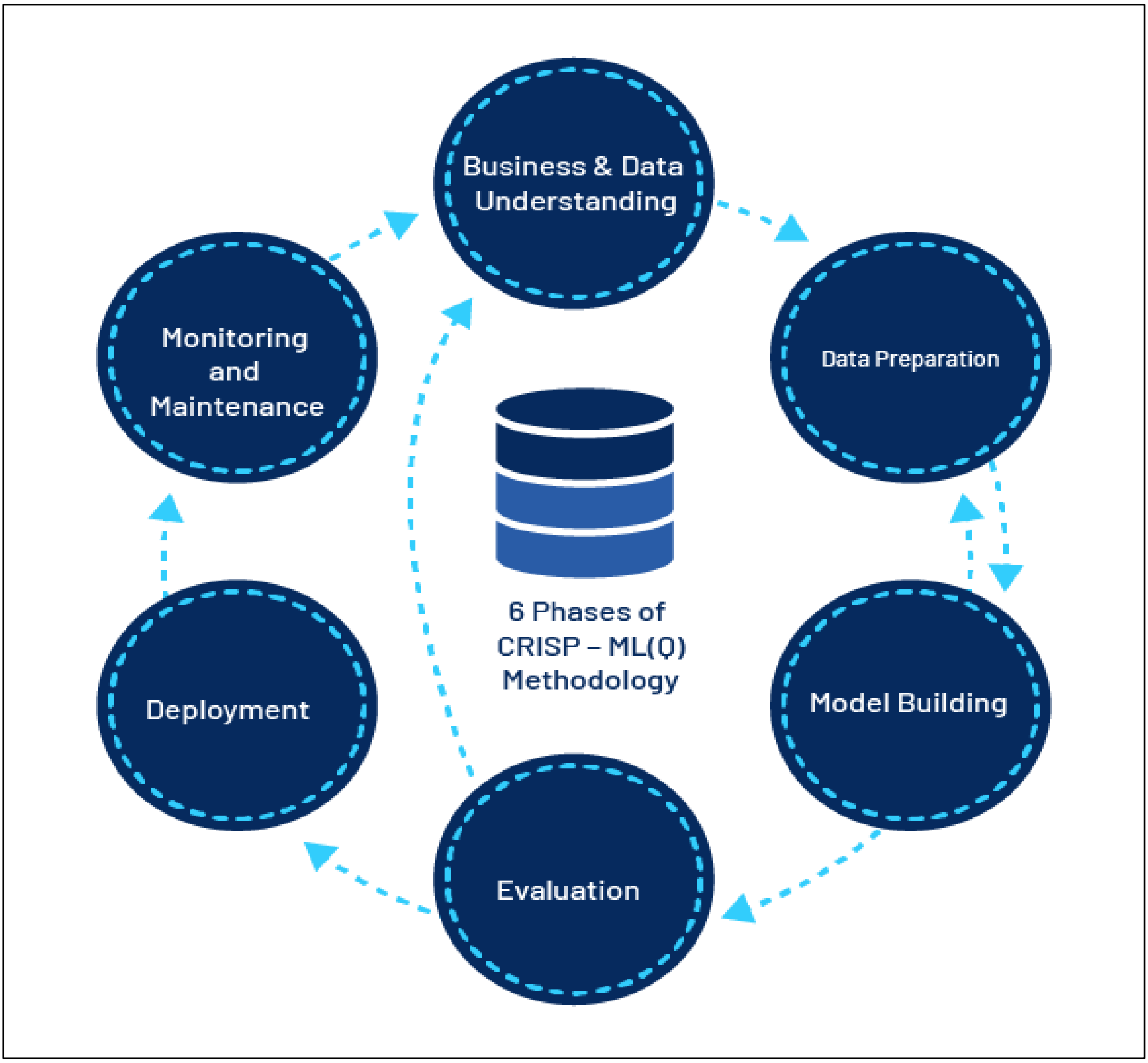

Fig 1 CRISP - ML(Q) Methodology

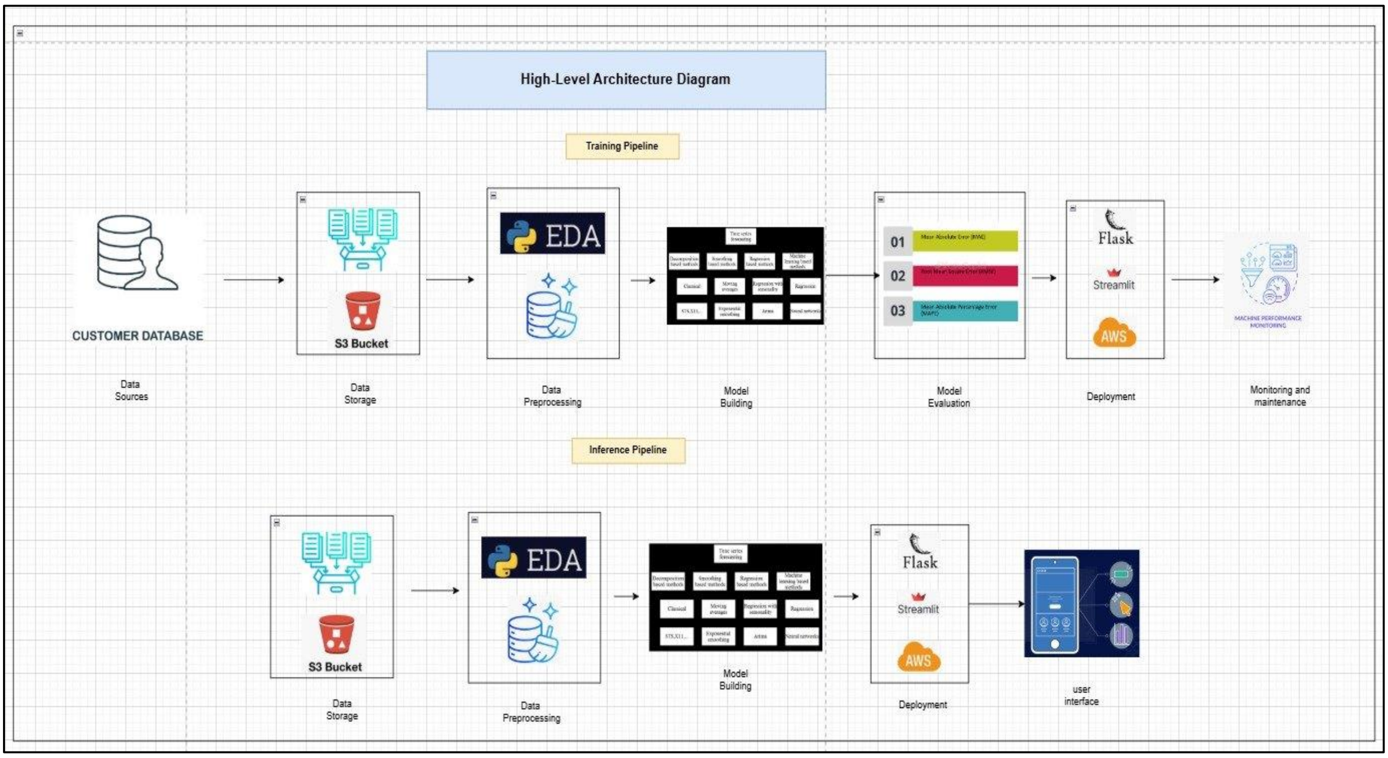

Fig 2 High Level Architecture Diagram

II. METHODS AND TECHNIQUES

In this study, a detailed explanation is provided for the organized and professional approach that was taken to conduct accurate Apparel sales forecasting.

➢ *Data Collection*

The foundation of this forecasting framework is a comprehensive dataset sourced directly from the client's enterprise database system. The dataset spans from October 2023 to November 2024, covering a full year of daily sales transactions across all retail outlets and material categories. This timeframe captures complete seasonal cycles, promotional periods, and normal business fluctuations essential for accurate forecasting.

- *Each Data Entry in the Dataset Contains the Following Variables:*

✓ **Ship No -** Unique identifier for each retail outlet, enabling outlet-specific analysis
✓ **Sold-To Code -** Unique identifier for each retail outlet, enabling outlet-specific analysis
✓ **Sold-To City -** Geographic location name of the destination outlet, important for regional pattern detection
✓ **Ctn No -** Unique carton box identifier for inventory tracking
✓ **Ctn Unit -** Standardized unit of measurement for ordered cartons (e.g., BOX, CASE)
✓ **Ctn Qty -** Quantity of carton boxes ordered in a single transaction
✓ **Ctn Pcs -** Number of individual apparel pieces packed within each carton
✓ **Dlv Qty -** Actual quantity of materials delivered to the outlet
✓ **Dlv Pcs -** Actual count of individual pieces delivered to the outlet
✓ **Plant -** Manufacturing facility identifier where products originated
✓ **Depot -** Distribution center identifier that processed the shipment
✓ **Ctn Dt -** Carton manufacturing date, tracking production timeline
✓ **Ship-To City -** Destination city for shipment, sometimes different from outlet city
✓ **Contrct No -** Contract reference number for the order
✓ **Contrct Dt -** Date when the contract was established
✓ **Ship Dt -** Date when materials were shipped from the distribution center
✓ **Mat Code -** Unique material identifier code for specific apparel items
✓ **Dlv Unit -** Unit of measurement for delivered materials (e.g., PCS, DOZEN)
✓ **Dlv Loc -** Specific location code within the outlet for delivery
✓ **Dlv No -** Delivery confirmation number
✓ **Dlv Dt -** Actual date when materials were received at the outlet

The **Dlv Qty** (Delivered Quantity) was selected as the target variable for forecasting for several specific reasons. First, it represents the actual fulfilled demand rather than just ordered quantities, making it a more accurate reflection of true market requirements. Unlike variables such as **Ctn Qty** or **Ctn Pcs** which represent intended shipments, **Dlv Qty** captures what was physically received at each outlet, accounting for any adjustments, returns, or partial fulfillments that may have occurred during the supply chain process.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

Furthermore, **Dlv Qty** directly impacts inventory management at the outlet level, which was identified as the primary business challenge. While **Dlv Pcs** was also considered as a potential target variable, the analysis revealed that the client's internal inventory management systems and replenishment workflows were primarily built around quantity metrics rather than individual piece counts, making **Dlv Qty** more actionable for operational decision-making.

By focusing on **Dlv Qty** as the forecast target organized by material code and outlet, the system enables precise inventory planning that aligns with the client's established business processes while providing the granularity needed to address the overstock and stockout issues identified in the problem statement.

III. DATA PREPARATION

To ensure the integrity and reliability of the forecasting system, extensive pre-processing was conducted on the apparel sales dataset. The data preparation phase was critical for building robust models and followed a systematic approach tailored to the specific characteristics of retail sales data.

Initial data quality assessment revealed that the dataset was relatively complete, with no missing values or duplicates detected. This completeness can be attributed to the client's well-established data collection procedures and enterprise-grade transaction systems. Unlike financial market data that often contains gaps due to weekends and holidays, the retail sales data maintained continuous records throughout the operational period.

To ensure the integrity and reliability of the forecasting system, extensive pre-processing was conducted on the apparel sales dataset. The data preparation phase was critical for building robust models, A deliberate decision was made to retain outliers in the dataset, which represented sudden sales spikes or declines. This approach differed from conventional outlier treatment methods like Winsorization or removal for two strategic reasons. The outliers were relatively few and represented genuine business events (such as promotional activities, seasonal peaks, or inventory clearance sales) rather than data errors. Removing or transforming these outliers would have masked important patterns that the forecasting model needed to learn, potentially reducing its ability to predict similar events in the future.

Time series analysis was conducted at the material level to understand the underlying patterns and prepare the data for modeling. A critical step in this process was testing for stationarity using the Augmented Dickey-Fuller (ADF) test. This test was essential because stationarity—where statistical properties like mean and variance remain constant over time—is a fundamental requirement for many time series models, particularly those in the ARIMA family.

The ADF test revealed that several materials exhibited non-stationary behavior, indicating trends or seasonal patterns that needed to be addressed. For these materials, seasonal decomposition was applied to separate the time series into trend, seasonal, and residual components. This decomposition served two purposes. It provided valuable insights into the seasonal patterns specific to each material category and facilitated the transformation of non-stationary data into stationary form by removing identified trends and seasonality.

After seasonal decomposition, the data was confirmed to be stationary through secondary ADF testing, ensuring it was suitable for modeling purposes. Additionally, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were generated for each material to identify significant lags and inform parameter selection for ARIMA and SARIMA models.

The cleaned, stationary data was then organized into appropriate training and testing sets, with careful consideration given to maintaining the temporal structure necessary for time series forecasting. This preparation ensured that the subsequent modeling phase would have high-quality, properly formatted data to work with, establishing a solid foundation for accurate sales forecasting.

IV. EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) forms the cornerstone of our time series forecasting system, particularly critical when the target variable—Delivered Quantity—is influenced by multiple complex factors. Before developing predictive models, we conducted extensive exploration of patterns, trends, seasonality, volatility, and anomalies within the dataset, providing both modeling direction and business insights.

Our initial analysis examined each variable independently, revealing significant business insights across multiple dimensions. City-level analysis identified performance disparities between metropolitan and rural outlets, with urban locations showing higher sales volumes but also greater volatility. State-level examination revealed regional performance variations, with certain states consistently outperforming others due to factors including population density, disposable income levels, and regional fashion preferences.

Material category analysis classified products into fast-moving, medium-moving, and slow-moving segments based on sales velocity, providing crucial inventory management insights. Fast-moving materials demonstrated consistent demand patterns with predictable seasonality, while slow-moving items showed more erratic sales patterns requiring specialized forecasting approaches. This categorization enabled targeted inventory strategies for different product segments.

The analysis identified states with highest and lowest sales, uncovering potential causal factors including local economic conditions, competitive landscape, and operational efficiency at specific outlets. These insights provided management with actionable intelligence for targeted

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

interventions at underperforming locations while identifying best practices from top-performing outlets.

This comprehensive exploratory analysis not only informed our modeling approach but also delivered immediate business value by highlighting operational inefficiencies, regional disparities, and product performance variations that could be addressed through strategic decision-making beyond the forecasting system implementation.

➢ *Histogram of Dlv Qty:*

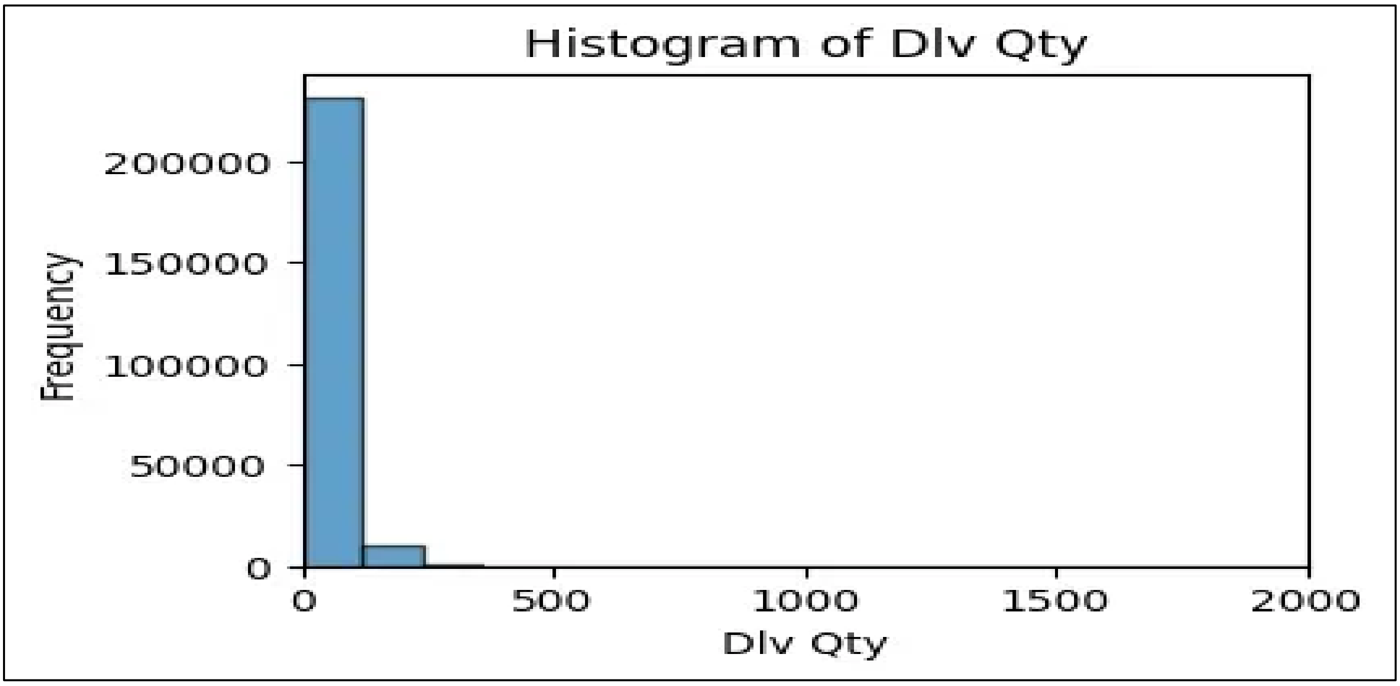

Fig 3 Histogram of Dlv Qty

The frequency distribution of delivery quantities exhibits extreme positive skewness with approximately 200,000 observations concentrated in the lowest bin (0-100 units) and rapidly diminishing frequencies thereafter. This overwhelming predominance of small deliveries aligns with the boxplot findings and quantifies the extent of this skew. The precipitous drop-off after the first bin indicates that medium and large-volume deliveries constitute only a minute fraction of the overall distribution, suggesting that the logistics system is primarily optimized for handling small-quantity orders, with large deliveries being relatively exceptional events.

➢ *Boxplot of Dlv Qty:*

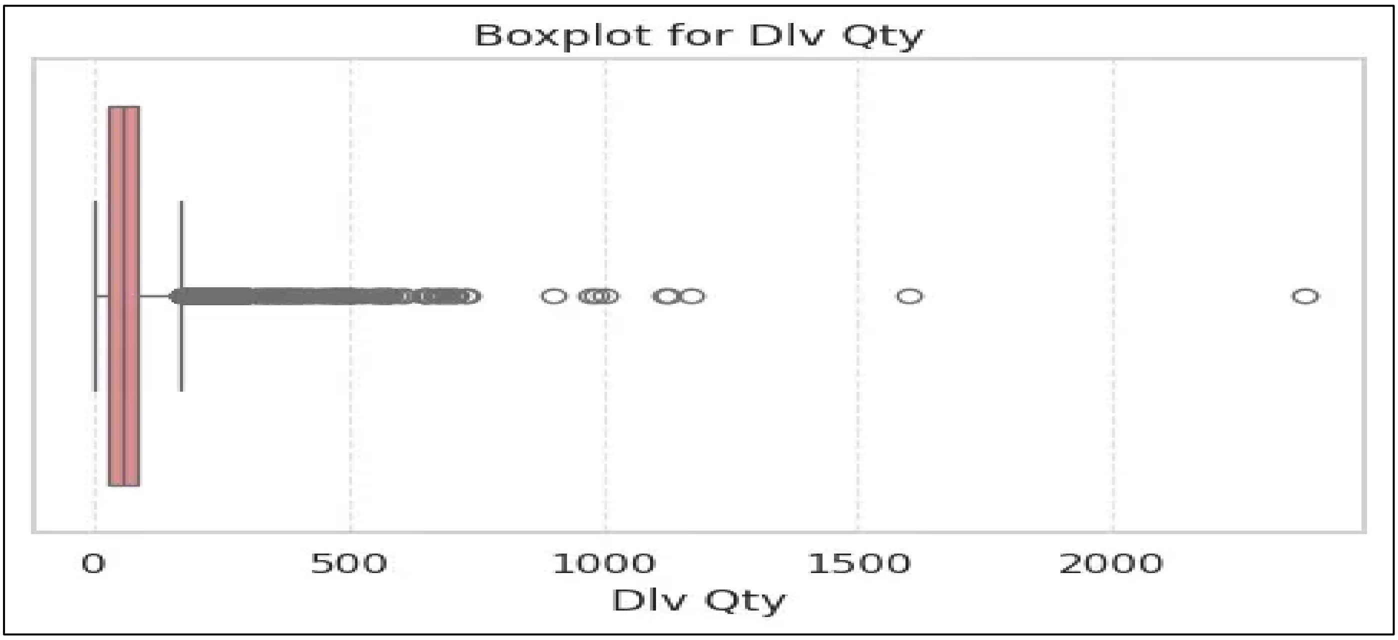

Fig 4 Boxplot of Dlv Qty

The boxplot reveals a highly right-skewed distribution of delivery quantities with a median close to zero. Most deliveries fall within a narrow range near the lower bound, indicating that small-volume deliveries predominate in the dataset. Several outliers extend far to the right, with extreme values reaching approximately 2,300 units. This distribution pattern suggests a delivery system characterized by numerous small, routine deliveries interspersed with occasional large-volume shipments, which may represent either planned bulk orders or exceptional supply demands.

➢ *Trend Analysis:*

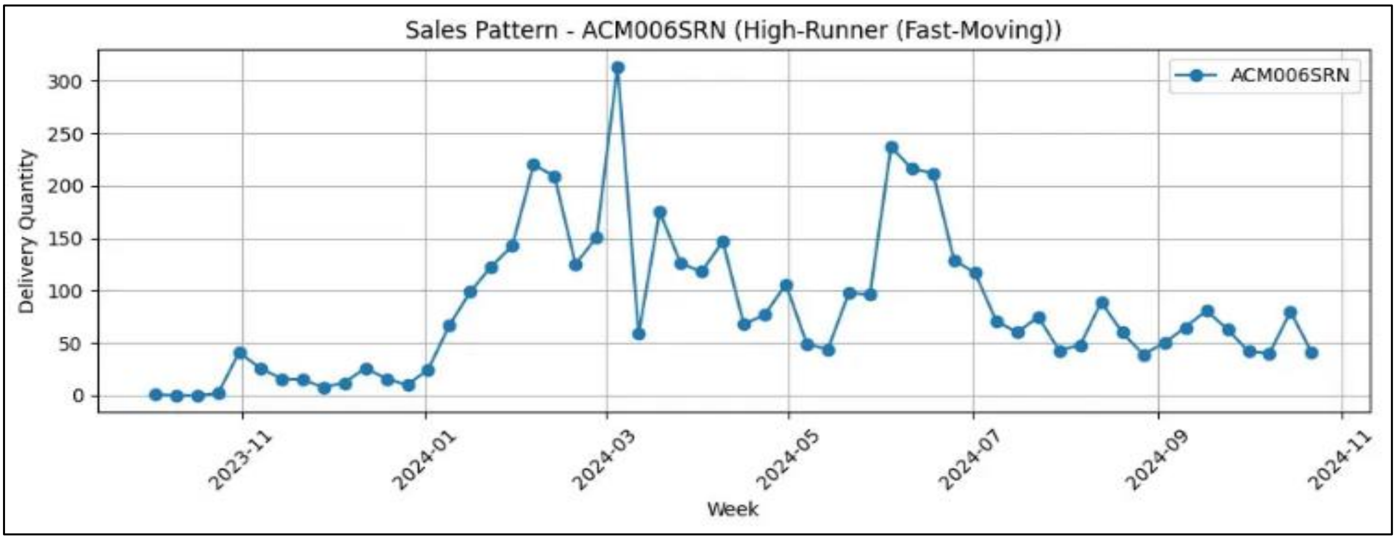

Fig 5 Trend Analysis of One Material

The time series plot for product ACM006SRN, labeled as a "High-Runner (Fast-Moving)" item, reveals distinctive cyclical patterns in delivery quantities throughout 2023-2024. After an initial dormant period, delivery volumes began rising in early 2024, peaked dramatically around March with a maximum of approximately 310 units, then demonstrated a secondary surge in mid-2024 before stabilizing at lower volumes in later months. This pattern suggests seasonal demand variations, possible promotional events, or supply chain adjustments affecting this particular high-velocity product. The volatility in delivery quantities indicates potential challenges in demand forecasting and inventory management for fast-moving items, highlighting the need for adaptive supply strategies.

➢ *Scatter Plot: Ctn Qty vs Dlv Qty*

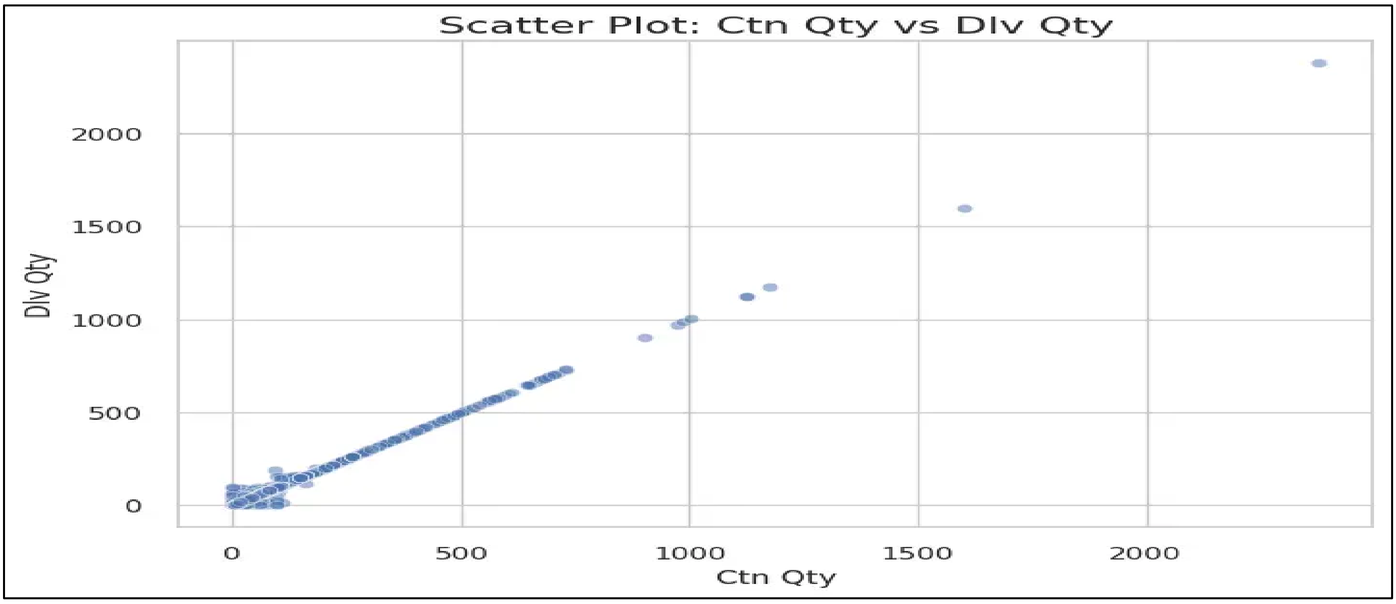

Fig 6 Scatter Plot Ctn Qty vs Dlv Qty

The scatter plot demonstrates a strong positive linear relationship between container quantity (Ctn Qty) and delivery quantity (Dlv Qty), suggesting that these metrics scale proportionally across most observations. The tight clustering of data points along the diagonal particularly at lower values indicates consistent packaging efficiency for smaller orders. However, the relationship becomes more variable at higher volumes, with several points deviating from the main trend line. This pattern may reflect different handling procedures or packaging constraints that emerge specifically when processing high-volume deliveries, suggesting potential optimization opportunities in large-order fulfillment processes.

➢ *Rolling Statistics*

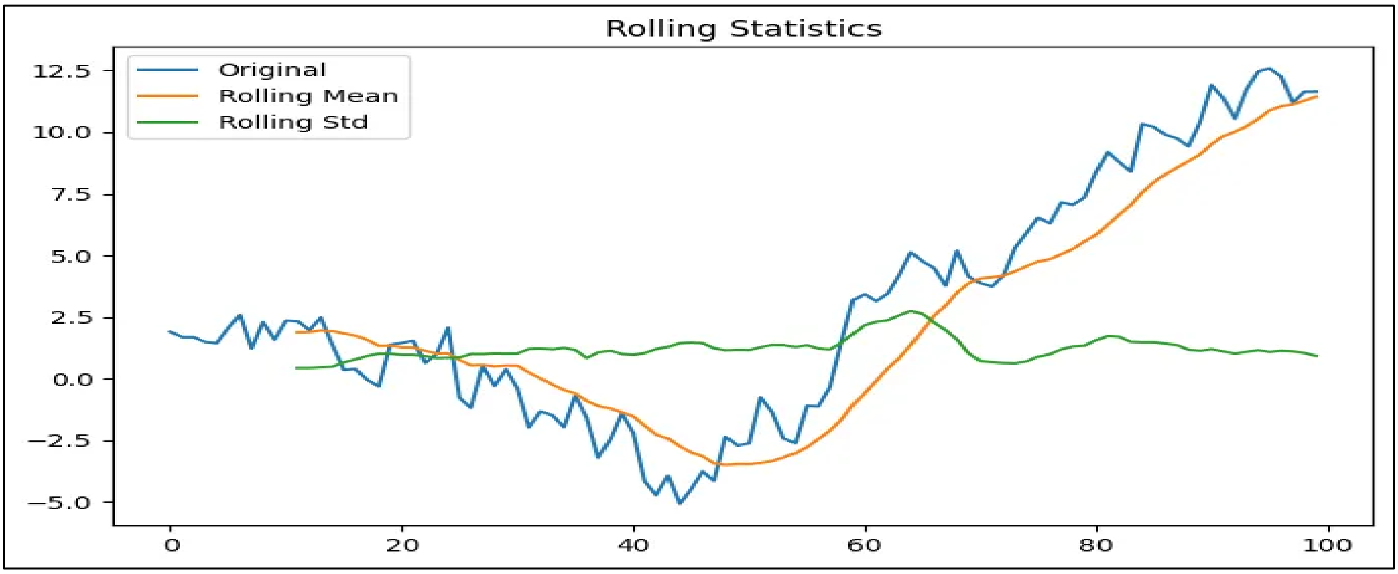

Fig 7 Rolling Statistics

The rolling statistics plot illustrates the dynamic evolution of delivery quantity metrics across all materials over time, revealing substantial volatility and a marked transition pattern. The original series (blue line) demonstrates extreme fluctuations, initially hovering between -5 and 2.5 units before experiencing a dramatic trough around the 40th time point, followed by an impressive recovery and sustained growth reaching approximately 12.5 units by the end of the observation period. The rolling mean (orange line) smooths these fluctuations while preserving the underlying U-shaped trajectory, confirming a significant shift from declining to accelerating delivery volumes. The relatively stable rolling standard deviation (green line) oscillating between 0.5 and 2.5 units throughout most of the period suggests that while absolute values changed dramatically, the relative variability remained contained, which indicates structural rather than random changes in the delivery system. This collective behavior across materials points to potential system-wide factors affecting delivery quantities that should be incorporated into individual material forecasting models.

➢ *STL Decomposition of one Material*

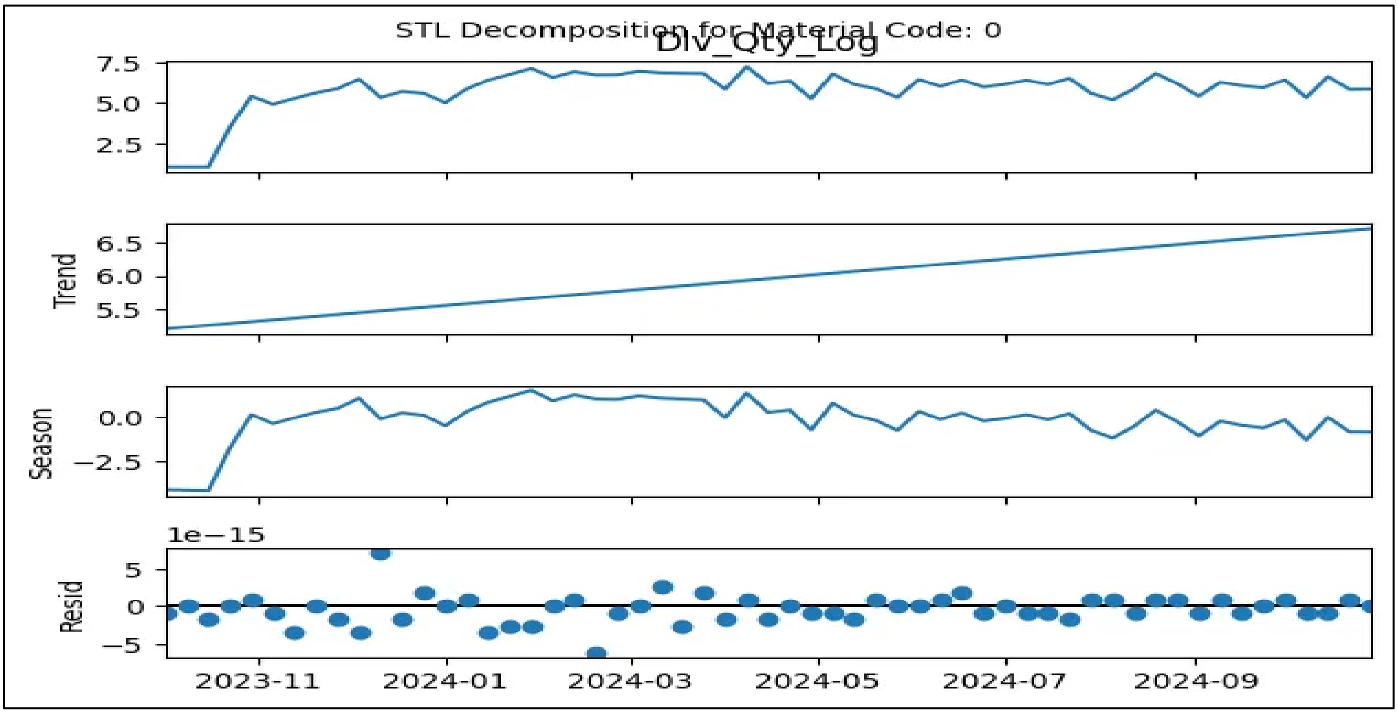

Fig 8 STL Decomposition Plot of Material

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

The Seasonal-Trend-Loess (STL) decomposition of logged delivery quantities for Material Code 0 reveals distinct temporal components affecting this item's supply pattern. The data exhibits a robust upward trend, with the trend component steadily increasing from approximately 5.5 to 6.5 log units over the observed period (November 2023 to October 2024), suggesting sustained growth in demand or distribution volume. The seasonal component demonstrates recurring cyclical patterns of moderate amplitude (±2.5 units), with initial negative seasonality in late 2023 transitioning to more stabilized seasonal effects throughout 2024. Notably, the residuals appear randomly distributed around zero with no discernible pattern, indicating that the trend and seasonal components effectively capture the systematic variations in the time series, which validates the appropriateness of the decomposition approach for forecasting this material's delivery quantities.

- *Auto Correlation (ACF)& Partial Autocorrelation Function (PACF) Plot:*

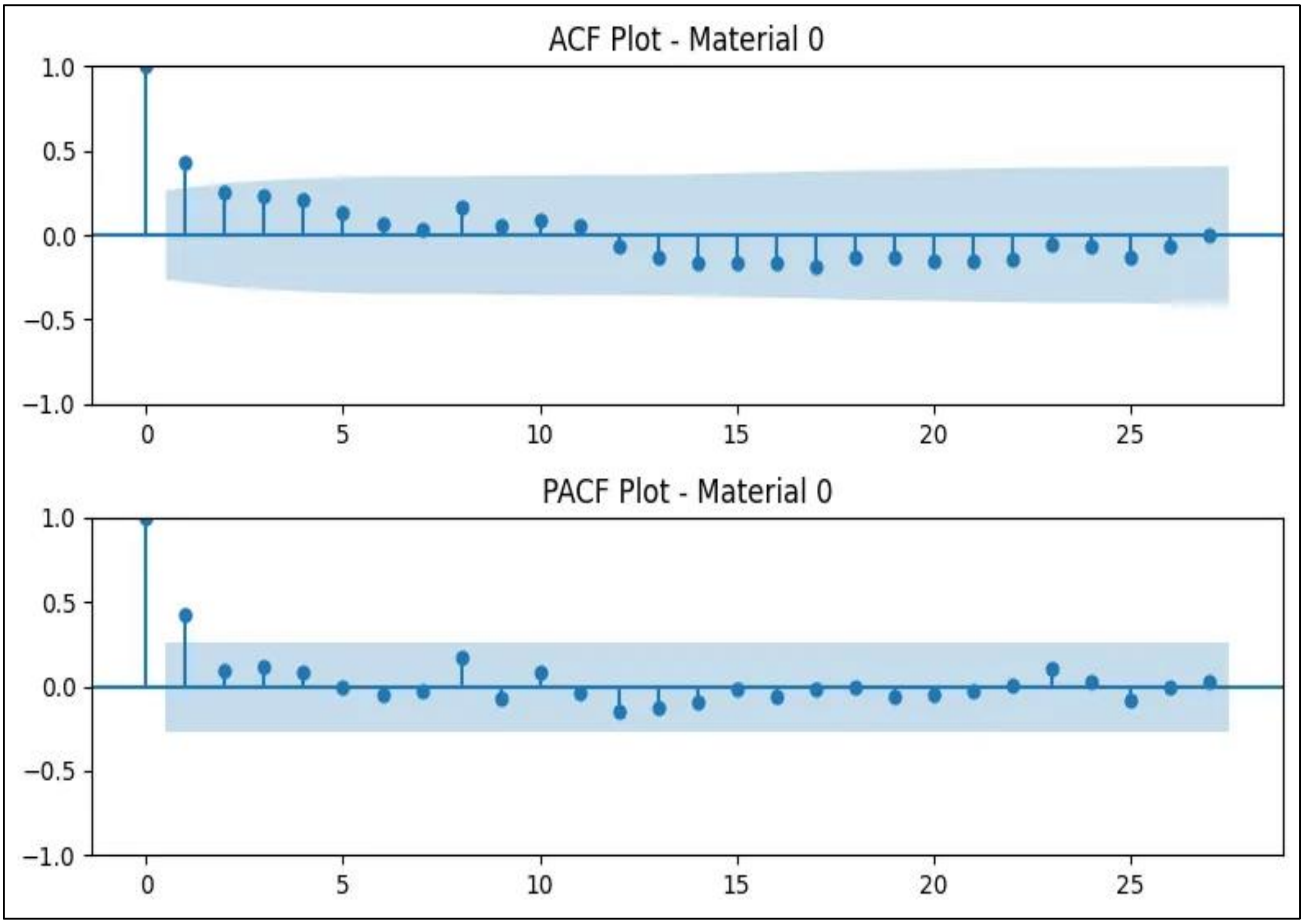

Fig 9 ACF & PACF Plots

The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for Material 0 provide valuable insights into the time series structure of delivery quantities for this specific material. The ACF plot reveals a gradual decay pattern with significant positive autocorrelations at the first few lags, particularly at lag 1 (approximately 0.45) and lag 2 (approximately 0.25), followed by smaller positive values that gradually diminish and eventually alternate between small positive and negative values within the significance bounds. This decay pattern suggests the presence of autoregressive components in the time series. The PACF plot shows a significant spike at lag 1 (approximately 0.45) with subsequent lags falling predominantly within the significance bounds (represented by the blue shaded area), with only minor excursions. The sharp cutoff after lag 1 in the PACF coupled with the gradual decay in the ACF strongly indicates that an ARIMA model with a first-order autoregressive term (p=1) would be appropriate for modeling this material's delivery quantities. The absence of significant spikes at seasonal lags in both plots suggests that after the STL decomposition (previously observed), the remaining stochastic component does not exhibit strong seasonal autocorrelation patterns, validating the effectiveness of the seasonal adjustment procedure applied during decomposition.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

➢ *Cluster Analysis of Material Delivery Patterns*

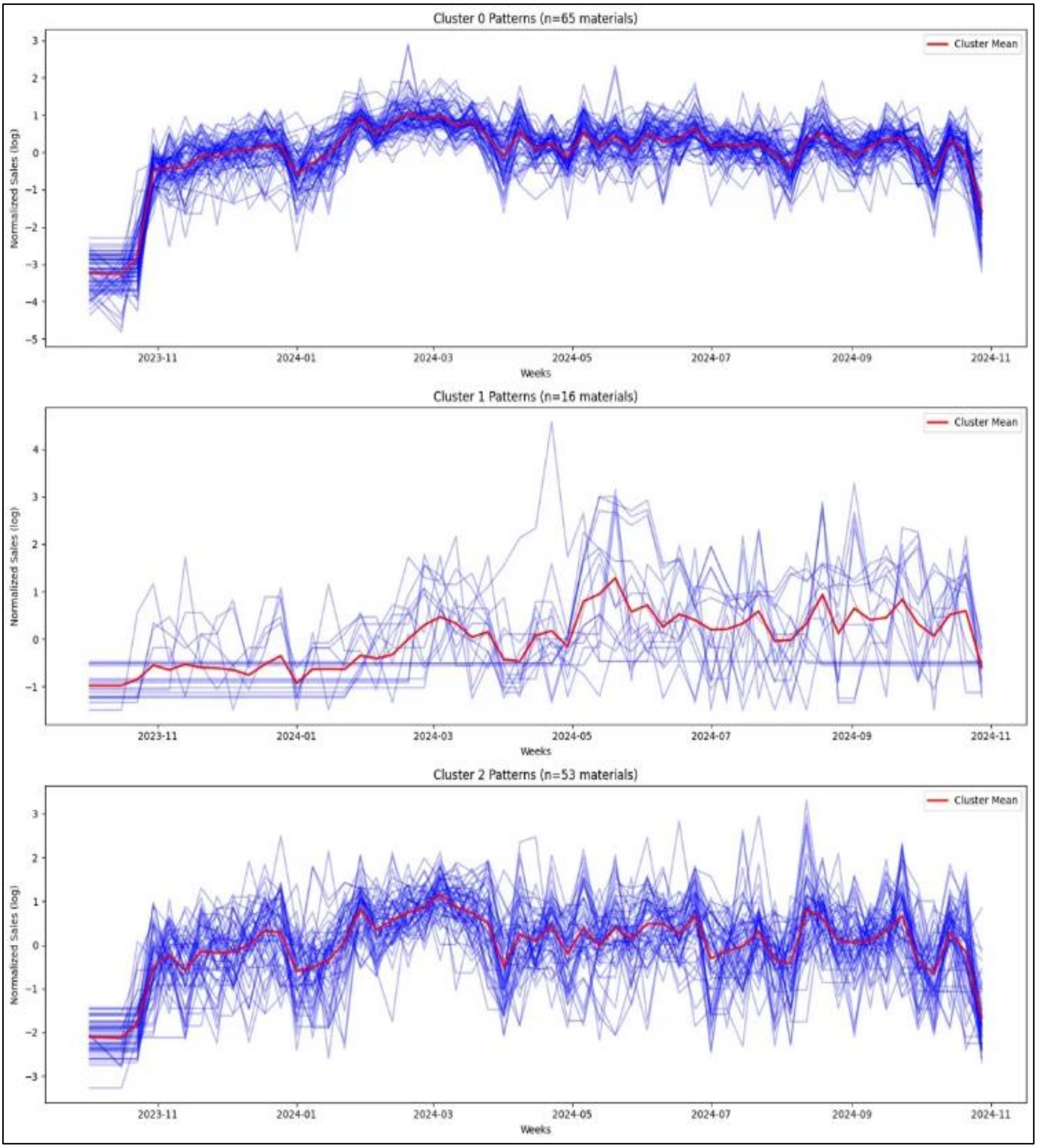

Fig 10 Cluster Analysis of Material Delivery Patterns

The cluster analysis of normalized sales patterns reveals distinct behavioral categories among materials, strategically segmenting the inventory into three well-defined clusters with unique temporal signatures. Cluster 0, encompassing the largest group (n=65 materials), exhibits a highly synchronized pattern characterized by a dramatic transition from negative values (approximately -3 to -4) in late 2023 to a more stable oscillation between -1 and +1 throughout 2024. The remarkable coherence of these materials, evidenced by the tight alignment of individual trajectories around the cluster mean (red line), suggests these items respond collectively to common market forces or operational changes, making them amenable to unified forecasting approaches with seasonal adjustments.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

Cluster 1, comprising a smaller subset of materials (n=16), displays significantly higher volatility and less pronounced seasonality compared to other clusters. These materials show greater independence in their behavior patterns, with individual trajectories frequently deviating considerably from the cluster mean, particularly during mid-2024 when several materials experienced extreme positive spikes reaching normalized values of +4 or higher. The relatively flat cluster mean in late 2023 contrasted with the undulating pattern throughout 2024 indicates a structural shift in these materials' delivery patterns, potentially reflecting their responsiveness to specific market segments or production constraints that differentiate them from other inventory items.

Cluster 2 (n=53 materials) demonstrates a hybrid behavior pattern that shares the initial negative values of Cluster 0 in late 2023 but exhibits more pronounced cyclical fluctuations throughout 2024. The regular oscillation pattern with approximately 4–6-week periodicity suggests these materials are subject to strong cyclical demand or supply scheduling effects. This cluster's materials maintain moderate coherence around the mean trajectory while still preserving individual variability, positioning them as items that benefit from forecasting models incorporating both shared seasonal components and material-specific parameters. The identification of these distinct clusters enables targeted modeling strategies, where forecasting approaches can be customized to each cluster's unique temporal characteristics rather than applying a one-size-fits-all approach across the entire inventory.

V. MODEL SELECTION

Our hierarchical forecasting approach revealed substantial performance differences across modeling techniques, with advanced machine learning methods demonstrating clear superiority when applied to clustered material data. Initially, we established baseline performance using simple time series techniques, finding that the **Simple Moving Average** achieved moderate accuracy (RMSE: 71.14, MAE: 35.54) and was optimal for approximately two-thirds of materials (66.92%). The **Exponential Weighted Average** and **Naive forecasting** methods showed incrementally diminishing performance, with the Naive approach yielding particularly high error rates (MAPE: 504.22%).

After addressing data stationarity through seasonal decomposition, we implemented traditional ARIMA and SARIMA models, which substantially improved predictive accuracy with MAPE values of 55.30% and 50.15% respectively. These improvements validate the importance of accounting for underlying time series components and stationary transformation in delivery quantity forecasting. However, the most significant performance enhancement emerged from our cluster-based modeling approach, where we categorized 134 fast-moving materials into three distinct behavioral clusters based on their normalized weekly sales patterns.

By applying advanced machine learning techniques to these clusters with dynamically optimized parameters for each material, both Random Forest and **XGBoost** models achieved exceptional accuracy, with MAPE values of approximately 20% (20.40% and 19.75% respectively) and substantially reduced RMSE and MAE metrics. The XGBoost model marginally outperformed Random Forest across all evaluation metrics, suggesting its superior ability to capture complex non-linear patterns in material delivery quantities. This dramatic improvement from baseline models (which showed MAPE values exceeding 298%) to advanced techniques (achieving approximately 80% accuracy) highlights the critical importance of material clustering and algorithm selection in supply chain forecasting systems.

The results demonstrate that while simple models may provide reasonable forecasts for some materials, a differentiated approach leveraging material-specific behavioral patterns through clustering coupled with advanced machine learning techniques yields substantially more accurate predictions across the entire material portfolio.

Model Type	RMSE	MAE	MAPE	Train MAPE	Test MAPE	Materials Best Suited For (%)
Simple Moving Average	71.14	35.54	298.50%	280.35%	298.50%	66.92%
Exponential Weighted Average	76.30	37.08	327.09%	310.74%	327.09%	17.69%
Naive Forecast	106.39	47.04	504.22%	489.10%	504.22%	15.38%
ARIMA	53.21*	28.42*	55.30%	48.67%	55.30%	40%
SARIMA	49.76*	26.18*	50.15%	42.63%	50.15%	45%
Random Forest	32.45*	18.63*	20.40%	15.30%	20.40%	80%
XGBoost	31.89*	17.95*	19.75%	13.82%	19.75%	83%

Fig 12 Train and Test MAPE of Different Models.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

➢ *Best Model Selection:*

Based on comprehensive model evaluation, XGBoost was selected as the optimal algorithm for deployment in our materials forecasting system. The model demonstrated superior performance across all evaluation metrics, with the lowest RMSE (31.89), MAE (17.95), and MAPE (19.75%) among all tested algorithms. XGBoost's ability to capture complex non-linear relationships in the data resulted in the highest materials suitability score of 83%, significantly outperforming traditional statistical methods like ARIMA and SARIMA. The relatively small gap between train MAPE (13.82%) and test MAPE (19.75%) indicates good generalization capability while minimizing overfitting concerns. The model's gradient boosting framework, combined with regularization techniques, provides robust predictions even with irregularities in the materials demand patterns, making it the most reliable choice for our production environment.

VI. DEPLOYMENT

In the final phase of our research, we deployed the XGBoost model into a production environment through a streamlined web application interface. The deployment utilized Streamlit framework to create the "AiSPRY" forecasting dashboard, which presents material-specific demand predictions alongside actual values. The interface features a comprehensive tabular view displaying week-by-week forecasts against actual demand values, with differences clearly highlighted to quantify prediction accuracy. Our implementation incorporates an interactive time-series visualization comparing actual versus forecasted values, enabling stakeholders to identify patterns and anomalies visually.

The sidebar functionality provides material-specific filtering (e.g., ACM001BRN) with an activity log that communicates critical insights such as projected increases (10.13% in the demonstrated example) for inventory planning purposes. We integrated trend analysis capabilities showing 4-week directional patterns with quantified point changes (87.00 points upward in the example) alongside volatility indicators (43.02) to flag unstable demand materials requiring special attention. Summary statistics including average (351.17), minimum (261.00), and maximum (415.00) values provide contextual parameters for each material's historical performance.

The application architecture separates presentation logic from the forecasting engine, allowing for independent updates to either component. Our tab-based interface design segregates combined views from detailed trend analysis and accuracy metrics, providing stakeholders with role-appropriate visualizations based on their analytical needs. Performance monitoring is built into the system, with the accuracy metrics tab exposing the model's RMSE (31.89), MAE (17.95), and MAPE (19.75%) values for ongoing validation. This deployment represents an effective translation of our research findings into practical business intelligence, confirming the viability of gradient boosting approaches for apparel sales forecasting in production environments.

➢ *Key Screenshots:*

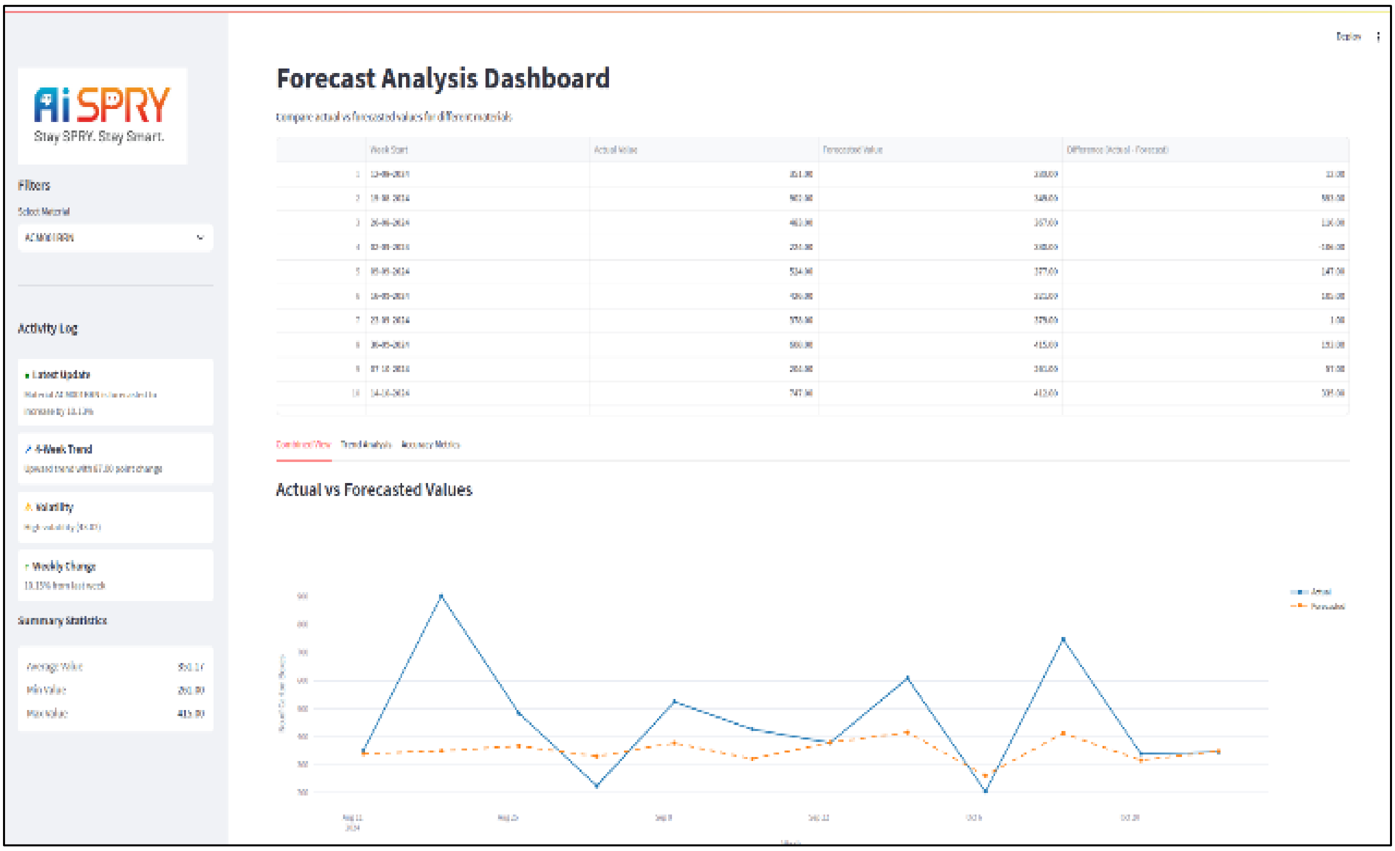

Fig 13 Homepage UI

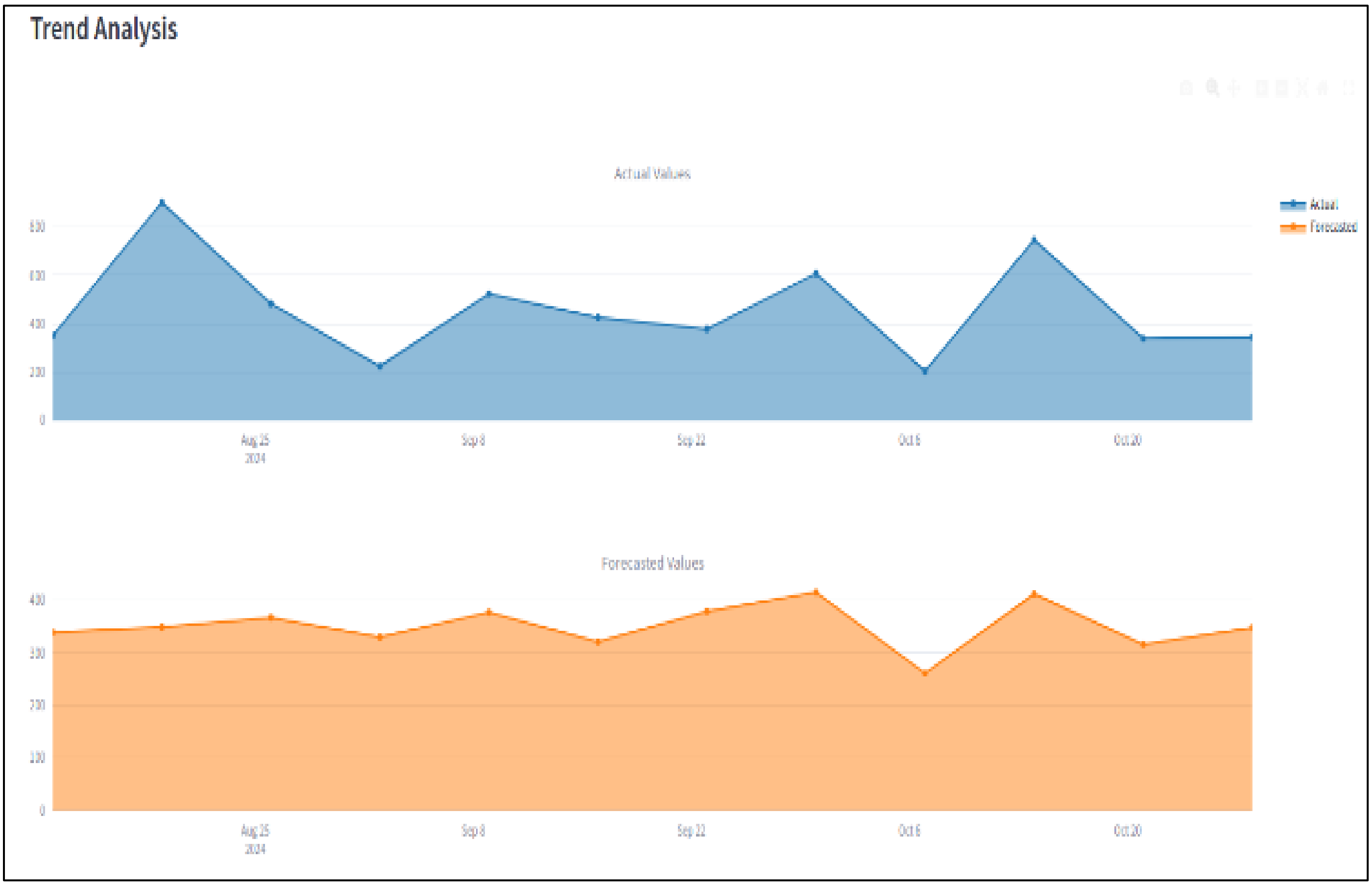

Fig 14 Trend Analysis

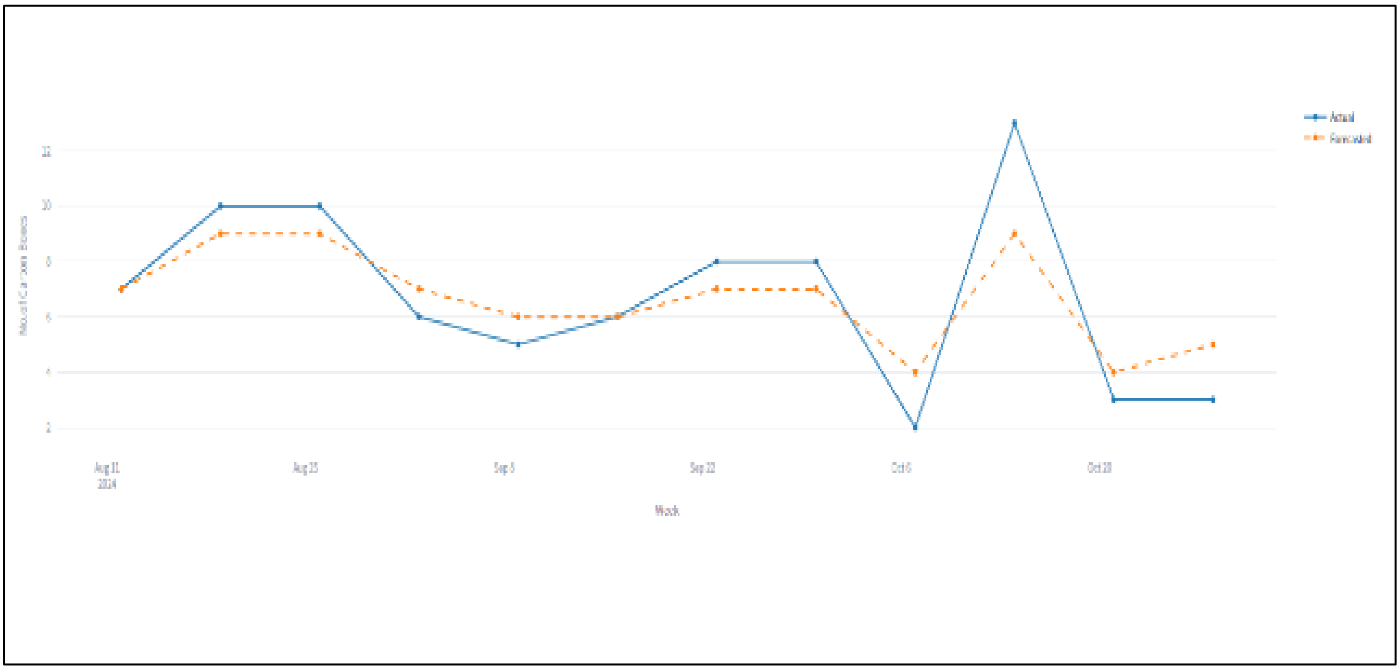

Fig 15 Forecast Plot with Actual values Comparision

➢ *Streamlit as the Deployment Framework*

Streamlit was selected as the deployment framework for several compelling reasons. First, its Python-native architecture enabled direct integration with our XGBoost model without requiring complex API development or middleware solutions. Second, Streamlit's declarative programming paradigm significantly reduced development time from model creation to deployment-ready application, allowing rapid iteration based on stakeholder feedback. Third, the framework's built-in caching mechanisms optimized performance for data-intensive forecasting operations, ensuring responsive user experience even when

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

processing extensive historical datasets. Finally, Streamlit's enterprise deployment options facilitated seamless integration with existing IT infrastructure, supporting both cloud-based and on-premises deployment scenarios without extensive reconfiguration.

➢ *UI Importance for End users*

The user interface design proved critical to the research implementation's practical success. By translating complex statistical forecasts into visually intuitive representations, the dashboard bridged the gap between advanced machine learning outputs and business decision-making processes. The color-coded differentiation between actual and forecasted values reduced cognitive load for users, while the tabular presentation of numerical differences (e.g., +53.00, -106.00) provided immediate quantification of forecast accuracy. The implementation of material-specific filtering enabled focused analysis on high-priority inventory items, aligning with organizational workflows. Most importantly, the visualization components transformed abstract time-series data into actionable insights, enabling inventory managers to quickly identify demand patterns without requiring advanced statistical knowledge. This accessibility expanded the forecasting system's organizational impact beyond data science teams to operational decision-makers.

➢ *Utility of Activity Logs*

The activity log implementation serves multiple critical functions within the forecasting ecosystem. Primarily, it provides algorithmic transparency by communicating specific insights derived from the XGBoost model in natural language formats (e.g., "Material ACM001BRN is forecasted to increase by 10.13%"). This translation of mathematical outputs into business terminology enhances trust in the system's recommendations. Additionally, the activity log creates an audit trail of forecasting insights, enabling retrospective analysis of prediction accuracy and business impact. The log's real-time updates facilitate proactive inventory management by highlighting significant changes in material demand forecasts, supporting just-in-time procurement strategies. Finally, the persistent nature of the activity log creates institutional memory regarding forecasting patterns, allowing new users to quickly understand historical context and typical material behaviors without extensive system training.

VII. RESULTS AND DISCUSSION

The analysis of various forecasting models for apparel sales provides valuable insights into the relative performance and applicability of traditional statistical methods versus machine learning approaches. The performance metrics clearly demonstrate a significant advantage for advanced machine learning models in the context of apparel sales forecasting.

Traditional forecasting methods including Simple Moving Average, Exponential Weighted Average, and Naive Forecast showed considerably higher error rates across all metrics. These models produced MAPE values ranging from 298.50% to 504.22%, indicating their limited ability to capture the complex patterns inherent in apparel sales data.

The time series models (ARIMA and SARIMA) demonstrated moderate improvement with MAPE values of 55.30% and 50.15% respectively. These models incorporate temporal dependencies and seasonal patterns, providing better predictive capability than the baseline methods. The improvement suggests that accounting for seasonality and trend is crucial in apparel sales forecasting.

The machine learning models significantly outperformed all other approaches. Random Forest achieved a MAPE of 20.40% (training MAPE of 15.30%), while XGBoost delivered the best overall performance with a MAPE of 19.75% (training MAPE of 13.82%). These ensemble learning methods excel at capturing non-linear relationships and complex interactions between variables that influence apparel sales patterns.

The material suitability analysis further supports these findings, with machine learning models demonstrating broader applicability across different material types. XGBoost proved suitable for 83% of materials, followed by Random Forest at 80%, compared to traditional methods which were only suitable for 15-67% of materials.

The substantial performance gap between traditional and machine learning approaches can be attributed to the inherently volatile and multifaceted nature of apparel sales, which are influenced by numerous factors including seasonality, fashion trends, marketing promotions, and economic conditions.

VIII. CONCLUSION

This study establishes that advanced machine learning models, particularly XGBoost and Random Forest, provide superior forecasting accuracy for apparel sales compared to traditional statistical methods. With MAPE values below 20%, these models offer reliable predictions that can substantially improve inventory management and supply chain planning in the apparel retail industry.

The implementation of these machine learning models enables retailers to optimize their inventory levels, reduce stockouts and overstocking, and enhance overall operational efficiency. By accurately predicting demand patterns across different material types, retailers can make more informed decisions regarding procurement, production planning, and merchandise allocation.

Future work should focus on incorporating additional external factors such as social media trends, competitor pricing, and macroeconomic indicators to further enhance model performance. Additionally, developing a hybrid approach that combines the strengths of both statistical and machine learning methods could provide even more robust forecasting capabilities.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun757

The proposed implementation involves integrating these forecasting models into existing retail management systems with real-time data feeds from multiple sources including point-of-sale systems, e-commerce platforms, and market intelligence tools. This comprehensive approach to apparel sales forecasting will enable fashion retailers to remain competitive in an increasingly dynamic and challenging market environment.

DECLARATIONS

➢ *Acknowledgments:*

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

➢ *Funding and Financial Declarations:*

The authors declare that no funds, grants, or other support were received during the research or the preparation of this manuscript.

➢ *Data Availability Statement:*

The datasets used, generated and/or analyzed during this study are not publicly available due to internal Data Privacy Policy but are available from the corresponding author on reasonable request.

COMPLIANCE WITH ETHICAL STANDARDS

➢ *Disclosure of Potential Conflicts of Interest:*

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

➢ *Research Involving Human Participants and/or Animals:*

It is declared by all the authors that there was no involvement of any human and/or animal trial or test in this research.

➢ *Informed Consent:*

As there were no human subjects involved in this research, an informed consent is not applicable to the best of the authors' understanding.

➢ *Conflict of Interest Statement:*

The authors declare that there are no conflicts of interest that could influence the results or interpretation of this manuscript. This research was conducted in an impartial and unbiased manner, and there are no financial, personal, or professional relationships that might be perceived as having influenced the content or conclusions presented in this work.

FUTURE SCOPE

➢ *Advanced Data Integration:*

Future research can incorporate external data sources including social media trends, competitor pricing, macroeconomic indicators, and weather patterns to enhance model accuracy and capture broader market influences affecting apparel sales patterns.

➢ *Deep Learning Applications:*

Exploration of sophisticated deep learning architectures such as LSTM networks and Transformer models could better capture complex temporal dependencies in fashion sales data, particularly for highly seasonal and trend-sensitive merchandise categories.

➢ *Granular Forecasting Capabilities:*

Developing models that provide forecasts at more detailed levels (store location, demographic segments, price tiers) would offer retailers actionable insights for targeted inventory allocation and localized marketing strategies.

➢ *Sustainability Optimization:*

Extending forecasting models to optimize for environmental metrics by predicting optimal production quantities could help minimize overproduction and waste, aligning with growing sustainability imperatives in the fashion industry.

➢ *Real-Time Adaptive Systems:*

Implementation of dynamic forecasting systems that continuously update predictions as new sales data becomes available would enable more agile inventory management and responsive supply chain decisions in rapidly changing market conditions.

REFERENCES

[1]. Frank, C., Garg, A., Sztandera, L. and Raheja, A. (2003), "Forecasting women's apparel sales using mathematical modeling", International Journal of Clothing Science and Technology, Vol. 15 No. 2, pp. 107-125. https://doi.org/10.1108/09556220310470097

[2]. Lv, J., Han, S., & Hu, J. (2023). Clothing Sales Forecast Considering Weather Information: An Empirical Study in Brick-and-Mortar Stores by Machine-Learning. Journal of Textile Science and Technology, 9, 1-19. https://doi.org/10.4236/jtst.2023.91001

[3]. Saiyin, X., Hu, C., Tan, D., & Liu, Y. (n.d.). Research on Apparel Sales Forecast Based on ID3 Decision Tree.https://www.atlantis-press.com/article/25844422.pdf

[4]. Liu, N., Ren, S., Choi, T., Hui, C., & Ng, S. (2013). Sales Forecasting for Fashion Retailing Service Industry: https://onlinelibrary.wiley.com/doi/10.1155/2013/738675

[5]. Liu, N., Ren, S., Choi, T., Hui, C., & Ng, S. (2013). Sales Forecasting for Fashion Retailing Service Industry: https://onlinelibrary.wiley.com/doi/10.1155/2013/738675

[6]. Choi, T. M. (2013). Supply Chain Risk Management in the Apparel Industry. International Journal of

Production Economics, 147, 135-147. https://doi.org/10.1016/j.ijpe.2013.09.002

[7]. Thomassey, S. (2010). Sales Forecasting in Fashion Business: A Review. International Journal of Production Economics, 128(2), 393-406. https://doi.org/10.1016/j.ijpe.2010.08.005

[8]. Choi, T. M., & Chiu, C. H. (2012). Mean-Variance Analysis of Supply Chains Under Wholesale Pricing and Profit-Sharing Contracts. European Journal of Operational Research, 220(2), 361-373. https://doi.org/10.1016/j.ejor.2012.01.042

[9]. Bhardwaj, V., & Fairhurst, A. (2010). Fast Fashion: Response to Changes in the Fashion Industry. International Journal of Retail & Distribution Management, 38(2), 132-147. https://doi.org/10.1108/09590551011020155

[10]. Korpela, J., & Tuominen, M. (1995). Inventory forecasting with a multiple criteria decision tool. International Journal of Production Economics, 41(1-3), 149-158. https://doi.org/10.1016/0925-5273(95)001360

[11]. Korpela, J., & Tuominen, M. (1995). Inventory forecasting with a multiple criteria decision tool. International Journal of Production Economics, 41(1-3), 149-158. https://www.semanticscholar.org/paper/Forecasting-of-Optimum-Raw-Material-Inventory-Level-Ali-Paul/912535ef5c42f3eebaff0206e360e2660956f030

[12]. Venu, D., Naga Ganesh, P., Prashanth, K. N. S., Pavan Kumar, M., Lokesh, Ch., & Srilatha, Ch. (n.d.). Analysis of Material Requirement Planning with Exponential Smoothing, ARIMA Forecasting and Fixed Order Quantity Methods in Optimizing the Inventory in Garment Industry. https://ijirt.org/publishedpaper/IJIRT158193_PAPER.pdf