Retail Analysis and Inventory Management

Manyamala Sunaina[1], Monisha A[1] , Ballupet Prakash Monal [2], Prarthana R[3] , Priyanka TR[3]

Rekha BS[3] (Professor)

[1] Department of Computer Science & Engineering, RV College of Engineering Bengaluru, India

[2] Department of Artificial Intelligence & Machine Learning, RV College of Engineering Bengaluru, India

[3] Department of Information Science & Engineering, RV College of Engineering Bengaluru, India

***Abstract* — Retail enterprises depend on effective inventory management and comprehensive sales analysis to improve their operational efficiency and profitability. This initiative presents a data-centric framework for retail analysis and inventory oversight, integrating association analysis to reveal connections between products and enhance cross-selling potential.**

The project encompasses several critical elements, including the examination of sales trends, the identification of product associations through methodologies such as market basket analysis, and the categorization of inventory according to turnover rates and demand patterns. Predictive analytics and demand forecasting techniques are employed to fine-tune stock levels, thereby reducing the risks of overstocking and stockouts. Furthermore, advanced visualization methods empower retailers to extract actionable insights, supporting informed strategic decision-making.

By merging association analysis, inventory optimization, and predictive modeling, this initiative offers a practical and scalable approach to tackling contemporary retail challenges, ultimately fostering sustainable growth and profitability.

Keywords — Retail Analysis, Inventory Management, Association Analysis, Market Basket Analysis, Demand Forecasting, Sales Trends, Inventory Optimization, Data Analytics, Predictive Modeling.

I. INTRODUCTION

In the dynamic and fiercely competitive retail sector, mastering inventory control and gaining valuable sales insights are essential to sustain profitability and keep customers happy. Retailers need to find a balance between ensuring they have enough stock to meet customer needs and avoiding surplus inventory that can result in higher expenses. Furthermore, gaining insights into customer buying behaviors and recognizing connections between products can boost sales and enrich the overall shopping journey.

Efficient inventory management guarantees that the appropriate products are readily accessible when needed, all the while reducing operational inefficiencies like stock shortages or excess inventory. Conversely, sales analysis is dedicated to unveiling consumer behavior trends, pinpointing top-performing products, and assessing the effectiveness of promotional strategies. These elements work in harmony to give retailers the ability to make well-informed decisions, enhance product availability, and boost overall store performance.

An essential aspect of sales analysis involves association rule mining, a method that reveals connections between items often purchased in tandem. For example, recognizing that customers who buy bread frequently also tend to purchase butter can provide valuable insights for improving product displays, implementing cross-promotions, and developing focused marketing initiatives. Utilizing such valuable insights allows retailers to increase revenue while improving customer satisfaction with tailored shopping experiences.

This project aims to design and implement a comprehensive **Retail Analysis and Inventory Management System** tailored for small- to medium-sized retail chains. The system integrates inventory optimization, sales trend analysis, and product association mining to provide actionable insights.The project incorporates modern technologies, such as, machine learning algorithms for forecasting, and interactive dashboards for visualizing sales patterns and inventory metrics. By addressing critical aspects of retail operations, the proposed system provides a scalable and cost-effective solution for retailers to enhance efficiency, reduce waste, and meet dynamic customer demands.

This paper explores the technical implementation of the system while discussing its broader implications in retail operations. By integrating inventory management with sales and product association analysis, the project highlights how data-driven decision-making can transform retail operations, particularly for businesses operating in diverse and competitive markets.

II. LITERATURE SURVEY

In [1], it presents a framework to help Walmart, Costco and Kroger to improve sales forecasting, market analysis and operational efficiency.It identifies CPI and wages as key sales predictors and reveals a mutualistic market relationship among the firms.

[2] demonstrates that advanced machine learning models, particularly XGBoost, significantly enhance Walmart sales forecasting accuracy. These models outperform traditional methods, capturing complex patterns with minimal bias and high fairness. [3] highlights how retail analytics uses big data and predictive models to optimize operations, personalize customer experience and enhance profitability.[3] demonstrates how tools like dynamic pricing, inventory management, and customer behavior analysis help retailers make data-driven decisions, improve efficiency, and stay competitive in the evolving market.

[4] demonstrates how predictive analytics, through the use of regression models, decision trees, and CRM systems, can optimize sales forecasting and inventory management in retail chains. By making data-driven decisions, retailers can ensure they meet customer demand while maintaining efficient operations. [5] showcases how applying data mining techniques like clustering and rule mining can help retailers better understand their customers and improve operational strategies. By segmenting stores based on customer purchasing behavior, retailers can make more **data-driven decisions** that enhance customer engagement, streamline inventory processes, and optimize marketing efforts across different regions.

[6] shows that while both LSTM and LGBM are effective in sales forecasting, **LGBM is the preferred model** due to its superior accuracy and efficiency. This research highlights the importance of selecting the right machine learning model for specific forecasting tasks. Future research could explore adding external factors and refining model parameters for further improvements.[7] examines the role of and possibilities for big data in retailing and show that it is improved data quality (‘better’ data) rather than merely a rise in data volumes that drives improved outcomes. Much of the increase in data quality comes from a mix of new data sources, a smart application of statistical tools and domain knowledge combined with theoretical insights. These serve also to effect better data compression, transformations, and processing prior to analysis.

III. METHODOLOGY

This study follows a systematic approach to analyze sales data, uncover patterns, and forecast future performance. The methodology encompasses several stages, each designed to address specific research objectives effectively.

1. Data Cleaning and Preprocessing:

The first stage focuses on ensuring the quality and consistency of the dataset. Missing values, duplicates, and incorrect data were identified and handled to maintain data integrity. Date formats were standardized, ensuring consistency across all time-related analyses. Outliers in critical fields, such as sales and revenue, were examined and removed where necessary to avoid skewing the results. This step ensures that the dataset is robust and reliable for subsequent analysis.

2. Seasonality and Trends Analysis:

Understanding temporal patterns in sales is crucial for effective forecasting. The data was grouped by time periods such as month, quarter, and year to aggregate total sales and profit. Time series decomposition techniques were applied to separate the data into its trend, seasonal, and residual components. Sales performance during specific events and holidays was compared to assess their impact. Finally, seasonal adjustments were made, and ARIMA models were utilized to forecast future sales, capturing both long-term trends and recurring seasonal patterns.

3. Sales Performance Analysis:

To evaluate the overall performance of sales across various dimensions, data was aggregated by category, subcategory, region, and city. Performance metrics such as total sales and profit were calculated and analyzed to identify top-performing categories and regions. Additionally, profitability was examined by computing profit margins, which were then visualized alongside sales data. This step provided valuable insights into the factors driving performance and highlighted areas for improvement.

4. Customer Segmentation:

Customer segmentation was conducted to categorize and understand different customer groups, enabling targeted marketing strategies. Data normalization techniques were applied to bring all features to a comparable scale. The Elbow Method and Silhouette Analysis were used to determine the optimal number of clusters. Subsequently, K-means clustering was employed to segment customers based on spending patterns and store preferences. This analysis allowed for the identification of high-value customer segments, enabling the development of loyalty programs and exclusive offers to enhance customer retention.

5. Association Rule Mining:

Association rule mining was used to uncover relationships between products, providing insights into customer purchasing behavior. The Apriori algorithm was applied to derive metrics such as support, confidence, and lift, which quantify the strength of associations between items. These findings were leveraged to design cross-selling strategies by bundling complementary products, thereby increasing the average transaction value and customer satisfaction.

6. Demand Forecasting:

The forecasting stage aimed to predict future sales and optimize resource allocation. The time series data was visualized to identify trends and patterns. Stationarity checks were conducted to ensure the applicability of ARIMA models for forecasting. Time series decomposition further enhanced the analysis by isolating trend, seasonality, and residual components. The demand for specific product categories and regions was forecasted, facilitating inventory optimization and promotion planning. These predictions help allocate resources efficiently during peak demand periods.

7. Data Visualization:

The final step involved presenting the insights gained through interactive visualizations. Tools such as Power BI and Tableau were used to develop dashboards that provide an intuitive representation of sales trends, performance metrics, and forecasted demand. These dashboards allow stakeholders to monitor key performance indicators (KPIs) and make informed decisions. Visualizations also serve as a critical tool for communicating results and enabling actionable strategies.

IV. IMPLEMENTATION

Tools and Technologies:

Programming language: Python

Libraries Used:

1. pandas:For data manipulation and analysis, including reading, filtering, and aggregating datasets.
2. matplotlib: For creating static, animated, and interactive visualizations in Python.
3. seaborn: For creating visually appealing statistical plots and handling complex visualizations easily.
4. sklearn (scikit-learn): For implementing machine learning algorithms, including regression, clustering, and preprocessing.
5. mlxtend: For association rule mining (e.g., Apriori algorithm) and additional machine learning utilities.
6. statsmodels: For statistical modeling and time series analysis, including ARIMA.
7. streamlit: For building interactive web applications and dashboards for data and machine learning models.
8. numpy:For numerical operations and efficient handling of large arrays and matrices.

Machine learning models Used:

1. Random Forest Regressor: Predicting Sales Trends
2. K-Means Clustering: Used for customer segmentation
3. Apriori Algorithm: Used for association rule mining

Web-Interface:

1. HTML5: For structuring the web pages.
2. CSS3: For styling and layout designs.
3. JavaScript: For adding interactivity and dynamic content loading.
4. JSON: To parse the data generated by Python scripts for dynamic visualization.
5. Chart.js/D3.js/Plotly.js: For rendering various graphs and charts
6. Bootstrap: For responsive design and pre-built UI components

Visualization: Tableau and Power BI

Customer Focused Analysis:

In Tableau, several customer-focused visualizations were created to derive actionable insights:

- Total Sales and Average Sales per Customer: Bar charts illustrating sales distribution and averages across all customers.
- Profit Margin per Customer: A scatterplot highlighting customers' sales and profit contributions.

- Retention and Loyalty Analysis: Line charts showcasing repeat purchases and loyalty trends over time.
- Customer Ranking by Sales and Profit: Dual-axis bar and line charts ranking customers by their sales and profit contributions.
- Discount Impact on Profitability: A heatmap showing the relationship between discounts offered and the resulting profit margins**.**

Python was utilized for data preprocessing and analysis. Key steps included:

1. Data Cleaning:
   - Converted order dates to a standard datetime format.
   - Handled missing values and outliers to ensure data integrity.
   - Sorted the dataset by order dates for time-based analyses.
2. Customer-Centered Metrics Calculation:
   - Total Sales and Profit: Calculated cumulative sales and profit per customer.
   - Average Sales and Discounts: Derived averages for sales and discounts to understand customer behavior.
   - Order Frequency and Retention: Computed the frequency of purchases and categorized customers as "Repeat Buyers" or "One-Time Buyers."
3. Cross-Selling Opportunities:
   - Analyzed product categories frequently purchased together by customers.
4. Profitability and Ranking:
   - Segmented customers based on profit margins and created rankings to highlight top performers.

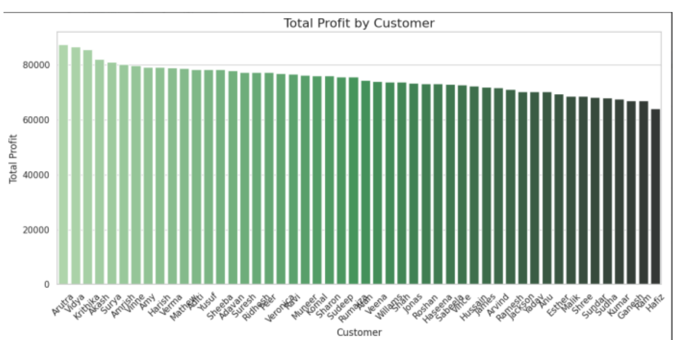

Fig.1. Bar graph showing the total profit by each customer

Fig.2. Graph showing the total sales by each customer

Product Analysis:

To understand the sales trends effectively, Power BI was employed to create dynamic and interactive dashboards. These visualizations included:

- Bar Charts: Highlighting the year-wise sales trends for each category and subcategory.
- Pie Charts:Displaying the proportion of profit contribution by each category.
- Tree Map: Offering a hierarchical view of sales distribution across subcategories.

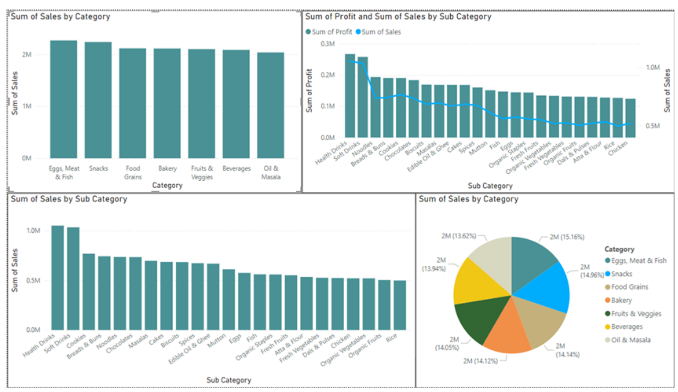

Fig.3. Dashboard of analysis based on category

These dashboards provided actionable insights, such as identifying the most profitable categories and subcategories and highlighting areas where high discounts impacted profitability.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Regional Analytics:

The analysis was performed using both Tableau for visualization and Python for data forecasting and processing. Below are the steps taken during the implementation:

In Tableau, the following visualizations were created to analyze city, regional, and category-level sales and trends:

- **City Total Sales and Profit:** A bar chart representing total sales and profit across various cities.
- **Regional Total Sales:** Aggregated sales data for each region using a stacked bar chart.
- **City Food Sales:** Focused analysis of food category sales in different cities.
- **Annual City-wise Sales:** Yearly breakdown of sales figures for each city.
- **Annual Regional Sales:** Regional comparison of annual sales figures.
- **Regional Food Sales:** Total food sales data for each region.
- **Seasonal Food City Sales:** Analyzed seasonal variations in food sales at the city level.
- **Customer Analysis for Regions:** Identified key customers and their contributions to regional sales.

Python was used to preprocess data and perform forecasting. The key steps included:

1. **Data Cleaning:**
    - Removed rows with Region = North to focus on other regional data.
    - Dropped irrelevant columns, such as State, to streamline the dataset.
2. **Category-wise Sales Analysis:**
    - Grouped data by Category and calculated total sales.
    - Identified the category with the highest sales for further analysis.
3. **Regional and City-level Analysis:**
    - Filtered data by City and Region to perform targeted sales comparisons.
    - Calculated cumulative sales and profit for cities and regions.

Fig.4. Graph showing the total sales and profit city wise

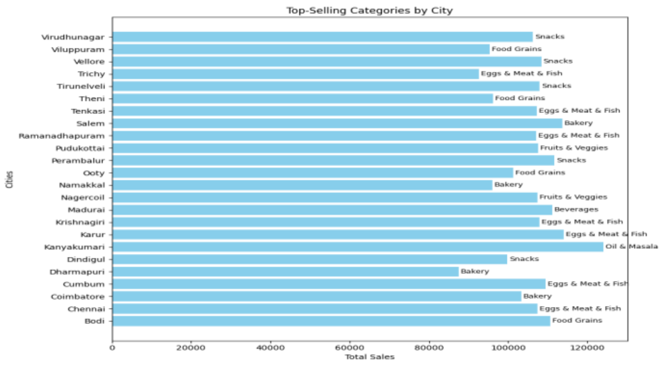

Fig.5. Graph showing the top selling categories city-wise

Association Rule Mining:

1. Dataset Preprocessing:
    - The dataset contained transaction data with items purchased by customers.
    - Transactions were transformed into a list of sets, where each set represented items purchased together.
    - Data cleaning involved removing missing values and ensuring consistent formatting.
2. Mining Frequent Itemsets:
    - The Apriori algorithm was employed to identify frequent item sets.
    - Parameters included a minimum support threshold to filter itemsets appearing in a significant fraction of transactions.
3. Generating Association Rules:
    - Rules were generated from frequent itemsets using confidence and lift metrics to measure their strength and reliability.
    - A minimum confidence threshold ensured only strong associations were considered.
4. Code Implementation:
    - Python libraries such as mlxtend were utilized for implementing Apriori and generating rules.
    - Key functions included:
        - apriori: To extract frequent itemsets.
        - association_rules: To derive rules from the frequent itemsets.

Frequent Itemsets:

- Examples of frequent itemsets identified:

    - {"Milk", "Bread"} with support = 0.25.
    - {"Eggs", "Bread", "Butter"} with support = 0.15.
  - These itemsets reveal commonly purchased combinations.

Association Rules:

  - Examples of generated rules:
    - Rule: {"Bread"} → {"Milk"}, Confidence = 0.6, Lift = 1.5.
    - Rule: {"Eggs", "Butter"} → {"Bread"}, Confidence = 0.7, Lift = 2.0.
  - Rules with high confidence and lift suggest strong and meaningful associations.

Sales Prediction:

The implementation starts with loading and preprocessing the dataset, converting dates to extract features such as year and month, and aggregating sales data based on category, city, year, month, and discount. The categorical columns are encoded using label encoding to make them suitable for a machine learning model. A random forest regressor is then trained based on the processed data to predict sales. Streamlit is used to create an interactive interface where users can input parameters such as category, city, year, month, discount, etc. and get the predicted sales based on their inputs.

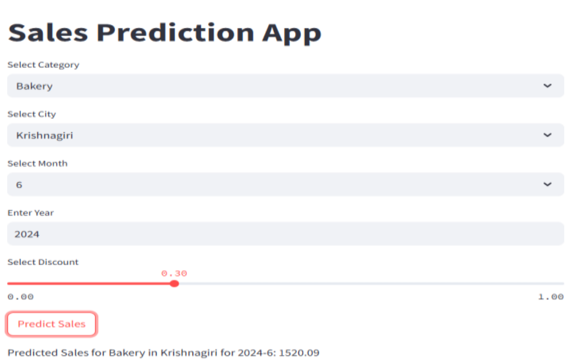

Fig.6. Streamlit application window for sales prediction

Front-End Architecture

The front-end is structured into four main categories of analysis, each linked to dedicated subpages for specific functionalities. The architecture ensures modularity, scalability, and ease of navigation. Below are the main components:

1. Home Page:

The landing page contains four primary sections, each representing one category of analysis:

  - Product Insights
  - Customer Profiles
  - Predictive Trends
  - Regional Analytics

Each section contains a list of cards, linking to subpages with detailed analysis.

2. Category-Specific Pages:

Each category has dedicated pages featuring visualizations such as pie charts, bar graphs, line charts, histograms, and data tables.

3. Dynamic Data Loading:

Python scripts process the dataset and produce JSON files. These JSON files are dynamically loaded into the front-end using JavaScript, ensuring that the visualizations reflect the latest data.

Integration with Back-End

1. Data Flow:
   - Backend Python scripts analyze the dataset and generate JSON files.
   - JSON files are fetched via AJAX requests and used for rendering charts dynamically.
2. Error Handling:
   - Validations ensure that missing or corrupt data does not crash the visualizations.
   - Fallback mechanisms display placeholder messages or alerts when data is unavailable.

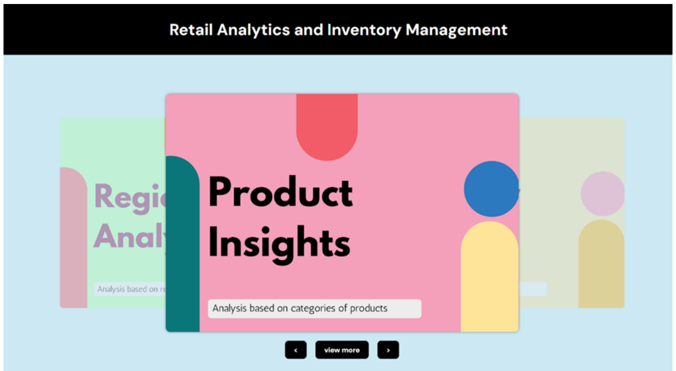

Fig.7. Home page of the website created

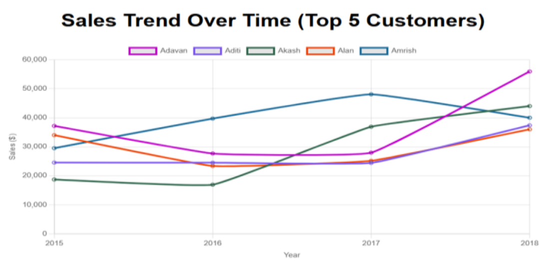

Fig.8. Line graph showing sales trend for top 5 customers

V. RESULTS AND DISCUSSION

Customer Focused Analysis:

Sales & Profit Analysis: Identified top customers by total sales, average sales, and profit margins, revealing key revenue generators and areas for strategic retention.

Retention & Loyalty: Highlighted growth in repeat buyers, especially during holidays, and long-term customers for targeted programs.

Customer Behavior: Flagged high-value buyers and cities like Vellore and Trichy for improved profit management.

Cross-Selling: Detected frequent purchases of complementary products, suggesting strong cross-selling potential.

Discount Impact: Noted reduced profitability from excessive discounts for certain customer segments.

Product Analysis:

The product analysis identifies the most and least profitable categories and subcategories, helping businesses understand performance trends. This insight allows for targeted strategies to boost profitability, optimize product offerings, and allocate resources effectively.Additionally, it enables future forecasting by analyzing historical trends, allowing businesses to make informed decisions and plan for long-term growth.

Regional Analytics:

City Sales & Profit: High-sales cities identified; some revealed low profit margins, indicating cost inefficiencies.

Regional Sales: Southern region led in sales, followed by the Western region.

Customer Analysis: Top customers in key regions flagged for targeted marketing.

Seasonal Trends: Holiday food sales peaked; steady growth predicted in metro areas over two years.

Profit Analysis: Balanced sales-to-profit cities marked as ideal for expansion.

Sales Prediction:

A predictive model was developed to forecast sales using features like region, category, and discounts.

The model demonstrated satisfactory accuracy in predicting future trends, supporting decision-making processes.

However, its accuracy can be enhanced with the inclusion of additional factors like customer demographics, seasonality, and market trends.

VI. CONCLUSION

In this study, an in-depth examination of sales data was performed to obtain insights into business performance, customer behavior, and product trends. Significant discoveries comprised recognizing top-performing areas, client groups, and subcategories, along with the effect of discounts on profit margins. A forecasting model was created to project future sales, showcasing its ability to aid in strategic planning and resource distribution.

The initiative emphasizes the significance of making decisions based on data to enhance business functions. Through the use of analytics and machine learning, companies can boost customer interaction, optimize product strategies, and increase profitability. Future initiatives ought to emphasize integrating more data sources, including demographic and seasonal elements, to improve the precision and relevance of insights even further.

VII. REFERENCES

[1] Chih-Hsuan Wang andYu-Wei Gu, "Sales Forecasting, Market Analysis, and Performance Assessment for US Retail Firms: A Business Analytics Perspective", August 2022 in the Journal of Applied Sciences

[2] Cyril Neba C, Gerard Shu F, Gillian Nsuh, Philip Amounda A, Adrian Neba F,"Advancing Retail Predictions: Integrating Diverse Machine Learning Models for Accurate Walmart Sales Forecasting",June 2024 in Asian Journal of Probability

[3] Sudeep B. Chandramana, "Retail Analytics: Driving Success in Retail Industry with Business Analytics",January 2017 in TIJ's Research Journal of Social Science & Management

[4] Sandhya Makkar, Arushi Sethi, and Shreya Jain, "Predictive Analytics for Retail Store Chain",2020 in International Journal of Data and Network Science

[5] Emrah Bilgic, O Cakir, M Kantardzic, y Duan and G Cao, "Retail analytics: Store segmentation using Rule-Based Purchasing behavior analysis",Retail analytics: Store segmentation using Rule-Based Purchasing behavior analysis

[6] Rasel Mahmud Jewel, Ahmed Ali Linkon, Mujiba Shaima, Badruddoza, Md Shohail Uddin Sarker,Rumana Shahid, Norun Nabi, Md Nasir Uddin Rana and Md Ahnaf Shahriyar, Mehedi Hasan,Md Jubayar Hossain, "Comparative Analysis of Machine Learning Models for Accurate Retail Sales Demand Forecasting", February 2024 in Journal of Computer Science and Technology Studies

[7] Eric T. Bradlow, Manish Gangwar,Praveen Kopalle1, Sudhir Voleti, "The Role of Big Data and Predictive Analytics in Retailing", 2024 in IEEE paper

[8] Dr. Mark van Rijmenam, "Nordstrom Case Study: Big Data in Retail", October 2013 in Dartmouth Center for Digital Strategies

[9] C. Vanessa Munoz Macas, J. Andres Espinoza Aguirre, R. Arcentales-Carrion, and M. Pena, "Inventory management for retail companies: A literature review and current trends," 2021 Second International Conference on Information Systems and Software Technologies (ICI2ST)

[10] P. Sridhar, C. R. Vishnu, and R. Sridharan, "Simulation of inventory management systems in retail stores: A case study," Materials Today: Proceedings, vol. 47

[11] J. L. Ruiz-Real, J. Uribe-Toril, J. C. Gázquez-Abad, and J. De Pablo Valenciano, "Sustainability and Retail: Analysis of Global Research," Sustainability, vol. 11, no. 1, p. 14, Dec. 2018