iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67646

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

A Hybrid Recommendation System for FMCG Middlemen: Integrating Content-Based and Collaborative Filtering

Ayush Kumar Singh[1], Chahat Gautam[2], Anand Kasyap[3], Supriya Singh[4]
[1, 2, 3, 4]*B.Tech Student, Department of Computer Science and Engineering, School of Engineering and Technology, CMR University*

Abstract: ***Middlemen are critical in the distribution of goods in the Fast Moving Consumer Goods (FMCG) sector. This paper describes a scheme where a hybrid recommendation system uses a combination of content based filtering and collaborative filtering through user ratings to recommend products to middlemen. The system allows the middleman to enter a product they deal with, then provides them with relevant recommendations. User ratings facilitate the recommendation process. Results from experiments show the product discovery capabilities of the system.***
Keywords: FMCG, Middlemen, Hybrid recommendation system

I. INTRODUCTION

In the world of fast-moving consumer goods (FMCG), which includes drinks, snacks, and majorly products with a shelf life, middlemen like wholesalers, distributors, and retailers play an important role. They are faced with multiple issues since they can decide whether to stock and market any product from a selection of numerous products competing for their attention. Their decision can have serious repercussions on their profitability so having a wide product portfolio is not an automatic solution to their challenges.

Whilst recommendation algorithms are universal in consumer platforms such as online shopping or streaming services, the majority are tailored to personal preferences rather than the complex, business-centric decisions middlemen have to make. Traditional approaches such as content-based methods of recommending items do it by simple feature comparisons like brand or category and user experience is not part of the equation. On the other hand, collaborative filtering, which depends on user ratings patterns, has its own shortcomings when it comes to new or less popular products.

To fill these gaps, we suggest a hybrid recommendation system that incorporates both content-based and collaborative filtering. It enables intermediaries to input a product they are currently working on which returns constantly updated recommendations from users' ratings. This approach mitigates both methodologies' common challenges, for instance, the cold-start problem, and facilitates improved decision-making with more precise suggestions regarding which products to acquire.

In subsequent subsections, we investigate literature on recommendation systems, outline our approach, and experiment with actual FMCG data. Then, we analyze the system's tangible effects, bounded scope, and other directions of research expected.

II. BACKGROUND

The benefits from Fast-Moving Consumer Goods (FMCG) goods sales such as food, beverages, and even personal care products are achieved in a very active manner with their retail outlets experiencing a faster rate of churn. As these products are purchased, used and consumed on a daily basis, the sheer volume of products available around bewilder both the end consumer and intermediaries such as wholesalers, distributors, and retailers.

Advanced recommender systems have surfaced in recent times to aid in the intricate product selection processes by making tailored selections and performing stock optimization. Therefore, the goal of this project is to develop a hybrid recommendation system which utilizes content-based filtering (that utilizes product features such as brand, category, ingredients, etc.) in conjunction with other filtering approaches that rely on users' rating and behavioral patterns. The primary objective of combining these approaches is to resolve common difficulties encountered such as the cold-start issue, data sparsity, and others to form better and more precise product recommendation paradigms in the FMCG industry.

III. LITERATURE REVIEW

Currently, research focused on Fast-Moving Consumer Goods (FMCG) areas emphasizes the increasing relevance of recommendation systems for enhancing customer relations and stock control. Basic content-based approaches depend on characteristics of a product to provide recommendations and work best in cases with insufficient user information, whereas a collaborative filtering approach uses the behavior of other users to recommend items based on their popularity. All of these methods have weaknesses: content-based filtering is too focused and lacks diversity, while collaborative filtering has problems with new and seldomly rated items.

Some scholars have attempted to construct hybrid models that integrate both approaches to increase coverage in terms of personalized preferences and general demand. Most literature also indicates some problems that remain unsolved such as sparsity of data, cold-starts, and real-time sentiment analysis. In addition, more recent work focuses on the application of sustainability indicators and lean approaches that address product requests from eco-sensitive and region-specific consumers as well as used in marketing product recommendations.

In response to these issues, the hybrid system outlined in this paper attempts to use social media data, alongside environmental data, in real time to incorporate both recommendation strategies in an effort to improve recommendation precision. This approach shall provide better automated decision support to FMCG distributors to help them determine the most suitable, economically viable, and environmentally friendly products to promote in the market.

IV. SYSTEM DESIGN AND METHODOLOGY

In this section, we describe the architecture and methodologies employed in our hybrid recommendation system. The system integrates multiple components, including dataset preparation, content-based filtering, collaborative filtering, and a hybrid approach, all orchestrated within a unified system architecture.

A. Dataset Preparation

Product data is sourced from an Excel dataset comprising attributes such as product name, type, and company. Preprocessing involves data cleaning—eliminating duplicate entries and handling missing values—and feature engineering. A combined feature is constructed by concatenating key product attributes, which is later used for text-based representation. Additionally, user ratings are collected and stored in a SQLite database. These ratings are normalized to a common scale (0–5) to ensure consistency when integrated with content-based measures.

B. Content-Based Filtering

The content-based filtering module leverages the engineered product attributes. Textual data is transformed into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. Cosine similarity is then computed between these vectors to quantify the similarity between products. The system identifies products with the highest similarity scores relative to a queried item, thereby recommending items with comparable attribute

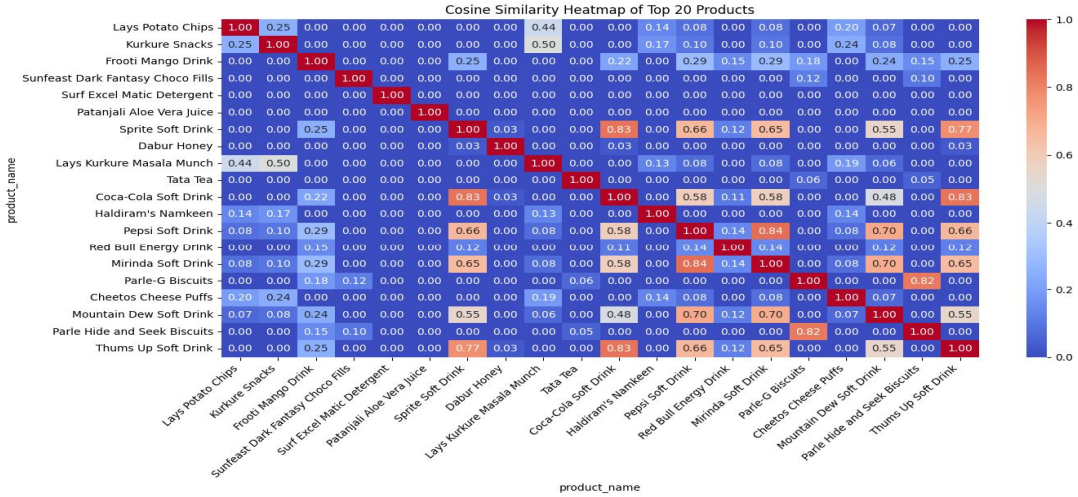


Fig. 1 Cosine Similarity Heatmap

C. Collaborative Filtering

To capture community-driven insights, the collaborative filtering module analyzes user ratings. For each product, average ratings are computed from stored user feedback. This module not only highlights popular products but also addresses issues like the cold-start problem—where products with limited interaction data are supplemented by baseline values or content-based insights.

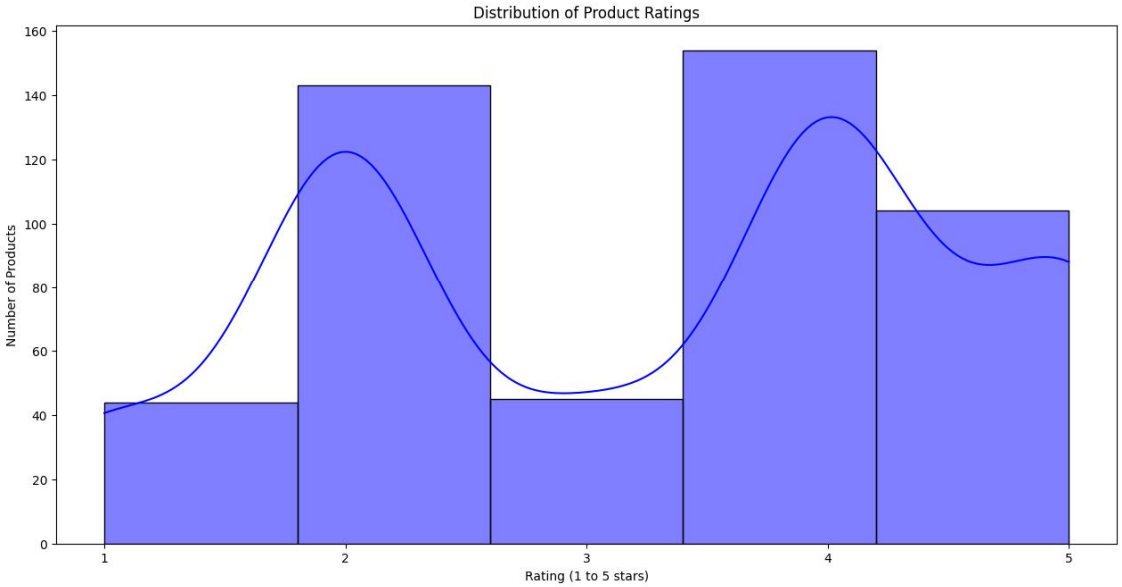


Fig. 2 Distribution of Product Ratings

D. Hybrid Approach

The hybrid approach amalgamates the strengths of content-based and collaborative filtering. A weighted score is computed for each product by combining:

1) A content score, derived from cosine similarity,

2) A rating score, based on normalized average user ratings.

A weighted sum (70% content similarity and 30% rating score) is used to generate a final hybrid score. This approach mitigates the individual limitations of both methods and yields a more balanced and accurate recommendation.

E. System Architecture

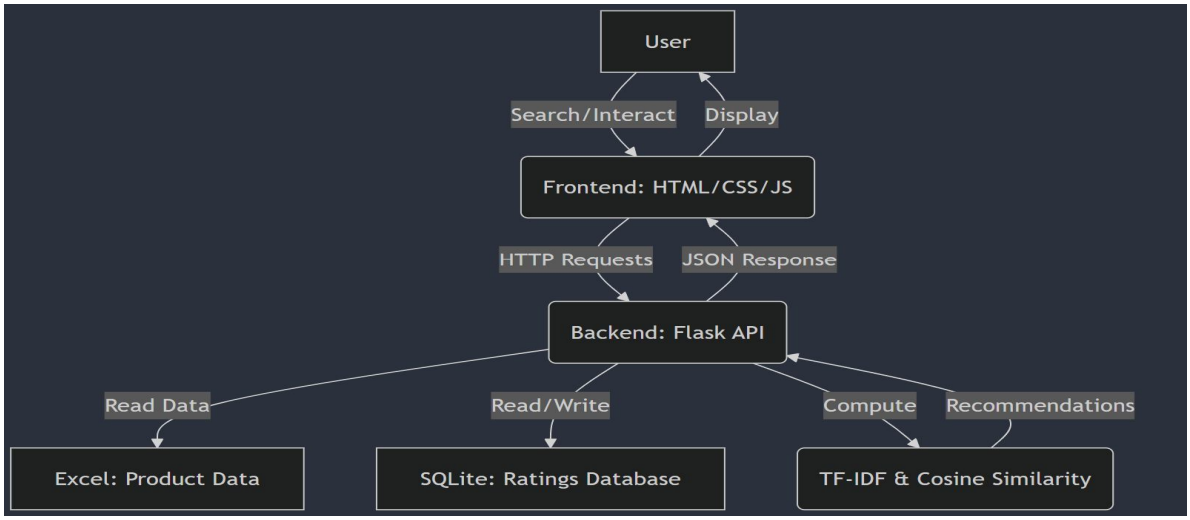


Fig. 3 System Architecture

V. IMPLEMENTATION

A. Environment Setup

The system is built in a virtual environment using Python on a Windows environment. Key libraries include:

1) Flask: For API development.
2) Pandas and NumPy: For data manipulation.
3) Scikit-learn: For TF-IDF vectorization and cosine similarity.
4) SQLite: For lightweight user ratings storage.
5) Matplotlib/Seaborn: For visualizations during evaluation.

B. System Modules

1) Dataset Preparation:
- Data Sources: An Excel file provides product information, and a SQLite database stores user ratings.
- Preprocessing: Duplicate entries are removed, and a combined feature (product name, type, company) is generated for text vectorization.

2) Content-Based Filtering:
- TF-IDF converts the combined product features into numerical vectors.
- Cosine similarity is computed between these vectors to identify similar products.

3) Collaborative Filtering:
- User ratings are aggregated and normalized from the SQLite database.
- This module provides a baseline popularity score to aid products with sparse rating data.

4) Hybrid Recommendation:
- A weighted scoring mechanism (e.g., 70% content similarity, 30% normalized rating) combines both filtering methods to produce final recommendations.

5) API Integration: The Flask application exposes several API endpoints that facilitate interaction with the recommendation system.

i. Index Endpoint (/):
   1. Renders a simple HTML home page for testing.

ii. Product Listing Endpoint (/products):
   1. Purpose: Lists products (ID and name) for search and selection.
   2. Functionality: Returns product data in JSON format.

iii. Recommendation Endpoint (/recommend):
   1. Purpose: Returns ranked product recommendations.
   2. Functionality:
      a. Accepts a product_name query parameter.
      b. Computes cosine similarity via content-based filtering.
      c. Incorporates collaborative ratings for refinement.
      d. Outputs a JSON array with hybrid scores and average ratings.

iv. Rating Submission Endpoint (/rate):
   1. Purpose: Accepts user product ratings.
   2. Functionality:
      a. Receives a POST request with a JSON payload (product IDs and ratings).
      b. Updates the SQLite database with new ratings.

VI. RESULT

The proposed hybrid recommendation system was evaluated on an FMCG dataset comprising 392 training entries and 98 test entries. The average product ratings from the training set were comparable to those from the test set, suggesting a consistent distribution across splits.

For evaluation, we employed Precision@5, Recall@5, and F1@5 metrics to assess the quality of the top-5 recommendations. Different products showed varied performance, with average scores across the test queries as follows:

- Average Precision@5: 0.68
- Average Recall@5: 0.50
- Average F1@5: 0.49

These results indicate that while the system is effective in recommending relevant items (as seen from the high precision in several cases), there is room for improvement in retrieving a larger fraction of all relevant items (recall).

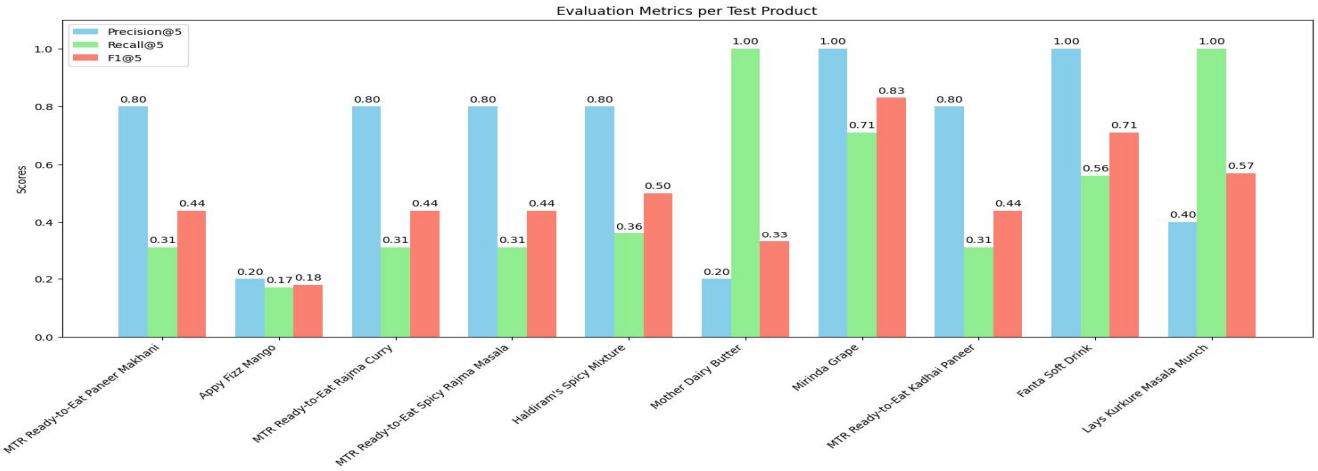


Fig. 4 Grouped bar chart displaying Precision@5, Recall@5, and F1@5 scores for each test product

VII. CONCLUSIONS

This paper presented a hybrid recommendation system designed for FMCG middlemen that combines content-based and collaborative filtering techniques. By leveraging TF-IDF vectorization and cosine similarity for product attribute analysis, along with user ratings for enhanced relevance, the system effectively addresses common challenges such as data sparsity and the cold-start problem. Evaluation using metrics such as Precision@5, Recall@5, and F1@5 demonstrated promising performance, with an average Precision@5 of 0.68, though improvements in recall are needed to capture a broader range of relevant products.

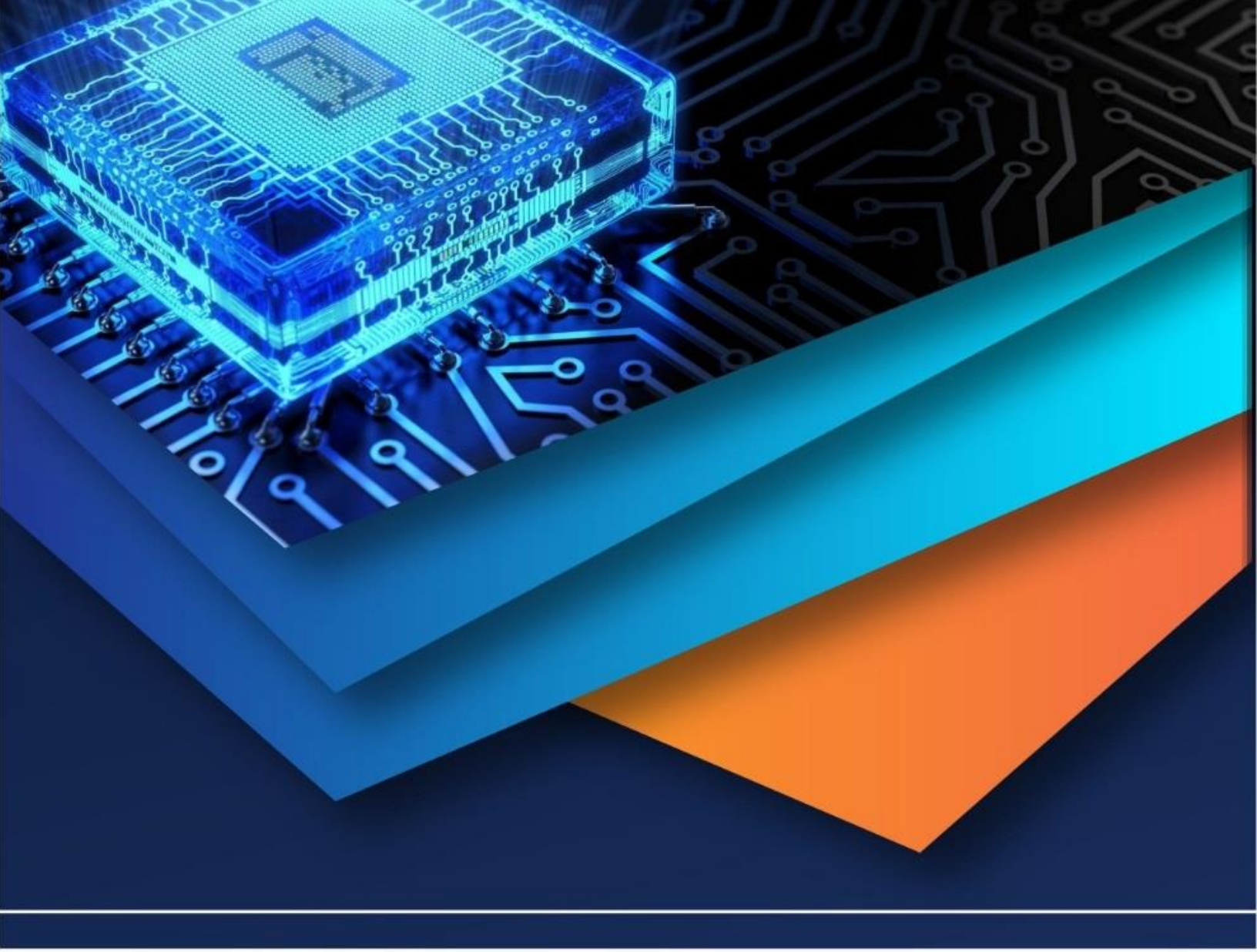
Overall, the hybrid approach provides a balanced framework for personalized product recommendations, facilitating more informed decision-making in the competitive FMCG market. Future work will explore the integration of additional features, such as sentiment analysis and real-time data, to further refine recommendation accuracy and improve system robustness.

VIII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Dr. Manjunath C R, Head and Professor in the AIML, Department of Computer Science and Engineering at CMR University, for his invaluable guidance and unwavering support throughout this project. We also extend our appreciation to the faculty and staff of the department for their insightful suggestions and continuous encouragement.

REFERENCES

[1] Prakash P. Rokade, Aruna Kumari D – Business recommendation based on collaborative filtering and feature engineering – aproposed approach - RESEARCH GATE.
[2] C. Shyamala Kumari; B S V Vidya Deepu; Dheeraj G; Vinay Kumar Sai Surampudi – FMCG Market Analysis for Wholesalers and Retailers Using Machine Learning - RESEARCH GATE
[3] Elcio Tarallo* Getúlio K. Akabane** Camilo I. Shimabukuro***- Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research - SCIENCE DIRECT
[4] Kai Lu, Yi Zhang, Lanbo Zhang, Shuxin Wang – Exploiting User and Business Attributes for Personalized Business Recommendation - ACM Digital Library
[5] Zeshan Fayyaz 1, Mahsa Ebrahimian 1, Dina Nawara 1, Ahmed Ibrahim 2 and Rasha Kashef Rasha Kashef – Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities – MDPI
[6] Akhmetova, Z. B.; Kim, I. A. - MODERN CONCEPTS OF INTERACTION WITH CONSUMERS IN THE FMCG SECTOR. - SCIENTIFIC JOURNALS AL-FARABI KAZAKH NATIONAL UNIVERSITY
[7] Kapilya Gangadharan, K. Malathi, Anoop Purandaran, Barathi Subramanian, Rathinaraja Jeyaraj – From Data to Decisions: The Transformational Power of Machine Learning in Business Recommendations - arXiv



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY