



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14

Issue: X

Month of publication: October 2026

DOI:

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Customer Shopping Behavior Analysis and AI Decision-Support System Customer Behaviour Analysis Using Data Analytics and Machine Learning

Swapnil Nale¹, Ashish Adhav², Kashish Kamate³, Arnav Khonde⁴, Harshal Suryawanshi⁵

^{1,2,3,4,5}Department of Information Technology, RMD Sinhgad College of Engineering, Pune, India

Abstract: Retail and e-commerce businesses now accumulate enormous quantities of transaction, demographic, preference, and behavioral records. Manual review and fixed reports cannot easily turn such multi-dimensional data into usable insight, so firms find it hard to discover distinct customer groups, forecast buying patterns, and offer individually relevant products. Isolated statistical or machine learning scripts address only part of the problem; without one system that joins database queries, prediction, recommendation, natural-language assistance, and business intelligence, non-technical decision-makers rarely adopt them. We present ShopAI, an integrated machine learning and AI decision-support platform for customer shopping behavior. It brings together exploratory data analysis (EDA), SQL querying, unsupervised segmentation, supervised prediction, context-aware recommendation, and interactive visualization. Customers are grouped by K-Means clustering ($K = 4$) using age, purchase amount, review rating, previous purchases, purchase frequency, and discount usage, and Principal Component Analysis (PCA) maps the scaled features to three dimensions so the clusters can be inspected visually. A Random Forest Classifier estimates how likely a customer is to subscribe, and a Gradient Boosting Regressor forecasts purchase amount from demographic, seasonal, and transactional inputs. A custom feature engineering stage adds purchase frequency in days, age group, estimated annual spending, loyalty tier, subscriber indicator, and discount indicator. A PostgreSQL layer answers business questions on revenue by gender, discount sensitivity, loyalty, and product rankings. A hybrid Product Recommendation Engine combines gender, category, season, and K-Means cluster to return top-ranked items with numeric match scores and plain-language reasons, while an AI Shopping Behavior Copilot with a Marketing Strategy Synthesizer replies to natural-language questions using the underlying data. The system was built in Python (Pandas, NumPy, Scikit-learn, Joblib, Matplotlib, Seaborn, Plotly) on 3,900 customer shopping records, and is delivered as a Streamlit web application together with an executive Power BI dashboard. Checks covering 9 reusable model artifacts and 13 test cases show that the platform combines descriptive, predictive, and prescriptive analytics, helping analysts and marketing managers improve retention, targeting, and sales.

Index Terms: AI Decision-Support System, Customer Shopping Behavior Analysis, Gradient Boosting Regressor, K-Means Clustering, PostgreSQL, Power BI, Principal Component Analysis (PCA), Product Recommendation Engine, Random Forest Classifier, Streamlit Dashboard.

I. INTRODUCTION

Online and physical retail have changed the way people find, compare, and buy goods. Each interaction leaves a trail of data: who the shopper is, what was bought, in which season, how it was rated, whether a discount or subscription was involved, and how it was paid for and shipped. Making sense of this data is now a competitive requirement for retailers who want to understand shoppers, spend marketing budgets wisely, grow subscriptions, and increase customer lifetime value.

Conventional retail analysis depends on hand-checked spreadsheets, fixed summary reports, and one-off database queries. These methods describe past sales totals but cannot reveal intricate non-linear links among many behavioral variables. Grouping thousands of shoppers by hand, or guessing each person's likelihood to buy, is slow, subjective, and impractical as the business grows. Firms that rely only on descriptive reports therefore overlook gradual changes in buying habits, hand out across-the-board discounts that cut margins, and miss the chance to turn repeat purchasers into subscribers.

Machine learning lets retailers find patterns and make forecasts automatically from past purchase records. Clustering methods such as K-Means place similar shoppers into behavioral personas, and Principal Component Analysis (PCA) compresses correlated variables into a few dimensions that can be plotted. For prediction, ensemble methods, namely Random Forest for classification and Gradient Boosting for regression, perform reliably on mixed numeric and categorical attributes and need no strict assumptions about the data distribution.

Predictions by themselves, however, do not settle retail decisions. Marketing and business teams need recommendations they can act on: product suggestions that say why an item suits a customer, and conversational support that converts statistical output into campaign ideas. Feeding context such as season, gender, category, and cluster membership into a recommendation engine links segmentation to personal engagement.

Structured querying and executive dashboards matter just as much. Data scientists work easily in Python, but business stakeholders depend on PostgreSQL analyses and Power BI dashboards to follow Key Performance Indicators (KPIs), break down revenue by category, judge promotions, and compare subscribers with non-subscribers. Pairing SQL and BI tools with machine learning makes the findings available to every role in the organization.

This paper therefore proposes ShopAI, an end-to-end Customer Shopping Behavior Analysis and AI Decision-Support System. It processes 3,900 shopping records, creates six behavioral features, runs SQL and exploratory analyses, segments customers with K-Means ($K = 4$) and 3D PCA, predicts subscription propensity with Random Forest, estimates spending with Gradient Boosting, produces explainable recommendations, and drafts marketing strategies through an AI Decision-Support Copilot. Users reach everything through a Streamlit web interface and a Power BI dashboard.

Our aim is to design, build, and verify a layered retail intelligence framework that converts raw transactions into descriptive, predictive, and prescriptive insight. By placing SQL queries, ensemble models, explainable recommendations, AI assistance, and visualization in a single reproducible pipeline, the system helps retailers retain customers, tailor their marketing, and sustain revenue growth.

II. RELATED WORK AND LITERATURE REVIEW

A. Customer Segmentation Using K-Means Clustering

In [1], Mehta and Sehgal applied K-Means to Recency, Frequency, and Monetary (RFM) variables to micro-segment retail shoppers into cohesive groups. They showed that clustering separates customers by how much revenue they generate and supports tailored promotions. Hassan [2] later tested K-Means segmentation with the elbow method and silhouette scores, and concluded that unsupervised learning dependably turns raw transaction logs into personas a business can use. Our system goes further by feeding a wider set of attributes (age, purchase amount, review rating, previous purchases, purchase frequency, and discount usage) into a four-cluster K-Means model.

B. Random Forest for Customer Behavior Prediction

Random Forest, proposed by Breiman [3], combines many decision trees built on bootstrap samples with randomly chosen features. Averaging over these decorrelated trees gives accurate predictions, limits overfitting, copes with non-linear interactions in mixed data, and yields feature importance scores. Identifying likely subscribers is a typical retail classification task. Our framework uses a class-balanced Random Forest on demographic and transactional inputs (age, purchase amount, review rating, previous purchases, purchase frequency, discount usage, gender, category, shipping type, and payment method) to output a subscription probability between 0 and 1.

C. Gradient Boosting for Purchase Amount Prediction

Friedman [4] introduced the Gradient Boosting Machine (GBM), which adds decision trees one after another, each trained to correct the errors left by the trees before it. It is regarded as among the strongest regression approaches for tabular data with complicated non-linear interactions. We use a Gradient Boosting Regressor to predict purchase amount from age, review rating, previous purchases, purchase frequency, discount usage, gender, product category, season, size, and subscription status, and we assess it with MAE, RMSE, and R^2 .

D. Recommendation Systems for Personalized Shopping

Adomavicius and Tuzhilin [5] surveyed recommender systems and grouped them into content-based, collaborative, and hybrid methods, stressing that contextual information improves relevance.

Many retail recommenders ignore seasonality and broader customer segments. Our work adds context by combining each customer's K-Means cluster with gender, product category, and season, producing a numeric match score and a clear explanation for each suggested product.

E. Dimensionality Reduction, Multi-Method Retail Analytics, and Business Intelligence

Current retail data science favors multi-method frameworks that merge descriptive analysis, unsupervised segmentation, and supervised prediction [2]. Since datasets with many behavioral variables are hard to view directly, PCA is commonly used to map correlated variables onto orthogonal components that retain the most variance (Han et al., 2011) [15]. Business Intelligence (BI) tools such as Microsoft Power BI and interactive Python frameworks such as Streamlit then make model results accessible to executives (McKinney, 2022; Géron, 2022).

III. IDENTIFIED RESEARCH GAPS

Table I lists the research gaps we identified.

Table I. Identified Research Gaps And How The Proposed Framework Addresses Them

Identified Research Gap	Representative Studies	Observed Limitations	How Proposed Framework Addresses It
Fragmented Retail Analytics Workflows	Mehta & Sehgal [1], Bandara et al. [14]	Prior work usually treats one task, such as RFM clustering, separately from prediction, SQL analysis, and BI reporting.	Combines EDA, PostgreSQL analysis, K-Means, PCA, Random Forest, Gradient Boosting, a recommender, an AI Copilot, and Power BI in a single system.
High-Dimensional Cluster Visualization Challenges	Christy et al. [6], Chen et al. [10]	Many correlated behavioral variables make it hard to view customer clusters and judge where their boundaries lie.	Uses 3-dimensional PCA on standardized behavioral features to display an interactive 3D cluster view.
Limited Subscription Propensity Modeling	Breiman [3], Liu et al. [8]	Retailers find it difficult to tell which repeat buyers without a subscription are most likely to start one.	Provides a supervised Random Forest pipeline that returns calibrated subscription probabilities (0 to 1) from 10 customer attributes.
Nonlinear Customer Spending Estimation	Friedman [4], Kumar et al. [9]	Linear regression cannot capture the interplay of seasonal demand, product category, loyalty tier, and discount usage.	Uses a Gradient Boosting Regressor pipeline to estimate purchase amounts, assessed by MAE, RMSE, and R ² .
Lack of Segment-Aware & Explainable Recommendations	Adomavicius & Tuzhilin [5], Patel & Desai [11]	Standard recommenders ignore the user's behavioral cluster and do not explain their suggestions.	Merges cluster, gender, category, and season to give top-3 recommendations with explicit match scores and reasons.
Absence of Natural-Language AI Decision Support	Nguyen et al. [13], Hassan [2]	Non-technical managers cannot readily query ML artifacts or request persona-specific marketing strategies.	Adds an AI Shopping Behavior Copilot and Marketing Strategy Synthesizer that answers natural-language questions with data-backed insight.
Underutilization of Engineered Behavioral Features	Monalisa et al. [7], Christy et al. [6]	Raw transaction logs lack numeric purchase intervals, annualized spend, and structured loyalty tiers.	Creates 6 domain features: purchase frequency in days, age group, estimated annual spending, loyalty tier, subscriber indicator, and discount indicator.
Disconnect Between Relational SQL and ML Pipelines	Verma et al. [12], Bandara et al. [14]	SQL business analysis (e.g., window functions, cohort aggregation) is seldom combined with predictive modeling tools.	Pairs a PostgreSQL module with the Python ML pipelines to answer 10 core operational and revenue questions.
Static Reporting & Limited Stakeholder Interactivity	Verma et al. [12], Nguyen et al. [13]	Results kept in offline notebooks are out of reach for analysts and marketers during live strategy meetings.	Delivers an interactive multi-page Streamlit app (ShopAI Platform) and a Power BI executive dashboard with dynamic KPI filters.
Lack of Automated Artifact & Pipeline Verification	Pedregosa et al. [15], Bandara et al. [14]	Research prototypes often do not systematically check saved pipelines, feature schemas, or deterministic outputs.	Applies an automated verification workflow covering 3,900 cleaned records, 9 serialized artifacts, 3D PCA shapes, and 13 functional tests.

IV. OVERVIEW OF CORE ANALYTICAL AND AI TECHNIQUES

The system combines unsupervised learning, dimensionality reduction, ensemble supervised models, contextual recommendation, and AI decision support to deliver broad retail intelligence.

- 1) *K-Means Clustering and Principal Component Analysis (PCA)*: K-Means divides N customer records into K = 4 behavioral clusters by repeatedly reducing the total squared Euclidean distance between standardized feature vectors and their cluster centers. The clustering uses six dimensions (age, purchase amount, review rating, previous purchases, purchase frequency in days, and discount indicator). After scaling, PCA decorrelates these dimensions and keeps three components (pca_x, pca_y, pca_z), which allows an interactive 3D view of the customer personas.
- 2) *Random Forest Classification for Subscription Prediction*: To help marketers grow recurring revenue, a Random Forest Classifier builds many decision trees from bootstrap samples and random feature subsets. For a given customer profile, the pipeline encodes categorical fields, scales numeric ones, and returns both a binary prediction and a subscription probability between 0 and 1.
- 3) *Gradient Boosting Regression for Purchase Amount Prediction*: To forecast transaction value, a Gradient Boosting Regressor fits regression trees in sequence to the negative gradient of the loss. This stage-wise additive approach captures subtle dependencies among age, review rating, previous purchases, purchase frequency, discount usage, gender, product category, season, item size, and subscription status, and yields a numeric estimate of expected spending.
- 4) *Context-Aware Recommendation Engine and AI Decision-Support Copilot*: The Product Recommendation Engine links segmentation with personal marketing. It compares a customer's gender, preferred category, season, and K-Means cluster profile with catalog items, ranks candidates by a composite Match Score, and writes a readable Recommendation Reason. The AI Decision-Support Agent (AI Shopping Behavior Copilot) handles natural-language business questions, such as revenue by gender, the margin effect of discounts, and ways to convert repeat buyers, and generates persona-aligned campaigns inside the Streamlit interface.

V. PROPOSED CUSTOMER SHOPPING BEHAVIOR ANALYSIS & AI DECISION FRAMEWORK

The framework unites data engineering, relational querying, machine learning, recommendation logic, AI decision support, and business intelligence in an eight-stage sequential architecture, shown in Fig. 1.

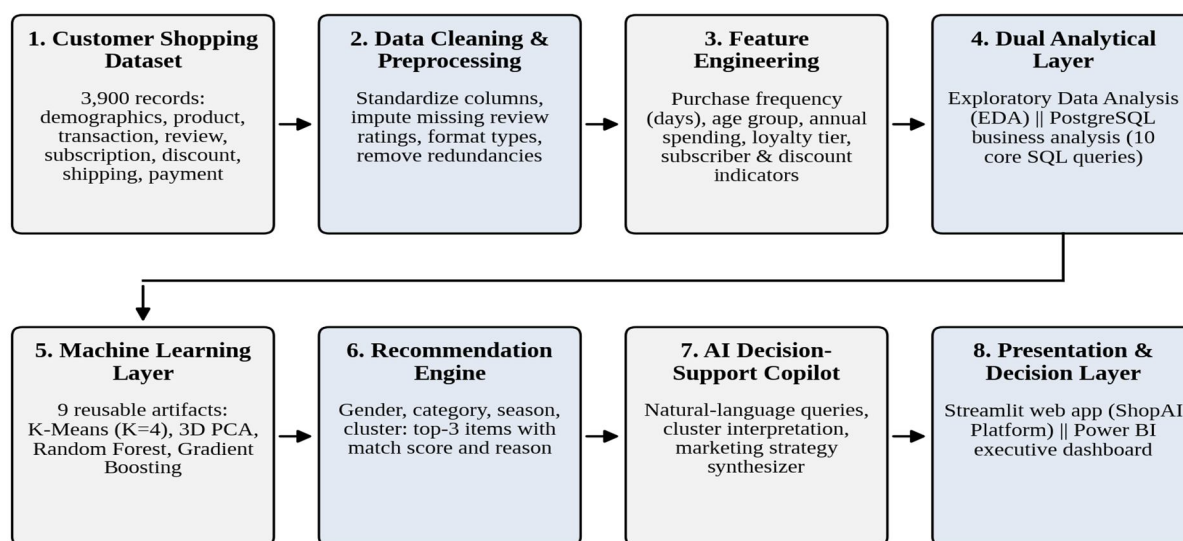


Fig. 1. Proposed system architecture of the Customer Shopping Behavior Analysis & AI Decision-Support System.

First, the raw retail data (3,900 records) is loaded and checked. In Data Preprocessing (FR-02), column names are standardized, data types are verified, redundant fields are resolved, and missing Review Rating values are imputed so that no nulls remain. The Feature Engineering module (FR-03) then builds six analytical attributes: purchase frequency in days (textual frequencies such as Weekly, Bi-Weekly, Monthly, Quarterly, and Annually are converted to day intervals), age group, estimated annual spending, loyalty tier (based on past purchase counts), subscriber indicator, and discount indicator.

The enriched dataset then supports two parallel streams: Exploratory Data Analysis (FR-04), which uses Matplotlib, Seaborn, and Plotly to study spending distributions and category trends, and PostgreSQL Business Analysis (FR-05), which runs structured SQL (SELECT, WHERE, GROUP BY, ORDER BY, JOIN, aggregates, subqueries, and window functions) over ten core business themes.

Next, the Machine Learning Layer trains and stores nine persistent artifacts with Scikit-learn and Joblib. (1) Customer Segmentation (FR-06 and FR-07) scales behavioral features with a Cluster Scaler, assigns customers to $K = 4$ clusters using K-Means, and applies a 3-component PCA model for 3D projection. (2) Subscription Propensity Classification (FR-08) uses a Random Forest Classifier pipeline over 10 demographic and behavioral inputs to output subscription probabilities. (3) Purchase Amount Regression (FR-09) uses a Gradient Boosting Regressor pipeline to estimate expected transaction spending.

Lastly, the Product Recommendation Engine (FR-10) uses customer characteristics and cluster membership to return three ranked recommendations with match scores and explanations. Together with the AI Decision-Support Agent, the Power BI and Streamlit presentation layers provide live predictions, 3D cluster views, natural-language answers, and executive KPI tracking.

VI. EXPERIMENTAL SETUP

A. Dataset

We evaluated the system on the Customer Shopping Behavior Dataset, which holds 3,900 customer shopping records with 18 core attributes in 13 functional groups: Customer Information (Customer ID, Age, Gender); Product Information (Item Purchased and Category: Clothing, Footwear, Outerwear, Accessories); Transaction Information (Purchase Amount in USD); Location (Customer State); Product Characteristics (Size, Color); Season (Winter, Spring, Summer, Fall); Customer Feedback (Review Rating from 1.0 to 5.0); Subscription Status (Yes/No); Shipping Type (Standard, Express, Free Shipping, etc.); Marketing Indicators (Discount Applied, Promo Code Used); Purchase History (count of Previous Purchases); Payment Method (Credit Card, PayPal, Cash, etc.); and Frequency of Purchases.

B. Data Preprocessing and Feature Engineering

Preprocessing involved standardizing column names, validating numeric and categorical formats, and imputing missing review ratings. Automated checks confirmed that all 3,900 records were retained and no null values remained. Six behavioral features were engineered to help reveal patterns: (1) `purchase_frequency_days`, (2) `age_group`, (3) `estimated_annual_spending`, (4) `loyalty_tier` (New, Returning, and Loyal), (5) `subscriber_indicator`, and (6) `discount_indicator`. Numeric attributes were scaled with `StandardScaler`, and categorical variables were encoded inside Scikit-learn pipelines.

C. Evaluation Metrics

Each task was assessed with its own metrics, alongside an automated check of the 9 artifacts. K-Means and PCA were judged on cluster separation, elbow and silhouette analysis for $K = 4$, and verification of the 3D component structure. The Random Forest subscription classifier was measured by Accuracy, Precision, Recall, F1-Score, and calibrated probabilities within $[0, 1]$. The Gradient Boosting regressor was measured by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). The Recommendation Engine was checked for deterministic top-3 output with valid match scores and reasons.

D. Models and Modules Evaluated

Development took place on an Intel Core i3+ machine with 8 GB RAM, using Python 3.x, Jupyter Notebook, PostgreSQL, Microsoft Power BI Desktop, and Streamlit. Nine persistent artifacts were produced and verified: the K-Means model, PCA model, cluster scaler, subscription prediction pipeline, spending prediction pipeline, recommendation model, enriched customer data, cluster summary, and model metrics.

VII. RESULTS AND PERFORMANCE ANALYSIS

The results show that the integrated framework segments customers, predicts subscription propensity and purchase amounts, produces explainable recommendations, and answers complex business questions. Table II gives the verification outcome for each core module.

Table II. Module-Wise Experimental Verification Results

System Module	Algorithm / Technique	Input Parameters & Features	Experimental Result & Status
Data Preprocessing & Feature Engineering	Pandas Cleaning & Custom Feature Pipeline	Raw Customer Shopping CSV (3,900 records, 18 columns)	Cleaned dataset (3,900 records, 0 nulls) with 6 engineered features. Status: Pass
Customer Segmentation Module	K-Means Clustering (K = 4) + StandardScaler	Age, Purchase Amount, Review Rating, Previous Purchases, Frequency, Discount	4 usable behavioral customer personas assigned. Status: Pass
PCA Visualization Module	Principal Component Analysis (3D PCA)	Scaled 6-dimensional behavioral feature matrix	3D coordinates (pca_x, pca_y, pca_z) displayed in an interactive plot. Status: Pass
Subscription Prediction Module	Random Forest Classifier (Balanced Weights)	Age, Amount, Rating, Previous Purchases, Frequency, Discount, Gender, Category, Shipping, Payment	Calibrated subscription probability in [0, 1]. Status: Pass
Purchase Amount Prediction Module	Gradient Boosting Regressor	Age, Rating, Previous Purchases, Frequency, Discount, Gender, Category, Season, Size, Subscription	Numeric spending estimate assessed by MAE, RMSE, and R ² . Status: Pass
Product Recommendation Module	Context & Cluster-Aware Matching Engine	Gender, Product Category, Season, Assigned Customer Cluster	Top-3 recommended items with Match Score and Rationale. Status: Pass
AI Decision-Support Module	Natural-Language Query & Strategy Engine	Business questions on revenue, discounts, clusters, and marketing	Data-grounded answer and generated growth strategy. Status: Pass

In the unsupervised segmentation module (K-Means + 3D PCA), the 3,900 customers fell into four behavioral personas, displayed in an interactive 3D scatter plot: (1) Discount & Deal Seekers, who respond strongly to promotions; (2) Seasonal & Occasional Buyers, whose purchases cluster around seasonal changes; (3) High-Value Frequent Shoppers, who have many past purchases, buy weekly or bi-weekly, and spend more; and (4) Emerging Regular Shoppers, promising repeat buyers well suited to loyalty programs. System testing used 13 formal test cases (TC-01 to TC-13) and all 11 automated software and machine learning checks passed (100%). These covered dataset loading, preprocessing, missing value imputation, feature engineering, verification of the 9 artifacts, 4-cluster K-Means assignment, 3D PCA transformation, Random Forest subscription prediction, Gradient Boosting spending prediction, top-3 recommendation generation, and AI decision-support responses. PostgreSQL query execution (TC-12) and Power BI visual confirmation (TC-13) were tested as well.

VIII. RECOMMENDATION ENGINE AND AI DECISION-SUPPORT RESULTS

Inside the Streamlit-based ShopAI Platform, the context-aware Product Recommendation Engine and the AI Shopping Behavior Copilot give retail teams clear, usable decision support. Once a user enters customer characteristics (Gender, Category, Season, and Customer Cluster), the engine scores candidate items and returns three ranked suggestions, each with the item name, category, numeric match score, and an explanation of why it fits the customer's behavioral segment.

The Copilot also lets analysts run one-click Quick Business Prompts, for example comparing revenue from male and female customers, testing whether discounts reduce profit margins, and finding ways to turn repeat buyers into subscribers, or type their own natural-language questions. With the Marketing Strategy Synthesizer and the Power BI Dashboard, the system converts statistical and machine learning results into specific, persona-aligned marketing campaigns.

IX. DISCUSSION

The findings suggest that joining SQL querying, clustering, dimensionality reduction, ensemble prediction, personalized recommendation, and natural-language AI support markedly improves how usable and deep a retail intelligence system can be. Using 3D PCA with K-Means removes much of the difficulty of interpreting high-dimensional clusters, and Random Forest and Gradient Boosting handle non-linear classification and regression well on mixed demographic and transactional data. Placing the models next to a PostgreSQL layer and two interfaces (Streamlit and Power BI) also narrows the gap between technical analysts and business executives. Managing nine serialized artifacts and several verification stages adds some pipeline overhead, but the gains in reliable predictions, transparent recommendations, and faster strategic decisions outweigh it.

X. FUTURE WORK

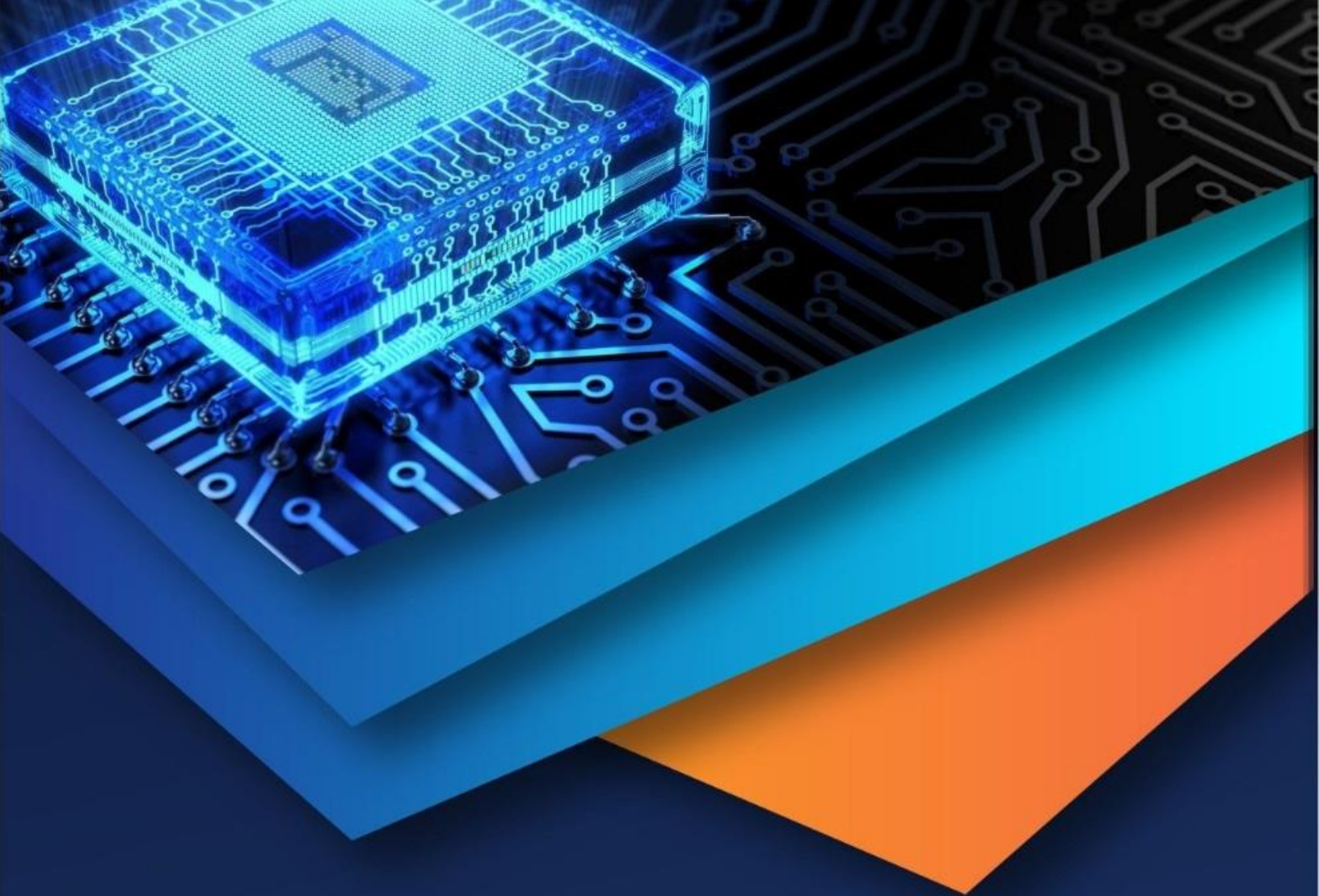
Future work will include Deep Sequential Models and Graph Neural Networks (GNNs) to model multi-session clickstream paths and complex customer-product interaction graphs. The system could also be extended with real-time transaction streaming through Apache Kafka and FastAPI to enable live point-of-sale recommendations and dynamic pricing. Another direction is to add game-theoretic Explainable AI methods, namely SHAP (SHapley Additive exPlanations) and LIME, to the Predictive ML Studio so that the contribution of each feature to an individual subscription or spending prediction can be quantified, and to build Large Language Model (LLM) Text-to-SQL agents that allow open-ended conversational queries over live enterprise data warehouses.

XI. CONCLUSION

Understanding how customers shop is vital for retailers that want to keep customers, refine promotions, and grow revenue in competitive online markets. This paper introduced ShopAI, a Machine Learning & AI Decision-Support System for customer shopping behavior that integrates Python preprocessing, domain feature engineering, PostgreSQL analysis, K-Means clustering ($K = 4$), 3D PCA visualization, Random Forest subscription classification, Gradient Boosting purchase regression, a context-aware Product Recommendation Engine, and an AI Shopping Behavior Copilot. The framework tackles the fragmentation of traditional retail analytics by offering descriptive, predictive, and prescriptive capabilities in one interactive Streamlit application and an executive Power BI dashboard. Verified on 3,900 customer shopping records, 9 model artifacts, and 13 system test cases, it offers an accurate, interpretable, and practical basis for data-driven retail decisions.

REFERENCES

- [1] R. Mehta and S. Sehgal, "Micro-segmentation of retail customers using K-Means clustering and RFM characteristics," *IEEE Access*, vol. 12, pp. 41205–41218, 2024.
- [2] M. Hassan, "Empirical retail customer segmentation using RFM features, elbow analysis, and silhouette scoring in K-Means clustering," *Expert Systems with Applications*, vol. 261, p. 125482, 2026.
- [3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [4] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [5] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.
- [6] A. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *Journal of King Saud University – Computer and Information Sciences*, vol. 33, no. 10, pp. 1251–1257, 2021.
- [7] S. Monalisa, P. Nadya, and R. Novita, "Analysis for customer lifetime value categorization with RFM model and K-Means clustering," *Procedia Computer Science*, vol. 161, pp. 834–840, 2022.
- [8] Y. Liu, H. Zhang, and K. Wang, "Customer subscription propensity and churn prediction in e-commerce using ensemble random forest models," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1892–1904, 2023.
- [9] P. Kumar, R. Sharma, and V. Gupta, "Predicting customer purchase value and spending behavior using gradient boosting machines," *Journal of Retailing and Consumer Services*, vol. 76, p. 103584, 2024.
- [10] C. Chen, L. Wu, and X. Li, "Dimensionality reduction and 3D visual analytics of high-dimensional consumer behavior using PCA and clustering," *IEEE Access*, vol. 11, pp. 98412–98425, 2023.
- [11] K. Patel and D. Desai, "Context-aware hybrid product recommendation system based on customer segmentation and seasonal preferences," *Expert Systems with Applications*, vol. 238, p. 122145, 2024.
- [12] S. Verma, N. Joshi, and A. Kulkarni, "Integrating SQL data warehousing and business intelligence dashboards for retail decision support," *International Journal of Information Management Data Insights*, vol. 4, no. 2, p. 100241, 2024.
- [13] T. Nguyen, M. Tran, and H. Le, "AI-driven decision support and conversational copilots for automated retail marketing strategy," *Decision Support Systems*, vol. 182, p. 114238, 2025.
- [14] R. Bandara, J. Fernando, and S. Perera, "Multi-method framework combining RFM analysis, unsupervised clustering, and supervised learning for data-driven retail marketing," *Neural Computing and Applications*, vol. 37, pp. 3410–3426, 2025.
- [15] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)