

JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA
PGDM / PGDM (M) / PGDM (SM)
FIFTH TRIMESTER (Batch 2024-26)
END TERM EXAMINATIONS, JANUARY 2026

SET-2

Course Name	Advanced Machine Learning	Course Code	20841
Max. Time	2 hours	Max. Marks	40 MM

INSTRUCTIONS:

a. All questions are compulsory to attempt

1. Read the following case and answer the questions given at the end:

ShopEase's Analytics Journey: From Data Chaos to Intelligent Decisions

ShopEase Retail Solutions is a fast-expanding digital-first retail company operating a mobile app and website that cater to urban consumers seeking convenience-driven purchases. The company offers a curated range of products across categories such as personal care, packaged foods, home utilities, and lifestyle accessories. With rapid growth over the last two years, ShopEase has accumulated vast amounts of customer data, including transaction records, browsing logs, loyalty activity, and customer feedback.

Despite healthy customer acquisition numbers, ShopEase's leadership has started noticing inconsistent customer retention patterns. While some customers continue purchasing regularly, others disengage after a short period of activity. Preliminary analysis suggests that disengagement is often preceded by subtle behavioral changes such as reduced app visits, declining order frequency, higher dependence on discounts, and negative service interactions. The management team wants to identify such customers before they disengage, so that timely interventions can be planned.

At the same time, the marketing team faces challenges in designing targeted campaigns. Customers display wide variations in shopping behavior—some make frequent small purchases, others shop occasionally but spend significantly more, and a few focus heavily on specific product categories. Past campaigns that treated all customers similarly have delivered poor conversion rates, prompting the team to explore data-driven grouping of customers based on behavioral similarity.

On the merchandising front, ShopEase wants to increase average order value. Although transaction logs reveal that certain products are often purchased together, these relationships are not immediately obvious to category managers. Management believes that uncovering hidden co-purchase patterns can help in designing smart bundles, in-app prompts, and personalized checkout recommendations.

ShopEase also invests heavily in personalization features within its app. However, customers frequently complain that suggested products are repetitive or irrelevant, especially for users with limited purchase history. The company wants to improve product suggestions by learning from patterns in customer interactions and similarities across users and products, even when explicit ratings are unavailable.

Finally, ShopEase receives thousands of customer reviews and feedback messages every week across app stores, in-app surveys, and social media. While star ratings provide a high-level overview, they fail to capture the underlying sentiment and emotional tone expressed in textual feedback. Leadership wants to systematically analyze these reviews to detect dissatisfaction early, identify recurring service issues, and feed these insights back into retention and marketing strategies.

To address these interconnected challenges, ShopEase's analytics team is tasked with building a robust, scalable, and interpretable machine learning-driven decision framework that supports business growth while maintaining customer trust.

Case Questions:

Q1. ShopEase has observed that some customers disengage after a brief period of activity, often preceded by reduced app usage, declining order frequency, and increased discount dependence.

(1+6+5 = 12 Marks)

- Identify the type of machine learning problem that best fits ShopEase's objective of proactively identifying such customers.
- Critically evaluate how at least two different classification algorithms could behave differently when applied to this problem, considering data characteristics and business constraints.
- Explain how the output of the chosen model can be operationalized into concrete retention strategies, rather than remaining a purely predictive exercise.

Q2. To improve personalization and campaign effectiveness, ShopEase analyzed customer behavioral data including purchase frequency, average order value, discount usage, and app engagement. The analytics team applied clustering techniques to group customers based on similarity in behavior.

The following outputs were obtained after running clustering analysis in Python.

Cluster Summary Output

Cluster	Number of Customers	Avg. Monthly Orders	Avg. Order Value (₹)	Discount Usage (%)	App Engagement Level
C1	1,200	8.5	650	12	High
C2	900	3.2	1,450	8	Medium
C3	700	6.1	420	35	High
C4	1,100	1.4	520	5	Low

Cluster Profile Labels (Post-Analysis Interpretation)

Cluster	Behavioral Interpretation
C1	Loyal high-frequency customers
C2	High-value occasional buyers
C3	Discount-driven active users
C4	Low-engagement dormant users

Assess the output and answer the following questions:

(4*3 = 12 Marks)

- Analyze the cluster summary and identify which cluster contributes most to revenue potential. Justify your answer using at least two behavioral indicators.
- Evaluate the business risk of treating Clusters C3 and C4 with the same marketing strategy, despite both showing lower average order value than C2.
- Design one targeted intervention strategy for any one cluster that balances customer experience and business profitability.

Q3. You have been shortlisted for the position of Machine Learning Analyst at ShopEase Retail Solutions, the organization described in the case study. During the interview, the panel asks you to demonstrate your conceptual clarity and business understanding by answering the following questions with reference to ShopEase's analytics challenges. **(4*4 = 16 Marks)**

1. ShopEase observed that a churn prediction model performs extremely well on historical customer data but shows degraded performance on new customers.
Using this scenario, explain the difference between overfitting and underfitting in machine learning, and discuss how each can impact ShopEase's decision-making.
2. ShopEase uses different analytical approaches to predict customer disengagement and to group customers based on behavioral similarity.
With reference to these use cases, compare supervised learning and unsupervised learning, highlighting when and why each approach is appropriate.
3. ShopEase's customer dataset includes variables such as total spend, number of app visits, discount percentage used, and days since last purchase, all measured on different scales.
Discuss the difference between standardization and normalization and justify when each should be applied while building models for ShopEase.
4. While analyzing customer behavior data, ShopEase's analytics team uses statistical tests to determine whether certain variables significantly influence customer disengagement.
Explain what it implies when p-values are high or low in this context, and how this interpretation can guide model design or feature selection decisions.