Page No.
Date

Module 5

Q1) Market basket analysis

=> (1) Market basket analysis is a data mining technique that identifies purchase patterns in retail setting by analyzing combinations of products bought together.

(2) Its primary goal is to uncover customer behaviour and their association among items enabling businesses to enhance sale strategies

(3) Association Rule in market basket analysis:

(a) Expressed as {If} → {then}, this rule highlights the relationship between products frequently bought together

(b) Eg: If a customer buys bread, then they are likely to buy milk or butter.

(4) Apriori Algorithm

(a) An algorithm used to identify frequent itemsets and derive association rules.

(b) It has three key components:

(i) Support: Measures the frequency of a product or itemset

$$\text{Support}(A, B) = \frac{\text{Transaction with both A \& B}}{\text{Total transaction}}$$

(ii) Confidence: Measures how often items in the consequent are purchased

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Transaction with both A \& B}}{\text{Transaction with A}}$$

(iii) Lift: Measures the strength of association

$$\text{Lift}(A \rightarrow B) = \frac{\text{Confidence}(A \rightarrow B)}{\text{Support}(B)}$$

A lift value > 1 indicates a strong association

Page No.
Date

(5) Benefits of market basket analysis:

(a) Enhanced customer understanding
(i) Identifies frequently purchased products and customer preferences.
(ii) Eg: Knowing that customers buying cereal also prefer milk.

(b) Sales growth
(i) Optimizes online recommendations and store layouts to boost sales.
(ii) Eg: Displaying chocolates near flowers on valentine's day.

(c) Better pricing strategies
(i) Develops pricing bundles based on product association.
(ii) Eg: offering discounts on combo packs of chips and sodas.

(6) Applications of market basket analysis

(a) E-commerce
(i) Enhance product recommendations and targeted marketing

(b) Finance
(i) Predict investor behavior for better portfolio recommendations.

(c) Retail
(i) Optimizing shelf arrangements and promotions

(Q2) Explain multilevel and multidimensional association rule mining.

=> (1) Data mining techniques like multilevel and multi-dimensional association rule mining helps to uncover hidden patterns in datasets.

(2) Multilevel association rule mining

(i) Multilevel association rule mining analyzes data across various levels of abstraction using concept hierarchies.

(ii) The key features are:

(a) Concept hierarchies

(1) Rules are mined at different levels of abstraction.

(2) Eg: Electronics → Laptops → HP Laptops

(3) There are two types of rules: high level rules and low level rules.

(b) Support threshold

(1) Uniform support: Same support value is applied across all levels.

Eg: If the minimum support is 5%, both laptops & electronics must meet the threshold.

(2) Reduced support: Lower support thresholds are used at lower levels of abstraction to capture less frequent but meaningful patterns.

(c) Methods of mining

(1) Level-by-level independent search: each level is mined independently.

(2) Level-cross filtering by parent node: A child node is only mined if the parent node is frequent

(B) Group-based support: Groups are formed based on expert input.
(i) Eg: If a customer buys an HP laptop, they are likely to buy an HP laptop bag.
(3) Multidimensional association rule mining
(i) Multidimensional association rules analyze relationship across multiple dimensions of data
(ii) Key features are:
(a) Attributes
(1) Qualitative attributes: Non-numeric attributes like product type, customer region, etc.
(2) Quantitative attributes: Numeric attributes like age or income.
(b) Discretization approaches
(1) Static discretization - Attributes are categorized into fixed intervals before mining.
(2) Dynamic discretization - Data is discretized during mining using clustering.
(c) Predicate sets
(1) Association rules are based on combinations of predicates.
(2) Eg: If a customer is aged '20-25' and earns '30k - 40k', they are likely to purchase a laptop.
(d) Methods of mining:
(1) Apriori algorithm
(2) Dynamic clustering
(iii) Eg: If a customer is aged '30-40' and resides in 'urban area', they are likely to purchase a 'smartphone'.