

# Association Rule Mining

Discovering Patterns in Data



Market Basket



Pattern Discovery



Data Insights

# What is Association Rule Mining?

## Definition

Finding frequent patterns, associations, or correlations among sets of items in large databases

## Rule Format

$X \rightarrow Y$  (If X occurs, Y is likely to occur)

Example: {Bread}  $\rightarrow$  {Butter}

## Key Metrics

### Support

Frequency of itemset in the dataset

$$\text{Support}(X) = \text{Count}(X) / \text{Total Transactions}$$

### Confidence

Conditional probability of Y given X

$$\text{Confidence}(X \rightarrow Y) = \text{Support}(X \cup Y) / \text{Support}(X)$$

### Lift

Strength of association between items

$$\text{Lift}(X \rightarrow Y) = \text{Confidence}(X \rightarrow Y) / \text{Support}(Y)$$

# Association Rule Mining Algorithms



## Apriori

### Level-wise Search

Breadth-first approach

### Candidate Generation

Generate and test candidates

### Pruning Strategy

Eliminate infrequent itemsets

Best for: Small datasets



## FP-Growth

### FP-Tree Structure

Compact data representation

### No Candidate Generation

Direct pattern extraction

### Faster Execution

Reduced database scans

Best for: Large datasets



## Eclat

### Vertical Data Format

TID-based representation

### **Set Intersection**

Compute support via intersection

### Memory Efficient

Lower memory requirements

Best for: Dense datasets

# Applications



## Market Basket Analysis

Discover product purchase patterns



## Cross-Selling Strategies

Identify product recommendations



## Fraud Detection

Identify unusual transaction patterns



## Medical Diagnosis

Find symptom-disease associations



## Web Usage Mining

Analyze user navigation patterns







## Product Placement

Optimize store layout strategies

# Advantages & Disadvantages





## Advantages

-  **Discovers Hidden Patterns**  
Finds non-obvious relationships in data
-  **Easy to Understand**  
Rules are interpretable and actionable
-  **Scalable**  
Handles large datasets efficiently
-  **Domain Independent**  
Applicable across various industries



## Disadvantages

-  **Generates Many Rules**  
May produce overwhelming number of patterns
-  **Trivial Rules**  
May include obvious or meaningless patterns
-  **Requires Large Datasets**  
Needs substantial data for meaningful patterns
-  **Parameter Sensitivity**  
Results depend on support/confidence thresholds