Market Basket Analysis

Understanding customer purchasing behavior through transaction analysis

Key Applications:

- Identifies frequent itemsets in transactions
- Improves cross-selling and upselling strategies
- Aids in designing targeted promotions

@AIinMinutes

Core Concepts

Understanding the building blocks of MBA

Key Terms:

- **Item**: Single product in a transaction (e.g., iPhone)

- **Itemset**: Group of items purchased together

- **Transaction**: Collection of items in a single purchase

- **Frequent Itemset**: Set of items appearing often together

- **Association Rule**: Relationship expressed as:

$$X(Phone) \rightarrow Y(Charger)$$

@AIinMinutes

Transaction Examples

Understanding through examples

Sample Transactions:

ID	Items Purchased
1	Phone, Charger, Phone-cover
2	Phone, Charger
3	Charger, Phone-cover
4	Phone, Phone-cover
5	Phone, Charger, Phone-cover

Itemset Examples: - Single-itemset: {Phone}, {Phone-cover}
- Two-itemset: {Phone, Charger}, {Charger, Phone-cover}
- Three-itemset: {Phone, Charger, Phone-cover}

@AIinMinutes

Support

Measures how frequently an itemset appears in transactions

Formula:

$$\text{Support}(X) = \frac{\text{Transactions containing } X}{\text{Total Transactions}}$$

Interpretation:

- **Higher Support**: Itemset is common (e.g., {Phone, Phone-case})

- **Lower Support**: Itemset is rare (e.g., {Phone, Mouse})

- Used to filter out infrequent combinations

- Analogous to relative frequency

@AIinMinutes

Confidence

Measures how often item Y is bought when item X is bought

Formula:

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X, Y)}{\text{Support}(X)}$$

Interpretation:

- If > 0.5: Y appears in more than 50% transactions containing X

- If $= 1$: Every transaction with X also contains Y

- Similar to conditional probability P(Y | X)

- Stronger when antecedent is rare

@AIinMinutes

Lift

Measures how much stronger the association is compared to random chance

Formula:

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

Interpretation:

- **Lift** $>$ **1**: Positive association
- **Lift** $=$ **1**: No association
- **Lift** $<$ **1**: Negative association

@AIinMinutes

Conviction

Measures how strongly the presence of one item implies the presence of another, based on how often the rule fails

Formula:

$$\text{Conviction}(X \rightarrow Y) = \frac{1 - \text{Support}(Y)}{1 - \text{Confidence}(X \rightarrow Y)}$$

Interpretation:

- **Conviction** $>$ **1**: X suggests Y more strongly (more predictive power)

- **Conviction** $=$ **1**: No predictive power

@AIinMinutes

Leverage

Measures the difference between observed and expected frequency

Formula:

$$\text{Leverage}(X \rightarrow Y) = \text{Sup}(X, Y) - (\text{Sup}(X) \times \text{Sup}(Y))$$

Interpretation:

- **Leverage** $>$ **0**: Items co-occur more often than expected

- **Leverage** $=$ **0**: Items co-occur exactly as expected

- **Leverage** $<$ **0**: Items co-occur less than expected

- Analogous to observed vs expected frequency in χ^2 *test*

@AIinMinutes