

# Lý thuyết bán chéo

## Lý thuyết bài toán áp dụng

### 1. Xây dựng bộ quy tắc Association rule

#### 1.1 Lý thuyết Association rule

Từ itemset, việc KH sử dụng đồng thời các sản phẩm, tính năng sẽ được xem xét tại từng tháng. Khoảng thời gian dữ liệu được sử dụng để phân tích là 03 tháng gần nhất. Trong đó, mỗi dữ kiện về một KH sử dụng sản phẩm/tính năng tại mỗi tháng là một bản ghi. Dữ liệu tổng hợp sẽ là dữ liệu từ ba tháng.

Mỗi quy tắc kết hợp sản phẩm (association rule) bao gồm danh sách sản phẩm tiền đề (antecedent) và danh sách sản phẩm hệ quả (consequent). Các quy tắc được đánh giá dựa trên ba hệ số đo lượng bao gồm support, confidence, lift. Công thức cụ thể cho association rule như sau:

Xét một quy tắc  $\{X\} \rightarrow \{Y\}$  với X là danh sách sản phẩm tiền đề và Y là danh sách sản phẩm hệ quả

- **Support:** Thể hiện tần suất xuất hiện của itemset trong dữ liệu quan sát, được tính theo công thức như sau.

$$\text{Support}(\{X\} \rightarrow \{Y\}) = \frac{\text{Tổng số bản ghi chứa } X \text{ và } Y}{\text{Tổng số bản ghi}}$$

Hệ số Support có giá trị nằm trong đoạn  $[0, 1]$ . Hệ số càng cao thể hiện số lượng quan sát đối có chứa bộ quy tắc càng lớn, thể hiện mức độ tin cậy của bộ quy tắc càng lớn và ngược lại (nếu hệ số này quá nhỏ, thì sẽ có rất ít bản ghi chứa bộ quy tắc đang xét, và khả năng bộ quy tắc được xuất hiện do ngẫu nhiên càng cao, tính tin cậy thấp).

- **Confidence:** Thể hiện tần suất xuất hiện của sản phẩm Y trong itemset, được tính toán với công thức như sau.

$$\text{Confidence}(\{X\} \rightarrow \{Y\}) = \frac{\text{Tổng số bản ghi chứa } X \text{ và } Y}{\text{Tổng số bản ghi chứa } X}$$

Hệ số Confidence có giá trị nằm trong đoạn  $[0, 1]$ , giá trị Confidence càng cao thể hiện mức độ liên kết giữa X và Y càng lớn. Trong đó, nếu giá trị này là 1 thì 100% khi KH sử dụng các sản phẩm X, sẽ sử dụng đồng thời sản phẩm Y.

- **Lift:** Hệ số này so sánh khả năng mua Y kèm điều kiện đã mua X với khả năng mua Y không kèm điều kiện, tức việc mua X khiến khả năng mua Y thay đổi như thế nào.

$$\text{Lift}(\{X\} \rightarrow \{Y\}) = \frac{\frac{\text{Số lượng bản ghi chứa } X \text{ và } Y}{\text{Số lượng bản ghi chứa } X}}{\frac{\text{Số lượng bản ghi chứa } Y}{\text{Tổng số bản ghi}}}$$

Hệ số Lift có giá trị nằm trong khoảng  $[0, +\infty)$ . Giá trị Lift bằng 1 thể hiện X và Y không có ảnh hưởng đến nhau:

- Giá trị Lift nhỏ hơn 1 thể hiện ảnh hưởng tiêu cực của hai nhóm hàng, khi khách hàng mua X thì xu hướng mua Y giảm đi.
- Giá trị Lift lớn hơn 1 thể hiện ảnh hưởng tích cực của hai nhóm hàng, khi mua X thì xu hướng mua Y tăng lên.

## 1.2. Kết quả mô hình

Kết quả đầu ra mô hình là một số các quy tắc giữa sản phẩm tiền đề và sản phẩm hệ quả tốt nhất thỏa mãn các tiêu chí sau:

- Số lượng các sản phẩm trong bộ quy tắc phải vừa đủ như đã đề cập ở trên
- Các sản phẩm phải tương đối thông dụng
- Các ngưỡng support, confidence, lift cao (so sánh tương đối với các bộ quy tắc còn lại)

## 2. Phương án đánh giá hiệu quả

Phương án đánh giá hiệu quả sẽ được thực hiện thông qua A/B testing, chiến dịch marketing về quảng bá sản phẩm tới khách hàng hiện hữu sẽ được thực hiện trên 2 tập khác nhau: tập KH được xác định theo đề xuất và tập KH ngẫu nhiên. Tỷ lệ chuyển đổi của hai tập KH được đo cụ thể như sau:

$$\text{Tỷ lệ chuyển đổi} = \frac{\text{Số lượng KH sử dụng sản phẩm}}{\text{Số lượng KH nhận quảng cáo/marketing}}$$

Việc đánh giá hiệu quả của việc áp dụng mô hình sẽ được thực hiện bởi việc so sánh tỉ lệ chuyển đổi của hai nhóm KH này.