

# 연관성 분석

20241982\_안동건

# CONTENTS

01      연관성 정의

---

02      코드 flow 소개

---

03      결과 확인

---

# 내가 생각하는 연관성 분석

특정한 조건절에 대해 다음 나올 결과를 예측하는 것

|              |  |
|--------------|--|
| 지지도(support) | - 특정 항목 집합이 전체 거래 중에서 얼마나 자주 등장하는지를 나타내는 지표                    |
| 신뢰도(support) | - A가 발생했을 때 B가 발생할 확률을 나타내는 지표                                 |
| 향상도(lift)    | - A가 발생했을 때 B가 발생하는 확률이,<br>B가 무작위로 발생하는 확률보다 얼마나 높은지를 나타내는 지표 |

# 코드 flow

|            |   |
|------------|---|
| 데이터 전처리    | 공백 제거 -> 공백 다시 살리기 -> 불용어 처리 -> 형태소 추출            |
| 연관성 분석     | 형태소 기반으로 지지도,향상도,신뢰도를 csv 파일로 저장                  |
| 결과 종합 및 예측 | 지지도,향상도,신뢰도 각각 0.2//0.4//0.4의 가중치를 부여하여 다음 형태소 예측 |

$$P(A \cap B) = \frac{\text{A와 B가 동시에 포함된 거래 수}}{\text{전체 거래 수}}$$

|    |                                    |
|----|------------------------------------|
| 수식 | 전체 거래 중 상품 A와 상품 B가 동시에 포함되는 거래 비율 |
| 목적 | 전체적인 구매 경향 파악                      |

$$\frac{P(A \cap B)}{P(A)} = \frac{\text{A와 B가 동시에 포함된 거래 수}}{\text{A를 포함하는 거래 수}}$$

|    |  |
|----|--|
| 수식 | 상품 A가 포함된 거래 중에서 상품 A, B가 동시에 포함되는 거래 비율 |
| 목적 | 연관성의 정도 파악 (예를 들어, A를 구매하면 B도 구매하는지)     |

$$\frac{P(A \cap B)}{P(A)P(B)} = \frac{\text{A와 B가 동시에 포함된 거래 수} * \text{전체 거래 수}}{\text{A를 포함하는 거래 수} * \text{B를 포함하는 거래 수}}$$

|    |  |
|----|--|
| 수식 | 상품 B가 포함된 거래 중에서 상품 A를 거래 후<br>상품 B를 거래한 비율            |
| 목적 | A가 발생했을 때 B가 발생하는 확률이,<br>B가 무작위로 발생하는 확률보다 얼마나 높은지 확인 |

# 계산식

---

$$\text{score} = w_s \cdot S + w_c \cdot \max(C_{A \rightarrow B}, C_{B \rightarrow A}) + w_l \cdot \max(L_{A \rightarrow B}, L_{B \rightarrow A})$$

이렇게 가중치 추가