

ARM

Association Rule Mining

구매 물품의 연관성

214683 장인환

데이터 설명



데이터 설명

- 마트 사용자중 구매 물품을 행에 표현한 형태
- 데이터는 7500개가 존재한다.
- 해당 데이터에서 연관 데이터를 출력하고한다.
- 연관성이 가장 큰 두 제품에 대한 연관성을 찾을 것이다.
- 상위 10종을 이용해 Support, Confidence, Lift 값을 출력할 생각이다.

1	shrimp	almonds	avocado	vegetables	green grape	whole wheat	yams
2	burgers	meatballs	eggs				
3	chutney						
4	turkey	avocado					
5	mineral water	milk	energy bar	whole wheat	green tea		
6	low fat yogurt						
7	whole wheat	french fries					
8	soup	light cream	shallot				
9	frozen vegetables	spaghetti	green tea				
10	french fries						
11	eggs	pet food					
12	cookies						
13	turkey	burgers	mineral water	eggs	cooking oil		
14	spaghetti	champagne	cookies				
15	mineral water	salmon					
16	mineral water						
17	shrimp	chocolate	chicken	honey	oil	cooking oil	low fat yogurt
18	turkey	eggs					
19	turkey	fresh tuna	tomatoes	spaghetti	mineral water	black tea	salmon
20	meatballs	milk	honey	french fries	protein bar		
21	red wine	shrimp	pasta	pepper	eggs	chocolate	shampoo
22	rice	sparkling water					
23	spaghetti	mineral water	ham	body spray	pancakes	green tea	
24	burgers	grated cheese	shrimp	pasta	avocado	honey	white wine
25	eggs						
26	parmesan	spaghetti	soup	avocado	milk	fresh bread	
27	ground beef	spaghetti	mineral water	milk	energy bar	black tea	salmon
28	sparkling water						
29	mineral water	eggs	chicken	chocolate	french fries		
30	frozen vegetables	spaghetti	yams	mineral water			
31	herb & pepper	tomato sauce	light cream	magazines			
32	mineral water	chocolate	avocado	eggs			

모델링 과정



전체 코드

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules

df = pd.read_csv("C:/Users/user/Desktop/수업관련/3학년 2025/1학기/인공지능/데이터/transactions.csv")

# 각 행을 리스트로 변경
transactions = df.apply(lambda row: row.dropna().tolist(), axis=1).tolist()

# One-hot encoding
te = TransactionEncoder()
te_data = te.fit(transactions).transform(transactions)
df_encoded = pd.DataFrame(te_data, columns=te.columns_)

# Apriori
frequent_itemsets = apriori(df_encoded, min_support=0.01, use_colnames=True)

# rule 생성
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=0.5)

rules = rules.sort_values(by="lift", ascending=False)
print(rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']])
```

모델링 과정

- pandas를 통해 데이터 가져오기
- 리스트 형태로 만들기
현재 csv의 형태를 리스트 형태로 구현
각 행은 현재 구매자의 구매 물품으로 구성
구매 물품에 대한 각 리스트 생성
- One - hot 인코딩 실행
무분별한 리스트를 수학적으로 표현
각 물품이 존재 여부를 통해 0과 1로 표현
- Apriori 알고리즘 사용
최소 0.01의 support 값을 가지고 있는 항목만 추출
- rule 설정
자주 등장하는 목록들에 대해서 Lift가 1.0을 이하는 없애기
내림차순으로 정리 (상위 10가지만 사용)

결과



결과

antecedents	consequents	support	confidence	lift
(herb & pepper)	(ground beef)	0.015998	0.323450	3.291994
(ground beef)	(herb & pepper)	0.015998	0.162822	3.291994

Antecedents : 조건, 해당 상품을 구매

Confidence : 조건이 참일 때 결과도 함께 등장하는 확률(조건부 확률)

Consequents : 결과, 해당 상품을 같이 구매

Lift : 두 항목이 얼마나 자주 등장하는가 (Confidence / Consequents)

Support : 전체 거래 중 두 상품이 나온 비율

- herb & pepper → ground beef

Support : 전체 거래 중 1.5998% 두 항목 동시에 등장

Confidence : pepper를 구매한 사람이 beef 구매할 확률 32.345%

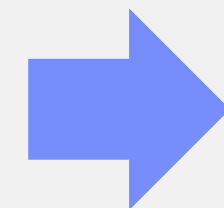
Lift : 3.291994배 더 자주 팔린다

- ground beef → herb & pepper

Support : 전체 거래 중 1.5998% 두 항목 동시에 등장

Confidence : beef를 구매한 사람이 pepper 구매 확률 16.2822%

Lift : 3.291994배 더 자주 팔린다



- beef와 pepper를 가까이 배치하면 조금 더 효율적인 방법이 될 수 있다.

- 역방향의 Confidence가 더 작은 걸 통해 beef의 판매량이 더 많다는 걸 알 수 있다.

결과

antecedents	consequents	support	confidence	lift
(herb & pepper)	(ground beef)	0.015998	0.323450	3.291994
(ground beef)	(herb & pepper)	0.015998	0.162822	3.291994
(spaghetti, mineral water)	(ground beef)	0.017064	0.285714	2.907928
(ground beef)	(spaghetti, mineral water)	0.017064	0.173677	2.907928
(spaghetti, mineral water)	(olive oil)	0.010265	0.171875	2.609786
(olive oil)	(spaghetti, mineral water)	0.010265	0.155870	2.609786
(frozen vegetables)	(tomatoes)	0.016131	0.169231	2.474464
(tomatoes)	(frozen vegetables)	0.016131	0.235867	2.474464
(shrimp)	(frozen vegetables)	0.016664	0.233209	2.446574
(frozen vegetables)	(shrimp)	0.016664	0.174825	2.446574

- herb & pepper → ground beef
- spaghetti, mineral water → ground beef
- spaghetti, mineral water → olive oil
- tomatoes → frozen vegetables
- shrimp → frozen vegetables



- 건강 식품 : frozen vegetables, tomatoes, shrimp
- 요리 : olive oil, spaghetti, mineral water



- 해당 되는 요소들을 비슷한 위치에 진열 혹은 광고에 활용한다면 유용한 정보가 될 수 있다.