



Estudo de associação de itens

Eric Ribeiro Fernandes

Objetivos do estudo



O presente estudo possui a finalidade de identificar quais itens em um supermercado são recorrentemente comprados juntos.



A técnica de mineração de dados utilizada foi o algoritmo de **Apriori**.



O *dataset* de estudo possui 539.574 observações.



Principais métricas do algoritmo de Apriori

Conceituação teórica das métricas de
Suporte, Confiança e Lift.

Suporte

- Suporte $(A \Rightarrow B) = P(A \cup B)$

- Definido pela fórmula acima o Suporte é uma métrica de **frequência** de ocorrência de determinado grupo de itens na base de dados.
- Ou seja, se em uma base de dados de 10 registros houverem 2 registros em que banana e maçã foram vendidos **juntos** o suporte desse conjunto é $\frac{2}{10}$, ou seja, 20% das transações totais tiveram banana e maçã.

Confiança

- **Confiança** ($A \Rightarrow B$) = $P(B|A)$

- Descrito pela fórmula acima a confiança representa a probabilidade de ocorrência de um valor, ou conjunto de valores, dado a ocorrência de um valor ou conjunto de valores.
- Por exemplo, das duas vezes que comprei banana e maçã também comprei morango uma vez, então a confiança que eu compre morango dado que comprei banana e maçã é $\frac{1}{2}$, ou seja, 50% das vezes.
- Logo, **SE** banana **E** maçã **ENTÃO** morango. (50% das vezes historicamente)
- **A ideia da métrica de confiança é mensurar a afirmação da preposição (SE...ENTÃO...).**

Lift

$$\text{Lift} = \frac{P(Y | X)}{P(Y)}$$

- Descrito pela fórmula acima o Lift é uma medida de correlação entre os os itens avaliados.
- Se **Lift > 1** significa dizer que o item Y possui **correlação positiva** com X, ou seja, há maior probabilidade de compra de Y dado a compra de X.
- Se **Lift < 1** significa dizer que o item Y possui **correlação negativa** com X, ou seja, há menor probabilidade de compra de Y dado a compra de X.



Pré-processamento e aplicação do Algoritmo

Criação da tabela de input

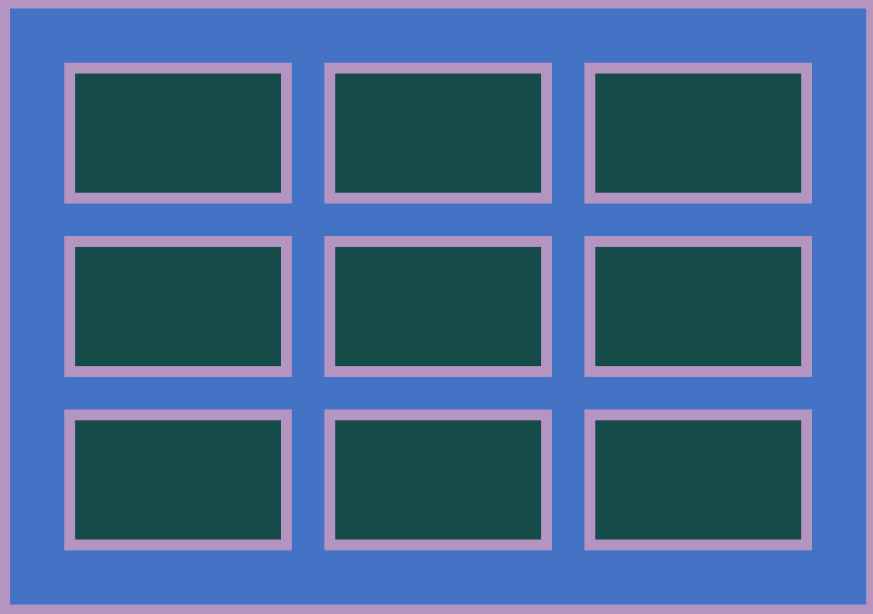




Tabela de Input

[illegible]

Biblioteca utilizada

- As bibliotecas utilizadas são descritas a seguir:

```
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

- A aplicação do algoritmo é descrita a seguir, sendo a cesta a tabela apresentada no slide anterior:

Aplicando o algoritmo de Apriori

[12] ▶ ▶≡ ML

```
# Frequência de itens
itens_frequentes = apriori(cesta, min_support = 0.02, use_colnames = True)
```

[13] ▶ ▶≡ ML

```
## Gerando regras e ordenando por confiança
regras = association_rules(itens_frequentes, metric = 'lift', min_threshold = 1)
regras.sort_values('confidence', ascending=False).head(50)
```

Resultados		antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage
174		(CREME DENTAL, DETERGENTES EM PO)	(SABONETES MASSA)	0.025943	0.076436	0.020654	0.796101	10.415230	0.018671
144		(CAFE A VACUO, FEIJAO PRETO)	(ACUCAR REFINADO)	0.026600	0.113110	0.020025	0.752826	6.655712	0.017016
162		(OLEOS, FEIJAO PRETO)	(ACUCAR REFINADO)	0.033079	0.113110	0.024590	0.743388	6.572272	0.020849
150		(CAFE A VACUO, OLEOS)	(ACUCAR REFINADO)	0.029484	0.113110	0.021187	0.718591	6.353036	0.017852
168		(SABONETES MASSA, OLEOS)	(ACUCAR REFINADO)	0.029388	0.113110	0.020503	0.697674	6.168116	0.017179
156		(OLEOS, DETERGENTES EM PO)	(ACUCAR REFINADO)	0.030919	0.113110	0.021378	0.691424	6.112852	0.017881
155		(ACUCAR REFINADO, DETERGENTES EM PO)	(OLEOS)	0.031151	0.086538	0.021378	0.686266	7.930268	0.018682
161		(ACUCAR REFINADO, FEIJAO PRETO)	(OLEOS)	0.036523	0.086538	0.024590	0.673278	7.780189	0.021430
166		(ACUCAR REFINADO, SABONETES MASSA)	(OLEOS)	0.030523	0.086538	0.020503	0.671742	7.762435	0.017862
173		(SABONETES MASSA, DETERGENTES EM PO)	(CREME DENTAL)	0.030796	0.060471	0.020654	0.670661	11.090622	0.018791
36		(AMACIANTE)	(DETERGENTES EM PO)	0.032053	0.079894	0.020859	0.650746	8.145072	0.018298
160		(ACUCAR REFINADO, OLEOS)	(FEIJAO PRETO)	0.040432	0.067893	0.024590	0.608181	8.957908	0.021845
85		(CREME DENTAL)	(SABONETES MASSA)	0.060471	0.076436	0.036523	0.603978	7.901725	0.031901
172		(SABONETES MASSA, CREME DENTAL)	(DETERGENTES EM PO)	0.036523	0.079894	0.020654	0.565494	7.078011	0.017736
143		(ACUCAR REFINADO, FEIJAO PRETO)	(CAFE A VACUO)	0.036523	0.090242	0.020025	0.548278	6.075659	0.016729
21		(FEIJAO PRETO)	(ACUCAR REFINADO)	0.067893	0.113110	0.036523	0.537950	4.756002	0.028844
148		(ACUCAR REFINADO, CAFE A VACUO)	(OLEOS)	0.040009	0.086538	0.021187	0.529552	6.119338	0.017724
154		(ACUCAR REFINADO, OLEOS)	(DETERGENTES EM PO)	0.040432	0.079894	0.021378	0.528736	6.617925	0.018148
149		(ACUCAR REFINADO, OLEOS)	(CAFE A VACUO)	0.040432	0.090242	0.021187	0.524003	5.806652	0.017538
46		(CENOURA)	(BATATA LISA)	0.050547	0.069533	0.026326	0.520822	7.490234	0.022811
75		(TOMATE VERMELHO/COMUM)	(CEBOLA GRANEL/EMB 1KG)	0.062603	0.069930	0.032518	0.519432	7.427902	0.028140
167		(ACUCAR REFINADO, OLEOS)	(SABONETES MASSA)	0.040432	0.076436	0.020503	0.507099	6.634278	0.017413

Obrigado!

Link do projeto: [https://github.com/Eric-Ribeiro-Fernandes/APRIORI PYTHON](https://github.com/Eric-Ribeiro-Fernandes/APRIORI_PYTHON)