

Utilizando o APRIORI para Cruzamento de Dados Sobre o Valor do Ouro pelos Anos

Pablo Leandro de Oliveira¹

Universidade do Estado de Minas Gerais (UEMG)

ABSTRACT: *This work was produced so that students can understand the functioning of the 'APRIORI' algorithm, as a means of understanding a database on the variation of the price of gold was used.*

RESUMO: Esse trabalho foi produzido para que os alunos possam entender o funcionamento do algoritmo 'APRIORI', como meio de entendimento foi utilizado um banco de dados sobre a variação do preço do ouro.

PALAVRAS-CHAVE: Gold, Ouro, Cotação, Flutuação.

1. INTRODUÇÃO

APRIORI é a nomenclatura designada a algoritmos que buscam '*linkar*' informações através do cruzamento de dados da base escolhida, seja ela as compras feitas por clientes ou acidentes de trânsito por exemplo. Vale ressaltar que nem toda base de dados é possível realizar o cruzamento de dados e obter alguma informação útil.

Nesse caso, utilizaremos a variação do ouro através dos tempos, com a utilização de 3 tipos de arquivos, que variam o tipo da amostragem da variação, indo desde variação por dia. Nos exemplos utilizados, foi usado o APRIORI na arquivo de variação diária.

	Data	Preco	Abertura	Maximo	Minimo	Vol	Mudanca %
0	Jan 01, 2021	1901.6	1901.6	1901.6	1901.6	-	-0.0065
1	Jan 04, 2021	1966.7	1934.7	1967.8	1934.7	0.34K	0.0342
2	Jan 05, 2021	1974.9	1965.3	1975.5	1965.3	1.41K	0.0042
3	Jan 06, 2021	1929.1	1975.9	1979.9	1926.5	1.13K	-0.0232
4	Jan 07, 2021	1932.8	1944.2	1946.7	1932.5	1.32K	0.0019

2. O QUE SE ESPERA

É esperado encontra alguma ligação entre ano e preço do ouro, o que pode ser ligado a estações e acontecimentos políticos, militares ou naturais, em maiores pesquisas.

3. METODO UTILIZADO

Como exemplo, temos abaixo um dos códigos testados na implementação:

```
import pandas as pd
from apriori_python import apriori

Gold_Daily= pd.read_csv("base\Gold_Daily.csv")

agrupar = Gold_Daily.groupby('Abertura')['Maximo'].apply(list)

##print(Gold_Daily.head())

freqItemSet, rules = apriori(agrupar, minSup=0.002, minConf=0.05)
print(len(agrupar))
print(rules)
```

Utilizando o 'groupby' houve a tentativa de encontrar uma ligação entre o dia de abertura e a máxima cotação do ouro no dia.

4. RESULTADO OBTIDO

Os resultados obtidos não foram muito consistentes em apresentar informações válidas, houveram muitas variações e inconsistências. Mas percebe-se que os anos entre 1770 e 1820 existe variações semelhantes e contínuas entre si.

```
[[{1810.3}, {1785.2}, 0.5], [{1817.5}, {1799.5}, 0.5], [{1779.2}, {1792.0}, 0.5], [{1798.8}, {1819.3}, 0.5], [{1813.3}, {1812.7}, 0.5], [{1802.1}, {1801.5}, 0.5], [{1802.1}, {1793.8}, 0.5], [{1823.7}, {1799.4}, 0.5], [{1799.4}, {1823.7}, 0.5], [{1785.2}, {1810.3}, 1.0], [{1799.5}, {1817.5}, 1.0], [{1875.1}, {1892.6}, 1.0], [{1892.6}, {1875.1}, 1.0], [{1772.6}, {1767.4}, 1.0], [{1767.4}, {1772.6}, 1.0], [{1794.8}, {1822.2}, 1.0], [{1822.2}, {1794.8}, 1.0], [{1791.0}, {1799.0}, 1.0], [{1799.0}, {1791.0}, 1.0], [{1800.9}, {1797.0}, 1.0], [{1797.0}, {1800.9}, 1.0], [{1792.0}, {1779.2}, 1.0], [{1760.0}, {1776.9}, 1.0], [{1776.9}, {1760.0}, 1.0], [{1802.4}, {1796.3}, 1.0], [{1796.3}, {1802.4}, 1.0], [{1819.3}, {1798.8}, 1.0], [{1812.7}, {1813.3}, 1.0], [{1824.3}, {1825.1}, 1.0], [{1825.1}, {1824.3}, 1.0], [{1801.5}, {1802.1}, 1.0], [{1809.0}, {1818.7}, 1.0], [{1818.7}, {1809.0}, 1.0], [{1793.8}, {1802.1}, 1.0]]
```

5. CONCLUSÃO

		Data	Preço	Abertura	Maximo	Minimo	Vol	Mudanca %
0	Jan 01, 2021	1901.60	1901.60	1901.60	1901.60	-	-	-0.0065
1	Jan 04, 2021	1966.70	1934.70	1967.80	1934.70	0.34K	-	0.0342
2	Jan 05, 2021	1974.90	1965.30	1975.50	1965.30	1.41K	-	0.0042
3	Jan 06, 2021	1929.10	1975.90	1979.90	1926.50	1.13K	-	-0.0232
4	Jan 07, 2021	1932.80	1944.20	1946.70	1932.50	1.32K	-	0.0019
5	Jan 08, 2021	1854.20	1935.80	1935.80	1850.00	1.11K	-	-0.0407
6	Jan 11, 2021	1869.60	1848.10	1867.80	1848.10	0.83K	-	0.0083
7	Jan 12, 2021	1862.40	1862.20	1866.00	1858.20	0.25K	-	-0.0039
8	Jan 13, 2021	1872.70	1875.10	1875.10	1871.20	0.25K	-	0.0055
9	Jan 14, 2021	1869.20	1856.60	1870.20	1856.60	0.15K	-	-0.0019
10	Jan 15, 2021	1846.40	1840.00	1842.00	1839.60	0.12K	-	-0.0122
11	Jan 17, 2021	1809.75	1826.80	1829.10	1803.60	-	-	-0.0198
12	Jan 18, 2021	1836.60	1809.65	1839.45	1809.65	-	-	0.0148
13	Jan 19, 2021	1855.00	1838.80	1858.80	1825.50	0.09K	-	0.0100
14	Jan 20, 2021	1880.50	1880.30	1880.30	1880.30	0.01K	-	0.0137
15	Jan 21, 2021	1879.50	1876.50	1879.40	1875.70	0.06K	-	-0.0005
16	Jan 22, 2021	1870.20	1852.90	1869.70	1852.90	0.04K	-	-0.0049
17	Jan 25, 2021	1869.60	1879.20	1879.20	1877.80	0.01K	-	-0.0003
18	Jan 26, 2021	1865.60	1868.60	1868.60	1857.00	0.01K	-	-0.0021
19	Jan 27, 2021	1859.50	1858.80	1860.80	1848.80	0.02K	-	-0.0033

Com a tentativa de retirar algum dado desta base de dados, o que eu esperava já se mostrou claro, para que se pudesse encontrar ligação entre a flutuação do preço do ouro, no mesmo arquivo deveria estar incluso mais informações, já que realmente temos apenas as variações e suas características.

Foi bem interessante utilizar o APRIORI e realmente tentar minerar dados, infelizmente o resultado não foi o esperado, mas todo resultado é um resultado válido. Posteriormente caso fosse necessário retirar dados desses arquivos seria interessante a adição de gráficos e tabelas (mas deixemos isto para outra ocasião).

6. REFERÊNCIAS

Base de Dados: <https://www.kaggle.com/nward7/gold-historical-datasets>