

Conjunto de dados de previsão de insuficiência cardíaca

Tarcisio Augusto Cabral Reis

Universidade do Estado de Minas Gerais

State University of Minas Gerais

Resumo. Esta pesquisa teve o intuito de analisar as informações relacionado a insuficiência cardíaca usando o al

1. INTRODUÇÃO

Este artigo tem o intuito de usar o banco de dados com o conjunto de informações de previsão de insuficiência cardíaca. Com isto, foi usado o algoritmo apriori para associar, com a intenção de busca por itens frequentes neste banco de dados que ocorrem simultaneamente em transações.

Para conseguir associar as informações do banco de dados foi necessário criar um código em python.

```
import pandas as pd
from apriori_python import apriori

heart = pd.read_csv('heart.csv')

agrupamento=
heart.groupby('Cholesterol')['ChestPainType'].apply(list)
print(heart.head())
freqItemSet, rules=apriori(agrupamento, minSup=0.000909,
minConf=0.5)
print(len(agrupamento))
print(rules)
```

As colunas analisadas para o uso do apriori foi em relação ao nível do colesterol e o tipo da dor causada por parada cardíaca, nesta associação encontramos as frequências demonstradas na tabela 1 e tabela 2 .

	Age	Sex	ChestPainType	...	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	...	0.0	Up	0
1	49	F	NAP	...	1.0	Flat	1
2	37	M	ATA	...	0.0	Up	0
3	48	F	ASY	...	1.5	Flat	1
4	54	M	NAP	...	0.0	Up	0

Tabela 1. Colunas da análise

```
[5 rows x 12 columns]
222
[[{'NAP', 'TA'}, {'ATA'}, 0.52],
```

```
[{'NAP', 'ASY', 'TA'}, {'ATA'}, 0.5217391304347826],
[{'NAP'} {'ATA'}, 0.5267857142857143],
[{'TA'}, {'ATA'}, 0.5384615384615384],
[{'ASY', 'TA'}, {'ATA'}, 0.5428571428571428],
[{'ATA'}, {'NAP'}, 0.5619047619047619],
[{'ATA', 'TA'}, {'NAP', 'ASY'}, 0.5714285714285714],
[{'NAP', 'ASY'}, {'ATA'}, 0.5862068965517241],
[{'TA'}, {'NAP', 'ASY'}, 0.5897435897435898],
[{'ATA', 'ASY'}, {'NAP'}, 0.6071428571428571],
[{'ATA', 'TA'}, {'NAP'}, 0.6190476190476191],
[{'ATA', 'ASY', 'TA'}, {'NAP'}, 0.631578947368421],
[{'TA'}, {'NAP'}, 0.6410256410256411],
[{'ASY', 'TA'}, {'NAP'}, 0.6571428571428571],
[{'NAP'}, {'ASY'}, 0.7767857142857143],
[{'ATA'}, {'ASY'}, 0.8],
[{'ATA', 'NAP'}, {'ASY'}, 0.864406779661017],
[{'TA'}, {'ASY'}, 0.8974358974358975],
[{'ATA', 'TA'}, {'ASY'}, 0.9047619047619048],
[{'NAP', 'TA'}, {'ASY'}, 0.92], [{"ATA", "NAP", "TA"}, {"ASY"},
0.9230769230769231]]
```

Tabela 2. Resultado da análise

Na segunda tabela encontram-se as análises, quando mostra qual tipo de dor sentida pelo ataque cardíaco e suas combinações para cada nível de colesterol. Percebe-se que a análise começa da menor frequência para a maior, na qual, quando as combinações de dor ‘NAP’, ‘ASY’, ‘TA’, {‘ATA’} tem a menor frequência (0.5217391304347826) e o que teve a maior frequência {‘NAP’, ‘TA’}, {‘ASY’}, 0.92], [{‘ATA’, ‘NAP’, ‘TA’} tem a maior frequência (0.9230769230769231). É notório que a dor ASY é a mais comum entre os pacientes que tiveram parada cardíaca com 54% das ocorrências.

2. CONCLUSÃO

Neste trabalho aborda o uso do algoritmo apriori para encontrar a frequência em relação entre o tipo da dor e o nível do colesterol e cheguei a conclusão Quando a combinação de dor 'NAP', 'ASY', 'TA', {'ATA'} tem a frequência mais baixa (0,5217391304347826) e a combinação de frequência mais alta {'NAP', 'TA'}, {'ASY'} , 0,92], [{‘ATA’, ‘NAP’, ‘TA’} tem a frequência mais alta (0,9230769230769231). A dor ASY é conhecida por ser a mais comum em pacientes com parada cardíaca, ocorrendo em 54% dos pacientes.

Este trabalho foi muito importante para aprofundar nos conhecimentos do algoritmo apriori e como relacioná-lo ao banco de dados. Devido a isto, adquire conhecimento sobre insuficiência cardíaca.

3. BIBLIOGRAFIA

<https://www.kaggle.com/fedesoriano/heart-failure-prediction>

https://docs.google.com/document/d/1M8EsiVKBh4_fwyrBH_Z0m4gNXIElXT_ieu5daF6apec/edit