

Mineração de Dados e Descoberta de Conhecimento

Profs. Heitor S. Lopes e Thiago H. Silva (UTFPR, 2022) - Exercício #5

A) Análise Associativa (I)



Objetivo: O dataset **bebê-baixo-peso** foi construído com base no Very Low Weighth Infants Dataset e registra dados de 671 bebês com peso ao nascer abaixo de 1600 gramas. Há dados sobre a mãe e o bebê incluindo alguns medicamentos utilizados, questões relacionadas ao parto, problemas diagnosticados no nascituro, e se o mesmo sobreviveu ou não ao parto. Na literatura médica, já são conhecidos alguns fatores que diminuem as chances de sobrevivência de bebês prematuros, por exemplo: bebês com extremo baixo peso (<1000 gramas), aqueles nascidos

muito prematuramente (<27 semanas de idade gestacional), e presença de comorbidades tais como pneumotórax ou complicações cardio-pulmonares resultantes da própria imaturidade pulmonar. Deseja-se, portanto, utilizar os métodos de regras de associação para descobrir associações relevantes que **não sejam óbvias**.

- Caso seja utilizado o Orange, é necessário reformatar o dataset para o formato "Basket" (mais informações [aqui](#) e [aqui](#)). Caso escolha o software Weka, utilize o algoritmo de associação *hotSpot* (instalável através do *Package Manager*), com *support=1* (ou muito próximo deste valor), *outputRules=True*, *maxBranchingFactor<5* e ajuste o consequente da regra para a variável *Dead* através da opção *target* e *targetIndex* (=1 ou 2 para *Died=no* ou *Died=yes*). Opcionalmente, pode ser utilizado Python.
- Em vez de regras genéricas, há interesse especial em descobrir regras que mostrem alguma associação entre o uso de *beta-methasone* pela gestante e a sobrevivência, ou não, do bebê-baixo-peso. Observe as regras encontradas pelo algoritmo e investigue se o resultado tem algum fundamento (Pesquise no Google!!).
- Fato importante: este dataset foi coletado na *Duke University Medical Center* no Estado da Georgia (USA), sendo este um dos estados americanos onde há uma grande concentração de afro-descendentes. Com isto em mente, obtenha regras de associação (com pelo menos 20% de suporte) que envolvam uma eventual relação entre raça e a prevalência, ou não, de morte de bebês-baixo-peso). Discuta as implicações sociais dos achados.

B) Análise Associativa (II)



Objetivo: O *dataset cafeteria-5000* foi construído com base no Extended Bakery Dataset e lista 5000 transações em uma cafeteria. Um total de 50 produtos são comercializados, sendo 40 comidas e 10 bebidas. Deseja-se utilizar os métodos de regras de associação para descobrir informações **não-óbvias** sobre o perfil de consumo dos clientes.

- a. Suponha que seja definido um “cliente fiel ao produto” como aquele que cada vez que vai ao estabelecimento consome os mesmos produtos. Utilizando o algoritmo **Apriori** ou **FP-Growth**, descubra regras que mostrem se o estabelecimento é frequentado por “clientes fiéis ao produto” ou não, considerando que tais regras se apliquem com grande confiança ($\geq 50\%$), no mínimo a 10% das transações.
- b. Por simples inspeção no *dataset*, verifique qual é a bebida e qual é a comida que são os campeões de venda. Com estas informações, considerando **apenas as transações que incluam uma comida e uma bebida**, encontre **regras de associação** com os máximos valores possíveis de suporte e acurácia para descobrir o seguinte: (1) qual a comida que mais acompanha a bebida campeã de vendas, (2) qual a bebida que mais acompanha a comida campeã de vendas, (3) quais os pares comida/bebida mais vendidos.
- c. Discuta como os achados anteriores poderiam ser úteis para subsidiar uma campanha de marketing para incentivar o consumo na Cafeteria e fidelizar os clientes.