

ATIVIDADE MINERAÇÃO DE DADOS 3

“Regra de associação sobre o dataset Dotto”

1 Descrição resumida da atividade desenvolvida

Esta atividade tem como objetivo principal proporcionar uma visão geral do processo de geração e análise das regras de associação dentro da área de Inteligência Artificial.

A proposta consiste basicamente no tratamento de uma base de dados fornecida (referente ao supermercado Dotto), implementação do algoritmo de geração de regras de associação apriori e análise das regras geradas na aplicação do método.

De forma direta, foi desenvolvido na linguagem Python de programação uma rotina dentro de um Jupyter Notebook que implementa os requisitos solicitados para avaliação na atividade. O código está disponível gratuitamente em [Micanga GitHub - Atividade 3](#).

Por fim, este documento busca discutir brevemente e analisar os resultados obtidos com o algoritmo desenvolvido sobre a base de dados fornecida de modo a classificar as regras de associações encontradas entre triviais e inovadoras.

2 Resultados obtidos

Nesta seção serão apresentados os resultados colhidos com o algoritmo Apriori de geração de Regras de Associação. Na primeira coluna da tabela temos a base da regra de associação (conjunto de itens **A**) e na segunda coluna temos a implicação da base (conjunto de itens **B** seguindo a forma $A \Rightarrow B, A \subset I, B \subset I, A \neq \emptyset, B \neq \emptyset, A \cap B = \emptyset$ com **I** sendo o conjunto de itens do dataset.

| Base | Adição |
|--|--------------|
| 'FERMROYAL' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'ABSSLIVRE' | 'LEITE_MOCA' |
| 'BOMBRIL', 'ACHOCNESCAU' | 'BISCNESTLE' |
| 'BISCNESTLE', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'MAIONESE_HELLMANN_S', 'ACHOCNESCAU' | 'BISCNESTLE' |
| 'MARGDORIANA', 'ACHOCNESCAU' | 'BISCNESTLE' |
| 'BOMBRIL', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'ACHOCNESCAU', 'ESPONJA_BOMBRIL' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'FEIJAO_TORRESAN', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'MAIONESE_HELLMANN_S', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'SAL_CISNE', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'SPO_OMO', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'VEJA_MUSO', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'ACUCAR_DA_BARRA' | 'LEITE_MOCA' |
| 'ATUM_COQUEIRO', 'ACUCAR_UNIAO' | 'LEITE_MOCA' |

| Base | Adição |
|--|--------------|
| 'GUARANA_ANTARTICA', 'ACUCAR_UNIAO' | 'COCA_COLA' |
| 'FERMROYAL', 'ACUCAR_UNIAO' | 'LEITE_MOCA' |
| 'GELATINA_ROYAL', 'ACUCAR_UNIAO' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'AGUA_SANITCANDURA' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'AGUA_SANITVAREK' | 'LEITE_MOCA' |
| 'AGUA_SANITVAREK', 'CREME_DE_LEITE_NESTLE' | 'COCA_COLA' |
| 'AGUA_SANITVAREK', 'CREME_DE_LEITE_NESTLE' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'AMACCOMFORT' | 'LEITE_MOCA' |
| 'ARROZ_PRATO_FINO', 'BISCNESTLE' | 'COCA_COLA' |
| 'ARROZ_PRATO_FINO', 'BISCNESTLE' | 'LEITE_MOCA' |
| 'ARROZ_PRATO_FINO', 'BOMBRIL' | 'LEITE_MOCA' |
| 'ARROZ_PRATO_FINO', 'CREME_DE_LEITE_NESTLE' | 'COCA_COLA' |
| 'ARROZ_PRATO_FINO', 'SAL_CISNE' | 'COCA_COLA' |
| 'ARROZ_PRATO_FINO', 'SPO_OMO' | 'COCA_COLA' |
| 'ARROZ_PRATO_FINO', 'CREME_DE_LEITE_NESTLE' | 'LEITE_MOCA' |
| 'ARROZ_PRATO_FINO', 'FARTRIGO_RENATA' | 'LEITE_MOCA' |
| 'ARROZ_PRATO_FINO', 'SAL_CISNE' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'ATUM_COQUEIRO' | 'LEITE_MOCA' |
| 'MAIONESE_HELLMANN_S', 'ATUM_COQUEIRO' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'BOMBRIL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'CALDO_MAGGI' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'CDCLOSEUP' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'CDSORRISO' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'GUARANA_ANTARTICA' | 'COCA_COLA' |
| 'BISCNESTLE', 'CREME_DE_LEITE_NESTLE' | 'LEITE_MOCA' |
| 'MACRENATA', 'CREME_DE_LEITE_NESTLE' | 'BISCNESTLE' |
| 'BISCNESTLE', 'ESPONJA_BOMBRIL' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'GELATINA_ROYAL' | 'BISCNESTLE' |
| 'BISCNESTLE', 'FEIJAO_BROTO_LEGAL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'FEIJAO_TORRESAN' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'FERMROYAL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'MARGDORIANA' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'OLEO_SOJA_SOYA' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'PAPEL_HIGPERSONAL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'SABLUX' | 'LEITE_MOCA' |
| 'LEITE_MOCA', 'TOALHA_PAPEL_MASCOT' | 'BISCNESTLE' |
| 'BISCTOSTINES', 'CREME_DE_LEITE_NESTLE' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'BISCTOSTINES' | 'LEITE_MOCA' |
| 'GELATINA_ROYAL', 'BOMBRIL' | 'COCA_COLA' |
| 'COCA_COLA', 'BOMBRIL' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'BOMBRIL' | 'LEITE_MOCA' |
| 'DETERGYPE', 'BOMBRIL' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'BOMBRIL' | 'LEITE_MOCA' |
| 'FEIJAO_BROTO_LEGAL', 'BOMBRIL' | 'LEITE_MOCA' |
| 'GELATINA_ROYAL', 'BOMBRIL' | 'LEITE_MOCA' |

| Base | Adição |
|--|-------------------------|
| 'MAIONESE_HELLMANN_S', 'BOMBRIL' | 'LEITE.MOCA' |
| 'MARGDORIANA', 'BOMBRIL' | 'LEITE.MOCA' |
| 'PAPEL_HIGPERSONAL', 'BOMBRIL' | 'LEITE.MOCA' |
| 'SAL_CISNE', 'BOMBRIL' | 'LEITE.MOCA' |
| 'VEJA_MUSO', 'BOMBRIL' | 'LEITE.MOCA' |
| 'CDCLOSEUP', 'COCA_COLA' | 'LEITE.MOCA' |
| 'CDCOLGATE', 'CREME_DE_LEITE_NESTLE' | 'COCA_COLA' |
| 'CDCOLGATE', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'SPO_OMO', 'CDSORRISO' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'FANTA' | 'COCA_COLA' |
| 'CREME_DE_LEITE_NESTLE', 'GUARANA_ANTARTICA' | 'COCA_COLA' |
| 'SAL_CISNE', 'CREME_DE_LEITE_NESTLE' | 'COCA_COLA' |
| 'CREME_DE_LEITE_NESTLE', 'SPO_OMO' | 'COCA_COLA' |
| 'DESINFPINHO', 'COCA_COLA' | 'LEITE.MOCA' |
| 'GUARANA_ANTARTICA', 'FANTA' | 'COCA_COLA' |
| 'LEITE.MOCA', 'FANTA' | 'COCA_COLA' |
| 'FARTRIGO_RENATA', 'GELATINA_ROYAL' | 'COCA_COLA' |
| 'FARTRIGO_RENATA', 'GUARANA_ANTARTICA' | 'COCA_COLA' |
| 'FEIJAO_BROTO_LEGAL', 'COCA_COLA' | 'LEITE.MOCA' |
| 'FERMROYAL', 'COCA_COLA' | 'LEITE.MOCA' |
| 'COCA_COLA', 'GELATINA_ROYAL' | 'LEITE.MOCA' |
| 'LEITE.MOCA', 'GUARANA_ANTARTICA' | 'COCA_COLA' |
| 'LEITE_PARMALAT', 'GUARANA_ANTARTICA' | 'COCA_COLA' |
| 'PAPEL_HIGNEVE', 'GUARANA_ANTARTICA' | 'COCA_COLA' |
| 'SPO_OMO', 'GUARANA_ANTARTICA' | 'COCA_COLA' |
| 'OLEO_SOJA_SOYA', 'COCA_COLA' | 'LEITE.MOCA' |
| 'SAL_CISNE', 'LEITE.MOCA' | 'COCA_COLA' |
| 'LEITE.MOCA', 'SCLIXO_PLASLIXO' | 'COCA_COLA' |
| 'COCA_COLA', 'VINAGRE_CASTELO' | 'LEITE.MOCA' |
| 'DETERGLIMPOL', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'FARTRIGO_RENATA', 'GELATINA_ROYAL' | 'CREME_DE_LEITE_NESTLE' |
| 'FARTRIGO_RENATA', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'FEIJAO_BROTO_LEGAL', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'FERMROYAL', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'GELATINA_ROYAL' | 'LEITE.MOCA' |
| 'LEITE.MOCA', 'GUARANA_ANTARTICA' | 'CREME_DE_LEITE_NESTLE' |
| 'CREME_DE_LEITE_NESTLE', 'LEITE_PARMALAT' | 'LEITE.MOCA' |
| 'MACRENATA', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'MAIONESE_HELLMANN_S' | 'LEITE.MOCA' |
| 'MARGDORIANA', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'PAPEL_HIGNEVE' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'PAPEL_HIGPERSONAL' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'SABLUX' | 'LEITE.MOCA' |
| 'SAL_CISNE', 'CREME_DE_LEITE_NESTLE' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'SPO_OMO' | 'LEITE.MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'SUCO_MAGUARY' | 'LEITE.MOCA' |

| Base | Adição |
|--|--------------|
| 'CREME_DE_LEITE_NESTLE', 'VEJA_MUSO' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'VINAGRE_CASTELO' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'DETERGLIMPOL' | 'LEITE_MOCA' |
| 'FERMROYAL', 'DETERGLIMPOL' | 'LEITE_MOCA' |
| 'DETERGLIMPOL', 'MAIONESE_HELLMANN_S' | 'LEITE_MOCA' |
| 'FERMROYAL', 'FARTRIGO_RENATA' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'GELATINA_ROYAL' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'MACRENATA' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'SPO_OMO' | 'LEITE_MOCA' |
| 'FEIJAO_BROTO_LEGAL', 'SPO_OMO' | 'LEITE_MOCA' |
| 'FERMROYAL', 'GELATINA_ROYAL' | 'LEITE_MOCA' |
| 'FERMROYAL', 'MACRENATA' | 'LEITE_MOCA' |
| 'FERMROYAL', 'MAIONESE_HELLMANN_S' | 'LEITE_MOCA' |
| 'FERMROYAL', 'PAPEL_HIGPERSONAL' | 'LEITE_MOCA' |
| 'GELATINA_ROYAL', 'MAIONESE_HELLMANN_S' | 'LEITE_MOCA' |
| 'MARGDORIANA', 'GELATINA_ROYAL' | 'LEITE_MOCA' |
| 'GELATINA_ROYAL', 'SPO_OMO' | 'LEITE_MOCA' |
| 'SPO_OMO', 'MAIONESE_HELLMANN_S' | 'LEITE_MOCA' |
| 'MAIONESE_HELLMANN_S', 'SUÇO_MAGUARY' | 'LEITE_MOCA' |
| 'OLEO_SOJA_SOYA', 'PAPEL_HIGPERSONAL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'COCA_COLA', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'CREME_DE_LEITE_NESTLE', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'MAIONESE_HELLMANN_S', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'PAPEL_HIGPERSONAL', 'ACHOCNESCAU' | 'LEITE_MOCA' |
| 'ARROZ_PRATO_FINO', 'BISCNESTLE', 'COCA_COLA' | 'LEITE_MOCA' |
| 'ARROZ_PRATO_FINO', 'CREME_DE_LEITE_NESTLE', 'COCA_COLA' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'COCA_COLA', 'BOMBRIL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'CREME_DE_LEITE_NESTLE', 'COCA_COLA' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'CREME_DE_LEITE_NESTLE', 'SPO_OMO' | 'COCA_COLA' |
| 'BISCNESTLE', 'FARTRIGO_RENATA', 'COCA_COLA' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'COCA_COLA', 'FEIJAO_BROTO_LEGAL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'COCA_COLA', 'GELATINA_ROYAL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'COCA_COLA', 'PAPEL_HIGPERSONAL' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'SAL_CISNE', 'COCA_COLA' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'COCA_COLA', 'SPO_OMO' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'FARTRIGO_RENATA', 'CREME_DE_LEITE_NESTLE' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'MACRENATA', 'CREME_DE_LEITE_NESTLE' | 'LEITE_MOCA' |
| 'BISCNESTLE', 'CREME_DE_LEITE_NESTLE', 'PAPEL_HIGPERSONAL' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'COCA_COLA', 'BOMBRIL' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'COCA_COLA', 'BOMBRIL' | 'LEITE_MOCA' |
| 'SAL_CISNE', 'COCA_COLA', 'BOMBRIL' | 'LEITE_MOCA' |
| 'FARTRIGO_RENATA', 'CREME_DE_LEITE_NESTLE', 'COCA_COLA' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'COCA_COLA', 'GELATINA_ROYAL' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'COCA_COLA', 'PAPEL_HIGNEVE' | 'LEITE_MOCA' |
| 'CREME_DE_LEITE_NESTLE', 'COCA_COLA', 'SPO_OMO' | 'LEITE_MOCA' |

3 Análise e discussão dos resultados

Analizando os resultados obtidos executando a rotina disponibilizada, foi escolhido se realizar um estudo sobre as regras de associações geradas com os seguintes parâmetros:

$$\text{Suporte} = 0.01 = 1\% \quad \text{Confiância} = 0.7 = 70\% \quad \text{Lift} = 3$$

Perceba que esses valores foram escolhidos por, além de apresentarem um número razoável de regras geradas ($153 \approx 9\%$ do número de transações no dataset), possuírem valores que o aluno acredita condizer com valores usados em estudos aplicados ao mundo real.

Analizando as regras geradas, o aluno acredita que o *dataset* represente um período especial e fora da rotina comum de compra/venda do mercado de Dotto. Este resultado de análise surge devido ao fato de grande parte dos *itemsets* **A** implicarem majoritariamente no *itemset* **B** = {LEITE_MOÇA}.

Interessantemente, um caso similar foi estudo em sala de aula e dirige o aluno a conclusão de que, como no caso mencionado, este período de análise refere-se a um período de promoção do produto “leite de moça” no mercado de análise.

Note que este fato impede uma boa análise e distinção das regras de associação criadas entre associações triviais ou inovadoras, pois tendência a resposta do algoritmo para casos que incluem esse evento especial. Porém, reitera-se que ainda é possível se realizar uma análise sobre os dados, mas isso necessitaria um maior tempo de análise tanto para ajuste dos parâmetros quanto para análise (manual) das associações criadas.

4 Conclusão

Neste documento, explora-se o desenvolvimento e aplicação de algoritmos para geração de regras de associação sobre conjuntos de dados reais referente a um mercado.

A partir do pré-processamento de uma base de dados fornecida para o projeto, foi implementado o algoritmo Apriori para Geração de Regras de Associação (AR-Apriori) e extraído, sobre parâmetros julgados relevantes, as regras de associações encontradas no conjunto de dados. Após análise, acredita-se que o conjunto de dados fornecido apresenta um caso especial de estudo pois as regras de associações encontradas possuem *itemsets* de base que implicam em um mesmo *itemset* de adição (leite de moça) – levando o aluno a crer que os dados referem-se a um período de promoção do item em questão. Este fato torna a análise de casos de importância (i.e., regras de associações que trazem resultados inovadores) mais difícil devido a criação de regras “inesperadas”.

Por fim, entende-se que o algoritmo implementado é importante para estudos de casos genéricos. Possuindo fácil e rápida implementação na linguagem escolhida, acredita-se que este possa implicar em grandes melhoras para vários processos que necessitam do respaldo e criação de regras de associação (e.g., mercados e otimização de estoques) para otimização, necessitando somente de uma análise cuidadosa e longa sobre o conjunto de dados e período estudado para classificação das regras geradas quanto a sua representatividade e inovação.