

Lista 5 - Inteligência Artificial

Nome: **Victor Ferraz de Moraes**

Matrícula: **802371**

Questão 1)

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

Suporte Mínimo: 0,3

Confiança: 0,8

ItemSet 1:

Leite: $2/10 = 0,2$

Café: $3/10 = 0,3$

Cerveja: $2/10 = 0,2$

Pão: $5/10 = 0,5$

Manteiga: $5/10 = 0,5$

Arroz: $2/10 = 0,2$

Feijão: $2/10 = 0,2$

Portanto, somente os itens seguintes serão selecionados:

Café: $3/10 = 0,3$

Pão: $5/10 = 0,5$

Manteiga: $5/10 = 0,5$

ItemSet 2:

Café e Pão: $3/10 = 0,3$

Café e Manteiga: $3/10 = 0,3$

Pão e Manteiga: $4/10 = 0,4$

ItemSet 3:

Café, Pão e Manteiga: $3/10 = 0,3$

Regras ItemSet 2 (Calculando Confiança):

Café -> Pão: $3/3 = 1$

Pão -> Café: $3/5 = 0,6$

Café -> Manteiga: $3/3 = 1$

Manteiga -> Café: $3/5 = 0,6$

Pão -> Manteiga: $4/5 = 0,8$

Manteiga -> Pão: $4/5 = 0,8$

Portanto, as regras do ItemSet 2 selecionadas será:

Café -> Pão: $3/3 = 1$

Café -> Manteiga: $3/3 = 1$

Pão -> Manteiga: $4/5 = 0,8$

Manteiga -> Pão: $4/5 = 0,8$

Regras ItemSet 3 (Calculando Confiança):

Café e Pão -> Manteiga: $3/3 = 1$

Café e Manteiga -> Pão: $3/3 = 1$

Pão e Manteiga -> Café: $3/4 = 0,75$

Café -> Pão e Manteiga: $3/3 = 1$

Pão -> Café e Manteiga: $3/5 = 0,6$

Manteiga -> Café e pão: $3/5 = 0,6$

Portanto, as regras do ItemSet 3 selecionadas será:

Café e Pão -> Manteiga: $3/3 = 1$

Café e Manteiga -> Pão: $3/3 = 1$

Café -> Pão e Manteiga: $3/3 = 1$

Conclusão:

As regras que atendem os requisitos de Confiança e Suporte são:

Café -> Pão
Café -> Manteiga
Pão -> Manteiga
Manteiga -> Pão
Café e Pão -> Manteiga
Café e Manteiga -> Pão
Café -> Pão e Manteiga

Calculando o Lift:

Lembrando que $\text{Lift} = \text{Confiança}(A \rightarrow B) / \text{Suporte}(B)$

Regra	Confiança	Lift
Café -> Pão	$3/3 = 1$	$1/0,5 = 2$
Café -> Manteiga	$3/3 = 1$	$1/0,5 = 2$
Pão -> Manteiga	$4/5 = 0,8$	$0,8/0,5 = 1,6$
Manteiga -> Pão	$4/5 = 0,8$	$0,8/0,5 = 1,6$
Café e Pão -> Manteiga	$3/3 = 1$	$1/0,5 = 2$
Café e Manteiga -> Pão	$3/3 = 1$	$1/0,5 = 2$
Café -> Pão e Manteiga	$3/3 = 1$	$1/0,4 = 2,5$

Questões 2/3/4)

 Apriori.ipynb

Questão 5)

A biblioteca funciona da seguinte maneira: antes de iniciar a geração de regras de associação, precisamos tratar os dados para que estejam em um formato binário. Isso significa que cada linha representa uma transação e cada coluna representa um item, com 1 indicando a presença do item na transação e 0 sua ausência. Podemos utilizar o Pandas para manipular os dados antes de aplicá-los no algoritmo.

A função apriori da biblioteca identifica itemsets com base em um suporte mínimo. Ela utiliza o algoritmo Apriori, eliminando aqueles que não atendem ao suporte mínimo.

Os parâmetros da Função apriori são:

- **df:** DataFrame contendo os dados no formato binário.
- **min_support:** Especifica o suporte mínimo (entre 0 e 1). Exemplo: min_support=0.5 significa que o itemset deve aparecer em pelo menos 50% das transações.
- **use_colnames:** Se True, usa os nomes das colunas ao invés de índices numéricos.
- **max_len:** Limita o comprimento do itemset.

Exemplo:

```
// Gerando itemsets frequentes com suporte mínimo de 0.1
frequent_itemsets = apriori(df, min_support=0.1, use_colnames=True)
print(frequent_itemsets)
```

As métricas utilizadas em association_rules são:

Support (Suporte): Frequência do itemset em todas as transações.

Confidence (Confiança): Probabilidade de ocorrer o consequente, dado o antecedente.

Lift: Medida de correlação entre antecedente e consequente (valores >1 indicam correlação positiva).

Leverage: Diferença entre o suporte observado e o esperado (indica se a relação é maior do que o esperado pelo acaso).

Conviction: Indicador da força da regra (quanto maior, mais forte é a regra).

Com tudo isso em mente, um código utilizando a biblioteca poderia ficar da seguinte maneira:

```
# Preparando os dados
data = {'leite': [1, 1, 1, 0, 0],
        'pão': [1, 1, 0, 1, 1],
        'manteiga': [0, 1, 0, 1, 1],
        'cerveja': [1, 0, 1, 1, 0]}
df = pd.DataFrame(data)

# Gerando itemsets frequentes
itemSets_frequentes = apriori(df, min_support=0.1, use_colnames=True)

# Extraindo regras de associação
regras = association_rules(itemSets_frequentes, metric="confidence", min_threshold=0.7)
print(regras[['antecedents', 'consequents', 'support', 'confidence', 'lift']])
```

Questão 6)

O artigo começa estabelecendo o contexto da mineração de regras de associação (ARM) como uma técnica essencial na mineração de dados, especialmente em um cenário onde grandes volumes de dados estão disponíveis. Os autores discutem a importância da ARM para descobrir relações significativas entre itens em conjuntos de dados transacionais. Eles ressaltam que, apesar da eficácia dos algoritmos de mineração, como o Apriori, a visualização das regras geradas é crucial para a interpretação e compreensão dos resultados. A introdução também destaca a necessidade de uma revisão abrangente das técnicas de visualização existentes, dada a fragmentação da literatura. Os objetivos da revisão são claramente definidos, incluindo a análise das metodologias, desafios e direções futuras para a visualização de ARM.

Em seguida, os autores fornecem uma visão geral do problema da mineração de regras de associação. Eles definem formalmente o problema, apresentando a notação matemática que descreve um conjunto de itens e um banco de dados de transações. A definição de uma regra de associação é apresentada como uma implicação entre dois conjuntos de itens, A e B , onde A representa o antecedente e B o consequente. É também introduzido as medidas de qualidade que avaliam a eficácia das regras, como suporte e confiança.

No terceiro capítulo do artigo, eles abordam a definição matemática da visualização de ARM, detalhando como as regras de associação podem ser representadas visualmente. Os autores discutem diferentes métodos de visualização e suas características, enfatizando a importância de representar as regras de maneira que facilite a interpretação. Eles exploram as vantagens e desvantagens de várias abordagens de visualização, como gráficos, histogramas e matrizes, e como cada uma pode ser aplicada para destacar diferentes aspectos das regras de associação.

O quarto capítulo se dedica a uma análise abrangente dos métodos tradicionais de visualização de regras de associação. Os autores categorizam as técnicas de visualização em diferentes grupos, destacando suas características, vantagens e desvantagens. Entre os métodos discutidos, estão gráficos de barras, diagramas de Venn, matrizes de co-ocorrência e gráficos de rede. Os gráficos de barras são apresentados como uma forma simples e eficaz de visualizar a frequência de regras, mas limitados na representação de múltiplas dimensões. Os diagramas de Venn, embora intuitivos, tornam-se complexos quando se tenta representar mais de três conjuntos. As matrizes de co-ocorrência são elogiadas por sua capacidade de mostrar relações entre itens, mas podem ser difíceis de interpretar em conjuntos de dados grandes. Os gráficos de rede, por outro lado, oferecem uma visualização dinâmica das interações entre itens, permitindo uma análise mais profunda das relações, mas exigem um maior esforço interpretativo.

No quinto capítulo, os autores exploram inovações recentes na visualização de regras de associação, destacando abordagens que vão além dos métodos tradicionais. Eles discutem o uso de técnicas de visualização baseadas em aprendizado de máquina, como mapas auto-organizáveis e algoritmos de clustering, que permitem uma representação mais eficaz de grandes volumes de dados.

Os mapas auto-organizáveis são apresentados como uma ferramenta poderosa para a visualização de dados complexos, permitindo que padrões emergentes sejam identificados de

forma mais clara. Os autores também mencionam a utilização de técnicas de visualização 3D, que oferecem uma nova dimensão para a análise de regras de associação, embora apresentem desafios em termos de interpretação e usabilidade. Além disso, é abordado a integração de visualizações com ferramentas de mineração de dados, permitindo que os usuários realizem análises em tempo real e ajustem suas visualizações com base nos resultados obtidos.

O sexto capítulo compila uma lista de sistemas e pacotes de software especializados em visualização de regras de associação. Os autores discutem várias ferramentas disponíveis no mercado, destacando suas características e funcionalidades. Entre os sistemas mencionados, estão ferramentas que oferecem visualizações interativas, suporte para grandes conjuntos de dados e integração com algoritmos de mineração de dados. Os autores enfatizam a importância de escolher a ferramenta certa com base nas necessidades específicas do usuário e no tipo de dados que estão sendo analisados. Eles também discutem a evolução das bibliotecas de visualização, como matplotlib em Python e ggplot2 em R, que permitem a personalização das visualizações, mas podem exigir um conhecimento técnico mais avançado. O sétimo capítulo aborda os principais desafios e limitações enfrentados na visualização de regras de associação. Os autores discutem como a complexidade dos dados e a quantidade massiva de regras geradas podem dificultar a interpretação e a extração de insights significativos. Um dos principais problemas destacados é a "explosão combinatória" de regras, onde o número de regras possíveis cresce exponencialmente com o aumento do número de itens e transações, tornando a visualização e a análise manual praticamente inviáveis.

Além disso, o capítulo menciona a dificuldade em representar visualmente múltiplas dimensões e atributos, o que pode levar a uma sobrecarga cognitiva para os usuários. Os autores também discutem a falta de interatividade em algumas visualizações, o que limita a capacidade dos usuários de explorar os dados de maneira dinâmica e intuitiva. A necessidade de desenvolver métodos que integrem visualizações interativas e que sejam adaptáveis a diferentes tipos de dados e usuários é enfatizada como uma área crítica para pesquisa futura.

No oitavo capítulo, os autores apresentam uma taxonomia abrangente dos métodos de visualização de regras de associação. Eles classificam as técnicas em categorias baseadas em diferentes critérios, como características das regras exibidas, foco da visualização e tipos de atributos. Essa estrutura taxonômica é útil para pesquisadores e praticantes, pois fornece um guia claro sobre as opções disponíveis e suas respectivas aplicações.

As categorias incluem métodos tradicionais, como gráficos de dispersão e matrizes de co-ocorrência, e métodos mais recentes, como visualizações baseadas em redes e técnicas de aprendizado de máquina. Os autores também discutem como cada método pode ser mais adequado para diferentes contextos e tipos de dados, destacando a importância de escolher a técnica de visualização certa para maximizar a compreensão e a eficácia da análise.

O nono capítulo fornece uma revisão detalhada dos trabalhos analisados ao longo do estudo. Os autores discutem as principais contribuições de cada trabalho, destacando as metodologias utilizadas, os resultados obtidos e as implicações para a prática de visualização de regras de associação.