

Prezados Alunos,

Neste relatório, vou detalhar as atividades que vocês deverão realizar para a construção de modelos associativos e preditivos utilizando um banco de dados fornecido. O objetivo é aplicar técnicas de mineração de dados para descobrir padrões interessantes e prever comportamentos dos clientes de um supermercado.

Descrição da Atividade

Vocês receberão um banco de dados categorizado que simula os dados de clientes de um supermercado, chamado "SuperMarketPlus". O banco de dados contém cinco variáveis categóricas:

1. **Faixa Etária:** Categorias de faixas etárias ('18-25', '26-35', '36-45', '46-55', '56-65', '66-75', '>75').
2. **Gênero:** 'Masculino', 'Feminino'.
3. **Frequência de Compras:** Frequência com que o cliente faz compras no supermercado ('Diária', 'Semanal', 'Quinzenal', 'Mensal').
4. **Valor Médio de Compra:** Faixas de valor médio gasto por compra ('<50', '50-100', '100-150', '150-200', '>200').
5. **Categoria de Produto Favorita:** Categoria de produto que o cliente mais compra ('Alimentos', 'Bebidas', 'Limpeza', 'Higiene Pessoal', 'Outros').

O arquivo CSV com o banco de dados está disponível aqui.

Objetivos da Atividade

1. **Modelos Associativos:**
 - Utilizar o algoritmo Apriori para descobrir regras de associação entre as variáveis.
 - Identificar as associações mais fortes e interpretá-las no contexto do supermercado.
2. **Modelos Preditivos:**
 - Construir um modelo de árvore de decisão para prever a categoria de produto favorita dos clientes com base nas demais variáveis.
 - Avaliar a performance do modelo de classificação e interpretar os resultados.

Passos para Realização da Atividade

1. Preparação dos Dados

- Importar o banco de dados para seu ambiente de trabalho.
- Verificar e tratar valores ausentes (se houver).
- Converter variáveis categóricas para o formato adequado, caso necessário.

2. Construção do Modelo Associativo

- Aplicar o algoritmo Apriori:

- Definir um suporte mínimo e confiança mínima para as regras de associação.
- Utilizar a biblioteca `mlxtend` ou similar para implementar o algoritmo.
- **Interpretar os Resultados:**
 - Identificar as principais regras de associação descobertas.
 - Discutir como essas regras podem ser utilizadas pelo supermercado para melhorar suas estratégias de marketing e atendimento ao cliente.

3. Construção do Modelo Preditivo

- **Preparar os dados para a classificação:**
 - Separar as variáveis independentes (Faixa Etária, Gênero, Frequência de Compras, Valor Médio de Compra) da variável dependente (Categoria de Produto Favorita).
 - Dividir o conjunto de dados em subconjuntos de treinamento e teste.
- **Construir e treinar o modelo de árvore de decisão:**
 - Utilizar a biblioteca `scikit-learn` ou similar.
 - Ajustar os hiperparâmetros para otimizar a performance do modelo.
- **Avaliar o Modelo:**
 - Calcular métricas de performance como precisão, recall e F1-score.
 - Visualizar a árvore de decisão e interpretar as regras extraídas pelo modelo.

4. Relatório Final

- **Documentar todo o processo:**
 - Explicar cada passo realizado, desde a preparação dos dados até a construção dos modelos.
 - Incluir gráficos e tabelas que ajudem a visualizar os resultados.
- **Conclusões:**
 - Discutir os insights obtidos a partir das regras de associação e do modelo preditivo.
 - Sugerir como o supermercado pode utilizar esses insights para melhorar suas estratégias de marketing e atendimento ao cliente.

Entrega

- O relatório completo deve ser enviado até [data de entrega].
- Certifiquem-se de que o relatório está bem estruturado e que todas as etapas foram claramente explicadas.
- Incluam todo o código utilizado em um anexo ou em um notebook Jupyter.

Recursos Adicionais

Para auxiliar na realização da atividade, vocês podem consultar os seguintes recursos:

- Documentação da biblioteca `mlxtend` para regras de associação:
`mlxtend.frequent_patterns`
- Documentação da biblioteca `scikit-learn` para árvores de decisão:
`sklearn.tree.DecisionTreeClassifier`

- Tutoriais e exemplos disponíveis na plataforma de aprendizado online da disciplina.

Procedimento para Realizar a Atividade

Preparação dos Dados

- **Importar o banco de dados** para seu ambiente de trabalho.
- **Verificar e tratar valores ausentes** (se houver).
- **Converter variáveis categóricas** para o formato adequado, caso necessário.

Construção do Modelo Associativo

1. **Aplicar o algoritmo Apriori:**
 - Definir um suporte mínimo e confiança mínima para as regras de associação.
 - Utilizar a biblioteca `mlxtend` ou similar para implementar o algoritmo.
2. **Interpretar os Resultados:**
 - Identificar as principais regras de associação descobertas.
 - Discutir como essas regras podem ser utilizadas pelo supermercado para melhorar suas estratégias de marketing e atendimento ao cliente.

Construção do Modelo Preditivo

1. **Preparar os dados para a classificação:**
 - Separar as variáveis independentes (Faixa Etária, Gênero, Frequência de Compras, Valor Médio de Compra) da variável dependente (Categoria de Produto Favorita).
 - Dividir o conjunto de dados em subconjuntos de treinamento e teste.
2. **Construir e treinar o modelo de árvore de decisão:**
 - Utilizar a biblioteca `scikit-learn` ou similar.
 - Ajustar os hiperparâmetros para otimizar a performance do modelo.
3. **Avaliar o Modelo:**
 - Calcular métricas de performance como precisão, recall e F1-score.
 - Visualizar a árvore de decisão e interpretar as regras extraídas pelo modelo.

Relatório Final

1. **Documentar todo o processo:**
 - Explicar cada passo realizado, desde a preparação dos dados até a construção dos modelos.
 - Incluir gráficos e tabelas que ajudem a visualizar os resultados.
2. **Conclusões:**
 - Discutir os insights obtidos a partir das regras de associação e do modelo preditivo.
 - Sugerir como o supermercado pode utilizar esses insights para melhorar suas estratégias de marketing e atendimento ao cliente.