

DADOS INTERMEDIÁRIO - Oficina 2

ALUNO: Herik Douglas Oliveira Reinaldo

Agora você vai praticar o que aprendeu neste conteúdo da aula de Fundamentos de Mineração de Dados. Imagine que você é um cientista de dados encarregado de resolver um problema complexo em um setor à sua escolha, como saúde, finanças, varejo ou outro de sua preferência.

Com base no que discutimos sobre Ciência de Dados, KDD, Mineração de Dados e o processo CRISP-DM, descreva como você aplicaria cada uma dessas etapas para abordar e solucionar esse problema.

Por exemplo, como você definiria o objetivo de negócio na fase de “Entendimento do Negócio”? Quais tipos de dados seriam necessários na fase de “Entendimento dos Dados” para alcançar esse objetivo? Que procedimentos você seguiria na “Preparação dos Dados” para garantir que eles estejam prontos para a análise? Quais algoritmos ou técnicas você usaria na “Modelagem” para extrair padrões e insights dos dados? E como você avaliaria os resultados para garantir que eles estão alinhados com os objetivos do negócio antes de implementá-los para gerar valor real?

Explique como cada conceito discutido se integra ao longo dessas fases e como sua aplicação prática no setor escolhido pode ajudar a resolver o problema de maneira eficaz e orientada por dados.

Inclua um exemplo de aplicação prática em um setor de sua escolha, explicando como cada um desses conceitos é utilizado para resolver um problema específico.

Setor: Saúde

Problema: Diagnóstico precoce do Câncer de Próstata Utilizando Mineração de Dados.

1. Entendimento do Negócio

Objetivo:

Melhorar o diagnóstico precoce do câncer de próstata por meio de análise de dados clínicos históricos, como idade, níveis de PSA, toque retal e histórico familiar. O objetivo principal é auxiliar médicos na tomada de decisão clínica mais rápida e precisa, reduzindo atrasos no diagnóstico.

2. Entendimento dos dados

Dados Necessários:

- Idade do paciente
- Nível de PSA e sua evolução
- Histórico familiar de câncer
- Exames anteriores e Biópsias se houver
- Confirmação de diagnóstico (variável - alvo)

3. Preparação dos Dados

Procedimentos:

- Remoção de registros inválidos
- Transformação de variáveis categóricas
- Normalização de variáveis numéricas
- Análise descritiva para entender a distribuição dos dados
- Tratamento de outliers
- Seleção de variáveis relevantes

4. Modelagem

Técnicas:

- Regressão logística
- Árvore de decisão
- KNN ou Naive Bayes
- Validação Cruzada
- Balanceamento de Classes

5. Avaliação

Métricas:

- Acurácia
- Precisão
- Recall
- F1 – Score
- Comparação entre modelos

6. Implantação

- Fornecer relatórios de apoio à decisão médica
- Adaptar a diferentes realidades hospitalares como por exemplo SUS, hospitais regionais, hospitais especializados em oncologia.

Temos nas etapas descritas acima os passos do CRISP-DM, mostrando como se trata de um caminho seguro na construção (planejamento, execução e avaliação) de conhecimento prático a partir da análise de dados. Dentro desse processo temos as técnicas de mineração que, depois de um processo cuidadoso de tratamento de dados, implicam em um passo de máxima importância na busca de padrões, tendências e informações úteis. Enfim a Ciência de Dados, além de abarcar os conceitos identificados anteriormente no projeto apresentado conta com, neste caso, o apoio dos conhecimentos médicos, evidenciando sua forte característica interdisciplinar.