

Projeto de Mineração de Dados

Entrega via classroom

Data limite: 08-02-2026



ATENÇÃO:

- O presente trabalho prático deve ser feito em grupo de 2 estudantes.
- A data limite é 08-02-2026 às 12:00 (sem tolerância).
- Resoluções copiadas implicará a anulação de ambos trabalhos e a respetiva reprovação na disciplina.
- O estudante pode pesquisar nas referências bibliográficas disponibilizada pelo professor, links indexados e outros materiais didáticos, se assim o fizer, deve referenciá-los.

Data: 08/02/2026 às 12H00



Não se esqueça de acessar o [link](#) em anexo para ocupar o seu tema antes de qualquer atividade. O tema pertence a quem ocupar primeiro.

Para garantir a variedade e profundidade, os projetos foram selecionados a partir de diferentes áreas, como saúde, finanças, comportamento do usuário, entre outros. Esses projetos cobrem uma ampla gama de áreas e utilizam diferentes técnicas de mineração de dados. As fontes de dados fornecem conjuntos de dados brutos que podem ser usados diretamente para análise ou treinamento de modelos.

1. Apresentação e Entrega dos PMD

Prezados(as) Estudantes,

Gostaria de formalizar as diretrizes para a **apresentação e entrega dos projetos de mineração de dados**, que deverão ser realizados em **duplas**. Abaixo estão as orientações e requisitos para a entrega e apresentação dos projetos, que refletem as expectativas acadêmicas para esta atividade.

Objetivo

O objetivo deste exercício é proporcionar uma experiência prática no uso de técnicas de mineração de dados e aprendizado de máquina, onde os alunos terão a oportunidade de aplicar metodologias analíticas a conjuntos de dados reais, realizando a implementação e análise de modelos preditivos ou descritivos.

Organização dos Grupos

Os projetos deverão ser realizados em **grupos de dois alunos**. Cada dupla será responsável por um projeto, sendo que ambos os membros devem estar envolvidos ativamente na implementação, análise e apresentação do projeto.

Etapas do Projeto

As etapas para o desenvolvimento do projeto incluem, mas não se limitam a:

1. Seleção e Preparação dos Dados:

- Baixar e pré-processar os dados brutos fornecidos (ou adquiridos, conforme especificado para cada projeto).
- Limpeza dos dados (remover ou tratar valores ausentes, normalização, transformação de variáveis, etc.).
- Definir claramente as variáveis independentes e dependentes para as tarefas de modelagem.

2. Análise Exploratória dos Dados:

- Realizar uma análise exploratória inicial (EDA) para entender as características dos dados, como distribuições, outliers e possíveis correlações entre as variáveis.
- Visualizar as distribuições dos dados e os padrões identificados.

3. Desenvolvimento e Implementação do Modelo:

- Escolher as técnicas de mineração de dados apropriadas (como algoritmos de classificação, regressão, agrupamento, etc.).
- Implementar o modelo de acordo com as metodologias indicadas no projeto.
- Ajustar os parâmetros e validar o desempenho do modelo, utilizando métricas apropriadas (como precisão, erro médio, matriz de confusão, etc.).

4. Análise e Interpretação dos Resultados:

- Avaliar a acurácia do modelo desenvolvido.
- Interpretar os resultados obtidos e discutir as implicações dos mesmos.
- Se necessário, realizar ajustes no modelo para melhorar sua performance.

5. Documentação:

- Redigir um relatório detalhado que descreva todo o processo do projeto, desde a escolha dos dados até a interpretação dos resultados finais.
- O relatório deverá ser claro, técnico e bem estruturado, com a explicação das escolhas metodológicas e análise dos resultados obtidos.

Entrega

A entrega final do projeto deverá incluir os seguintes artefatos:

1. Relatório Técnico:

- O relatório deve abordar de maneira detalhada todas as etapas do projeto, incluindo:
 - Descrição dos dados utilizados.
 - Técnicas de pré-processamento aplicadas.
 - Descrição do modelo implementado e justificativa de sua escolha.
 - Análise dos resultados obtidos, com gráficos e métricas de desempenho.
 - Discussão sobre possíveis melhorias ou limitações do modelo.

2. Código Fonte:

- O código fonte completo e comentado em **Python** (ou outra linguagem definida para o projeto), contendo a implementação do pré-processamento, construção do modelo e avaliação de resultados.
- Certifique-se de que o código seja bem estruturado, com explicações claras sobre o funcionamento de cada parte do processo.

3. Apresentação em PowerPoint:

- A dupla deverá preparar uma apresentação de **10 a 15 minutos** para demonstrar o trabalho realizado. A apresentação deve incluir:
 - Uma introdução ao problema e aos dados.
 - A metodologia escolhida para a solução.
 - Principais resultados obtidos e insights gerados.
 - Discussão sobre a validade dos resultados e possíveis melhorias.
 - Conclusões gerais do projeto.
- As apresentações serão realizadas para toda a turma, e cada dupla deverá estar preparada para responder a perguntas sobre o projeto.

Avaliação

A avaliação será feita com base nos seguintes critérios:

1. **Qualidade do Relatório Técnico:** Clareza e profundidade da análise e explicações.
2. **Desempenho do Modelo:** Acuracidade, interpretação dos resultados e justificativas apresentadas.
3. **Apresentação:** Clareza, concisão e qualidade da comunicação durante a apresentação.
4. **Código Fonte:** Organização e documentação do código, além da implementação correta dos algoritmos e técnicas.

Observações Finais

Reforçamos a importância do trabalho em equipe, onde ambos os membros devem colaborar ativamente em todas as etapas do projeto, desde a coleta e análise de dados até a implementação do modelo e a apresentação final. O trabalho colaborativo é um dos pontos-chave desta atividade.

Em caso de dúvidas ou dificuldades durante o desenvolvimento dos projetos, não hesitem em entrar em contato para orientação. O sucesso deste projeto depende da aplicação prática das técnicas aprendidas e da capacidade de comunicar de forma clara os resultados obtidos.

Aguardamos ansiosos pelas apresentações e pelo material de entrega, que deverão refletir o esforço e aprendizado ao longo do desenvolvimento do projeto.

Atenciosamente,
Bongo Cahisso
Docente, Mestre
ISPTEC, Mineração de Dados

2. Projetos

Atenção: A lista de projectos abaixo é uma sugestão para grupos que desejam trocar de tema. Dois grupos não podem ter o mesmo tema (isso resulta em reprovação).

1. Análise de Sentimentos de Comentários de Filmes

- **Objetivo:** Analisar o sentimento em comentários de filmes para prever a popularidade de filmes.
- **Técnicas:** Análise de Sentimentos, Processamento de Linguagem Natural (PNL).
- **Metodologias:** Classificação (positiva, negativa, neutra).
- **Artefatos:** Visualizações de sentimentos, modelo de classificação.
- **Link para dados:** [IMDB Movie Reviews Dataset](#)

2. Previsão de Demanda de Energia

- **Objetivo:** Prever a demanda de energia em diferentes períodos do dia para otimização de recursos.
- **Técnicas:** Regressão, Séries Temporais.
- **Metodologias:** Modelos ARIMA, Redes Neurais Recorrentes (RNN).
- **Artefatos:** Modelo de previsão, gráficos de séries temporais.
- **Link para dados:** [UCI Energy Consumption Data](#)

3. Detecção de Fraudes em Cartões de Crédito

- **Objetivo:** Detectar transações fraudulentas em um conjunto de dados de cartões de crédito.
 - **Técnicas:** Análise de Outliers, Classificação (SVM, Árvores de Decisão).
 - **Metodologias:** Detecção de Anomalias, Algoritmos Supervisionados.
 - **Artefatos:** Algoritmo de classificação, relatório de precisão.
 - **Link para dados:** [Kaggle - Credit Card Fraud Detection](#)
-

4. Previsão de Preços de Casas

- **Objetivo:** Prever os preços de imóveis com base em atributos como localização, tamanho e características.
 - **Técnicas:** Regressão Linear, Árvores de Decisão.
 - **Metodologias:** Modelos supervisionados, validação cruzada.
 - **Artefatos:** Modelo preditivo, gráficos de erro.
 - **Link para dados:** [Kaggle - House Prices](#)
-

5. Classificação de Imagens de Alimentos

- **Objetivo:** Classificar imagens de alimentos em diferentes categorias (ex. frutas, vegetais, etc.).
 - **Técnicas:** Redes Neurais Convolucionais (CNN), Transfer Learning.
 - **Metodologias:** Aprendizado Profundo, Classificação.
 - **Artefatos:** Modelo de classificação de imagens, matriz de confusão.
 - **Link para dados:** [Kaggle - Food101 Dataset](#)
-

6. Análise de Tweets para Tendências Sociais

- **Objetivo:** Analisar tweets para identificar tendências e tópicos populares.
 - **Técnicas:** Processamento de Linguagem Natural, Análise de Sentimentos, Topic Modeling.
 - **Metodologias:** TF-IDF, LDA (Latent Dirichlet Allocation).
 - **Artefatos:** Gráficos de tópicos, modelo de análise de sentimentos.
 - **Link para dados:** [Kaggle - Twitter Data](#)
-

7. Análise de Comportamento de Compras em E-commerce

- **Objetivo:** Estudar o comportamento de compras dos consumidores em um site de e-commerce.
 - **Técnicas:** Análise de Cluster, Recomendação, Associação.
 - **Metodologias:** Algoritmos de associação (Apriori, FP-growth).
 - **Artefatos:** Regras de associação, perfil do cliente.
 - **Link para dados:** [UCI Online Retail Dataset](#)
-

8. Classificação de Imagens Médicas para Diagnóstico

- **Objetivo:** Diagnosticar doenças com base em imagens médicas (ex. raios-X).
 - **Técnicas:** Redes Neurais Convolucionais (CNN).
 - **Metodologias:** Aprendizado Supervisionado, Classificação de Imagens.
 - **Artefatos:** Modelo classificador, métricas de precisão e recall.
 - **Link para dados:** [NIH Chest X-ray Dataset](#)
-

9. Análise de Dados de Trânsito Urbano

- **Objetivo:** Estudar padrões de tráfego para otimização do fluxo urbano.

- **Técnicas:** Séries Temporais, Análise de Cluster.
 - **Metodologias:** Modelagem preditiva, análise espacial.
 - **Artefatos:** Gráficos de tráfego, modelo preditivo de congestionamento.
 - **Link para dados:** [NYC Traffic Data](#)
-

10. Previsão de Séries Temporais de Vendas

- **Objetivo:** Prever as vendas de um produto em um determinado período.
 - **Técnicas:** Regressão, ARIMA, Redes Neurais.
 - **Metodologias:** Modelagem de Séries Temporais, Forecasting.
 - **Artefatos:** Previsões de vendas, erro de previsão.
 - **Link para dados:** [UCI Retail Data](#)
-

11. Análise de Dados de Saúde Pública

- **Objetivo:** Estudar a prevalência de doenças em diferentes regiões geográficas.
 - **Técnicas:** Regressão, Análise Espacial.
 - **Metodologias:** Modelagem Estatística, Geoprocessamento.
 - **Artefatos:** Mapas de prevalência, análise de correlação.
 - **Link para dados:** [Kaggle - Health Data](#)
-

12. Detecção de Padrões de Crédito Pessoal

- **Objetivo:** Analisar padrões de crédito e prever o risco de inadimplência.
 - **Técnicas:** Classificação, Árvores de Decisão.
 - **Metodologias:** Modelos de Risco de Crédito, Supervised Learning.
 - **Artefatos:** Modelo de risco de crédito, matriz de confusão.
 - **Link para dados:** [UCI Credit Approval Dataset](#)
-

13. Análise de Dados de Atividades de Usuários em Aplicativos

- **Objetivo:** Estudar o comportamento dos usuários em aplicativos móveis.
 - **Técnicas:** Análise de Cluster, Segmentação de Usuários.
 - **Metodologias:** Análise Comportamental, Machine Learning.
 - **Artefatos:** Segmentos de usuários, relatórios de comportamento.
 - **Link para dados:** [Kaggle - App Store Dataset](#)
-

14. Análise de Padrões de Consumo de Energia em Edifícios

- **Objetivo:** Identificar padrões de consumo energético em edifícios comerciais.
 - **Técnicas:** Análise de Séries Temporais, Classificação.
 - **Metodologias:** Machine Learning supervisionado.
 - **Artefatos:** Modelo preditivo de consumo, visualizações de padrões.
 - **Link para dados:** [UCI Building Energy Dataset](#)
-

15. Análise de Dados de Transações em Blockchain

- **Objetivo:** Analisar transações em uma rede blockchain para detectar padrões ou anomalias.
- **Técnicas:** Análise de Rede, Detecção de Anomalias.
- **Metodologias:** Análise de Grafos, Algoritmos de Anomalia.

- **Artefatos:** Detecção de padrões de transações, redes de transações.
 - **Link para dados:** [Bitcoin Blockchain Data](#)
-

16. Análise de Sentimentos em Avaliações de Produtos

- **Objetivo:** Analisar avaliações de produtos em e-commerce para identificar sentimentos dos consumidores (positivos, negativos, neutros).
 - **Técnicas:** Análise de Sentimentos, Processamento de Linguagem Natural (PNL).
 - **Metodologias:** Classificação de texto, Modelos supervisionados (ex: Naive Bayes, SVM).
 - **Artefatos:** Modelo de análise de sentimentos, gráficos de distribuição de sentimentos.
 - **Link para dados:** [Amazon Product Review Dataset](#)
-

17. Previsão de Rendimento de Culturas

- **Objetivo:** Prever o rendimento de culturas agrícolas com base em variáveis como clima, tipo de solo e uso de fertilizantes.
 - **Técnicas:** Regressão, Séries Temporais, Árvores de Decisão.
 - **Metodologias:** Previsão, Modelagem estatística, Análise de séries temporais.
 - **Artefatos:** Modelo preditivo, gráficos de previsão de rendimento.
 - **Link para dados:** [UCI Crop Data](#)
-

18. Análise de Tráfego de Sites Web

- **Objetivo:** Analisar padrões de tráfego de websites para prever picos de acessos e identificar comportamentos de usuários.
 - **Técnicas:** Análise de Séries Temporais, Clusterização, Algoritmos de Associação.
 - **Metodologias:** Análise de dados temporais, Análise preditiva.
 - **Artefatos:** Relatórios de tráfego, gráfico de tendências, segmentos de usuários.
 - **Link para dados:** [Web Traffic Dataset](#)
-

19. Detecção de Doenças Cardiovasculares

- **Objetivo:** Prever a probabilidade de ocorrência de doenças cardiovasculares com base em características como pressão arterial, colesterol e idade.
 - **Técnicas:** Classificação, Redes Neurais, Máquinas de Vetores de Suporte (SVM).
 - **Metodologias:** Classificação binária (doente/não, doente), Machine Learning supervisionado.
 - **Artefatos:** Modelo de previsão de doença, matriz de confusão, métricas de desempenho.
 - **Link para dados:** [UCI Heart Disease Dataset](#)
-

20. Análise de Comportamento de Usuários em Redes Sociais

- **Objetivo:** Analisar comportamentos e padrões de interação de usuários em redes sociais (curtidas, comentários, posts).
 - **Técnicas:** Análise de Redes Sociais, Análise de Texto, Clusterização.
 - **Metodologias:** Segmentação de usuários, extração de características de texto.
 - **Artefatos:** Visualizações de redes sociais, grupos de usuários, gráficos de atividade.
 - **Link para dados:** [Twitter Data - Kaggle](#)
-

3. Itens a serem entregues pelos grupos

Os estudantes devem entregar os seguintes itens para cada projeto de mineração de dados. Estes são os artefatos típicos e as apresentações que ajudam a comunicar o trabalho realizado:

1. [25%] Relatório Técnico Completo

O relatório deve ser bem estruturado e conter as seguintes seções:

- **Introdução:** Descrição do problema e objetivo do projeto.
- **Descrição dos Dados:** Explicação sobre os dados brutos, suas variáveis e como eles foram obtidos.
- **Pré-processamento dos Dados:** Descrição dos passos realizados para limpar e preparar os dados para análise.
- **Técnicas e Metodologias Utilizadas:** Detalhamento das técnicas de mineração de dados aplicadas, como algoritmos de classificação, regressão, etc.
- **Análise dos Resultados:** Avaliação do desempenho do modelo (ex. precisão, recall, F1-score) e insights extraídos dos dados.
- **Conclusão:** Resumo dos principais resultados, possíveis melhorias e futuras direções de pesquisa.
- **Referências:** Citações de artigos, livros ou fontes online utilizadas no desenvolvimento do projeto.

2. [35%] Código/Ficheiros de Implementação

- O código utilizado para a execução dos experimentos e modelagem dos dados. Isso pode ser entregue em arquivos .ipynb (Jupyter Notebooks) ou .py (Python scripts), dependendo da linguagem utilizada. O código deve ser bem documentado, com explicações claras sobre cada etapa do processo.

3. [5%] Apresentação em PowerPoint (ou equivalente)

Uma apresentação de **10-15 minutos** que cubra os principais pontos do projeto. A apresentação deve incluir:

- **Introdução** ao problema e objetivo.
- **Visão geral dos dados** e o processo de pré-processamento.
- **Metodologias e técnicas** utilizadas na análise.
- **Principais resultados** e insights.
- **Conclusões e possíveis melhorias** no modelo.
- **Demonstração rápida** do código (opcional, caso o tempo permita).
- **Perguntas e respostas** ao final.

4. [10%] Gráficos e Visualizações de Resultados

Gráficos que ajudem a ilustrar os resultados obtidos, como:

- **Matriz de confusão** (para problemas de classificação).
- **Curvas ROC e gráficos de precisão-recall** (para avaliar a performance do modelo).
- **Gráficos de dispersão, histogramas e boxplots** para explorar os dados.
- **Gráficos de séries temporais** (se aplicável).
- **Mapas de calor** (quando relevante para análise espacial).

5. [10%] Documentação de Instruções

- Instruções sobre como executar o código, incluindo dependências, como configurar o ambiente de desenvolvimento, bibliotecas necessárias, e qualquer outro detalhe relevante. Isso pode ser feito em um arquivo README.md ou documento equivalente.

6. [15%] Dashboard Interativa para Teste do Modelo e Visualização de Resultados

○ Funcionalidades da Dashboard:

1. Teste do Modelo Inteligente:

- A dashboard deve permitir que o utilizador insira novos dados (ou selecione dados de um conjunto de teste) e teste o modelo desenvolvido.
- O modelo deve fornecer uma previsão em tempo real, com base nos dados inseridos, mostrando os resultados gerados pelo modelo de forma clara.
- O feedback de desempenho do modelo, como **precisão**, **erro médio** ou outras métricas relevantes, deve ser apresentado após o teste.

2. Gráficos Resumo:

- A dashboard deve exibir gráficos interativos que mostrem de forma visual e intuitiva os resultados do modelo e as análises realizadas.

- Exemplos de gráficos podem incluir:
 - **Distribuições de variáveis:** para analisar as características dos dados.
 - **Matriz de Confusão:** se o modelo for de classificação, mostrando as previsões versus os resultados reais.
 - **Gráfico de Acurácia ou Erro ao Longo do Tempo:** caso o modelo tenha sido ajustado em diferentes fases ou iterações.
 - **Análise de Correlação:** para visualizar como as variáveis estão inter-relacionadas.
 - A dashboard deve ser interativa, permitindo ao utilizador selecionar variáveis ou parâmetros específicos para gerar gráficos e insights personalizados.
3. **Design da Dashboard:**
- A interface deve ser intuitiva e fácil de usar, permitindo que o utilizador realize testes no modelo e veja os resultados de forma clara e rápida.
 - A aparência e usabilidade devem ser consideradas, proporcionando uma experiência amigável para que qualquer utilizador, com ou sem conhecimento técnico, consiga compreender os resultados.