

# Aprendizado de máquina - atividade discente supervisionada 1

Prof. Mozart Hasse

## **LEIA ATENTAMENTE TODAS AS INSTRUÇÕES ATÉ O FINAL DA ÚLTIMA PÁGINA. CADA PALAVRA CONTA!**

Use exclusivamente os dados da base pública fornecida pelo professor. Descubra tipos, faixas de valores e distribuições através de análise exploratória. Dados a usar: Emissões de gases de efeito estufa no Brasil segundo a versão da basedosdados.org:

<https://basedosdados.org/dataset/9a22474f-a763-4431-8e3d-667908a1c7ab?table=104c6201-b0e7-47aa-b858-83252e2b149f> na aba download, botão download da tabela.

**Ajustes adicionais poderão ser necessários dependendo do algoritmo escolhido.**

**O objetivo da sua equipe é PREVER OS VALORES TOTAIS de TODAS AS LINHAS do campo emissao a partir das linhas dos demais campos fornecidos EXCLUSIVAMENTE QUANDO O gás é o Óxido Nitroso (N2O). Fica a seu critério ajustar formatos ou remover campos, desde que cada previsão possa ser mapeada para a linha de origem. Não é permitido enriquecer os dados com bases externas.**

**A solução deste problema DEVE OBRIGATORIAMENTE ser buscada usando um algoritmo de aprendizado de máquina que não envolva serviços externos nem algoritmos relacionados a redes neurais artificiais (execução 100% na máquina local).**

## Critérios de avaliação

**Organização e clareza do código: (40% da nota)**, incluindo comentários com a justificativa para as escolhas feitas na análise e na otimização. Durante a execução o código deve rodar EXCLUSIVAMENTE EM UMA MÁQUINA SEM CONEXÃO COM A INTERNET.

**Qualidade do tratamento de dados e análise exploratória: (30% da nota)** o que inclui tudo o que for feito antes da aplicação do algoritmo.

**Acurácia do modelo em relação ao baseline: (30% da nota):** O professor usará seu baseline com um algoritmo extremamente simplista e comparará os resultados medindo os quadrados das distâncias entre as previsões e os resultados concretos. A nota será proporcional ao ganho de desempenho na base de testes em relação à taxa de acertos do algoritmo de baseline.

## Instruções para entrega

**O trabalho deve ser entregue em UM arquivo ZIP contendo:**

- o notebook Python ou apenas o código-fonte necessário para executar o tratamento de dados, a análise exploratória e o algoritmo;
- O arquivo de saída CSV com os dados de teste fornecidos pelo professor, acrescidos de uma coluna com as respostas previstas pelo seu algoritmo para as linhas correspondentes. A única diferença entre o arquivo fornecido pelo professor e o resultado da equipe deve ser as colunas com as previsões.

## Observações gerais

O trabalho pode ser feito em equipes de até 4 alunos. A EQUIPE TODA É IGUALMENTE RESPONSÁVEL PELO SUCESSO DO TRABALHO.

CUIDADO: aqui está se avaliando tanto o resultado gerado quanto o código e algoritmo escolhidos. Apresente um código compreensível por todos os membros da equipe, especialmente quanto aos parâmetros escolhidos.

**Não use** este documento com ferramentas de IA pois o professor vê isso como uma atitude, hum... *contraproducente*. Ao pedir algo a uma IA, explique o que você precisa com as **SUAS** palavras.

**É TERMINANTEMENTE PROIBIDO compartilhar arquivos entre equipes. Qualquer tentativa de fazer isso implicará na atribuição de nota ZERO a TODOS os membros de TODAS as equipes envolvidas. Casos suspeitos passarão por prova de autoria, portanto todos os membros da equipe devem saber como o código funciona e os motivos de cada escolha feita.**

Cabe lembrar que compartilhar caminhos bem e mal sucedidos é permitido. O único cuidado é compartilhar também os **eventuais valores de parâmetros de configuração, que devem ter uma diferença de no mínimo 10% para mais ou para menos em pelo menos um parâmetro da implementação para cada equipe.**