

# Aprendizado de máquina - atividade discente supervisionada 2

Prof. Mozart Hasse

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

Use pelo menos os dados da base pública indicada pelo professor. Descubra tipos, faixas de valores e distribuições através de análise exploratória.

### **Base de Dados**

Dataset: **Top 100 Songs & Lyrics By Year (1959–2023)**

Fonte: Kaggle

Contém aproximadamente 6.500 músicas com letras completas. Link a seguir:

[https://www.kaggle.com/datasets/brianblakely/top-100-songs-and-lyrics-from-1959-to-2019?phase=FinishSSORegistration&returnUrl=/datasets/brianblakely/top-100-songs-and-lyrics-from-1959-to-2019/versions/3?resource=download&SSORegistrationToken=CfDJ8KvMat0eHzhGoPokVBGB7D26cb9ANijr9pEWDhFRnr8Qd6B2sv7Tmtap\\_nl6pkTDED1v3A0Wld-5wpxa7JM6DUQG-uxcLfN2lWwwokZuEoUfQWbn92PVpDoOiLTgxvBY0eqT58y1Ph4QDp8-q4mrRwbaJ05dgoGdfPn76mCCrE9IS5UmDcD9wnTraSIZjCPNeaNcd5WREaXDBbpgKDw-EzqJazak4-V8ssbQrtTYbucZFXinDCZKvSbiZTeTddEQfu6-0aXSdAIS\\_gw-eVHeLuOAFvcojkSeOG7z7XQx4VjID9Dhhs8mbDuTEp5BQPHsIdPawstQadgH48bx2ba4MIT16g&DisplayNa me=Mozart%20Hasse](https://www.kaggle.com/datasets/brianblakely/top-100-songs-and-lyrics-from-1959-to-2019?phase=FinishSSORegistration&returnUrl=/datasets/brianblakely/top-100-songs-and-lyrics-from-1959-to-2019/versions/3?resource=download&SSORegistrationToken=CfDJ8KvMat0eHzhGoPokVBGB7D26cb9ANijr9pEWDhFRnr8Qd6B2sv7Tmtap_nl6pkTDED1v3A0Wld-5wpxa7JM6DUQG-uxcLfN2lWwwokZuEoUfQWbn92PVpDoOiLTgxvBY0eqT58y1Ph4QDp8-q4mrRwbaJ05dgoGdfPn76mCCrE9IS5UmDcD9wnTraSIZjCPNeaNcd5WREaXDBbpgKDw-EzqJazak4-V8ssbQrtTYbucZFXinDCZKvSbiZTeTddEQfu6-0aXSdAIS_gw-eVHeLuOAFvcojkSeOG7z7XQx4VjID9Dhhs8mbDuTEp5BQPHsIdPawstQadgH48bx2ba4MIT16g&DisplayNa me=Mozart%20Hasse)

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

**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 e que aplique algoritmos de redes neurais artificiais (execução 100% na máquina local). O professor está ciente que isso pode comprometer significativamente o resultado concreto obtido no caso de LLMs, mas o que se está avaliando é predominantemente o método escolhido (exemplo: prompts) e todo o processo realizado.**

Cada equipe deve escolher UM dos temas a seguir:

- Detecção de conteúdo inapropriado: misoginia e/ou violência contra a mulher em Letras de Músicas Populares (1959–2023) (fonte inspiradora, [o link não vale como referência direta](#))
- Detecção de conteúdo inapropriado: depressão e/ou validação ou incitação ao suicídio em Letras de Músicas Populares (1959–2023) ([fonte inspiradora, o link não vale como referência direta](#))
- Detecção de conteúdo inapropriado: racismo, homofobia, discurso de ódio ou preconceito relacionado ao público LGBTQIAPN+ em Letras de Músicas Populares (1959–2023) ([fonte inspiradora, o link não vale como referência direta](#))
- Detecção de conteúdo inapropriado: estímulo ou validação de relacionamentos tóxicos em Letras de Músicas Populares (1959–2023) ([fonte inspiradora, o link não vale como referência direta](#))

## Objetivo

Aplicar técnicas de redes neurais e/ou LLMs (Large Language Models) para construir um modelo que **classifique músicas** (com base principalmente ou exclusivamente nas letras) de acordo com a presença de conteúdo vinculado ao tema escolhido pela equipe.

## Tarefa Principal

Criar um **modelo de classificação automática** que analisa letras de músicas e atribui uma **pontuação de intensidade do conteúdo inapropriado na letra da música**. A pontuação deve ser **contínua, com valor entre 0 e 1**, onde 0 descreve uma música que não tem ABSOLUTAMENTE NADA (nem insinuação) de

conteúdo inapropriado e 1 onde a letra descreva conteúdo flagrantemente criminoso com perigo à ordem pública de acordo com a legislação atual (talvez algumas músicas do século passado não tenham dificuldade em cair nesta classificação, dado o teórico avanço da sociedade no período).

## Etapas sugeridas para o projeto

### 1. Exploração do dataset

- Análise temporal: evolução da linguagem ao longo dos anos.
- Metodologia de classificação: escolha dos critérios, coerência da pontuação e intensidade, compatibilidade dos critérios com a pontuação atribuída às músicas classificadas manualmente, completude dos mecanismos de detecção de conteúdo inapropriado, etc.
- Frequência de palavras relacionadas a temas sensíveis.

### 2. Construção de um conjunto rotulado

- Curadoria manual: cada grupo deve rotular manualmente pelos próprios critérios ao menos 30 músicas.
- Uso de dicionários de termos ofensivos ou de qualquer forma violentos ou inadequados para apoio, incluindo sinônimos, expressões e analogias. Pode-se usar bases públicas específicas com estes termos para apoio.

### 3. Abordagem com modelos

Os alunos devem escolher ao menos uma dentre **as abordagens abaixo**:

- LLMs como serviço LOCAL (OLLAMA)** — com prompt engineering envolvendo modelos PEQUENOS (0.5B a 7B), de acordo com a capacidade da máquina, ao menos na versão final.
- Redes neurais recorrentes (RNNs, LSTM, GRU) ou CNNs para texto**
- Transformer simples (ex: DistilBERT)**
- Outra sugestão envolvendo redes neurais e validada pelo professor.**

### 4. Análise dos resultados

- Comparações entre anos: a linguagem e o teor nocivo das músicas piorou ou melhorou ao longo do tempo?
- Visualizações: ranking das músicas mais problemáticas, heatmaps por ano, nuvem de palavras...

## Resultados esperados

- Um **relatório técnico** explicando:
  - Escolhas metodológicas
  - Preprocessamento, bases complementares usadas e tratamentos específicos realizados

- Técnicas de rotulagem (**no mínimo uma referência** que CLARAMENTE apoie a escolha da gradação ou classificação, incluindo página ou parágrafo da fonte original)
- Abordagem de modelagem
- **Limitações éticas e técnicas, incluindo, mas não se limitando a:**
  - analogias não triviais ou limitações na pontuação ou contexto da letra que foram identificados pela equipe
  - ênfase dada pelo tom ou ritmo da música, que não tem como ser analisados nesta atividade
  - poder de interpretação de vocabulário e contexto do modelo e/ou LLM escolhidos
  - até onde é possível identificar conteúdo inadequado, destacando quais conteúdos não tem como ser identificados pelos critérios adotados
  - outras limitações identificadas pela equipe
- Um **ranking das músicas** NÃO-ROTULADAS MANUALMENTE com score mais alto no tema escolhido.
- Um notebook com a **apresentação final** (storytelling) das visualizações e insights de forma organizada e sustentado ou direcionando alguma conclusão.
- Todas as referências seguindo as normas da ABNT

## Critérios de avaliação

**Organização e clareza do código: (30% 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: (30% da nota)** o que inclui tudo o que for feito antes da aplicação do algoritmo e o notebook final com os achados e as conclusões.

**Coerência e funcionalidade do modelo (40% da nota):** O modelo deve demonstrar um mecanismo de classificação plausível e tão bem fundamentado quanto possível, capaz de diferenciar casos evidentes positivos ou negativos, o que deve ser feito usando-se como base dados rotulados pelos membros da equipe, com critérios encontrados em publicações recentes sobre o assunto e usadas nas referências e processos citados pela própria equipe.

## 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 eventual tratamento de dados, a análise exploratória e o algoritmo;
- OUTRO notebook python com gráficos demonstrando a análise da série histórica quanto aos fatores analisados, além de fatores ou bandas de destaque nos quesitos de avaliação/classificação propostos pela equipe.
- Documento PDF contendo o processo e metodologia usados, assim como as referências, TUDO seguindo as normas da ABNT.

## Observações gerais

O trabalho pode ser feito em equipes de até 6 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 (exceto os próprios dados indicados pelo professor). 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.**