

# ES413 - Classificador de Estilos Musicais

João Marcelo de Souza Ferreira - jmsf3  
Victor Pessoa Diniz - vpd

April 3, 2025

## 1 Introdução

Este relatório apresenta o desenvolvimento de um classificador de estilos musicais utilizando técnicas de processamento de sinais e aprendizado de máquina. O objetivo do projeto é extrair características relevantes de arquivos de áudio e treinar um modelo capaz de identificar automaticamente o estilo musical de uma música. O dataset utilizado é o GTZAN, composto por músicas de 10 gêneros diferentes.

## 2 Metodologia

### 2.1 Pré-processamento

No pré-processamento, o sinal de áudio é dividido em pequenos segmentos chamados *frames*, que permitem capturar a evolução temporal do áudio e facilitar a extração de características.

### 2.2 Extração de Características

Foram escolhidas características em três domínios distintos: tempo, frequência e tempo-frequência. Essas características foram selecionadas por parecerem significativas para os áudios analisados.

#### 2.2.1 Domínio do Tempo

- **Amplitude Enveloping:** Captura as variações globais da amplitude do sinal de áudio ao longo do tempo, representando a "forma" do áudio.
- **Signal Energy (RMS):** Representa a amplitude média do sinal, útil para identificar variações de intensidade ao longo do áudio.
- **Zero-Crossing Rate (ZCR):** Mede quantas vezes o sinal muda de polaridade (positivo para negativo ou vice-versa), indicando a percussividade do áudio.

### 2.2.2 Domínio da Frequência (Espectrais)

- **Spectral Centroid:** Indica o "centro de gravidade" do espectro de frequências, relacionado ao "brilho" do som.
- **Spectral Bandwidth:** Mede a dispersão das frequências ao redor do centroide, ajudando a diferenciar sons concentrados de sons mais espalhados.
- **Spectral Contrast:** Quantifica a diferença entre os picos e os vales no espectro de um sinal de áudio. Captura a variação de energia entre componentes harmônicos e não harmônicos.
- **Spectral Flatness:** Quantifica quão "tonal" ou "ruidoso" é o sinal. Valores próximos a 1 indicam sons parecidos com ruído, enquanto valores baixos sugerem tonalidades puras.
- **Spectral Rolloff:** Identifica a frequência abaixo da qual está concentrada 85% da energia do sinal, detectando a predominância de graves ou agudos.
- **Band Energy Ratio:** Quantifica a distribuição relativa de energia entre duas regiões de frequência de um sinal de áudio, divididas por uma frequência de corte.

### 2.2.3 Domínio do Tempo-Frequência

- **Coefficientes MFCC:** Simulam a percepção humana de som usando uma escala logarítmica (Mel). Os primeiros coeficientes capturam informações gerais de energia, enquanto os demais representam detalhes espectrais mais refinados.

## 2.3 Modelo de Classificação

Foram testados diferentes modelos para a classificação dos estilos musicais, incluindo:

- Regressão Logística
- Random Forest
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Gradient Boosting

Foram testados Regressão Logística, Random Forest, SVM, KNN e Gradient Boosting. O modelo SVM foi escolhido por maximizar a margem entre as classes e lidar bem com dados de alta dimensionalidade.

## 3 Resultados e Discussão

Os dois melhores modelos foram **SVM** e **Random Forest**, cujas acurácias são apresentadas na Tabela 1.

O SVM apresentou melhor desempenho devido à sua habilidade de encontrar hiperplanos ótimos de separação. A matriz de confusão, ilustrada na Figura 1, indica os erros mais comuns entre gêneros.

Modelo	Acurácia
SVM	78%
Random Forest	69%

Table 1: Acurácias dos modelos testados.

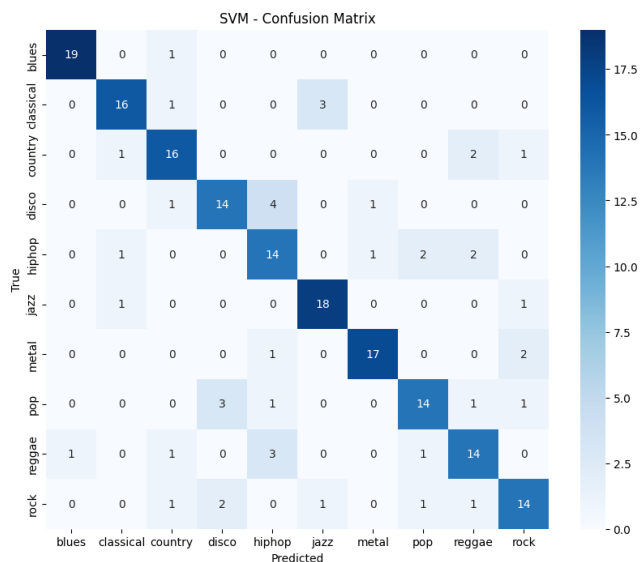


Figure 1: Matriz de confusão do classificador SVM.

## 4 Conclusão

O classificador de estilos musicais utilizou características extraídas de sinais de áudio para treinar modelos de aprendizado de máquina. O SVM obteve a melhor acurácia (78%), seguido pelo Random Forest (69%).

Os principais desafios foram a escolha de características e a limitação do dataset. Trabalhos futuros podem explorar redes neurais profundas e técnicas avançadas de engenharia de características para melhorar a precisão do modelo.