

Relatório da Análise de Faixas Musicais no Spotify

Resumo do Projeto

Este projeto teve como objetivo realizar a análise e categorização de faixas musicais a partir de dados extraídos via **API do Spotify for Developers**. As etapas principais incluíram:

1. **Coleta de Dados:** Extração de informações de cinquenta faixas, como duração, popularidade, nome do artista, nome do álbum e data de lançamento.
2. **Transformação de Dados:**
 - I. Conversão da duração das faixas para minutos.
 - II. Criação de categorias como curta, padrão e longa para a duração e baixa, moderada e alta para a popularidade.
 - III. Geração de uma nova variável alvo chamada **classe_faixa**, combinando as categorias acima.
3. **Pré-processamento:**
 - I. Remoção de classes com apenas uma ocorrência (evitando problemas de generalização).
 - II. Divisão entre treino e teste com estratificação.
 - III. Normalização dos dados com **StandardScaler**.
4. Treinamento com o classificador **Naive Bayes Gaussiano**.
5. Avaliação do modelo com métricas robustas como **Acurácia, Precision, Recall e F1-Score**.
6. Visualização de resultados com **Matriz de Confusão e Gráficos**.

Métricas de Avaliação

Relatório de Classificação:				
	precision	recall	f1-score	support
padrão_alta	1.00	0.83	0.91	6
padrão_baixa	1.00	1.00	1.00	2
padrão_moderada	0.67	1.00	0.80	2
accuracy			0.90	10
macro avg	0.89	0.94	0.90	10
weighted avg	0.93	0.90	0.91	10

Análise das Métricas

Acurácia

A **Acurácia** representa a proporção de acertos em relação ao total de previsões. Neste caso, o modelo acertou 9 em cada 10 (90%).

Precision

A **Precision** mede a quantidade de acertos entre as previsões feitas para uma classe específica:

1. **padrão_alta: 1.00** – todas as previsões feitas para esta classe estavam corretas.
2. **padrão_baixa: 1.00** – todas as previsões feitas para esta classe estavam corretas.
3. **padrão_moderada: 0.67** – houve erros ao prever essa classe, sinalizando possível confusão com outras.

Recall

O **Recall** avalia o quanto o modelo conseguiu identificar corretamente os exemplos reais de uma classe:

1. **padrão_alta: 0.83** – perdeu 1 instância.
2. **padrão_baixa e padrão_moderada: 1.00** – todas as instâncias foram detectadas corretamente.

F1-Score

O **F1-score** é uma média harmônica entre Precision e Recall — ideal para avaliar o equilíbrio entre eles. Os valores altos (acima de 0.80) indicam que o modelo mantém um bom desempenho de forma geral.

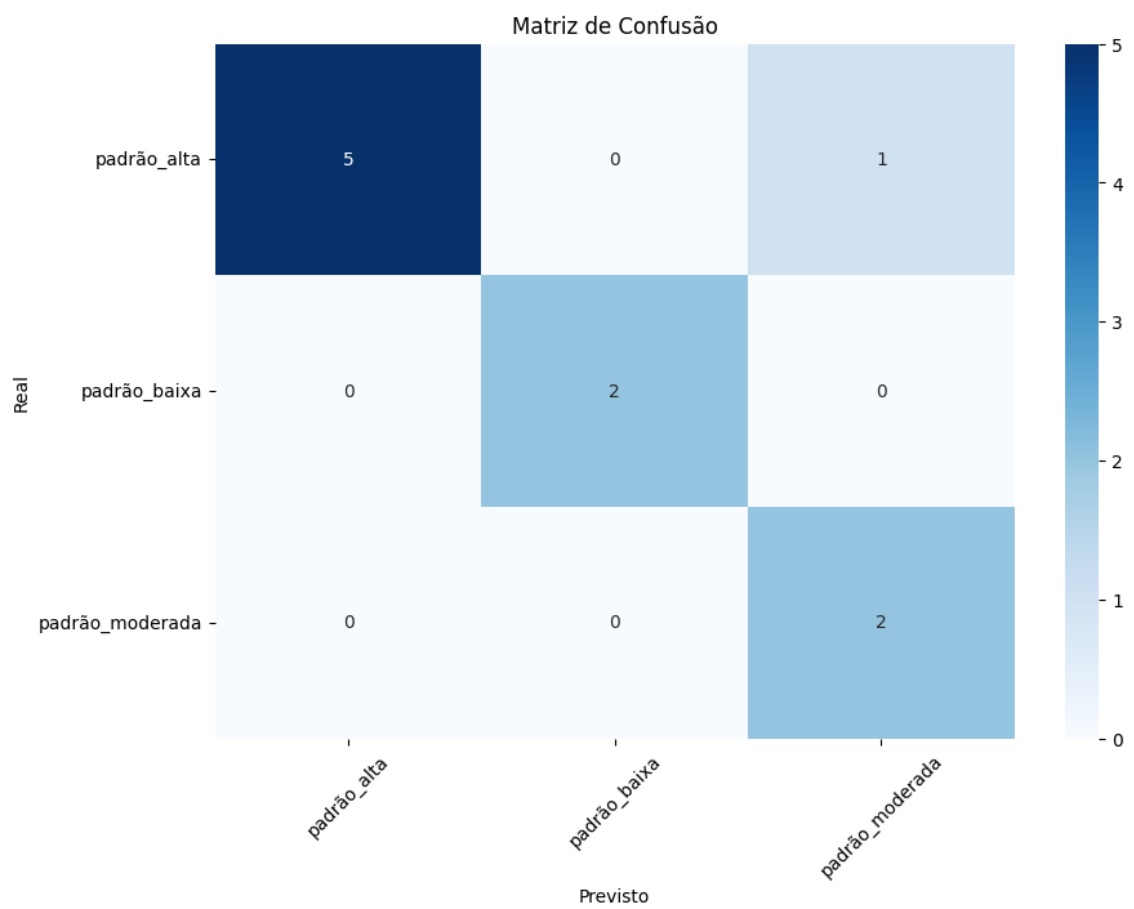
Média Ponderada vs. Média Macro

Tipo de Média	Descrição	Interpretação
Macro Avg	Média simples das métricas para cada classe (trata todas igualmente).	Ideal para comparar performance em classes desbalanceadas.
Weighted Avg	Média ponderada pelo número de exemplos em cada classe.	Reflete o desempenho geral levando em conta o tamanho das classes.

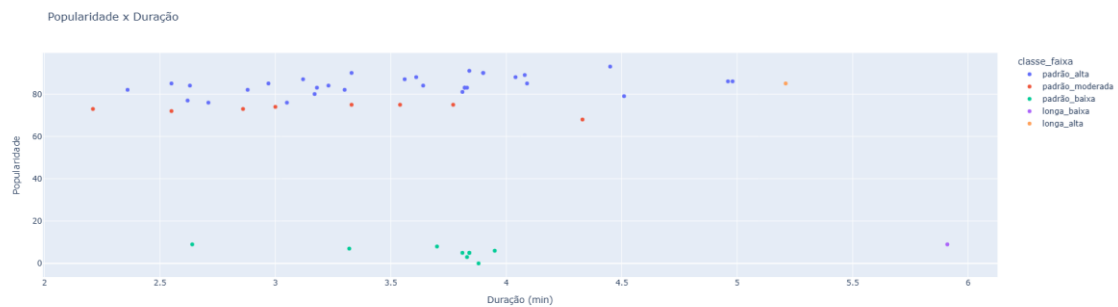
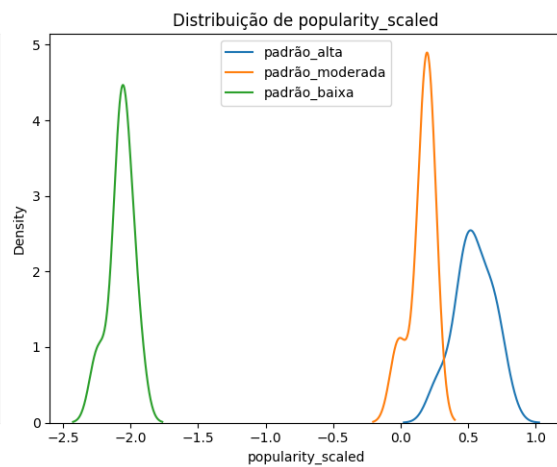
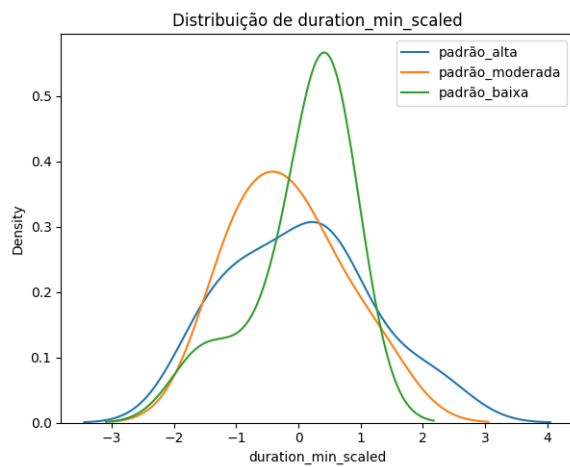
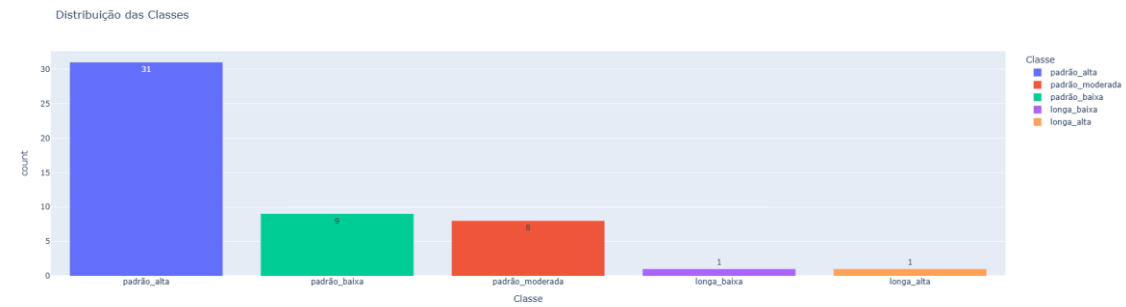
No projeto:

- 1. **Macro F1-score = 0.90** → O modelo tem desempenho semelhante entre as classes.
- 2. **Weighted F1-score = 0.91** → A média geral é ligeiramente influenciada pela predominância da classe padrão_alta.

Matriz de Confusão



Visualizações gerais e gráficos



Considerações Finais

1. O modelo **Naive Bayes** mostrou desempenho promissor, com **alta acurácia (90%)** e **excelente balanceamento entre precisão e recall** para a maioria das classes.
2. A **classe padrão_moderada** apresentou menor precision, sugerindo possíveis confusões — isso pode ser explorado em futuras versões do projeto com mais dados ou outros algoritmos.

3. A aplicação de pré-processamento, filtragem de classes raras e normalização foram fundamentais para garantir um desempenho consistente.
4. A **matriz de confusão** confirmou que o modelo raramente erra — e quando erra, geralmente é em classes com baixa representatividade.