

# Relatório de Neuro Análise - Classificação de Artistas

## 1. Descrição do Modelo:

Utilizamos o **Gaussian Naive Bayes (GNB)** para classificar os artistas em 8 categorias distintas de popularidade. O modelo foi treinado utilizando um conjunto de dados que contém informações sobre as características dos artistas e as suas respectivas classificações de popularidade.

## 2. Métricas de Avaliação:

Para avaliar a performance do modelo, utilizamos as seguintes métricas de classificação: **precisão, recall, f1-score e suporte**. A seguir, apresentamos um resumo detalhado das métricas para cada classe.

### 2.1 Relatório de Classificação:

Relatório de Classificação do Modelo Naive Bayes

	precision	recall	f1-score	support
Alta Popularidade	0.83	1.00	0.91	10
Artistas de Nicho	0.98	0.96	0.97	328
Descobertas Locais	0.99	0.97	0.98	230
Emergentes Globais	0.75	0.88	0.81	43
Estrelas Internacionais	0.75	0.75	0.75	8
Mainstream Regional	0.97	0.93	0.95	402
Nova Cena	0.89	0.92	0.91	92
Promissores	0.84	0.93	0.88	149
Superestrelas	0.00	0.00	0.00	2
accuracy			0.94	1264
macro avg	0.78	0.82	0.79	1264
weighted avg	0.94	0.94	0.94	1264

### 3. Interpretação das Métricas:

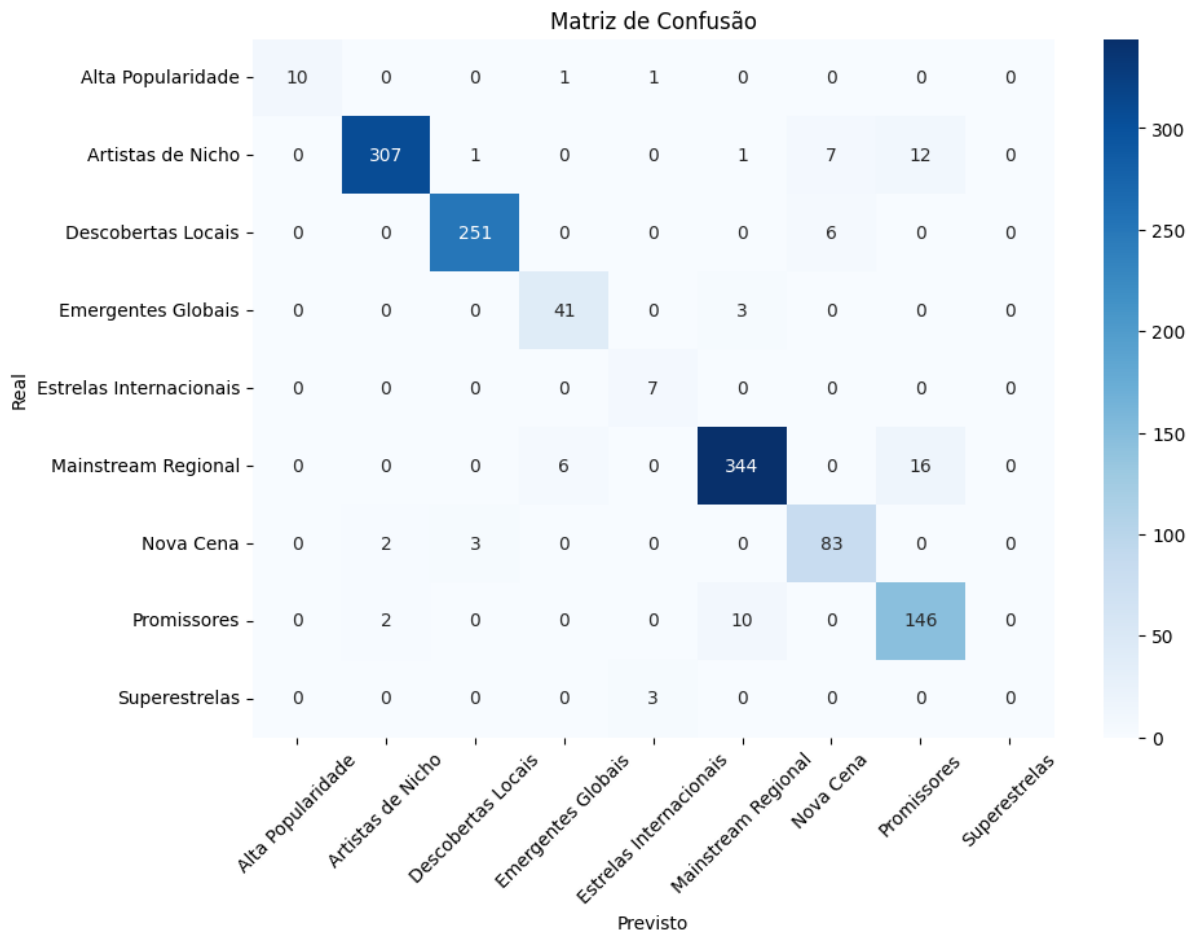
- **Alta Popularidade:** Embora a precisão para essa classe seja boa (0.83), o modelo obteve um **recall perfeito** (1.00), indicando que ele conseguiu identificar corretamente todos os casos de alta popularidade entre os poucos exemplos disponíveis.
- **Artistas de Nicho:** Esta classe obteve uma excelente performance, com **precisão** de 0.98 e **recall** de 0.96, o que significa que o modelo conseguiu identificar quase todos os artistas de nicho e classificou corretamente a maioria deles.
- **Descobertas Locais:** A classe "Descobertas Locais" também teve um ótimo desempenho, com **precisão** e **recall** superiores a 0.97, evidenciando que o modelo foi eficaz ao prever esse tipo de artista.
- **Superestrelas:** A classe "Superestrelas" apresentou um desempenho muito baixo, com **precisão** e **recall** de 0.00. Isso ocorre porque há poucos exemplos dessa classe no conjunto de dados (apenas 2 instâncias), o que dificulta a capacidade do modelo de prever corretamente essa classe.

### 4. Matriz de Confusão:

A **matriz de confusão** foi gerada para ilustrar como o modelo se comportou na classificação das diferentes faixas de popularidade. A matriz de confusão visualiza os erros cometidos, permitindo uma análise mais detalhada dos tipos de erros que o modelo fez.

**Matriz de Confusão:** A matriz de confusão do modelo, que mostra o número de previsões corretas e incorretas para cada classe, é apresentada na figura a seguir.

- As classes com alta taxa de erro podem ser analisadas para entender onde o modelo pode ser melhorado.



## 5. Insights:

- O modelo demonstrou um bom desempenho geral, com uma **acurácia** de 94%, o que indica que o modelo está classificando corretamente a maioria dos artistas nas faixas de popularidade.
- Algumas classes com poucos exemplos (como "Alta Popularidade" e "Superestrelas") mostraram menor desempenho, sugerindo que seria útil aumentar o número de exemplos dessas classes para melhorar a robustez do modelo.
- A classe "Superestrelas" apresentou um grande desafio para o modelo, provavelmente devido à baixa representatividade dessa classe nos dados de treinamento. A inclusão de mais exemplos pode ajudar a melhorar a classificação dessa classe.

## 6. Visualizações:

O gráfico da **distribuição das classes reais vs. previstas** foi criado para comparar as distribuições das classes no conjunto de dados real e no conjunto de previsões feitas pelo modelo. Esse gráfico ajudou a visualizar o equilíbrio entre as classes e possíveis áreas de melhoria.

