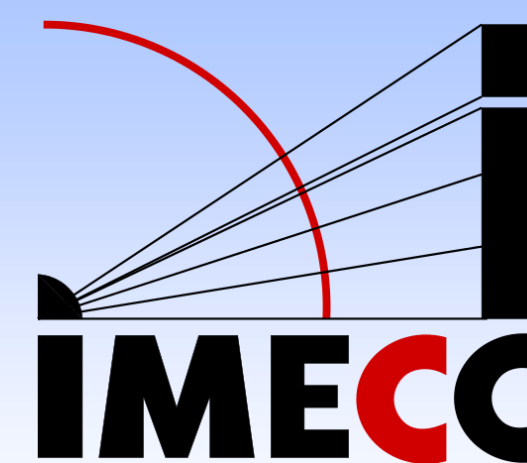




Bruno Peixoto, Beatriz Vasconcelos, Rodolfo Lindemute, Thaís Pietro  
Departamento de Estatística / IMECC - Unicamp



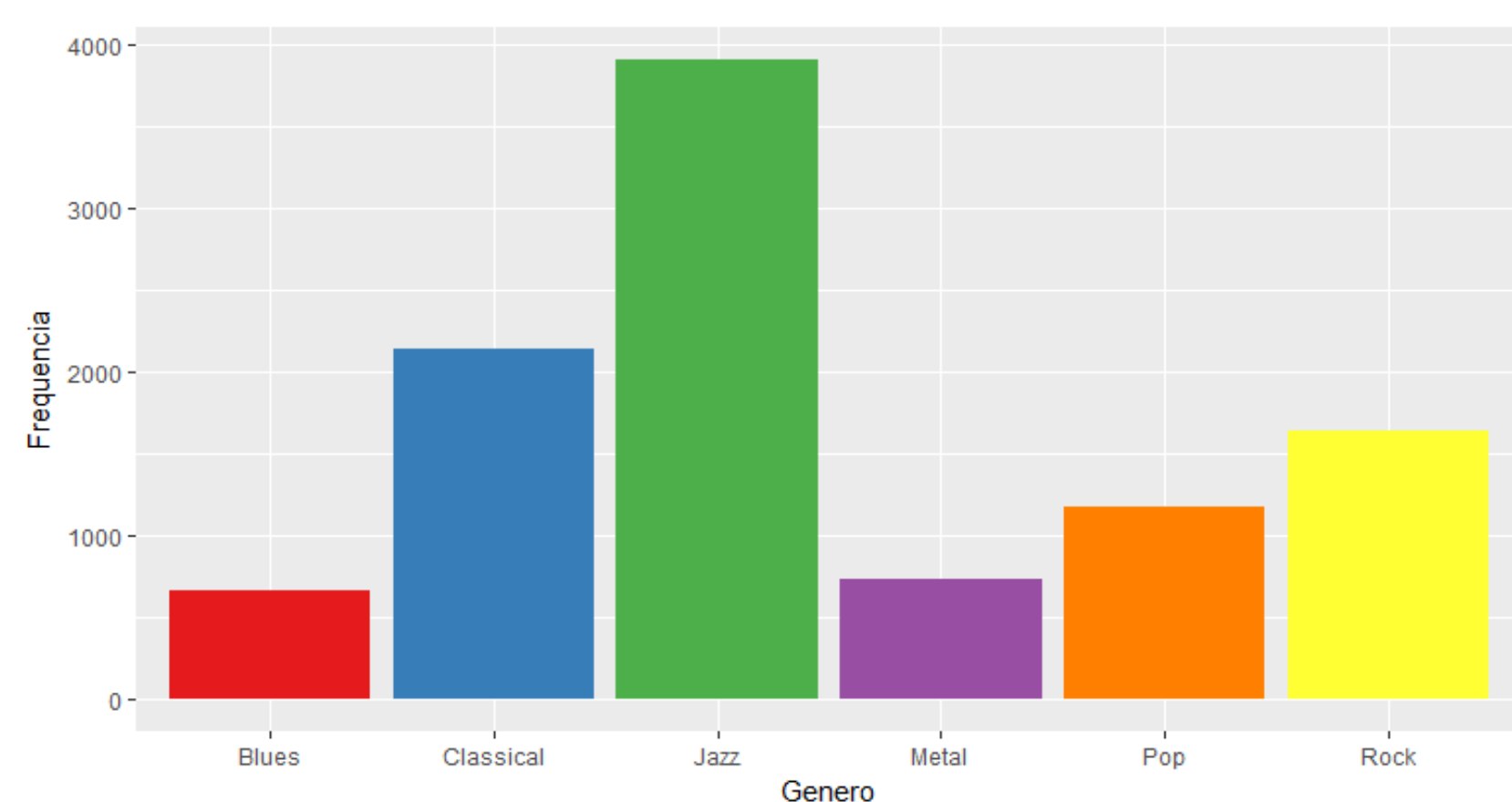
## Introdução

A classificação de músicas quanto ao seu gênero tem potencial de aplicação em diversos serviços de streaming, como no Spotify e no Youtube, por exemplo, essa classificação é utilizada na recomendação e gestão de playlists. O que possibilita a aplicação de técnicas de Machine Learning na classificação de áudio são parametrizações como as tratadas no trabalho de Kostek (2001). Oferecendo clareza e eficiência na descrição de dados baseado em quantidades compactas de informação. Neste trabalho utilizou-se o banco Music Genres (Kostek et. al, 2011) disponível na plataforma TunedIT. O banco contém fragmentos de músicas que foram classificados quanto ao gênero e parametrizados. O objetivo é classificar os gêneros musicais de cada fragmento, através técnicas de Machine learning, visando atingir a maior acurácia.

## Dados

191 variáveis preditoras referentes aos fragmentos de 15 a 20 músicas de 60 intérpretes pertencentes à 6 gêneros. Os dados foram divididos em bancos de treino e teste com 12495 e 10296 observações.

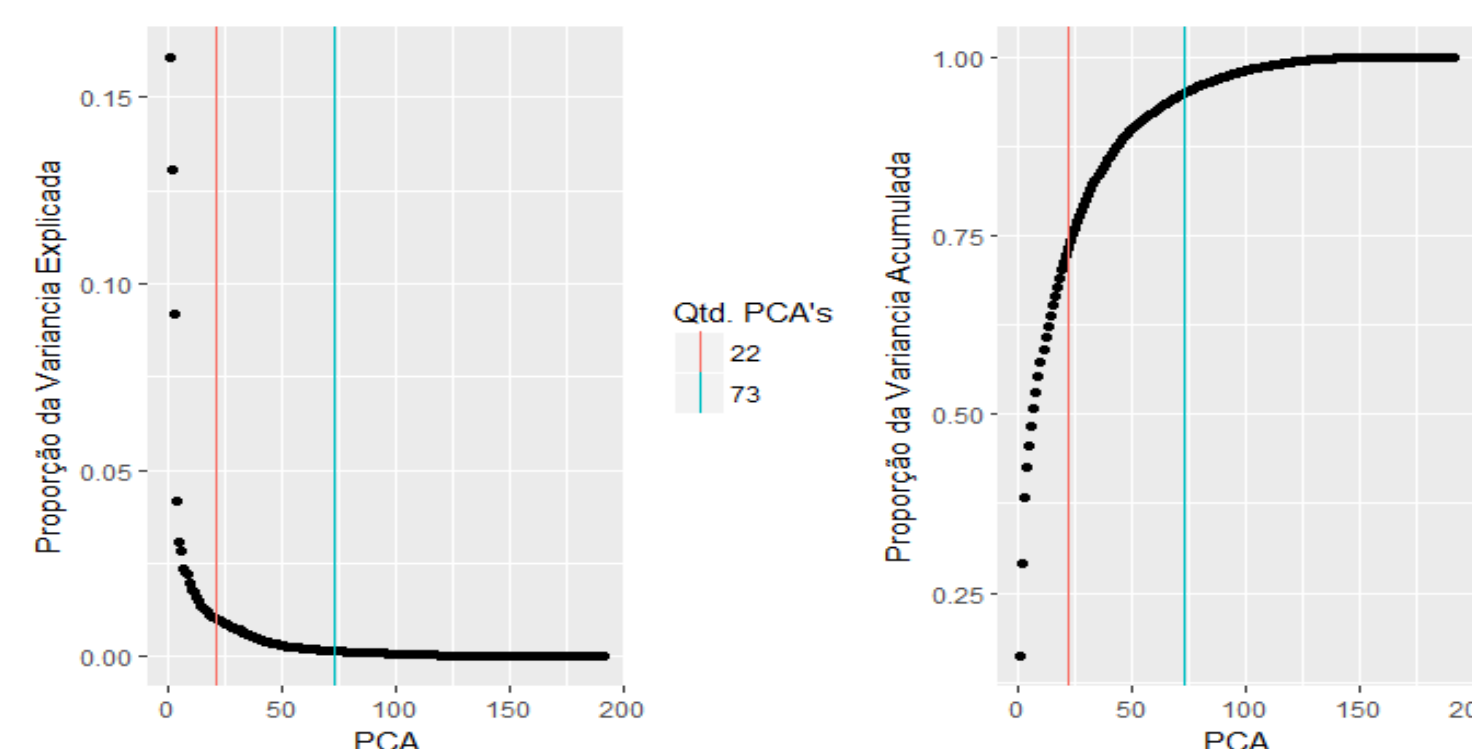
Frequência de observações por gênero



## Metodologia

### • PCA

São obtidas componentes principais, através de uma transformação linear nas variáveis, de forma que a primeira componente tenha a maior variância possível e as subsequentes tenham a maior variância possível desde que não correlacionadas com as anteriores. Assim o conjunto resultante concentra a variabilidade nas primeiras componentes.



### • LDA

Identifica uma combinação linear das características que descrevem ou separam duas ou mais classes de objetos ou eventos.

### • KNN

Método de classificação não-paramétrico que consiste em classificar uma nova observação a partir da classificação da maioria de seus k vizinhos mais próximos.

### • XGBoost

É uma técnica que combina Gradient Boosting com técnicas de Regularização. O Gradient Boosting ajusta uma série de modelos simples de forma que um modelo tenta diminuir o erro do anterior, minimizando uma função de perda dada, ao final todos os modelos são combinados. No XGBoost adiciona-se uma função de regularização a função perda, o que penaliza a complexidade do modelo, evitando-se sobreajuste.

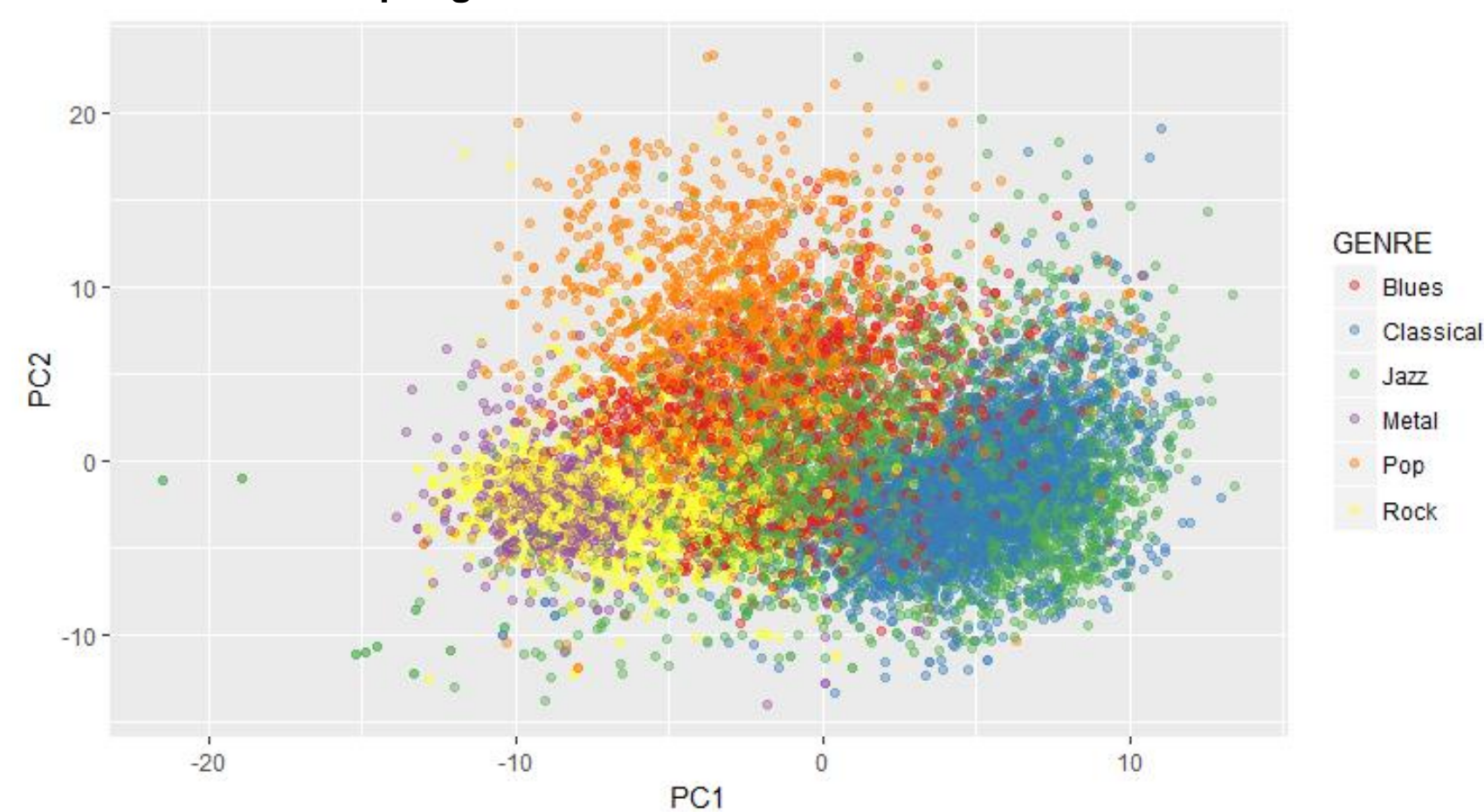
### • Floresta Aleatória

Envolve a implementação de múltiplas sub árvores, que combinadas fazem uma predição mais precisa a partir dos valores das variáveis preditoras.

## Resultados

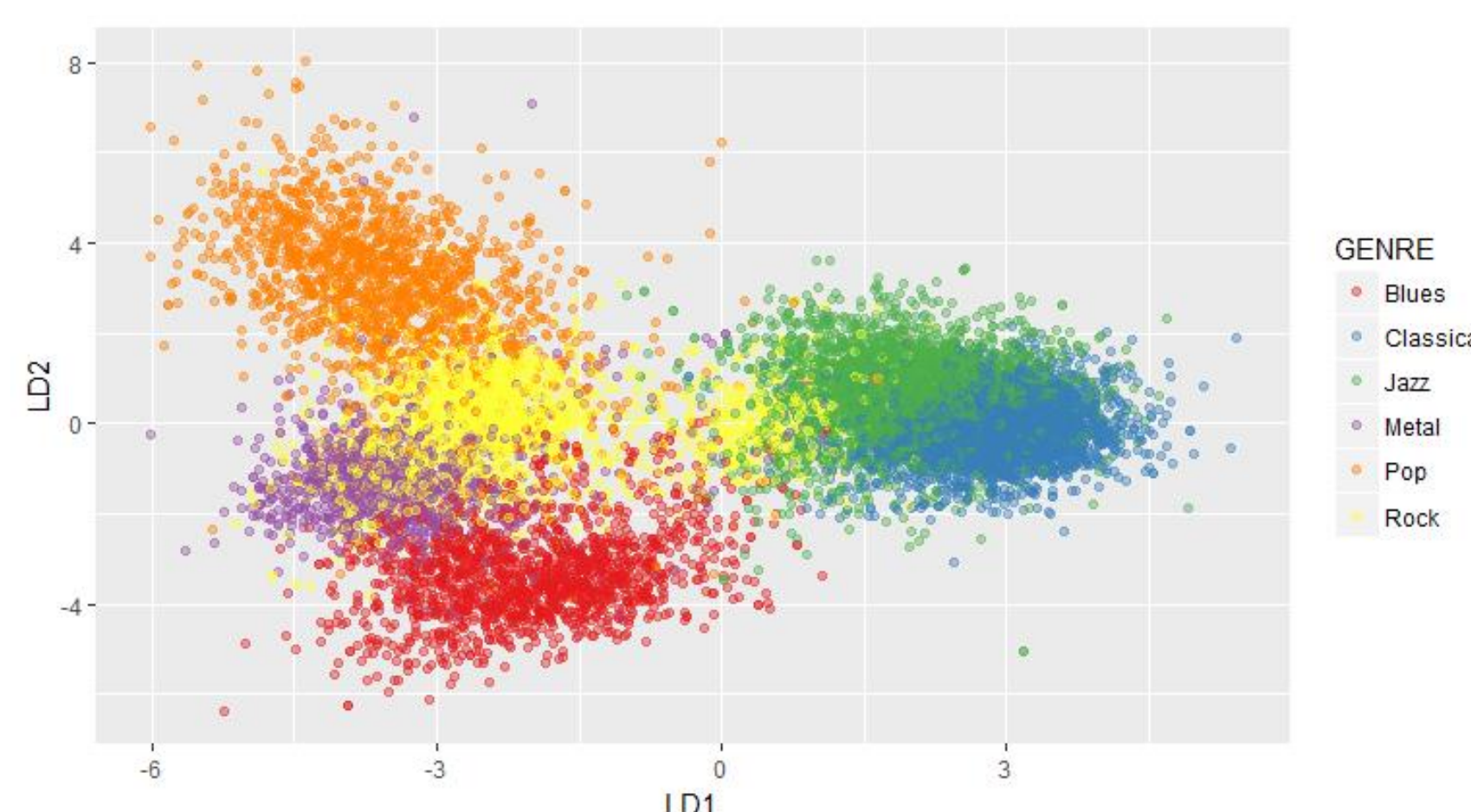
Inicialmente, foram aplicadas as técnicas de redução de dimensão por PCA e LDA.

Primeira e segunda componentes principais por gênero musical nos dados de treino



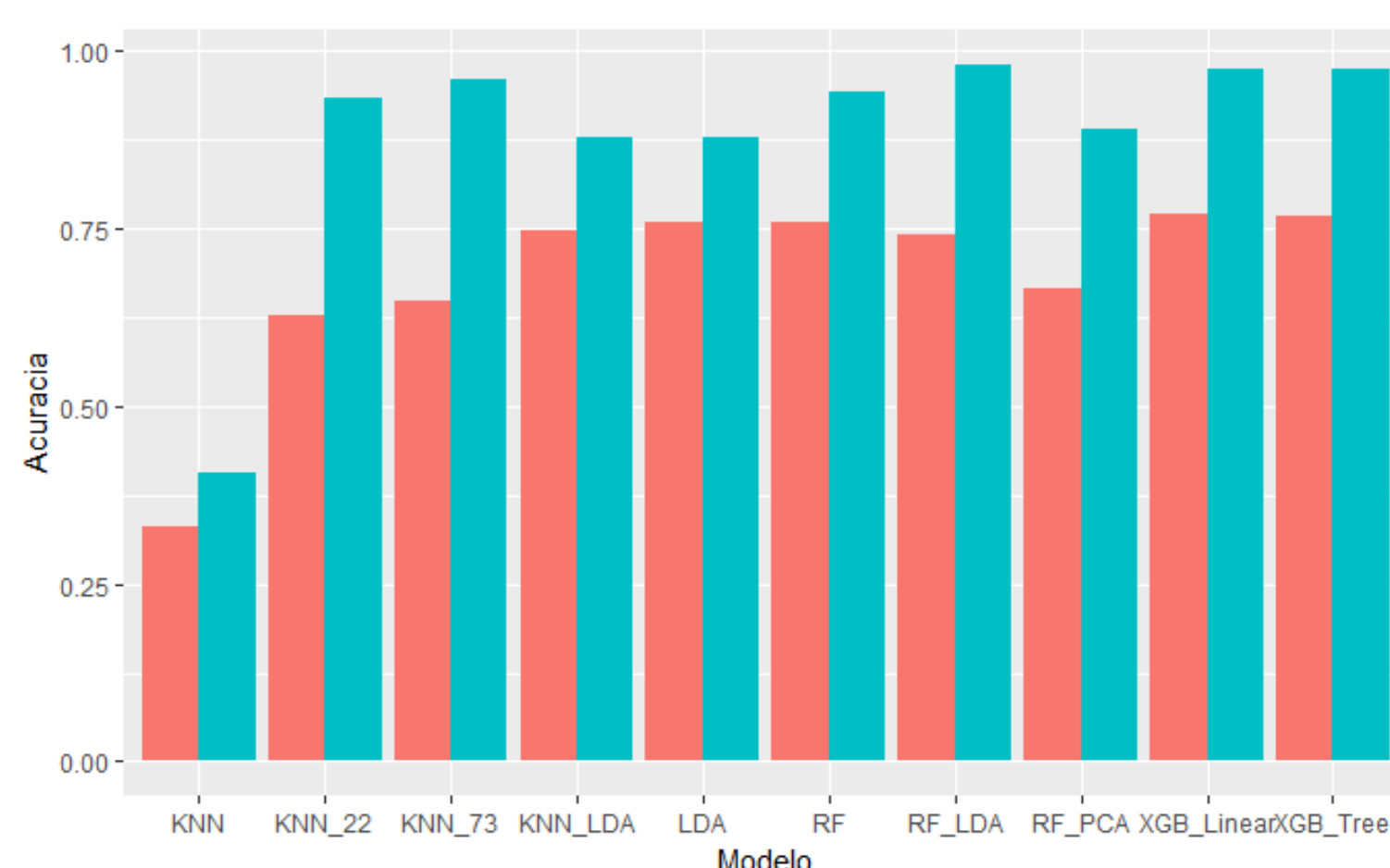
A Análise Discriminante Linear foi aplicada como técnica de redução de dimensão e também como classificador. Esse método de redução de dimensão apresentou resultados piores do que quando usado apenas como classificador.

Primeiro e segundo vetores da Análise Discriminante Linear



Observando a acurácia de cada modelo, a maioria dos ajustes obtiveram performances similares.

Acurácia dos modelos ajustados



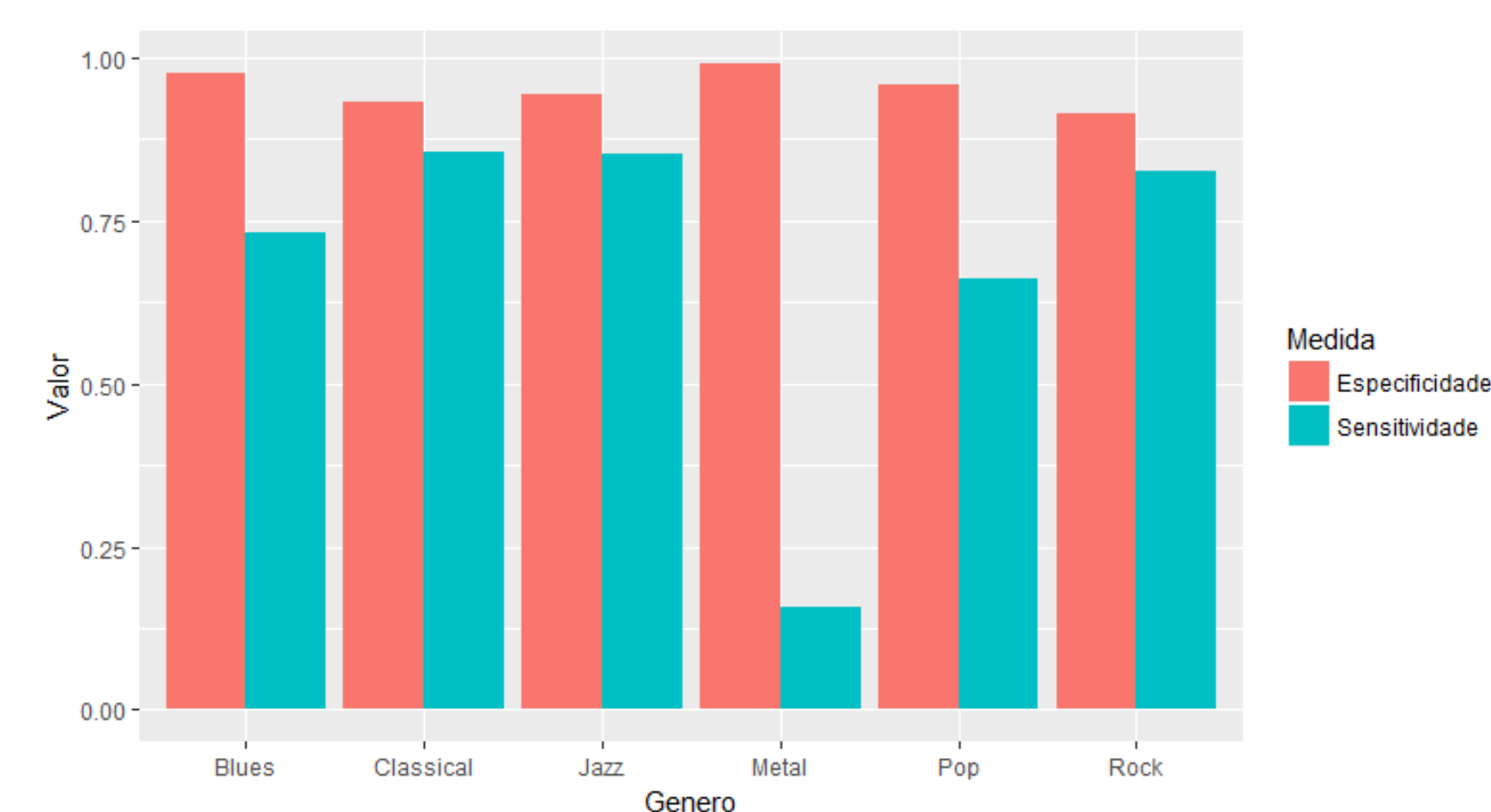
Modelo	Teste	Treino
KNN - 22 Componentes	0.63	0.93
KNN - 73 Componentes	0.65	0.96
KNN - LDA	0.75	0.88
KNN	0.33	0.40
XGBoost - Árvore	0.77	0.97
XGBoost - Linear	0.77	0.97
LDA	0.76	0.88
Floresta Aleatória	0.76	0.94
Floresta AL - PCA	0.67	0.89
Floresta AL - LDA	0.74	0.98

A grande maioria dos modelos obtiveram uma acurácia próxima de 90% para o banco de treino, com exceção do modelo KNN ajustado no banco sem redução de dimensão. Esse resultado evidencia a queda do desempenho à medida que a dimensionalidade cresce.

## Conclusões

O melhor modelo ajustado foi o XGBoost nos dados sem redução de dimensão, com acurácia de 76,89% nos dados de teste. Observa-se uma dificuldade de classificação para Metal, devido a proximidade desse gênero com o Rock.

Sensitividade e Especificidade por Gênero musical



O problema evidenciado com a sensibilidade do ajuste pode ser observado melhor pela matriz de confusão, onde pode-se verificar a má classificação dos gêneros, principalmente o Metal. A presença desse erro de classificação pode ser explicado pela quantidade reduzida de observações para esse gênero.

Matriz de Confusão para o modelo XGBoost nos dados de teste

	Blues	Classical	Jazz	Metal	Pop	Rock
Rock -	2	3	8	5	270	1350
Pop -	208	0	0	43	778	147
Metal -	22	0	2	115	97	499
Jazz -	16	532	3330	0	7	21
Classical -	0	1832	308	0	0	2
Blues -	491	13	37	37	8	96

A acurácia do modelo obtido com o XGBoost é satisfatória quando comparada com os resultados apresentados na competição TunedIT, de onde o banco de dados foi retirado. Considerando o resultado dos 144 participantes que submeteram os projetos, a acurácia obtida com esse modelo traria a 36ª posição na competição.

## Referências

Chen, T., He, T., "Higgs Boson Discovery with Boosted Trees" *JMLR: Workshop and Conference Proceedings*, 42:69-80, 2015

ISMIS 2011 Contest: Music Information Retrieval.  
Disponível em:  
<[http://tunedit.org/challenge/music-retrieval/genresfbclid=IwAR0PDNuUSyGs5xluV31lV\\_6yw5kzDLt63-leZlhoEpfBRQGp-ITGt1-ZNus](http://tunedit.org/challenge/music-retrieval/genresfbclid=IwAR0PDNuUSyGs5xluV31lV_6yw5kzDLt63-leZlhoEpfBRQGp-ITGt1-ZNus)>. Acesso em: 14/11/2018 às 19h47.

Kotek, B., Czyzewski, A., "Representing musical instrument sounds for their automatic classification", *Journal of the Audio Engineering Society. Audio Engineering Society*, 49(9):768-785 · September 2001