



# Algoritmos De Classificação Aplicados Em Recomendações De Faixas Para *Playlists*

Rodrigo Cavalcanti Loreto

30 de setembro de 2022



# Sumário

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

**1** Introdução

**2** Metodologia

**3** Resultados

**4** Conclusões

**5** Referências



# Introdução

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

A revolução digital na indústria musical afetou muito mais do que só as mídias através das quais consumimos música.



Figura: Coleção em mídia física.



# Introdução

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

As coletâneas tinham basicamente 2 limitadores: duração e catálogo da gravadora.

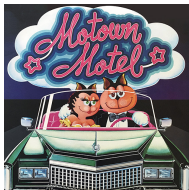


Figura: Coletâneas de diferentes temas em mídias diferentes.



# Introdução

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

As coletâneas ganharam um nome novo de “*Playlist*” e os “especialistas” agora são algoritmos.



**Figura:** Celular acessando a plataforma de streaming Spotify.



# Justificativa

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

Hoje, *playlists* têm pouquíssimas limitações e são uma das principais formas de divulgação da indústria da música.

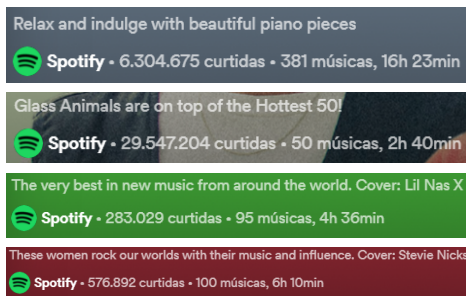


Figura: *Playlists* do Spotify sem limites de duração.



# Justificativa

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

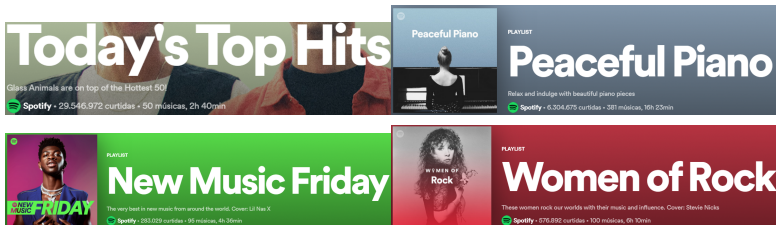


Figura: Alguns exemplos de *playlists*.



# Métodos

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

- Método paramétrico
  - Regressão Logística.
- Métodos não paramétricos
  - Árvores de Decisão.
  - *Random Forest*.





# Regressão Logística

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

No centro da regressão logística está a tarefa de estimar o log odds de um evento, ou seja, é uma função logística para modelar a probabilidade do sucesso. Matematicamente, a regressão logística estima uma função de regressão linear múltipla definida por:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (1)$$

Sendo  $\pi = P(Y = 1)$  e  $Y \sim \text{Bernoulli}(1, \pi)$



# Regressão Logística

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

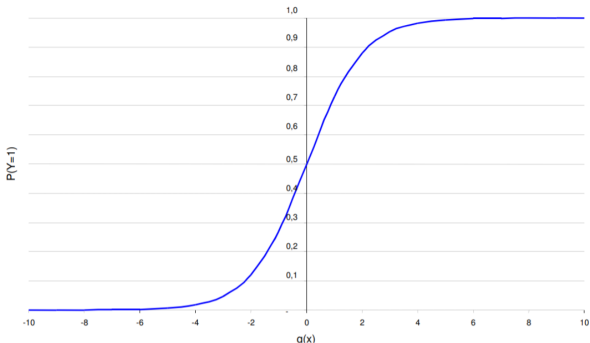


Figura: Exemplo de curva de regressão logística.

Fonte:(MIGUEL, 2020)



# Árvore de Decisão

Uma árvore de decisão é um modelo preditivo de aprendizado supervisionado que utiliza um conjunto de regras binárias (0,1) para calcular um alvo. Sua aplicação serve para classificação (variável-alvo categórica) ou para regressão (variável-alvo contínua), funcionando para variáveis de entrada e saída categóricas e contínuas. Sujeito a *overfit* com facilidade.

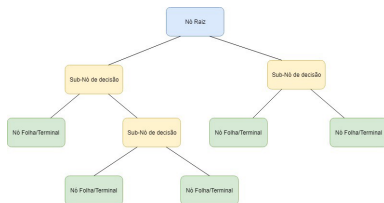


Figura: Exemplo de árvore de decisão.

(STANKEVIX, 2019)



# Random Forest

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

Combinação de preditores de árvores que controla o *overfit* através da aleatoriedade da seleção de variáveis das várias árvores.

Pode exercer a função tanto de classificação quanto de regressão. Este método neste trabalho foi utilizado apenas para classificação.



# *Data Splitting e Validação Cruzada*

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

- *Data Splitting*
- *K-folds*
- *Leave One Out*



# Seleção de *Playlists*

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

As *playlists* escolhidas para este estudo são editoriais sem personalização. Elas são criadas manualmente por editores do Spotify.

- Alone Again
- Beast Mode
- Life Sucks
- Piano Relaxante
- Power Hour
- Spooning



# Escolha dos pares

- 1 O primeiro par de *playlists* escolhido foi o de maior quantidade de músicas e que aparentou ser o mais distinto. As *playlists* escolhidas para esta primeira análise foram: Piano Relaxante e Beast Mode.
- 2 Em seguida, o par que indicou maior semelhança: Alone Again e Life Sucks.
- 3 E, como terceiro par, *playlists* de propostas nem muito diferentes nem muito similares. As *playlists* escolhidas foram: Power Hour e Spooning.
- 4 Por último, o desafio maior. Estudar as 6 *playlists* anteriores em conjunto.



# Variáveis

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

- O banco utilizado no estudo tem como variáveis as variáveis criadas e utilizadas pela plataforma Spotify;
- O banco de dados foi obtido através da ferramenta de API disponibilizada pelo próprio Spotify;
- A análise e tratamento dos dados foi feita com o software R;





# Variáveis utilizadas

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

Tabela: Variáveis agrupadas

Grupo	Variável	Tipo	Escala
"Mood"	Danceability	Float	Entre 0 e 1.
	Energy	Float	Entre 0 e 1.
	Tempo	Integer	Valores positivos.
	Valence	Float	Entre 0 e 1.
"Properties"	Instrumentalness	Float	Entre 0 e 1.
	Loudness	Float	Entre -60 e 0.
	Speechiness	Float	Entre 0 e 1.
"Context"	Acousticness	Float	Entre 0 e 1.
	Liveness	Float	Entre 0 e 1.



# Variáveis

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

Tabela: Variáveis sem grupo

Variável	Tipo	Escala
Key	Integer	Entre 0 e 11.
Mode	Integer	Entre 0 e 1.
Duration_ms	Integer	Valores positivos.
Popularity	Float	Entre 0 e 100.
Playlist	Factor	Categórica.

A variável *Playlist* foi criada para identificar a qual *playlist* a faixa pertence.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

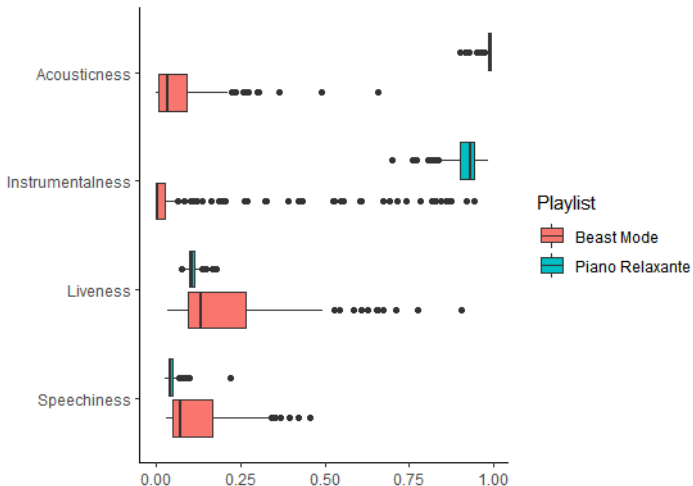


Figura: Primeiro grupo de variáveis Beast Mode vs Piano Relaxante.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

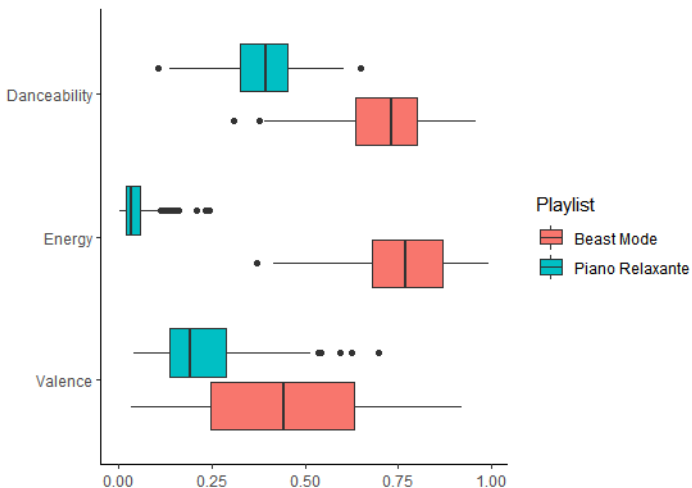


Figura: Segundo grupo de variáveis Beast Mode vs Piano Relaxante.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

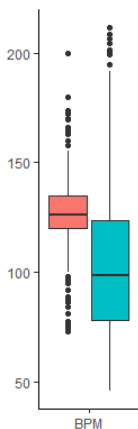


Figura: Duração  
Beast Mode vs  
Piano Relaxante.

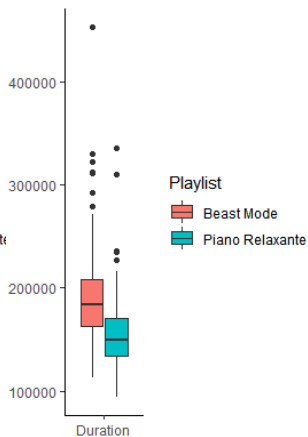


Figura: Tempo  
(BPM) Beast Mode  
vs Piano Relaxante.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

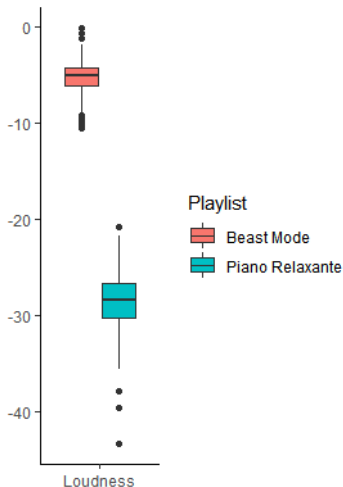


Figura: Loudness Beast Mode vs Piano Relaxante.



# Acurácia de Previsões

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

**Tabela:** Melhores resultados validação cruzada por método.

CV	Método	Acurácia	Kappa
<i>K-folds</i>	Regressão logística	0.9781	0.9522
<i>K-folds</i>	<i>Random forest</i>	0.9763	0.9488
<i>LOO</i>	Regressão logística	0.9799	0.9566
<i>LOO</i>	<i>Random forest</i>	0.9781	0.9527



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

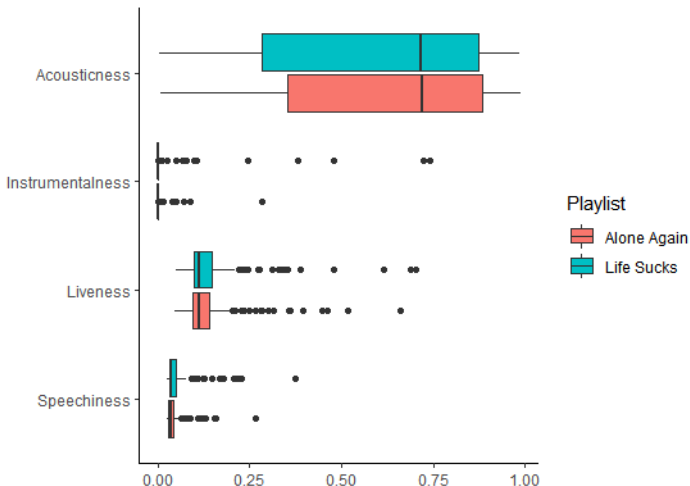


Figura: Primeiro grupo de variáveis Alone Again e Life Sucks.





# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

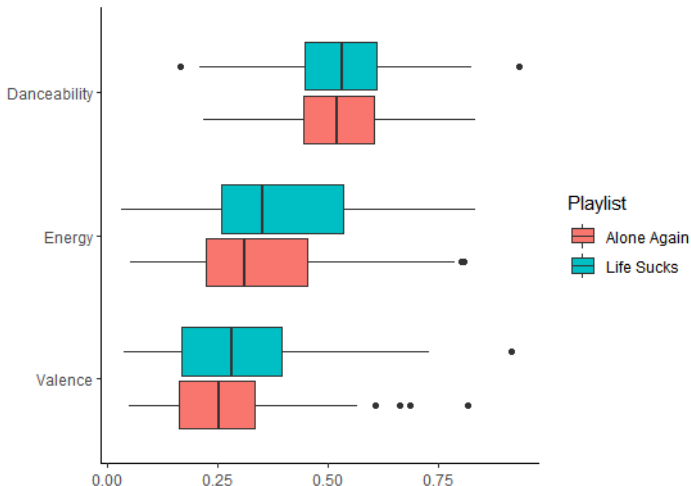


Figura: Segundo grupo de variáveis Alone Again e Life Sucks.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

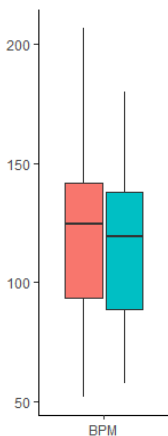


Figura: Duração  
Alone Again e Life  
Sucks.

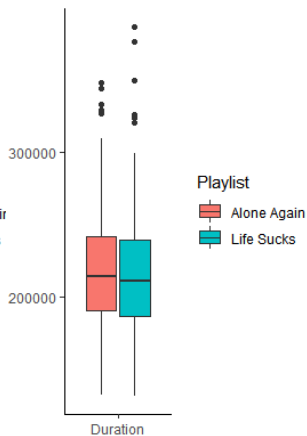


Figura: Tempo  
(BPM) Alone Again  
e Life Sucks.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

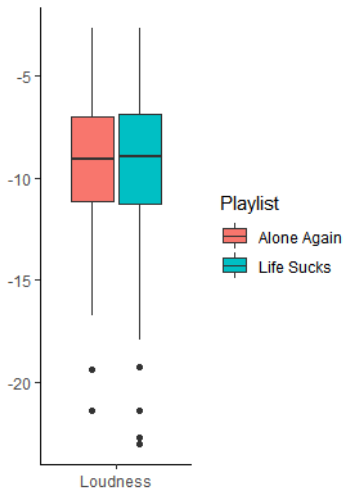


Figura: Loudness Alone Again e Life Sucks.



# Acurácia de Previsões

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

**Tabela:** Melhores resultados validação cruzada por método.

CV	Método	Acurácia	Kappa
<i>K-folds</i>	Regressão logística	0.6000	0.1282
<i>K-folds</i>	<i>Random forest</i>	0.4800	-0.0860
<i>LOO</i>	Regressão logística	0.5886	0.1048
<i>LOO</i>	<i>Random forest</i>	0.4371	-0.1618



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

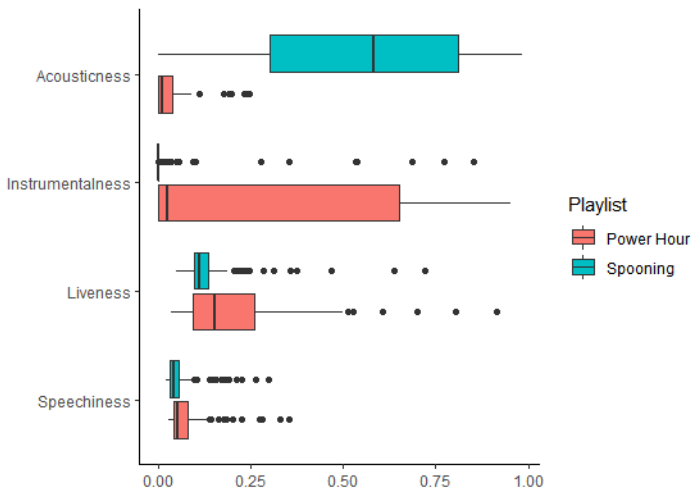


Figura: Primeiro grupo de variáveis Power Hour vs Spooning.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

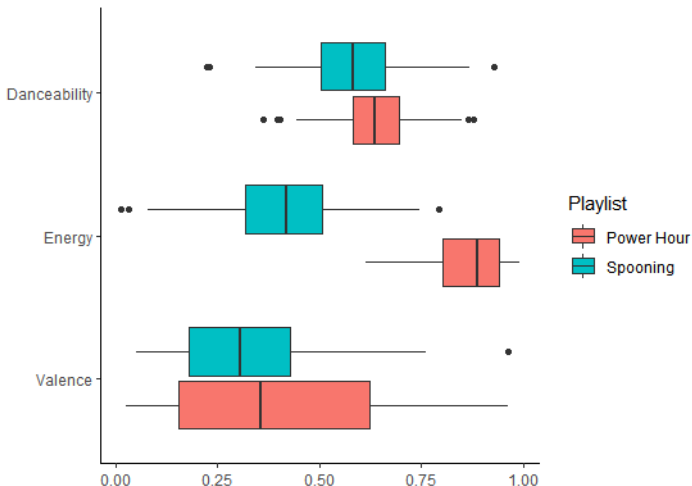


Figura: Segundo grupo de variáveis Power Hour vs Spooning.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

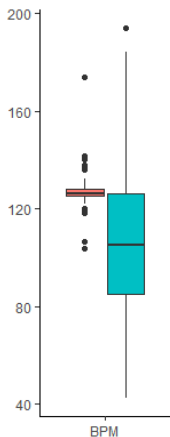


Figura: Duração  
Power Hour vs  
Spooning.

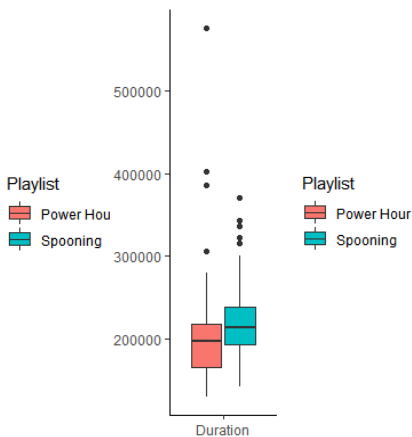


Figura: Tempo  
(BPM) Power Hour  
vs Spooning.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

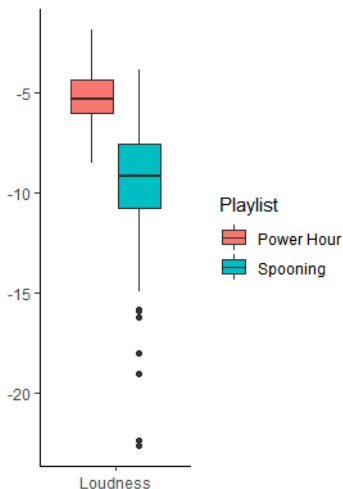


Figura: Loudness Power Hour vs Spooning.





# Acurácia de Previsões

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

**Tabela:** Melhores resultados validação cruzada por método.

CV	Método	Acurácia	Kappa
<i>K-folds</i>	Regressão logística	0.9320	0.8595
<i>K-folds</i>	<i>Random forest</i>	0.9640	0.9240
<i>LOO</i>	Regressão logística	0.9400	0.8756
<i>LOO</i>	<i>Random forest</i>	0.9640	0.9249



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

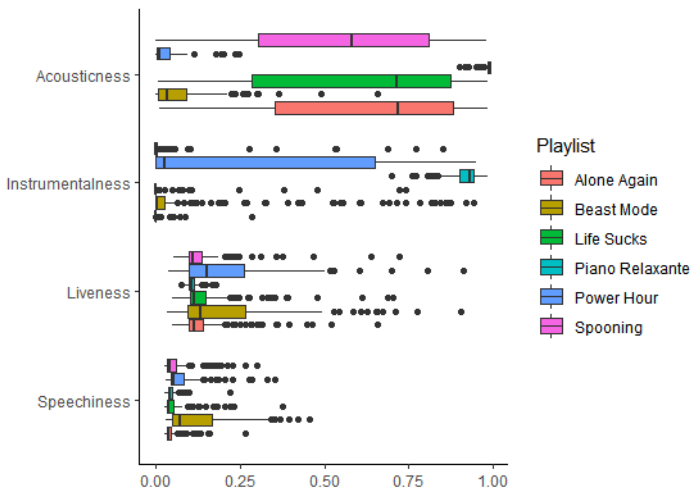


Figura: Primeiro grupo de variáveis todas as *playlists*.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

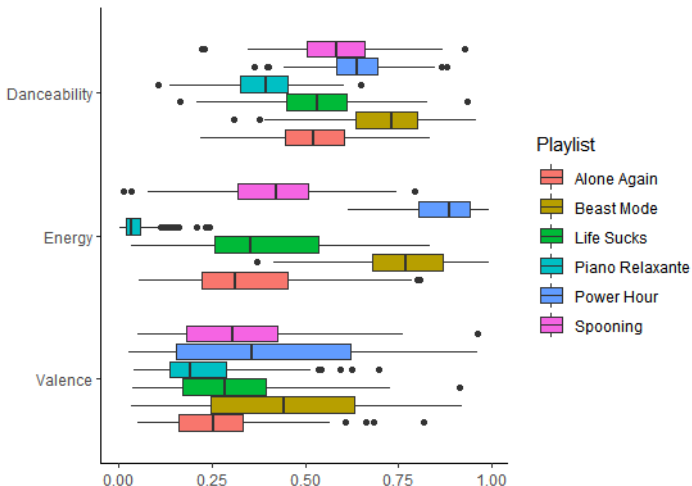


Figura: Segundo grupo de variáveis todas as *playlists*.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

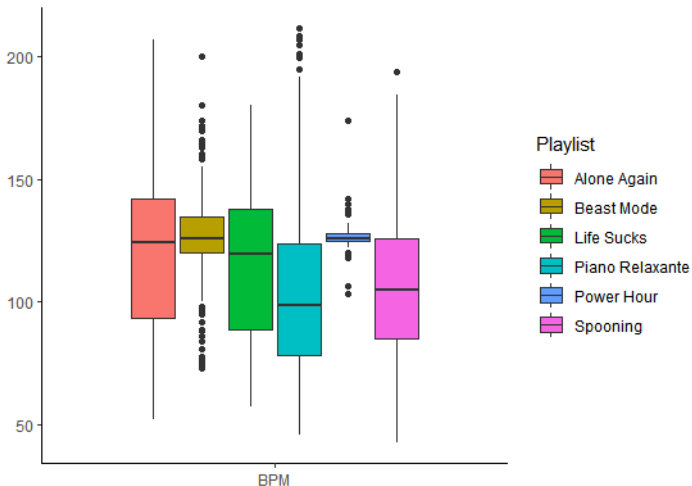


Figura: Duração de todas as *playlists*.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

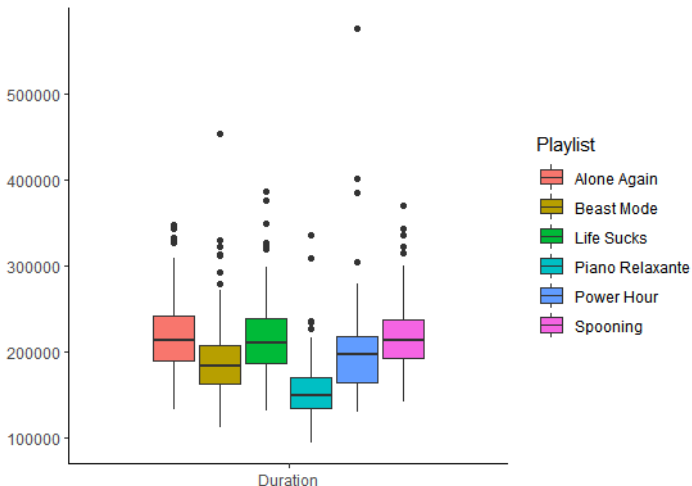


Figura: Tempo (BPM) de todas as *playlists*.



# Análise Descritiva

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

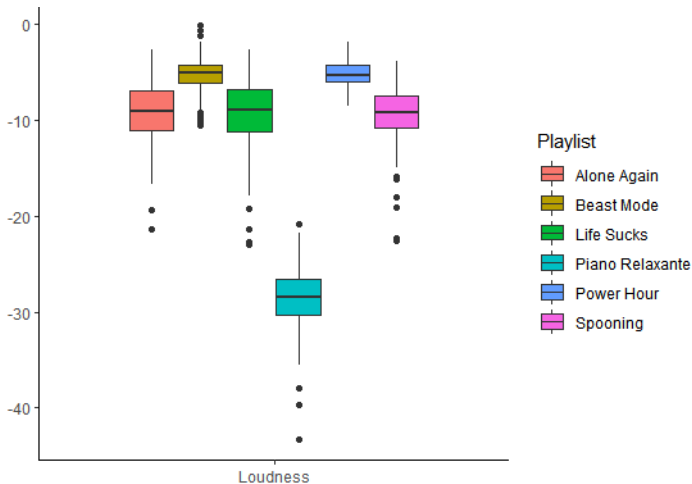


Figura: Loudness de todas as *playlists*.



# Previsões Regressão Logística

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
Playlists

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

Tabela: Previsão para Alone Again vs Outras.

	Alone Again	Beast Mode	Life Sucks	Piano Relaxante	Power Hour	Spooning
Alone Again	9	0	9	0	0	1
Outras	36	59	51	99	28	37

Tabela: Previsão para Beast Mode vs Outras.

	Beast Mode	Alone Again	Life Sucks	Piano Relaxante	Power Hour	Spooning
Beast Mode	43	0	1	0	15	1
Outras	16	45	59	99	13	37

Tabela: Previsão para Life Sucks vs Outras.

	Life Sucks	Alone Again	Beast Mode	Piano Relaxante	Power Hour	Spooning
Life Sucks	11	15	0	0	0	3
Outras	49	30	59	99	28	35



# Previsões Regressão Logística

Tabela: Previsão para Piano Relaxante vs Outras.

	Piano Relaxante	Alone Again	Beast Mode	Life Sucks	Power Hour	Spooning
Piano Relaxante	99	1	0	1	0	0
Outras	0	44	59	59	28	38

Tabela: Previsão do modelo 2 para Power Hour vs Outras.

	Power Hour	Alone Again	Beast Mode	Life Sucks	Piano Relaxante	Spooning
Power Hour	15	0	9	0	0	0
Outras	13	45	50	60	99	38

Tabela: Previsão do modelo 1 para Spooning vs Outras.

	Spooning	Alone Again	Beast Mode	Life Sucks	Piano Relaxante	Power Hour
Outras	38	45	59	60	99	28





# Previsões *Random Forest*

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Tabela: Previsão *random forest*

	Alone Again	Beast Mode	Life Sucks	Piano Relaxante	Power Hour	Spooning
Alone Again	9	0	23	0	0	9
Beast Mode	4	43	7	0	18	3
Life Sucks	27	4	22	0	0	13
Piano Relaxante	0	0	0	99	0	0
Power Hour	0	9	0	0	10	0
Spooning	5	3	8	0	0	13

Introdução

Metodologia

Resultados

Conclusões

Referências





# Conclusões


- De maneira geral, os métodos de regressão logística e *randomforest* obtiveram resultados muito interessantes. Performaram muito bem nos pares 1 e 3, mas muita dificuldade com o par 2.
- Possível necessidade de outras variáveis.
- É possível que 2 *playlists* tenham propostas semelhantes e suas músicas sejam intercambiáveis, tornando a distinção entre elas ao acaso ou incorreta.





# Referências


 IZBICKI, R.; SANTOS, T. M. dos. *Aprendizado de máquina: uma abordagem estatística*. [S.l.: s.n.], 2020. ISBN 978-65-00-02410-4.

 MCINTIRE, G. *A Machine Learning Deep Dive into My Spotify Data*. Disponível em: <<https://opendatascience.com/a-machine-learning-deep-dive-into-my-spotify-data/>>. Acessado em 16 set. 2021.

 MIGUEL, T. *Arvore de Decisão em R*. 2020. Disponível em: <<https://aprenderdatascience.com/regressao-logistica/>>. Acessado em 13 set. 2021.

 SPOTIFY. *Spotify for developers*. Disponível em: <<https://developer.spotify.com/documentation/web-api/>>. Acessado em 28 ago. 2021.

 STANKEVIX, G. *Arvore de Decisão em R*. 2019. Disponível em: <<https://medium.com/@gabriel.stankevix/arvore-de-decis%C3%A3o-em-r-85a449b296b2>>. Acessado em 16 set. 2021.

 TEAM, R. C. et al. *R: A language and environment for statistical computing*. Vienna, Austria, 2013.



# Conclusões

Algoritmos De  
Classificação  
Aplicados Em  
Recomenda-  
ções De  
Faixas Para  
*Playlists*

Rodrigo  
Cavalcanti  
Loreto

Introdução

Metodologia

Resultados

Conclusões

Referências

# Obrigado!