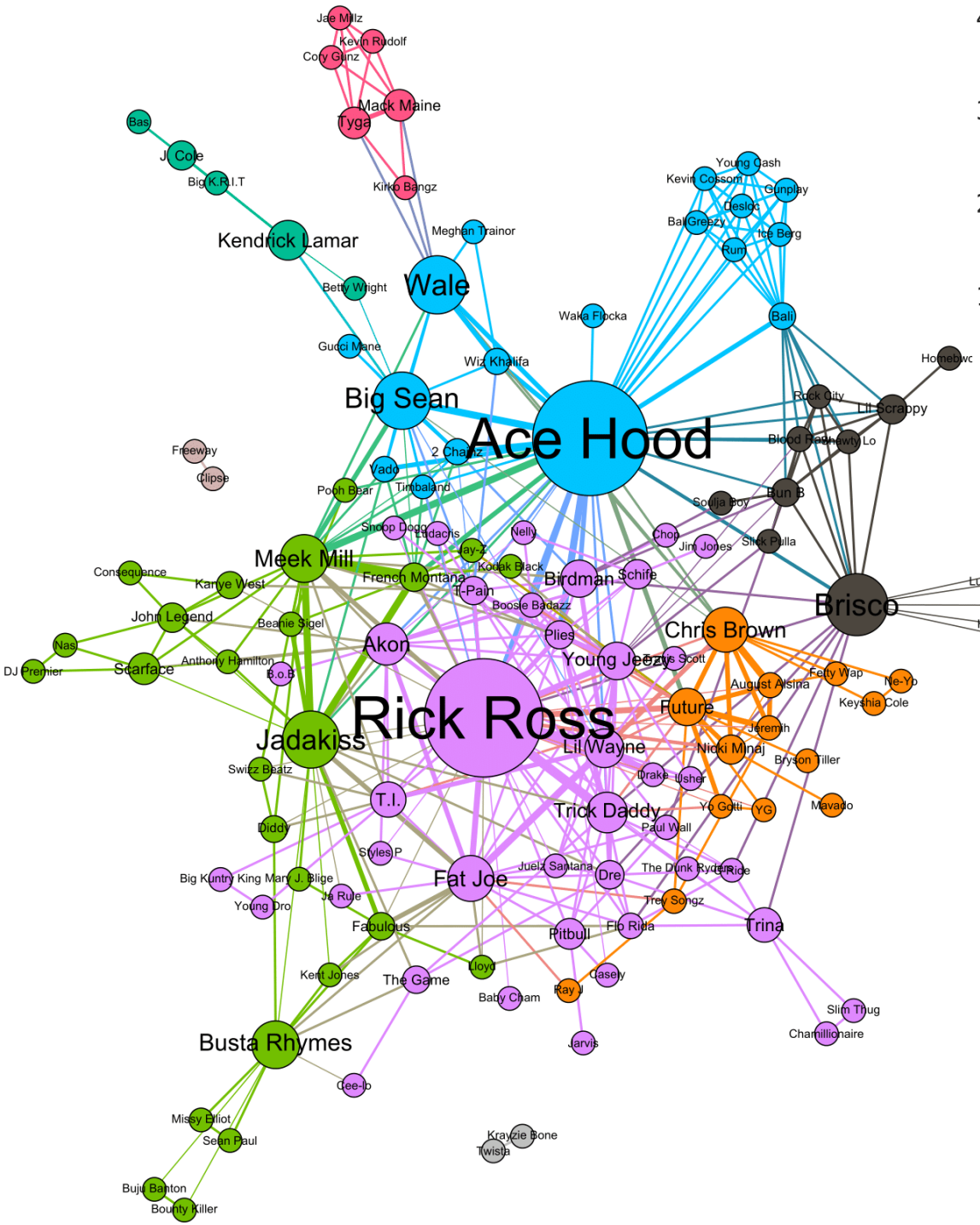
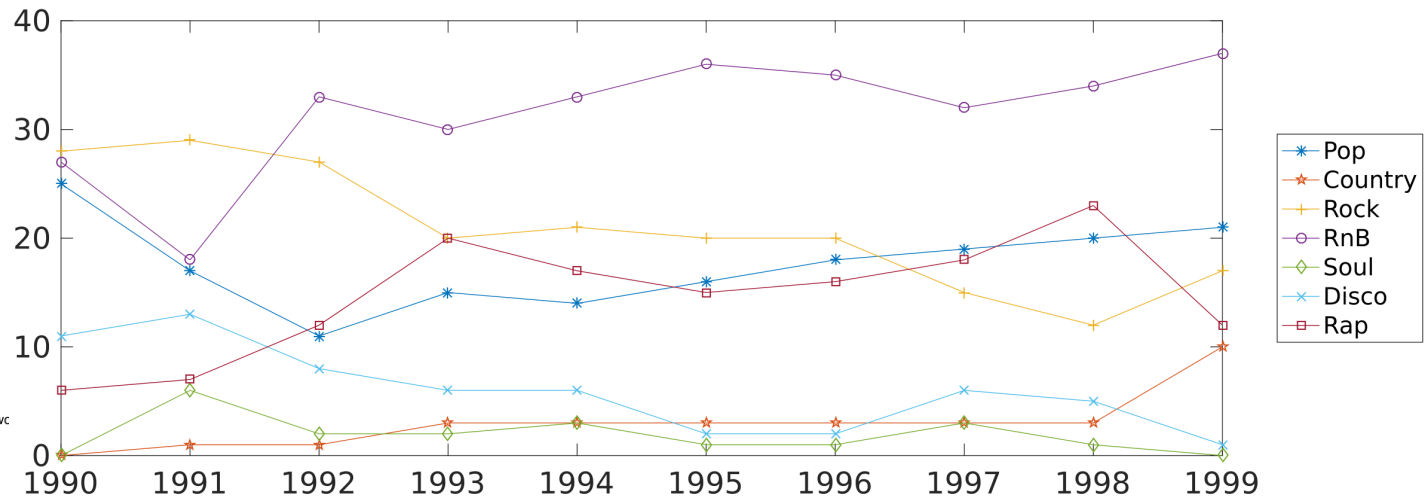
Previsão de Popularidade de Músicas em Plataformas de Streaming

Carlos Vicente Soares Araujo
vicente@icomp.ufam.edu.br


Hit Song Science


Spotify®


Jae Millz
Kevin Rudolf
Cory Gunz
Mack Maine
Tyga
Bas
J. Cole
Big K.R.I.T
Kirko Bangz
Kendrick Lamar
Meghan Trainor
Wale
Betty Wright
Gucci Mane
Wiz Khalifa
Big Sean
Ace Hood
2 Chainz
Freeway
Clipse
Vado
Pooh Bear
Timbaland
Shop Dogg
Eddachis
Nelly
Jay-Z
Kodak Black
Meek Mill
French Montana
T-Pain
Birdman
Consequence
Kanye West
John Legend
Beanie Sigel
Boosie Badazz
Akon
Nas
DJ Premier
Scarface
Anthony Hamilton
B.o.B
Rick Ross
Jadakiss
Swizz Beatz
T.I.
Diddy
Styles P
Big Kuntry King
Mary J. Blige
Young Dro
Ja Rule
Fabulous
Fat Joe
Busta Rhymes
Kent Jones
The Game
Lloyd
Baby Cham
Ray J
Dee-lo
Missy Elliot
Sean Paul
Buju Banton
Bounty Killer
Krayzie Bone
Twista
Jarvis
Slim Thug
Chamillionaire
Trina
Pitbull
Flo Rida
Casely
Trey Songz
Juelz Santana
Dre
The Dunk Ryders
Ride
Trick Daddy
Paul Wall
Lil Wayne
Drake
Usher
Nicki Minaj
Jeremih
Bryson Tiller
Mavado
Yo Gotti
YG
Future
August Alsina
Fetty Wap
Ne-Yo
Keyshia Cole
Young Jeezy
Travis Scott
Chris Brown
Plies
Young Cash
Kevin Cossom
Gunplay
Desloc
Ball Greezy
Ice Berg
Rum
Bali
Waka Flocka
Homebwc
Rock City
Lil Scrappy
Blood Raw
Lawrence Lo
Bun B
Slick Pulla
Jim Jones
Chop
Souja Boy
Brisco
Dela
Lunch Money
P.M.
C.O.
Hennessy
Dirty Red


40
30
20
10
0
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
Pop
Country
Rock
RnB
Soul
Disco
Rap

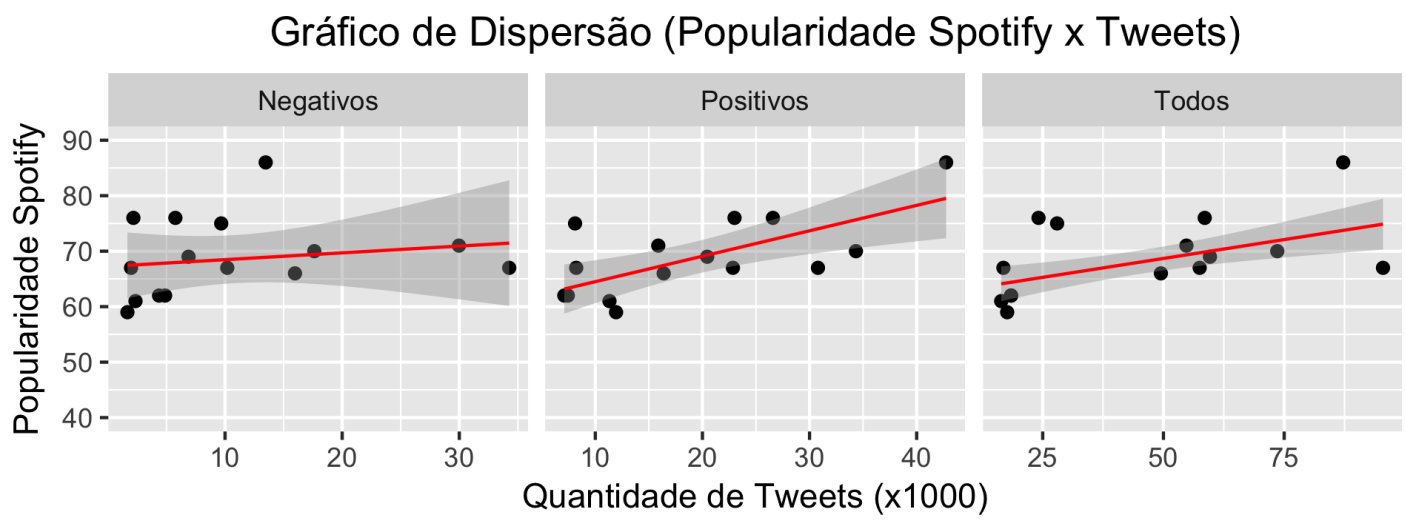


Gráfico de Dispersão (Popularidade Spotify x Tweets)
Negativos
Positivos
Todos
Popularidade Spotify
90
80
70
60
50
40
10
20
30
10
20
30
40
25
50
75
Quantidade de Tweets (x1000)


Spotify® for Developers
curl://

Playlists


```
        "available_markets": [],
        "disc_number": 1,
        "duration_ms": 730066,
        "episode": false,
        "explicit": false,
        "external_ids": {
          "isrc": "FR2X41475057"
        },
        "external_urls": {
          "spotify":
"https://open.spotify.com/track/5o3jMYOSbaVz3tkgwhELSV"
        },
        "href":
"https://api.spotify.com/v1/tracks/5o3jMYOSbaVz3tkgwhELSV",
        "id": "5o3jMYOSbaVz3tkgwhELSV",
        "is_local": false,
        "name": "Is",
        "popularity": 0,
        "preview_url": null,
        "track": true,
        "track_number": 21,
        "type": "track",
        "uri": "spotify:track:5o3jMYOSbaVz3tkgwhELSV"
```


developer.spotify.com/console/get-playlist-tracks/

Audio Features


```
{
  "danceability": 0.696,
  "energy": 0.905,
  "key": 2,
  "loudness": -2.743,
  "mode": 1,
  "speechiness": 0.103,
  "acousticness": 0.011,
  "instrumentalness": 0.000905,
  "liveness": 0.302,
  "valence": 0.625,
  "tempo": 114.944,
  "type": "audio_features",
  "id": "11dFghVXANMlKmJXsNCbNl",
  "uri": "spotify:track:11dFghVXANMlKmJXsNCbNl",
  "track_href":
"https://api.spotify.com/v1/tracks/11dFghVXANMlKmJXsNCbNl",
  "analysis_url": "https://api.spotify.com/v1/audio-
analysis/11dFghVXANMlKmJXsNCbNl",
  "duration_ms": 207960,
  "time_signature": 4
}
```

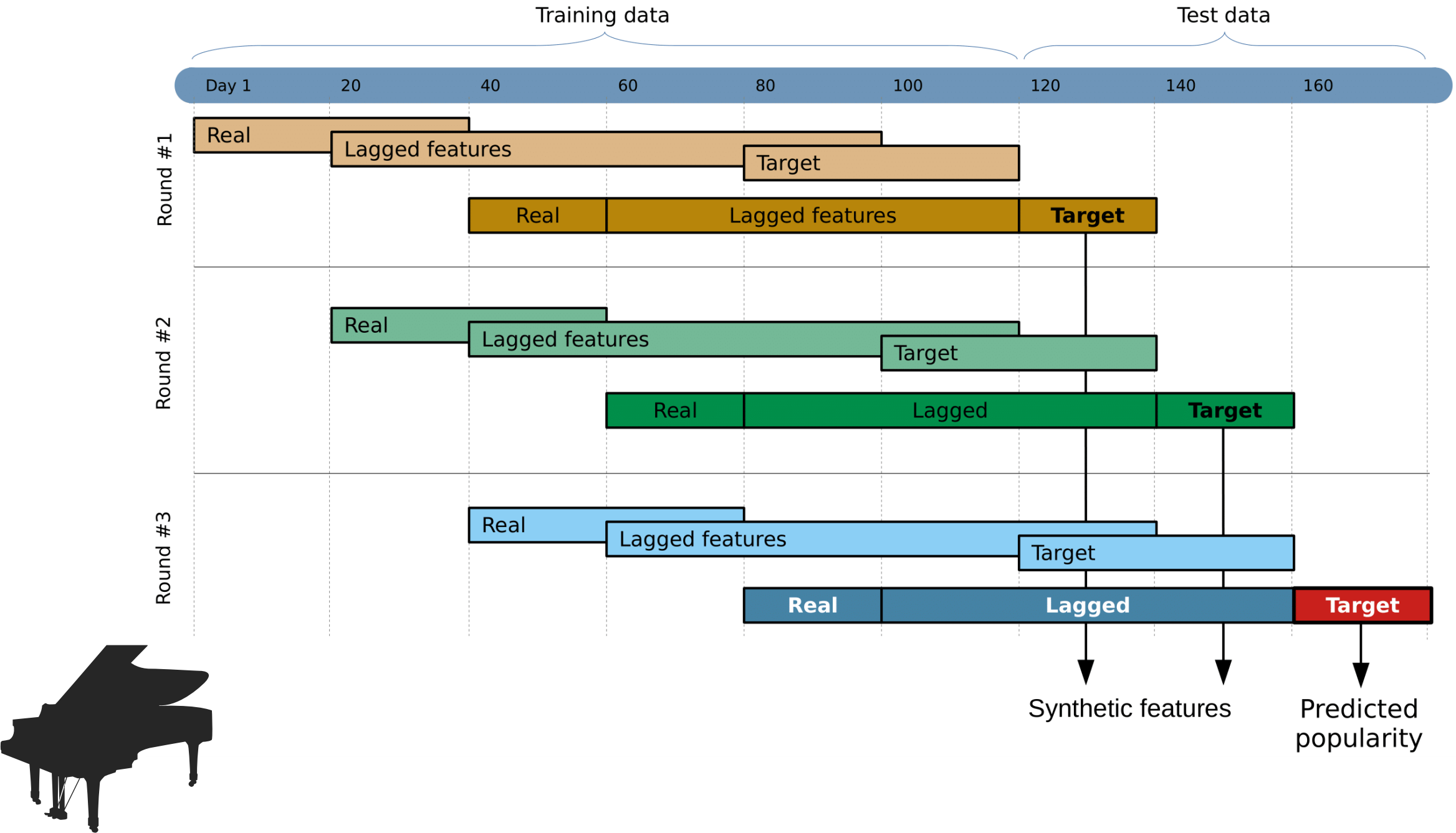

developer.spotify.com/console/get-audio-features-several-tracks/

Campos das Playlists

- Posição no Ranking
- Artistas Participantes
- Duração da Música
- Nome da Faixa
- Link para MP3
- Data do Ranking
- Data de Lançamento
- Explicitude
- Popularidade

Audio Features

- Danceability
- Key
- Mode
- Acousticness
- Liveness
- Tempo
- Energy
- Loudness
- Speechiness
- Instrumentalness
- Valence
- Time Signature


Training data
Test data
Day 1
20
40
60
80
100
120
140
160
Round #1
Real
Lagged features
Target
Real
Lagged features
Target
Round #2
Real
Lagged features
Target
Real
Lagged
Target
Round #3
Real
Lagged features
Target
Real
Lagged
Target
Synthetic features
Predicted popularity

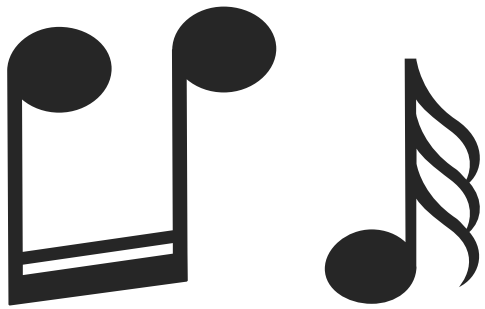
```
dtop = dtop.drop('Artistas', axis = 1)
dtop = dtop.drop('Nome da Faixa', axis = 1)
dtop = dtop.drop('Data do Ranking', axis = 1)
dtop = dtop.drop('Data da Música', axis = 1)
dtop = dtop.drop('Duração', axis = 1)
```

```
dtop = dtop.drop('key', axis = 1)
dtop = dtop.drop('loudness', axis = 1)
dtop = dtop.drop('tempo', axis = 1)
dtop = dtop.drop('time', axis = 1)
dtop = dtop.drop('mode', axis = 1)
```

```
dtop['Posição'] = np.where(dtop['Posição']>0, 1, 0)
```

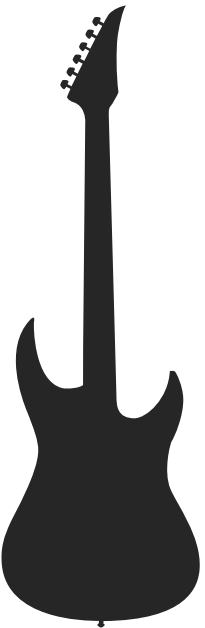
```
colunastop = list(dtop.columns.values)
colunastop.remove('Posição')
```

```
dtop['danceability'] = np.where(dtop['danceability']>0.5, 1, 0)
dtop['energy'] = np.where(dtop['energy']>0.5, 1, 0)
dtop['speechiness'] = np.where(dtop['speechiness']>0.33, 1, 0)
dtop['acousticness'] = np.where(dtop['acousticness']>0.5, 1, 0)
dtop['instrumentalness'] = np.where(dtop['instrumentalness']>0.5, 1, 0)
dtop['liveness'] = np.where(dtop['liveness']>0.8, 1, 0)
dtop['valence'] = np.where(dtop['valence']>0.5, 1, 0)
```

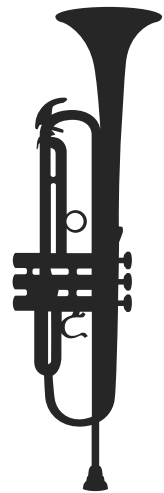


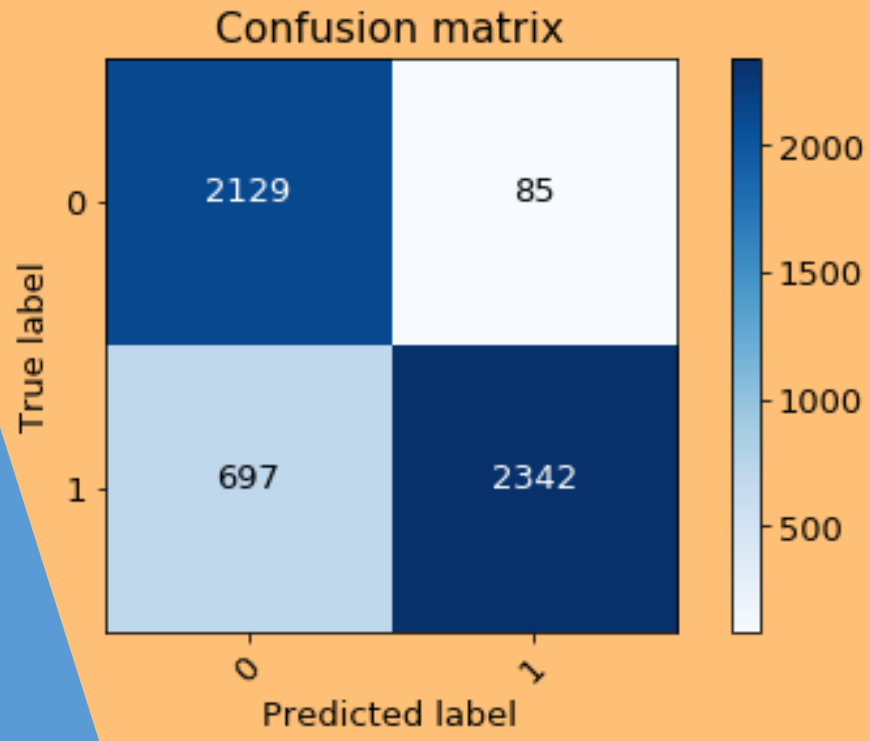
```
dtop.head(5)
```

	Posição	Explicitude	Popularidade	danceability	energy	speechiness	acousticness	instrumentalness	liveness	valence
0	1	1	100	1	1	0	0	0	0	1
1	1	0	97	1	1	0	0	0	0	0
2	1	0	99	1	1	0	0	0	0	1
3	1	0	92	1	0	0	1	0	0	1
4	1	1	84	1	1	0	0	0	0	0

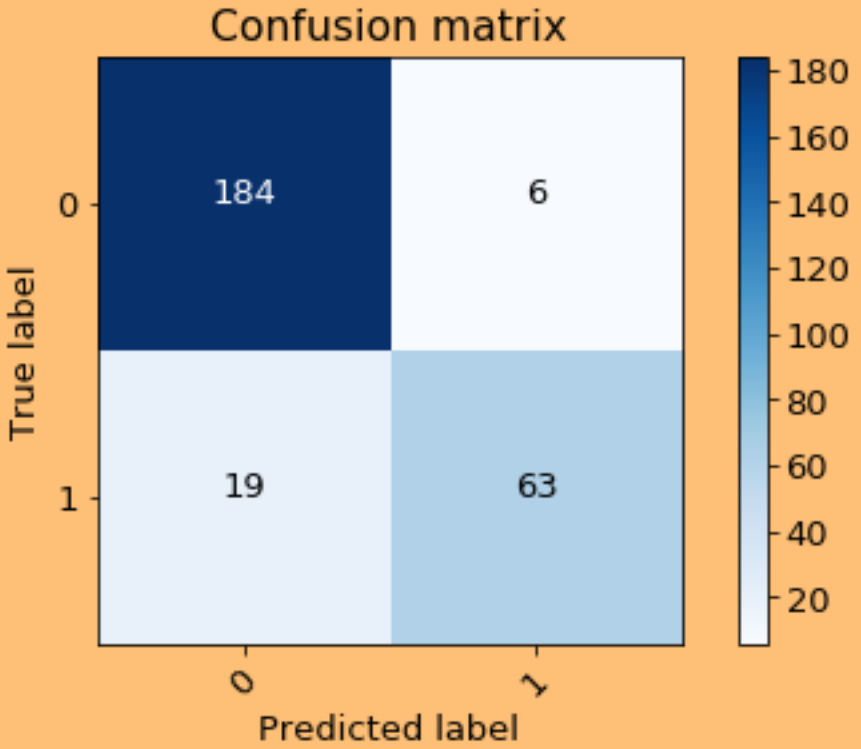


scikit learn



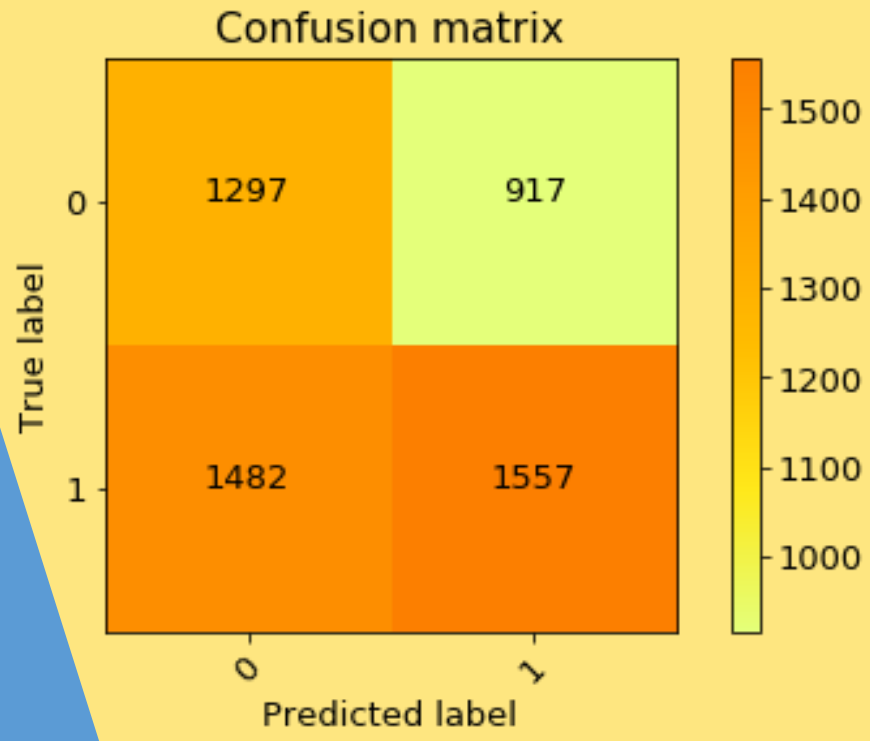


Previsão com Repetição de Músicas
Usando SVM c/ kernel RBF

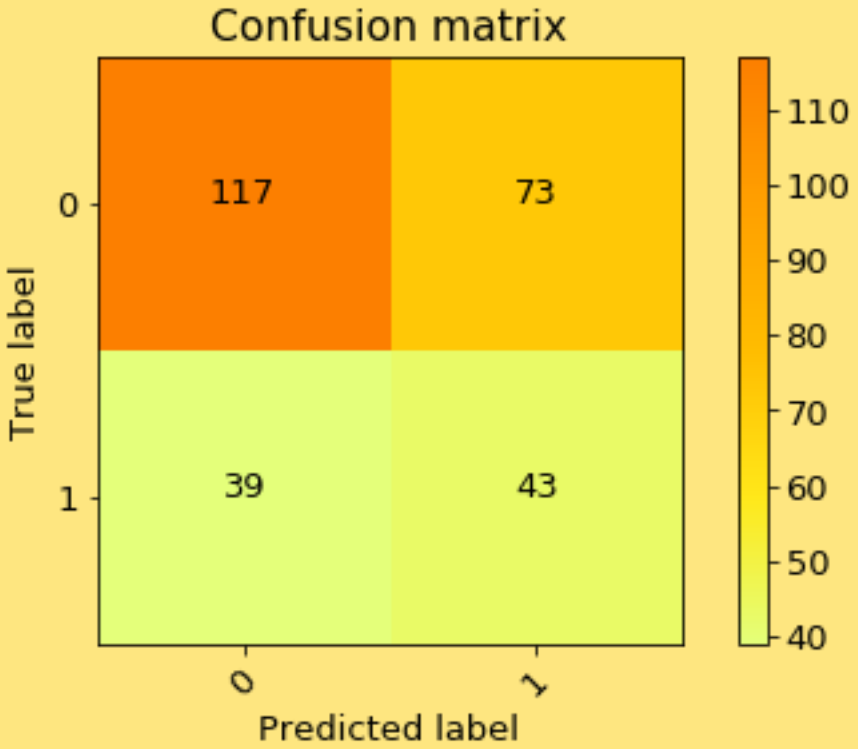


Previsão sem Repetição de Músicas
Usando SVM c/ kernel RBF

Usando Reiman e Örnell



Previsão com Repetição de Músicas
Usando KNN

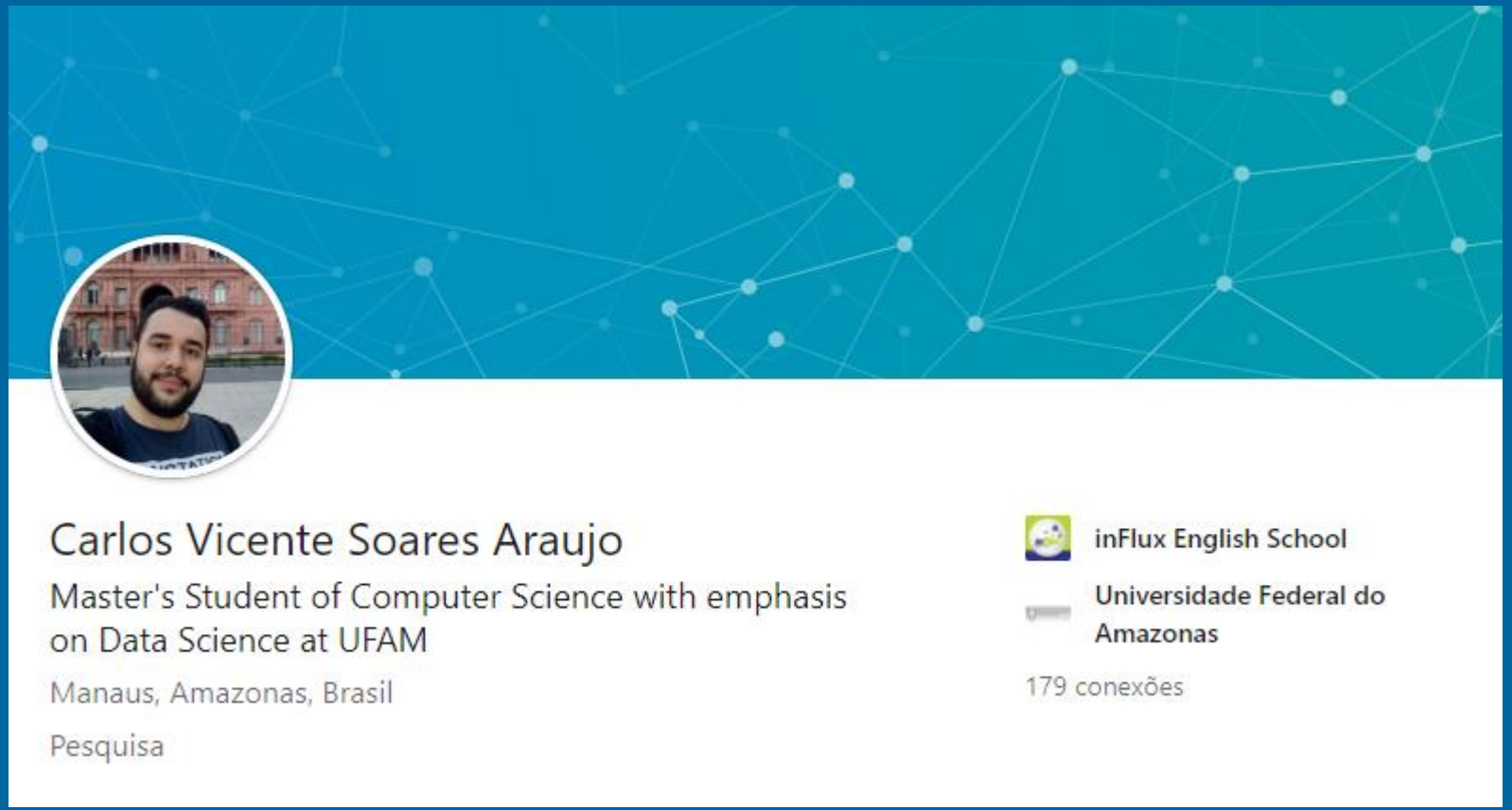


Previsão sem Repetição de Músicas
Usando SVM c/ kernel RBF

Uma acurácia melhor pode não significar um resultado melhor.

Araujo, Carlos
III PyData Manaus

LinkedIn



https://www.linkedin.com/in/carlosvsaraujo/

Obrigado!

vicente@icomp.ufam.edu.br

bit.ly/carlos-artigos