

Classificador de gêneros musicais pela letra

Lukas Iepsen, Patrick Guimarães

Dataset

Definimos 7 categorias para explorarmos:

- Axé
- MPB
- Sertanejo
- Bossa-nova
- Forró
- Gospel
- Samba

Foram coletados, em média, 900 exemplos de cada estilo.

Após um processamento para eliminarmos letras que estão em outras linguagens temos a seguinte distribuição:

```
In [12]: df.label.value_counts()
```

```
Out[12]: sertanejo      895  
         forro         891  
         mpb          889  
         samba        886  
         gospelreligioso 874  
         bossa-nova     800  
         axe           764  
         Name: label, dtype: int64
```

Resultados

Modelo	Resultado (valor médio)
LSTM Bidimensional (nível de palavras)	50%
XGBoost	48%
Random Forest	45%
Decision Tree	38%
MLP	15%

Planos para o futuro (muito muito próximo)

- Usar LSTM no nível de caracteres
- Testar o uso de uma CNN
- Testar em arquiteturas conhecidas (e.g. CharCNN)
- Testar categorias separadamente
- Analisar comportamento das classes no treinamento (proximidade)