

Anatomy of a song

Alex & Inès

Can lyrics define a music genre?





TABLE OF CONTENTS

- 01** INTRODUCTION
- 02** DATA PREPARATION
- 03** MODELLING
- 04** CONCLUSION



01

Introduction

Musical Characteristics





Musical Characteristics

Hip Hop

- Heavy bassline
- Flow
- 85 - 115 BPM

Rock

- Strong beat
- Screaming or shouting
- 110 - 140 BPM

Pop

- Simple melody
- Rhythms made to dance
- 100 - 130 BPM

Lyrical Characteristics





Lyrical Characteristics

Hip Hop

- Poverty
- Violence
- Women

Rock

- Drug
- Intimacy
- Rebellion &
Freedom

Pop

- Love
- Friendship
- Empowerment



02

Data Preparation

Data Set





Dataset

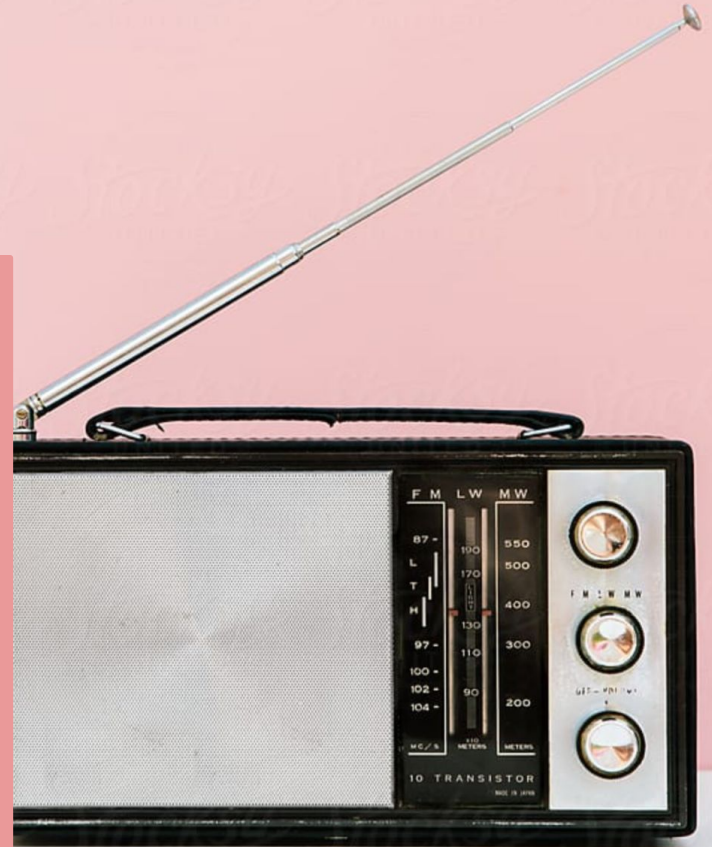
Where?

- Found on Kaggle
- Based off data scraped from a Brazilian music portal ([Vagalume](#))

What?

- 6 music genres
- 2 main languages
- 3242 artists
- 209522 songs

Data Cleaning & Processing



Data cleaning

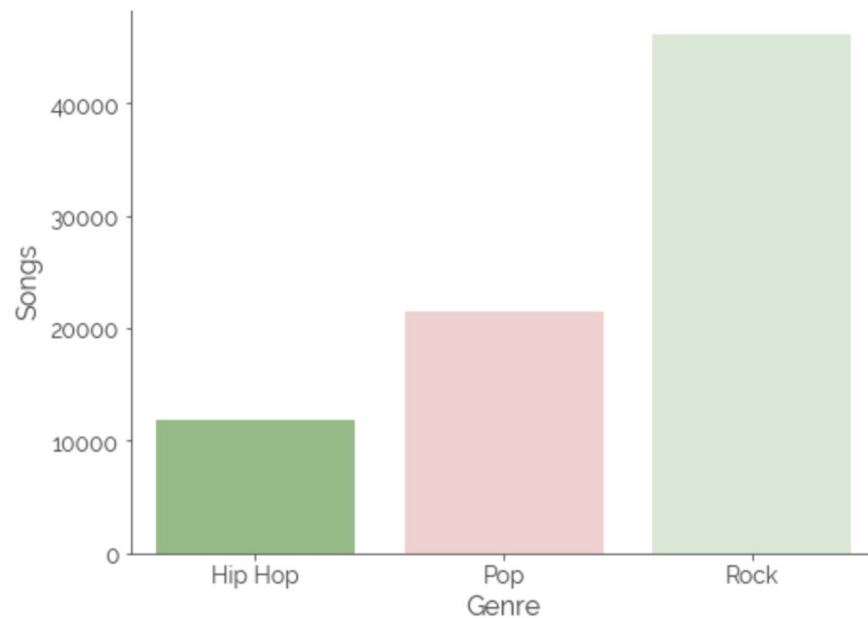
- Songs in English
- 3 genres: Hip Hop, Pop and Rock
- Artists with a single main genre
- 79452 songs



Genre Repartition

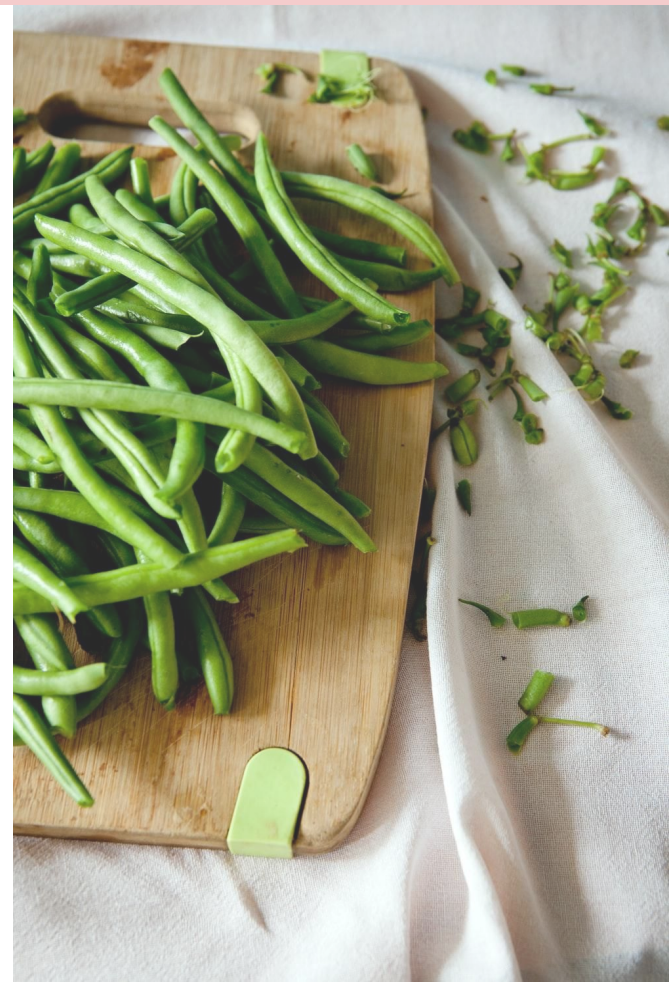
- Hip Hop ~ 15 %
- Pop ~ 27 %
- Rock ~ 58 %

The dataset is unbalanced, in favour of Rock songs.



Natural Language Processing

- Remove and keep track of trailing apostrophes (eg. tryin')
- Remove any brackets and its content (eg. [Chorus])
- Remove multiple letters (eg. yeeeeaaah -> yeah)
- Basic cleaning
- Stemming / Lemmatization
- Words vectorization



- Most common words in Hip Hop music
- Does not include the top 200 words in Pop and Rock
- Top 10 words include: explicit words, hit, cuz'
- 8 of the top 10 words are explicit



© 2010 Pearson Education, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage or retrieval system, without permission in writing from Pearson Education, Inc.

- Most common words in Pop music
- Does not include the top 200 words in Hip Hop and Rock
- Top 10 words include: kiss, touch, beautiful, hurt



- Most common words in Rock music
- Does not include the top 200 words in Hip Hop and Pop
- Top 10 words include: dead, alive, woman, child





Average number of words

Music Genre	Average number of words
All	123
Hip Hop	233
Pop	124
Rock	94

In average, Hip Hop lyrics have:

- 190 % more words than Pop lyrics
- 250 % more words than Rock lyrics

Rock lyrics have 25 % fewer lyrics than average



Feature Enrichment

Top Ten Words

- 3 binary markers
- 1 for each genre
- Marker = 1 if any word from the 10 most common words for each genre is present in the lyrics.

Trailing Apostrophe

- Track how many there are per song
- In average, 5 times more trailing apostrophes in Hip Hop songs



03

Modelling

Basic models



Random Forest

- GridSearch to find best hyper parameters
- 99 % accuracy on the training test
- Only 75 % accuracy on the testing set
- Overfitting suspected





Logistic Regression - Model Selection

Stemmer

- Porter
- Snowball
- Lancaster

Solver

- Limited-memory BFGS
(default)
- Newton CG
- SAGA

Penalizer

- L1
- L2 (default)
- Elastic Net

Logistic Regression - Results

- 76 % accuracy on the training set
- 74 % accuracy on the testing set
- 61% of the misclassified songs were Pop songs



Neural Networks



Sequences

What is a sequential approach to text transformation?

['apple', 'banana', 'car'] → [1, 18, 66]

['ice', 'flower', 'number'] → [22, 109, 888]

['food', 'apple', 'flower'] → [555, 1, 109]

Advantages:

- Interpretability
- Count and evidence
- Streamline ease

RNN

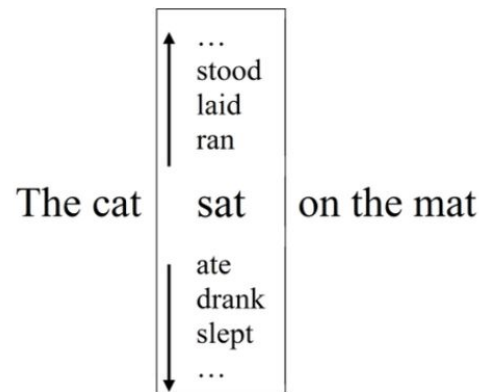
What is RNN?

- Type of NN exhibiting dynamic, bidirectional behaviour where nodes can exchange information back and forth to reinforce learning.
- Requires maintaining logs of past epochs in order to travel back 'in time'.

Why RNN?

- Why not Word2Vec, N-grams, Skip-grams?

There is no *context* to this sentence





Multiple Layer Perceptron (MLP)

What is *SLP*? - Single Layer Perceptron Neural Network

- Simplest form of NN: One input, one output and one hidden layer.
- Implicit forward propagation in its basic format
- Limited choices for activation and other hyperparameters

MLP = More layers!

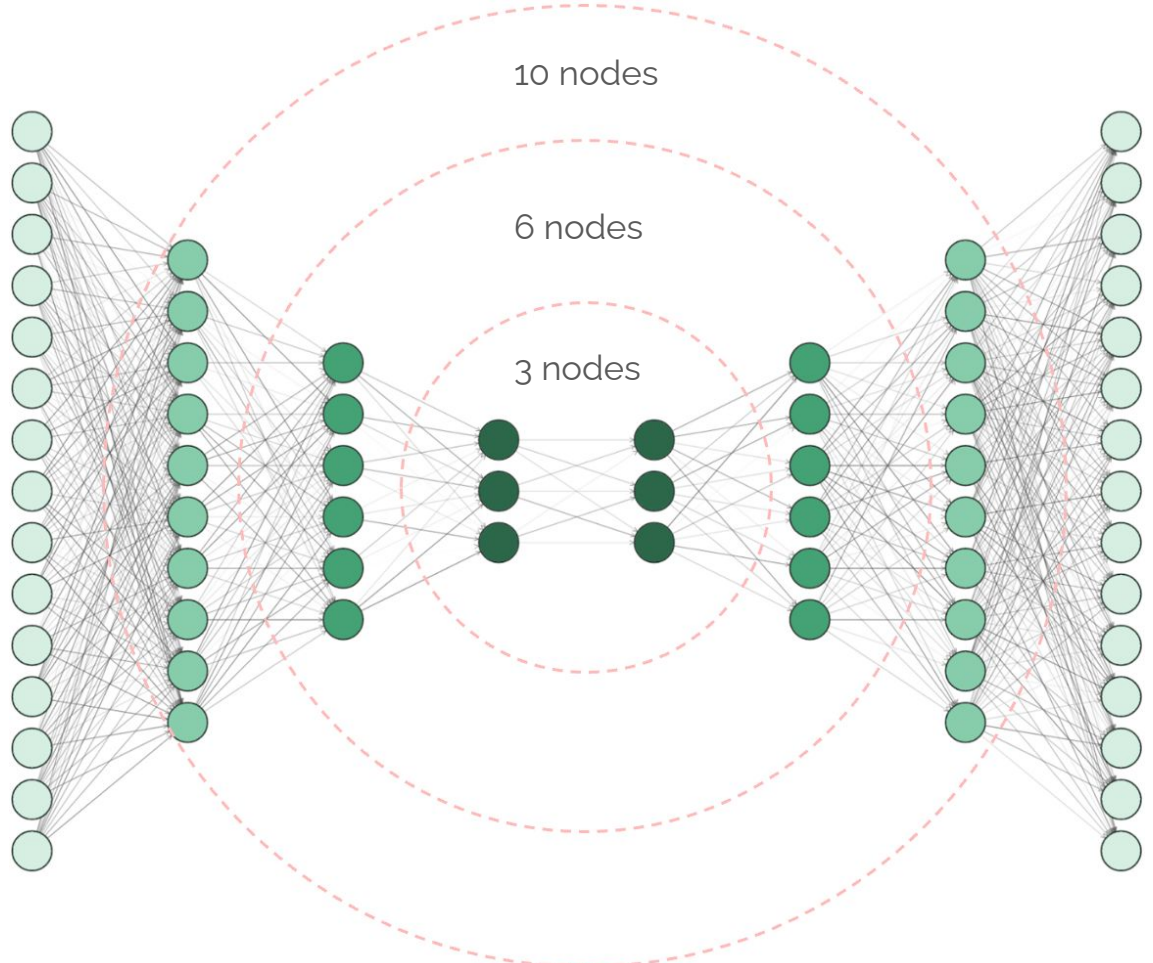
- Free choice of architecture
- Multiple hyperparameter possibilities
- Different activations for each layer



MLP - parameters

Hyperparameter selection:

1. Activation function:
 - ReLu (Rectified Linear Unit) : $f(x) = \max(0, x)$
 - Logistic: $f(x) = 1/(1+\exp(-x))$
 - Tanh: $f(x) = \tanh(x)$
2. Learning rate - Constant, **adaptive**, inverse
3. Solver - **Adam**, LBFGS
4. Architecture - symmetrical, **inverse symmetrical (AE-like)**, stiff, diverse, randomized, ...



Data formatting

Format required: NxM array

Problem: Not all songs have the same number of words!

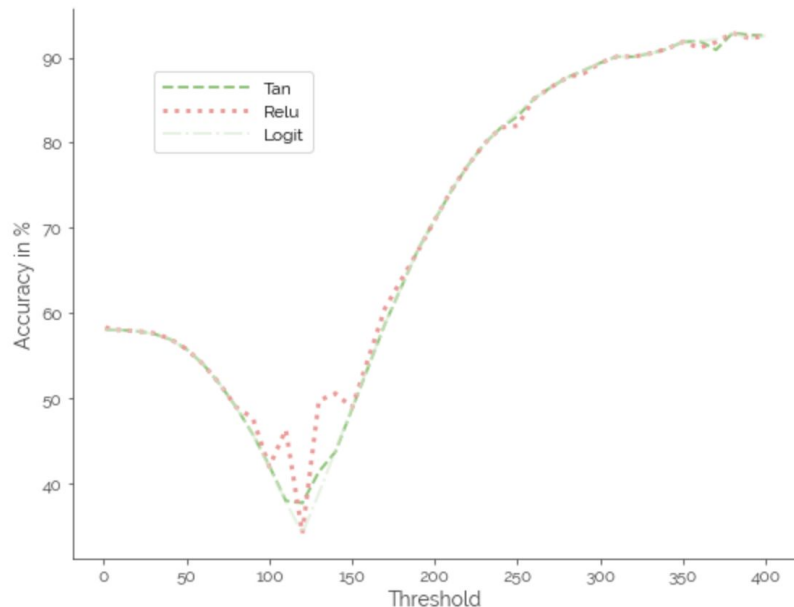
Solution:

- Use only songs with a number of words above a chosen threshold, M.
- For those songs, use only the first M words.

Result: We generate a performance spectrum based on the choice of thresholds.

```
08 02 22 97 38 15 00 40 00 75 04 05 07 78 52 12 50 77 91 08
49 49 99 40 17 81 18 57 60 87 17 40 98 43 69 48 04 56 62 00
81 49 31 73 55 79 14 29 93 71 40 67 53 88 30 03 49 13 36 65
52 70 95 23 04 60 11 42 69 24 68 56 01 32 56 71 37 02 36 91
22 31 16 71 51 67 63 89 41 92 36 54 22 40 40 28 66 33 13 80
24 47 32 60 99 03 45 02 44 75 33 53 78 36 84 20 35 17 12 50
32 98 81 28 64 23 67 10 26 38 40 67 59 54 70 66 18 38 64 70
67 26 20 68 02 62 12 20 95 63 94 39 63 08 40 91 66 49 94 21
24 55 58 05 66 73 99 26 97 17 78 78 96 83 14 88 34 89 63 72
21 36 23 09 75 00 76 44 20 45 35 14 00 61 33 97 34 31 33 95
78 17 53 28 22 75 31 67 15 94 03 80 04 62 16 14 09 53 56 92
16 39 05 42 96 35 31 47 55 58 88 24 00 17 54 24 36 29 85 57
86 56 00 48 35 71 89 07 05 44 44 37 44 60 21 58 51 54 17 58
19 80 81 68 05 94 47 69 28 73 92 13 86 52 17 77 04 89 55 40
04 52 08 83 97 35 99 16 07 97 57 32 16 26 26 79 33 27 98 66
88 36 68 87 57 62 20 72 03 46 33 67 46 55 12 32 63 93 53 69
04 42 16 73 38 25 39 11 24 94 72 18 08 46 29 32 40 62 76 36
20 69 36 41 72 30 23 88 34 62 99 69 82 67 59 85 74 04 36 16
20 73 35 29 78 31 90 01 74 31 49 71 48 86 81 16 23 57 05 54
01 70 54 71 83 51 54 69 16 92 33 48 61 43 52 01 89 19 67 48
```

Performance analysis across thresholds



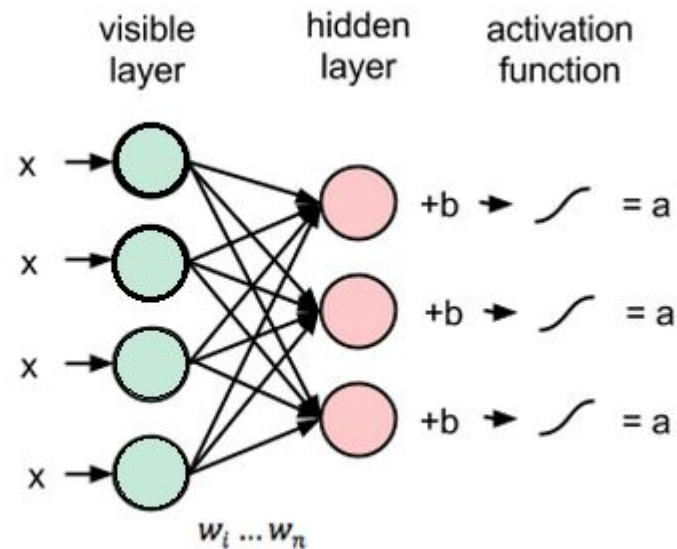
Behaviour:

- Constant ~58% until roughly M=7-8.
- Drop to minimum of ~36% accuracy at around M=120.
- Spike increase up to ~82% accuracy until the M=260 mark.
- Slower, continuous increase to maximum recorded of ~92% around M= 380 - 400

Why? More [here](#).

Results

- MLP performs best on larger songs
- The model can be improved using more data.
- Other possible models:
 - RBM: Useful in both topic modelling and classification
 - Note: NO topic modelling.





04

Conclusion

Next Steps

Models Improvement

Feature Enrichment

- Check for same words repetition within lyrics

Logistic Regression

- Error analysis: 61% of the misclassified songs were pop songs

Neural Networks

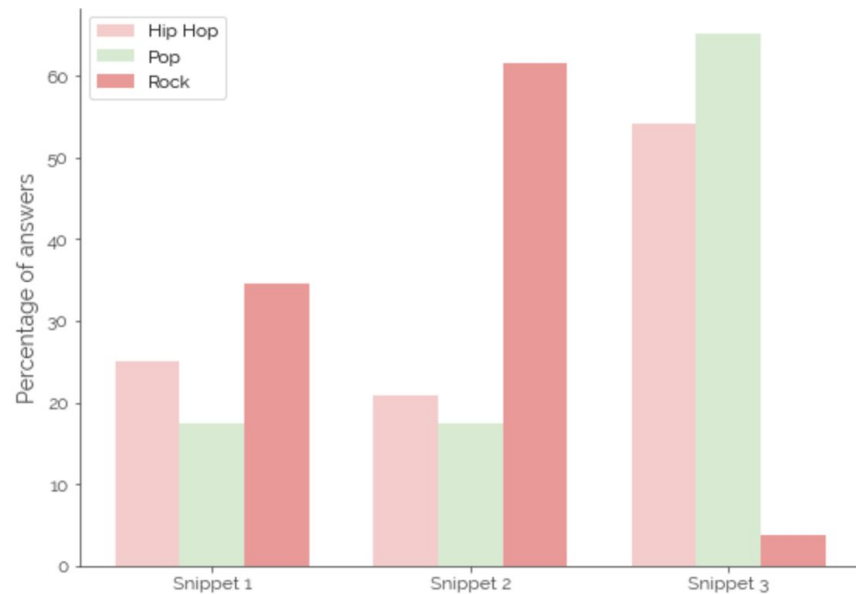
- More data



Demonstration



Did you beat the machine?





Thank you