21/11/2021

The Journey of Breaking Down Song Lyrics Using **NLP & Clustering**

Presented by:

Ahad Saad Almutairi

Renad Fawaz Albishri

OUTLINE

Introduction & Tools

Work Flow

About Data & EDA

NLP Process

Topic Modeling

Classification Model

Clustering

Scatter Text

Sub-Topics Using CorEx

Recommender System

INTRODUCTION

The main focus of this project is break down and analyze the under meaning of song lyrics using Natrual Langauge Processing and Clustering and to devide them into topics for further investigations.

TOOLS

scikit learn

python™

CoreX

NumPy

pandas

seaborn

NLTK

Python

GENSIM
topic modelling for humans

WORK FLOW

Songs Dataset from **Kaggle**

Analyze lyrics using **NLP**

Label song topics using **Topic Modeling**

Build classification model by making **topics** as new **Y**

Cluster documents into best fitted clusters **= 5**

Generate **subtopics** for all **6 main** topics using **CorEx**

Build a **recommender system** for singers & associated topics

DATASET

About the Data

- Source: Kaggle
- Song Dataset:
  200K rows 5 columns
- Artist Dataset
  3K rows 6 columns

Chosen Sample

- 20,000 rows

Chosen Sample Criteria

- English songs
- Songs with 20 character

EXPLORATORY DATA ANALYSIS

Null Vlaues

Doesn't contain any nulls

Duplicates (23101)

Removed all duplicates

New Column

Represents the length of the song

Unneeded Columns

Removed unneeded columns like Song and Singer Link.

NATURAL LANGUAGE PROCESSING

Text Cleaning

- Tokenization
- Vectorization (TFIDF)
- Lemmatization
- Removed digits from texts
- Made all text lowercase
- Got rid of all non-english characters
- Removed all punctuations

Extension of Stop Words

Added to the original English stop words common words said in songs.

- Like and Need
- humming and screams: oooh, ooo, ahh
- Explicit language

TOPIC MODELING

NMF

Topic 0
life, heart, world, eye, dream, light, live, fall, end, die
Topic 1
baby, cry, hold, gone, shake, heart, lord, lovin, maybe, kiss
Topic 2
man, little, rock, said, home, boy, woman, roll, old, around
Topic 3
time, mind, last, mine, friend, find, wait, every, gone, life
Topic 4
away, day, walk, far, run, stay, fade, gone, today, turn
Topic 5
feel, give, think, believe, something, nothing, always, everything, really, much

LSA

Topic 0
time, baby, away, feel, day, life, heart, world, think, eye
Topic 1
baby, little, rock, woman, said, man, shake, roll, lord, lovin
Topic 2
rock, man, little, roll, boy, said, woman, old, everybody, big
Topic 3
time, think, mind, feel, give, really, much, believe, mine, last
Topic 4
away, time, day, gone, run, far, walk, home, stay, christmas
Topic 5
away, feel, give, rock, something, think, chorus, run, walk, roll

LDA

Topic 0
baby, man, little, time, said, home, rock, around, ay, think
topic 1
holiday, careless, boom, wah, satellite, prisoner, ing, groovy, divide, messed
Topic 2
time, away, feel, life, heart, world, eyes, us, day, find
Topic 3
worlds, heroes, waits, santa, jane, collide, dancer, sale, woe, digging
Topic 4
johnny, bigger, hoo, hallelujah, glad, lee, kissing, gravity, arpet, peter
Topic 5
wherever, rider, runaway, owe, honky, ghosts, voodoo, sally, gal, dum

6 TOPICS

- Reality: Topic 0
- Heartbreak: Topic 1
- Coming of Age - Growing up: Topic 2
- Friendship: Topic 3
- Home: Topic 4
- Desire: Topic 5

WORD CLOUD

For each topic

For all documents

WORD FREQUENCY

CLASSIFICATION MODEL

Solving existing issues before model building:

- Imbalanced classes issue solved by using SMOTE

  Class 0: 5527, Class 1: 1288, Class 2: 6850, Class 3: 1565, Class 4: 2176 and Class 5: 1367

Splitting Data:

- Data was split by making the newly found topics from topic modeling as Y & the lyrics as X

Logistic Regression

- Training Score:0.9948
- Validation Score:0.9276
- Test Score: 0.9271
- Error Rate: 0.0728

01

Random Forest

- Training Score:1.0
- Validation Score:0.7804

02

Decision Tree

- Training Score:1.0
- Validation Score:0.8552

03

Gradient Boosting

- Training Score:0.9287
- Validation Score:0.8106

04

Multinomial

- Training Score:0.8660
- Validation Score:0.6673

05

CONFUSION MATRIX & PREDICTION

You've Got a Friend in Me
Toy Story

```
You've got a friend in me You've got a friend in me When the road looks rough ahead And you're miles and miles From y
our nice warm bed You just remember what your old pal said Boy, you've got a friend in me Yeah, you've got a friend i
n me You've got a friend in me You've got a friend in me You got troubles, I've got 'em too There isn't anything I wo
uldn't do for you We stick together and see it through 'Cause you've got a friend in me You've got a friend in me .
The topic of this song is:  ---> : ( Friendship )
```

CLUSTERING

Using KMeans Clustering

- Best number of clusters is 5
- Used PCA

CLUSTERING

Using DBSCAN Clustering

CLUSTERING

Using TSNE

WORD FOR EACH CLUSTER

WORD FOR EACH CLUSTER

SCATTER TEXT

Multi-Category

Only plotted the most popular topics among singers

- Coming of Age - Growing up
- Reality

CorEx

Derived sub-topics by using Corex and applied anchors to each document in the main topic

- Reality: Life & Die
- Heartbreak: Cry & Heart
- Coming of Age - Growing up: Little & Old
- Friendship: Friend
- Home: Away & Home
- Desire: Think & Believe

- Reality: Reality of death
- Heartbreak: Loss of someone
- Coming of Age - Growing up: Local band
- Friendship: The end of a friendship
- Home: Home abandonment
- Desire: Repentment

RECOMMENDER SYSTEM

Recommender system for singers

- Used SVD
- Recommends possible collabs between similar singers
- Since 311, AC/DC and Alanis Morissette are similar based on the topics they sing about. A collab would be successful

RECOMMENDER SYSTEM

Recommender system for singers & topics

- Used SVD
- Recommends possible topics between similar singers and topics
- Since 311, AC/DC and Alanis Morissette are similar based on the topics they sing about. They should try out similar topics like: Friendship, Home, Reality and Desire

RECOMMENDER SYSTEM

After building the recommender system for both singers and topics. It suggests the following:

- Singers who usually sing about Coming of Age - Growing up they should try to sing about Friendship & Home next

- Singer Alanis Morissette is most similar to Singer 311 and Alanis Morissette should try out the topics 311 usually sings about and they are: Coming of Age - Growing up, Home, Reality, Friendship and Desire

Thank you for listening

Any questions?

Back