

Music classification

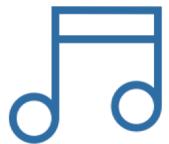
Module 4 project

Paul-Renaud Raymond
June 2019

What – analytically – makes a song qualify as a particular genre?

Objective: analysis of music text data to develop and test a classification model to predict the genre of a song based on its lyrics

Executive Summary



Analysis of lyrics
from **music**
listening history
over three years



Typical **stop words**
dominate the
frequency and
n_grams distribution



**High class
imbalance** with the
'Genre' feature;
predominately Hip
Hop / Rap



Model performance
effected by class
imbalance; **poor
predictions of
minority classes**



Resampling and
Deep Learning
among techniques
to **improve model
performance to
89%**

Exploratory data analysis: Persona listening metrics



Persona A

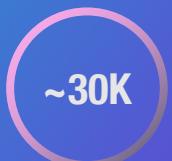
XXXXX
XXXXX
XXXXX
XXXXX



Persona:

Paul-Renaud Raymond

33ish years
Black / Haitian
Miami, FL
Graduate education
\$100K + Income
Liberal



played
songs



times
Drake
played



times
'Electricity'
played



Different
genres



Persona C

XXXXX
XXXXX
XXXXX
XXXXX

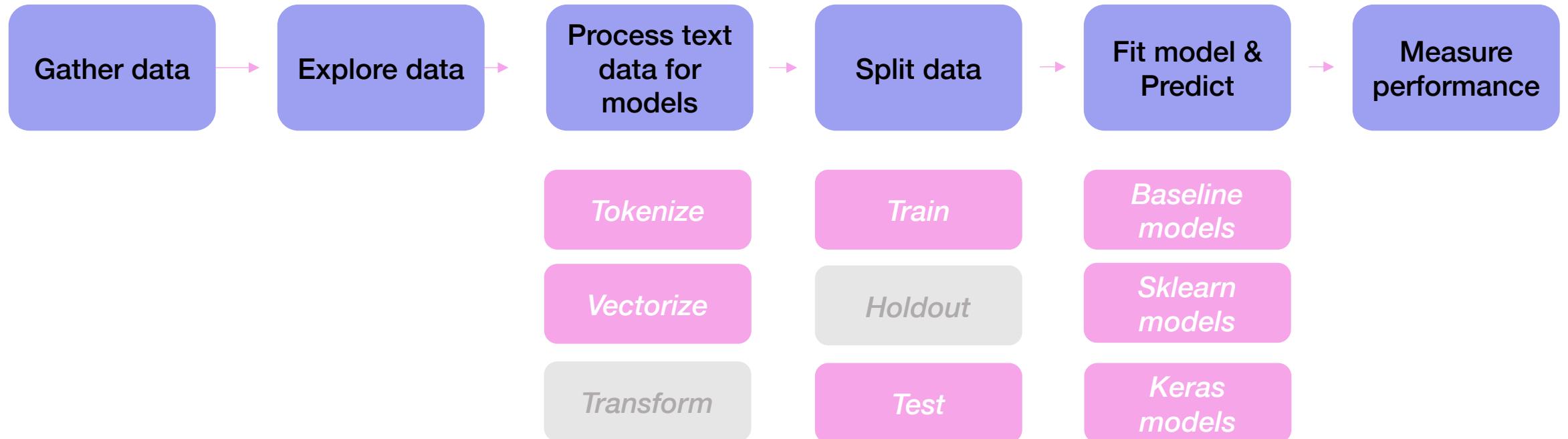


Business use cases

- Cohort analysis of music activity (with time-series analysis)
- Build persona profiles from user music activity, available to third-party purchasers
- Improve automation of feature classification ('Genre')
- Leverage persona and listening history to identify influential personalities, messaging
- Leverage persona and listening history to identify influential messaging

What genre is the appropriate distinction for this song?

Work flow for Deep Learning Natural Language Processing analysis



For select models

NLP Classification: Performance across models

	Accuracy	Precision	Recall	F-1	N-size
Baseline NB	44%	19%	44%	27%	105
Baseline RF	51%	35%	51%	40%	102
SMOTE NB	85%	89%	85%	82%	439
SMOTE RF	91%	91%	91%	91%	439
Undersampled NB GridSearch	14%	30%	14%	13%	51
Undersampled RF GridSearch	12%	5%	12%	7%	51
LSTM	49%				41
GRE	49%				41
Bidirectional LSTM	89%				41

Various modifications and further classifications to follow

Next steps	Approach
Create binary classifications (Hip/Hop / Not Hip/Hop)	<ul style="list-style-type: none">• Create new binary feature to determine performance of classification of hip hop songs and non-hip hop songs
Implement time-series and test for statistical significance	<ul style="list-style-type: none">• Determine whether results and insights are altered with consideration of time periods