Hip Hop Subgenre Classification

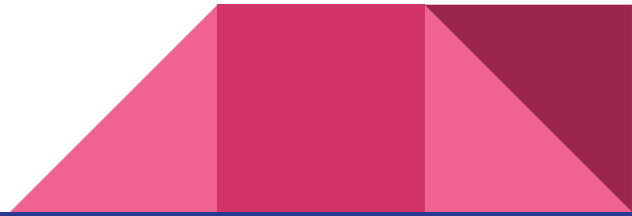
Temporal vs Persistent Song Features

“In the simplest terms, if Spotify notices that two of your favorite songs tend to appear on playlists along with a third song you haven’t heard before, it will suggest the new song to you.”

Adam Pasick after interviewing Mathew Ogle, Spotify’s Product Director (source)

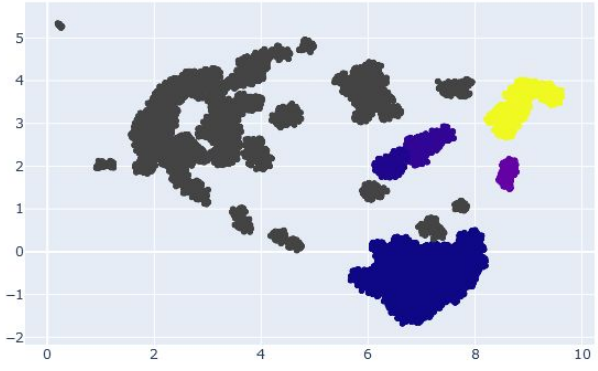
Recommendation Systems

- Largely based on user-generated data
- Likelihood of being recommended is based on notoriety or methods of outreach
- Data poor
  - New streaming service
  - Brand new music
  - Independent Artists
- Increase the granularity of analysis to better recommend solely based on content

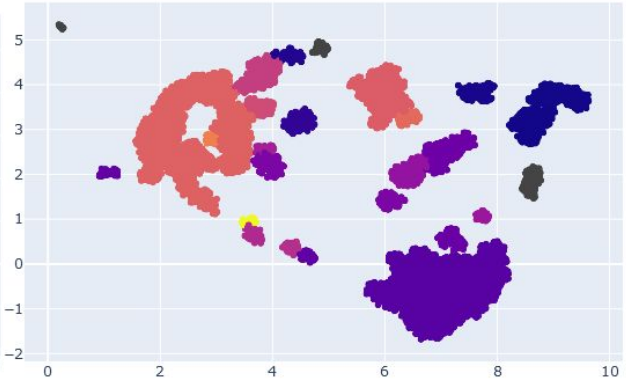


Genre Density

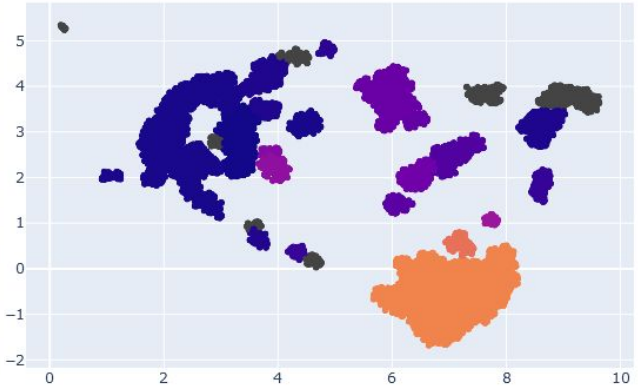
Classical



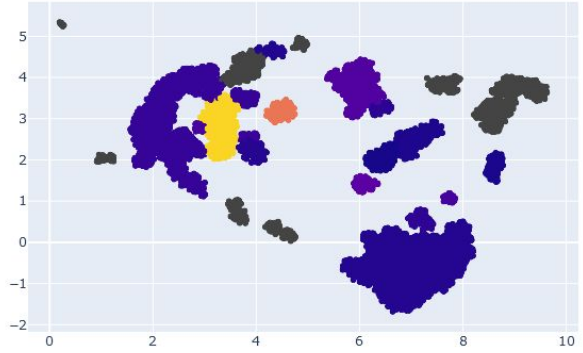
Pop



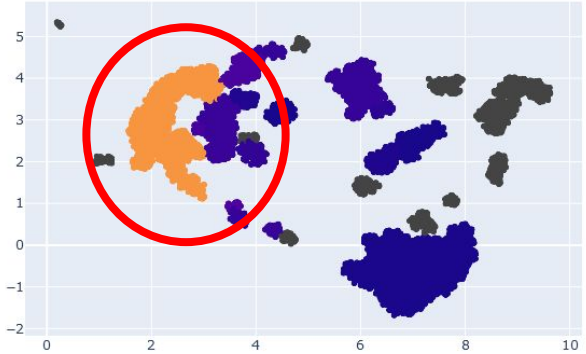
Rock



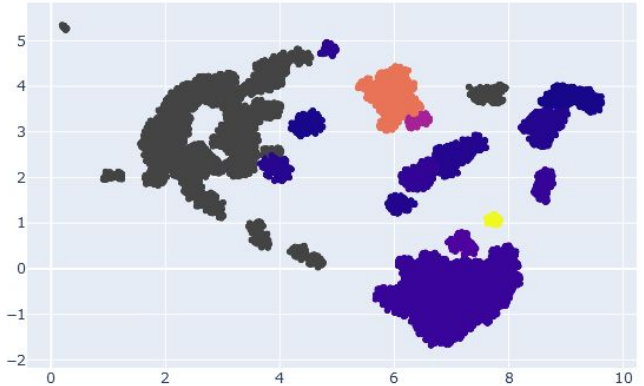
Latin



Rap

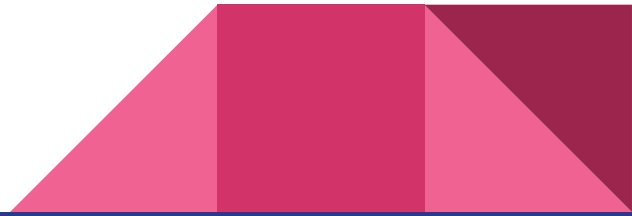


Indie-Folk



Hypothesis

If we leverage the power of LSTM (long short term memory), we can preserve temporal nature of the data to prevent information loss which should allow us to segment the data further and increase the granularity of analysis.



The Data

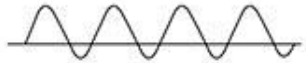
- Hip Hop subgenres chosen based on this article
  - Alternative Hip-Hop, Battle Rap, Conscious Rap, Crunk, East Coast Hip-Hop, Gangsta Rap, Hyphy, Southern Rap, West Coast Hip-Hop, Trap Music
  - Later consolidated to maintain class balance
- Lyrics gathered by using the Lyrics Genius Web API
  - Over 10,000 originally gathered
  - Filtered down to 5,000 after cleaning
- Songs tagged by artist genre according to Last.FM and Spotify using their respective Web APIs
  - Searching genre tag by song in Last.FM is too messy
  - Impossible to search genre tag by song in Spotify
- Audio Analysis data gathered from Spotify
  - 12 dimensional timbre vector sequences of varying lengths.

Timbre

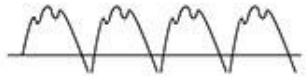
API documentation:

- Timbre is the quality of a musical note or sound that distinguishes different types of musical instruments, or voices.
- Timbre vectors are best used in comparison with each other.

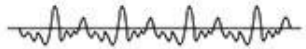
Tuning fork

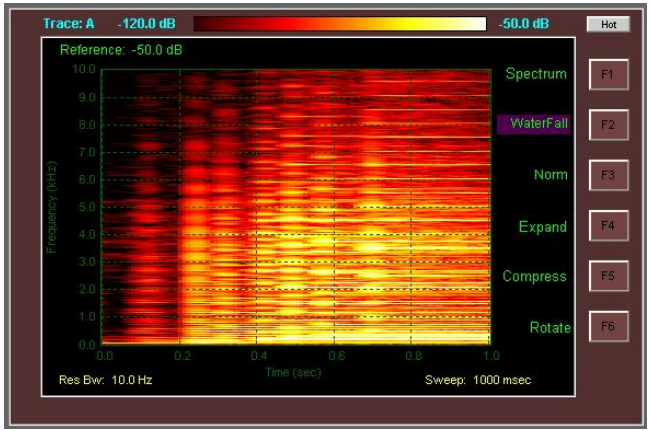


Flute



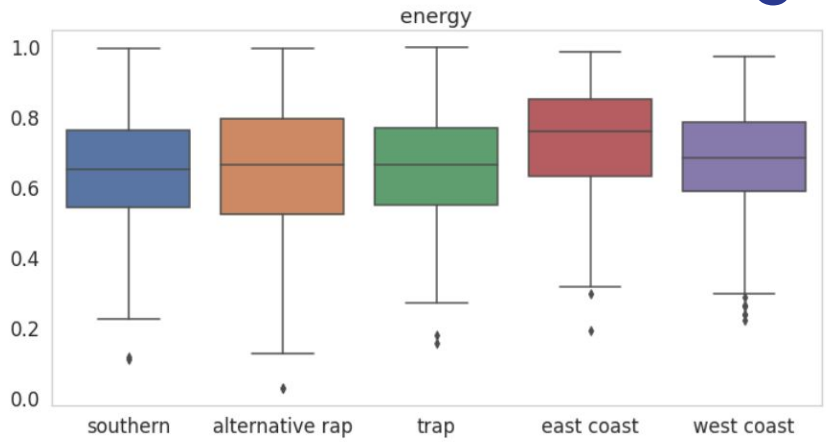
Voice

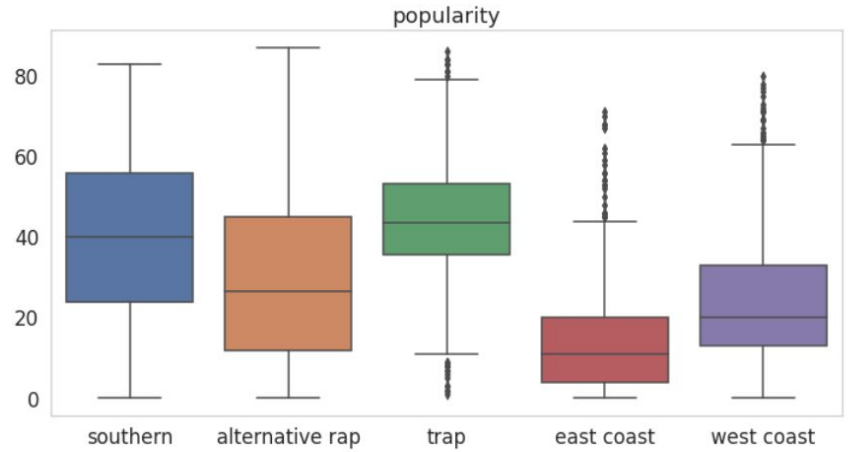


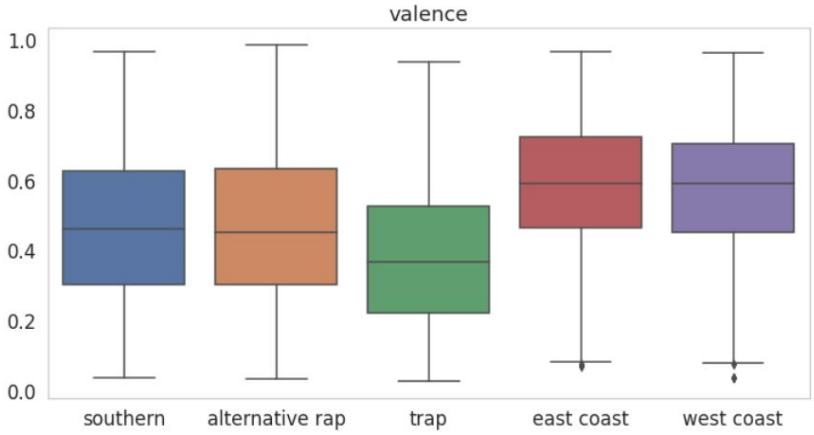


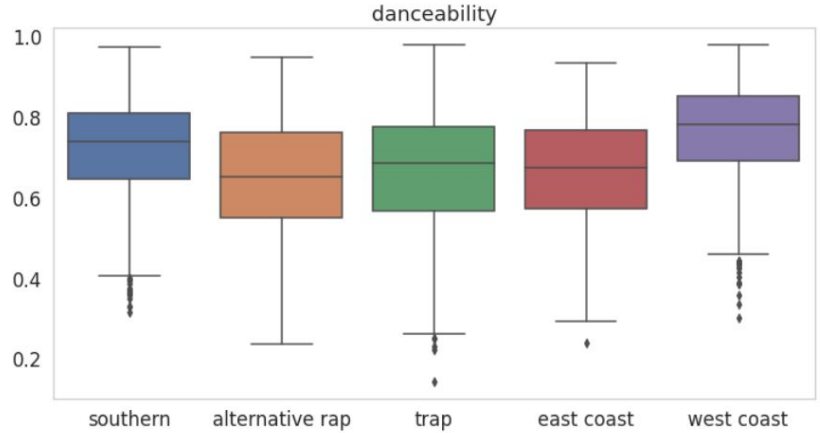
EDA

Song Features



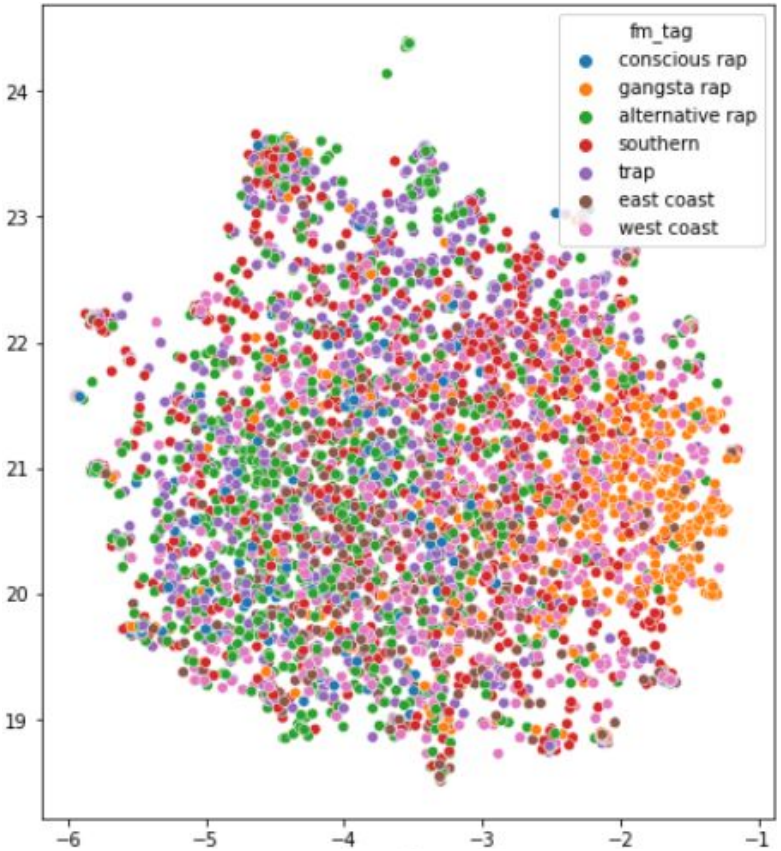




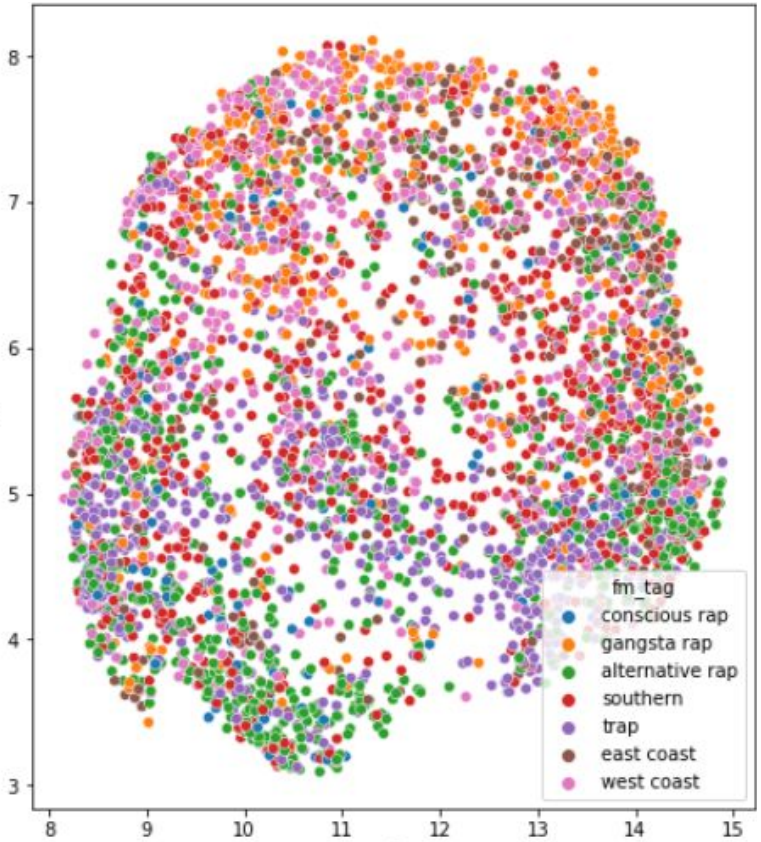


love
need
never
feel
one
say
way
keep
right
take
don
think
money
say
baby
right
tell
love
one
never
come
take
put
give
need
one
come
never
life
take
love
God
still
right
real
say
way
yo
ayy
love
feel
put
need
say
tell
come
hit
take
think
right
hit
one
yo
come
G
keep
take
put
hood
still
fool
give
life
love
never
way
take
game
hit
one
come
keep
say
u
money
right
put
still
need

TF-IDF UMAP



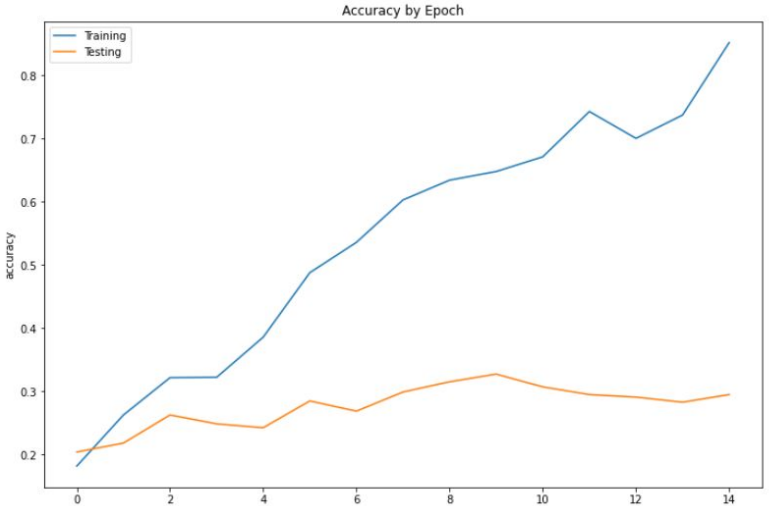
Timbre Covariance UMAP

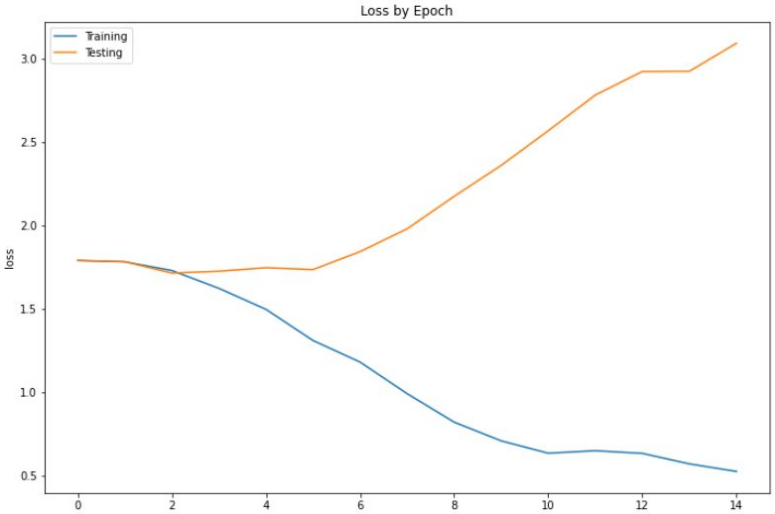


If order matters, then LSTM will be able to differentiate the classes

Simple NLP LSTM

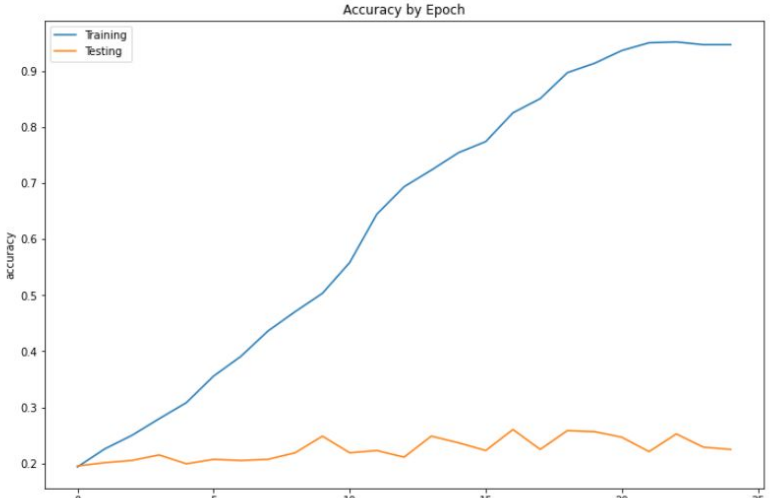
- Lyrics cleaned, stemmed, one hot encoded, padded, and fed into an embedding layer
- 6 classes equally represented
- ~2500 observations
- Longest sequence ~800
- Categorical_crossentropy loss function
- Adam optimizer
- Custom Gensim Word2Vec model yielded similar results

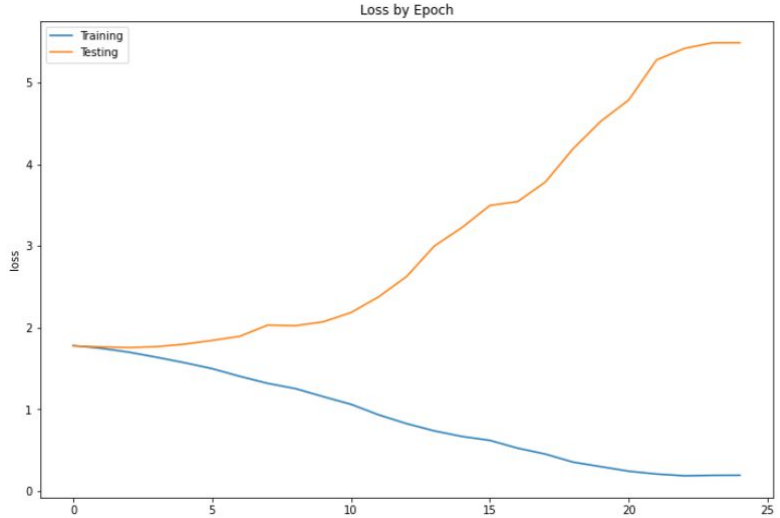




Timbre Sequence LSTM

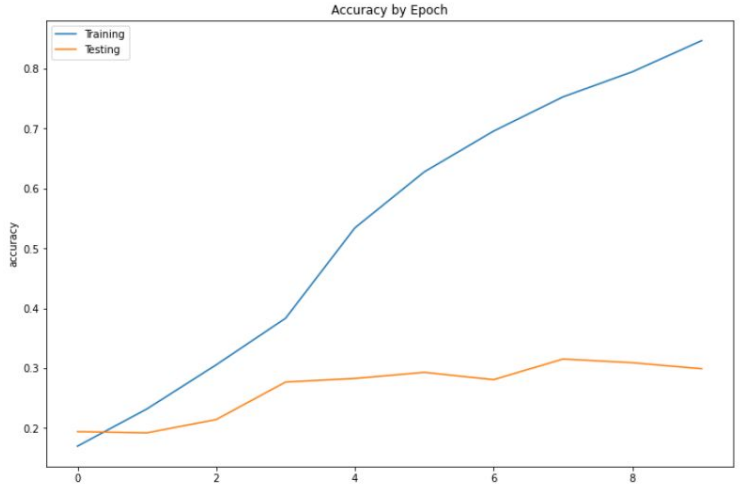
- Vector sequences padded and fed into double LSTM
- Longest sequence ~2300 vectors long
- Worse results than NLP

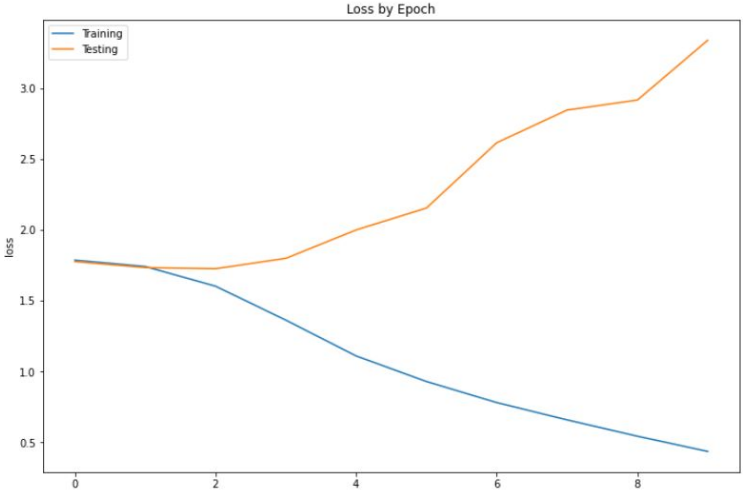


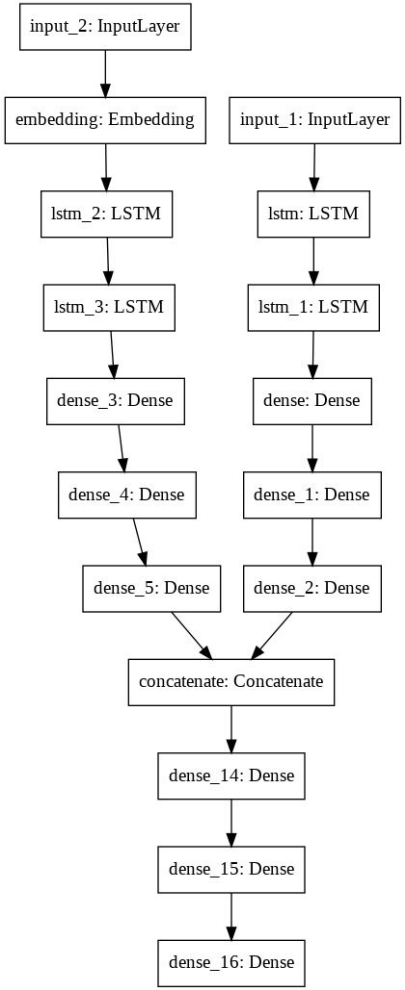


Dual input LSTM

- Marginally better than either on their own
- Model results imply a poor quality of data



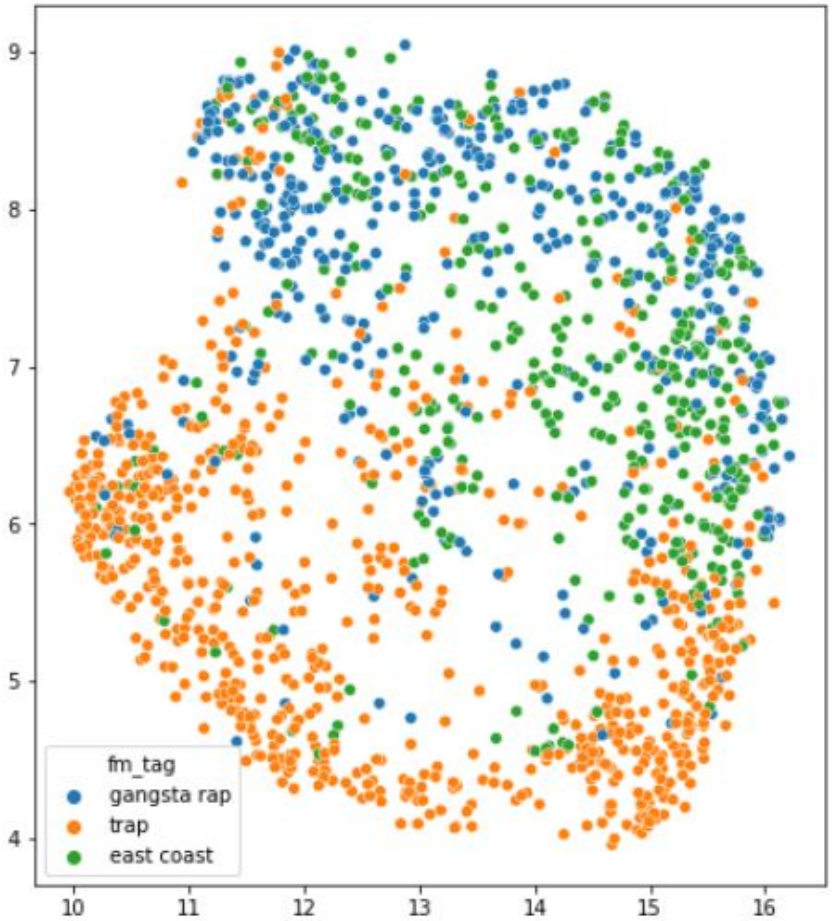




Back to the Drawing Board

- 4 different types of tagging strategies
  - Spotify vote tagging and consolidation
    - ‘cali rap’ ⇒ ‘west coast’
  - Last.FM artist tag
  - Spotify NaNs filled by Last.FM
  - Last.FM NaNs filled by Spotify (previously used)
- Filter out highest variance classes using UMAP
- Test models and check confusion matrices

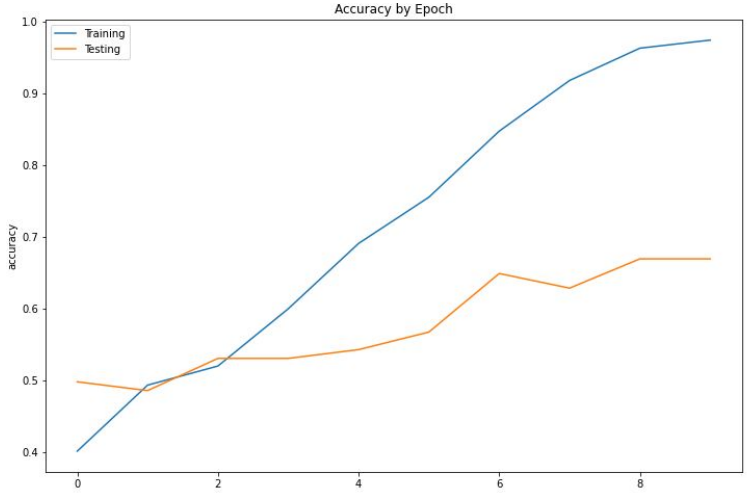
Timbre Covariance UMAP

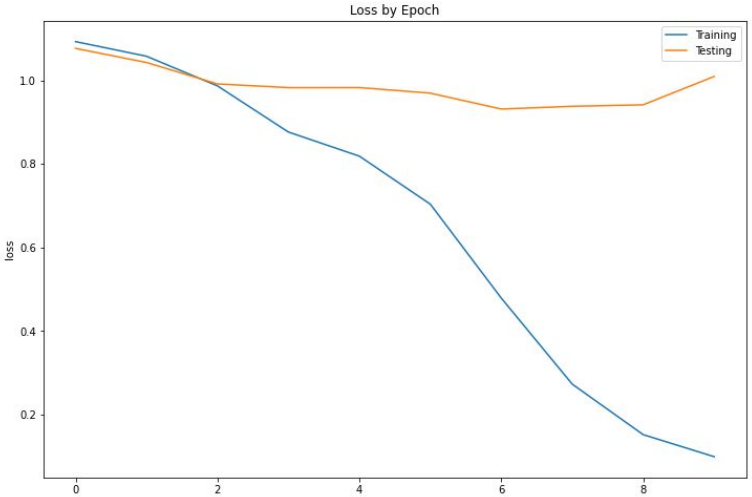


Multi Input Results

- ~1200 observations
- Using the Last.FM tags filled by Spotify
- Classes ‘west coast’, ‘trap’, ‘alternative rap’ yielded similar results
    - Solely Spotify vote tagging

	trap	east coast	gangsta rap
trap	0.69	0.21	0.12
east coast	0.15	0.62	0.19
gangsta rap	0.16	0.16	0.69

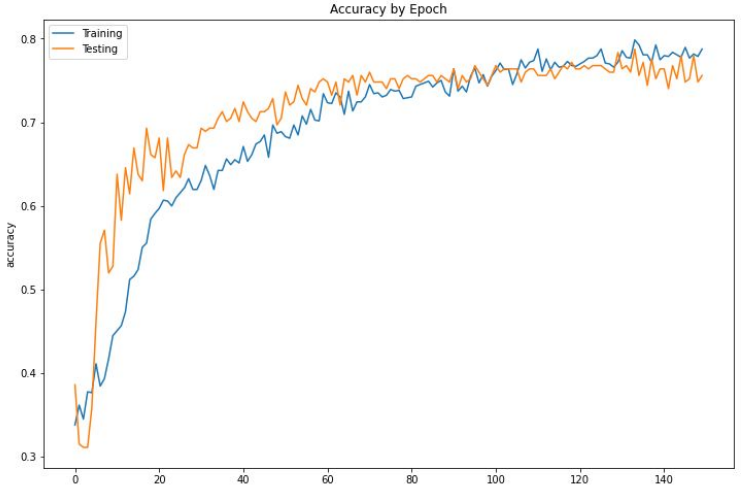


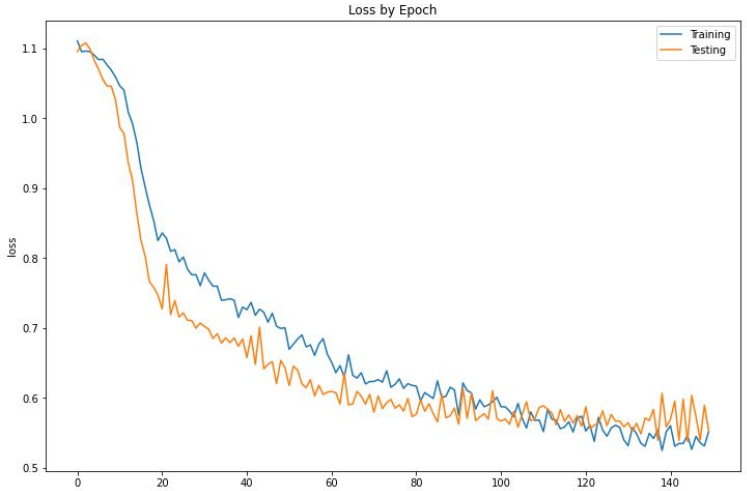


Persistent Features

- Dense neural network
- 90 features
- Variance-covariance matrices of the same vector sequences
- Categorical_crossentropy loss
- Adam optimizer
- Much faster
  - 150 epochs
- Supervised learning yielded similar results
  - Random Forest, XGBoost
  - Same for Tf-idf n-gram range (2,5)

	trap	east coast	gangsta rap
trap	0.93	0.05	0.10
east coast	0.05	0.69	0.24
gangsta rap	0.02	0.26	0.66





Persistent Features 6 classes

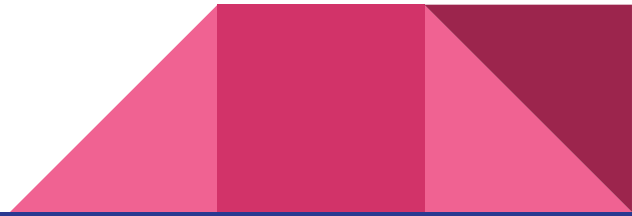
- ~0.5 validation accuracy
- Prediction probabilities provides a kind of ‘genre profile’
- Take the mean grouped by artist to get an overall artist ‘genre profile’

artist	alternative rap	southern	trap	west coast	east coast	gangsta rap
Above The Law	0.04	0.13	0.01	0.21	0.21	0.40
Berner	0.20	0.14	0.13	0.22	0.15	0.16
Compton's Most Wanted	0.04	0.04	0.01	0.30	0.11	0.49
E-40	0.07	0.31	0.08	0.33	0.05	0.16
Eshon Burgundy	0.13	0.09	0.01	0.18	0.45	0.14
ILL Bill	0.08	0.10	0.07	0.15	0.36	0.24
Kanye West	0.19	0.26	0.13	0.11	0.26	0.05
MC Ren	0.04	0.11	0.01	0.21	0.19	0.44
Requiem Inc.	0.19	0.16	0.17	0.17	0.20	0.10
South Central Cartel	0.06	0.12	0.06	0.24	0.17	0.35
The Click	0.04	0.09	0.01	0.38	0.06	0.43
YNW Melly	0.20	0.21	0.47	0.08	0.02	0.02
twikipedia	0.16	0.17	0.45	0.11	0.04	0.07

title	popularity	artist	alternative rap	southern	trap	west coast	east coast	gangsta rap
Hope	83	XXXTENTACION	0.166409	0.115354	0.674799	0.030976	0.005632	0.006830
Suicidal (Remix) [feat. Juice WRLD]	81	YNW Melly	0.132681	0.372167	0.374395	0.091054	0.013066	0.016637
POWER	78	Kanye West	0.197127	0.254508	0.453488	0.046055	0.026065	0.022756
The Way Life Goes (feat. Oh Wonder)	78	Lil Uzi Vert	0.270295	0.273906	0.298878	0.111623	0.023226	0.022071
10 Freaky Girls (with 21 Savage)	76	Metro Boomin	0.276578	0.235419	0.340459	0.091222	0.024010	0.032312

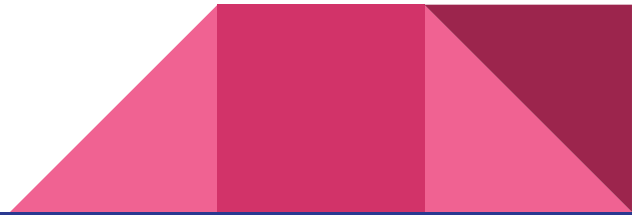
Conclusions

- Preserving the order doesn’t seem to be relevant when classifying genre
- A supervised model provides a framework to plug new data in faster and compare to existing observations.
- Reliably labelling a song by subgenre may not be possible



Moving Forward

- Try more combinations of multi input networks
  - Using all four - Timbre LSTM, NLP LSTM, Timbre Covariance, and TF-IDF
- Spend more time fitting a better Word2Vec
- Address overfitting of Random Forest and XGBoost
- Try larger umbrella genres instead of specific subgenres
- Quantify variance in the classes instead of UMAP
- Guess the key (Ab, G#, etc.) to seamlessly intertwine songs into the recommendation queue
- More data



Questions?

Thank you