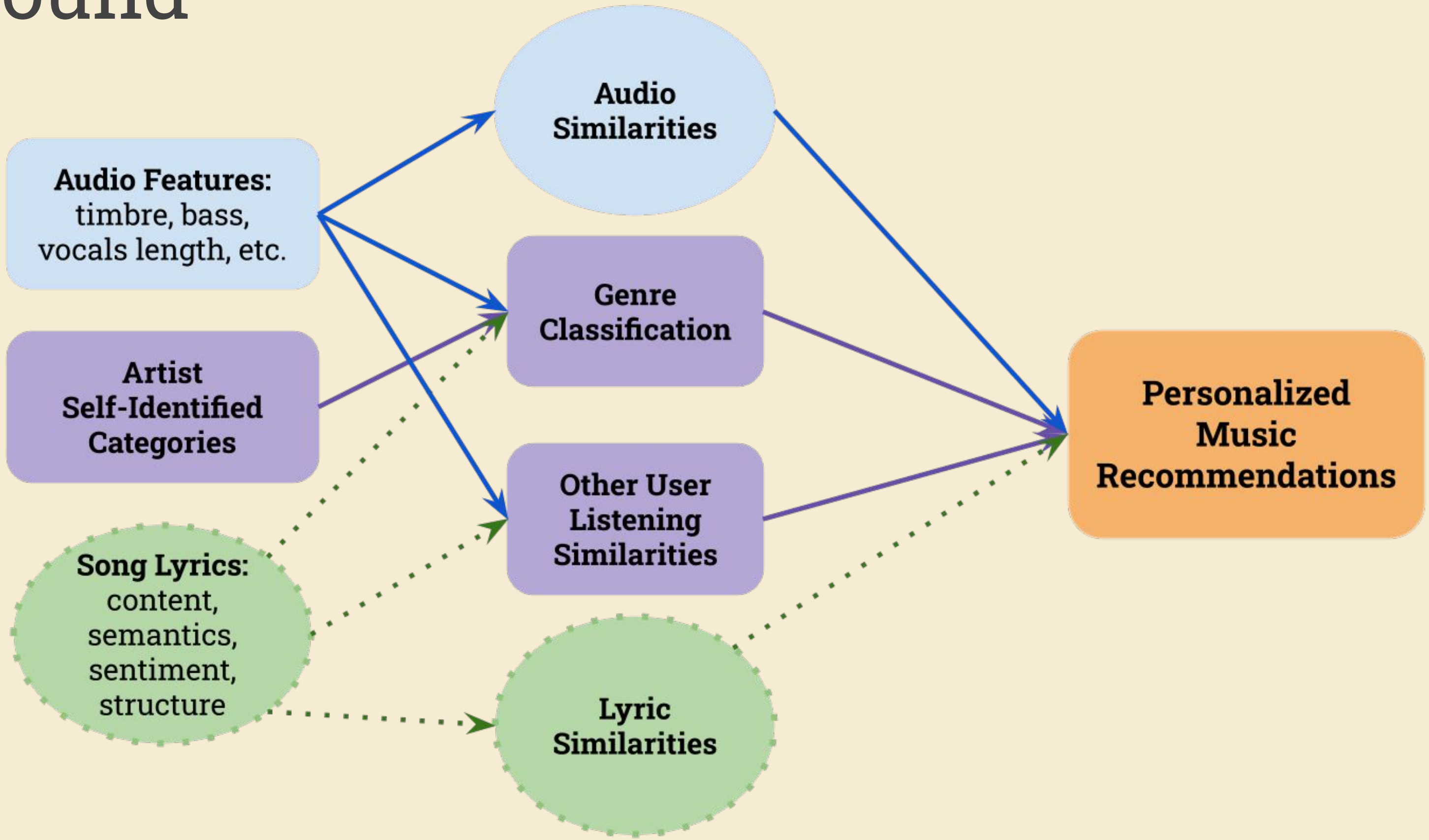
Beyond The Beat: Leveraging Lyric Content and Sentiment to Classify Songs Into Genres

Courtney Maynard

December 5th, 2023

Background


Audio Features: timbre, bass, vocals length, etc.
Artist Self-Identified Categories
Song Lyrics: content, semantics, sentiment, structure
Audio Similarities
Genre Classification
Other User Listening Similarities
Lyric Similarities
Personalized Music Recommendations

Past Works: Lyric Sentiments

- ➢ One study detailing combining lyric content with lyric sentiments
  - Utilized Affective Norms For English Words (ANEW) to classify emotional valence of songs as positive or negative
  - Bag of Words method on only 2500 words → direct matches with ANEW database
  - **No signal for the song genre**
- ➢ Poetry generation research utilizes ‘SentiStrength’
  - Estimates the negative and positive sentiment strength of a segment of text
  - Scaled from -4 (extremely negative) to 4 (extremely positive)

The text 'I love Natural Language Processing!'has positive strength **4**

Approximate classification rationale: I love[3] Natural [proper noun] Language [proper noun] Processing [proper noun] ![+1 punctuation emphasis]

Past Works: Embeddings and Architecture

- Custom and pre-trained GloVe Embeddings
- Word2Vec
  - With simple embeddings
  - With TF-IDF vector embeddings
- A majority of studies utilize bag - of - words
  - **Does not grasp the structural information of the song, only generalizes the content of the song**

- Naive Bayes, Support Vector Machines, Decision Tree Classifiers, Logistic regression, Multi-Layered Neural Networks, etc.
- LSTM and HAN for maintaining song structure

Research Gaps

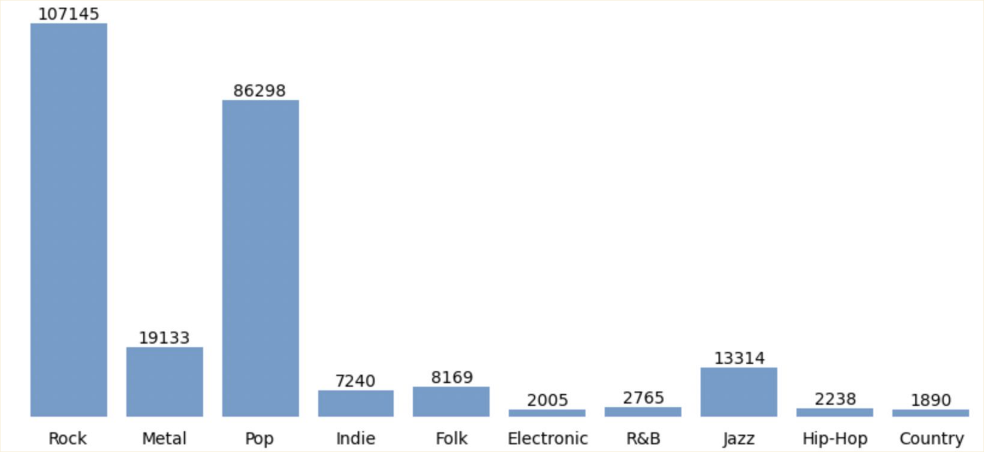
1. Number of Genres and Intact Corpus
   - ➢ A non bag-of-words approach with large dataset
2. Maintaining Lyric Structure during classification
   - ➢ LSTM and HAN model architectures
3. Custom Word2Vec Model on a Music-Specific Corpus
   - ➢ Line - level word2vec embeddings
4. Ensemble Model Combing Lyric Content with Lyric Semantics
   - ➢ Specifically with a non bag-of-words approach
   - ➢ Line-level analysis of both feature sets
   - ➢ SentiStrength comprehensive sentiment analysis

Dataset

Number of Songs	290,183
Languages	34
Unique Artists	1,152
Unique Genres	10

Number of Songs	250,197
Languages	English
Unique Artists	1,071
Unique Genres	10

- From Kaggle
  - Scraped from Spotify API, AZLyrics, three other Kaggle Datasets, and a Textract Hackathon dataset



Data Pre-Processing

- Cleaning Data
  - Duplicates
  - Lyrical Structure Markers
  - Weblinks to artists or song lyrics
  - Song attributions
- 1500 random songs per genre
  - Consistent representation of each category (no oversampling / undersampling)
  - Decreased set for computational complexity

```
trainSetRock.head()
```

	Artist	Song	Genre	Language	Lyrics
78168	franz ferdinand	swallow smile	Rock	en	I rise and curse the waking day\nCurse the gri...
155890	ry cooder	big bad bill is sweet william now	Rock	en	In the town of Louisville they got a man they ...
52293	daryl hall	let me be the one	Rock	en	We've been through love time & time again\nTri...
130715	mullins, shawn	we run	Rock	en	in the town that i was born in\nthere rides a ...
169162	stray cats	reckless	Rock	en	Well I met you in the backroom\nYou had lipsti...

Custom Word2Vec Embeddings

- Custom model based on corpus of 200,000+ songs
    - 1. Split all songs into lines
    - 2. Tokenized each line of the songs
    - 3. Train and save of embeddings
        - Window-size of 10 words
        - 100 dimensional vector
- Kept potential stopwords, punctuation
- Applied Word2Vec model to create line-level embeddings for songs
    - Average of embeddings to create document representation
    - TFIDF vectorization to create document representation

Custom Word2Vec Embeddings: Example

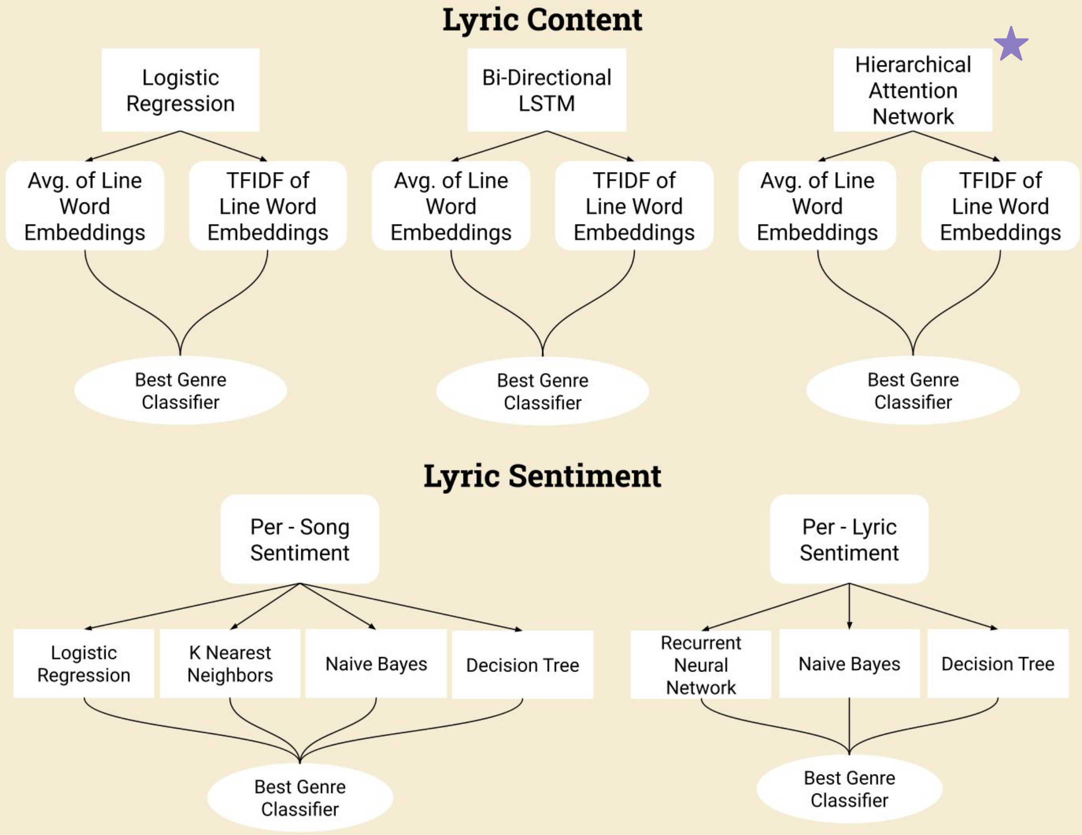
(Metal) Radio - Theatre of Tragedy

```
'Electric broadcast\nThe new transmission waves\nTurn the dial\nReceive the news\nTransistor radio\nWe've tuned in to the ether melody\nThe deadpan voice I
want to hear\nIt's bright and clear and full of energy\nReceive the tunes\nMusic won't stop\nElectro pop\nCommercial workshop\nI want your time, I need your
time to make a rhyme\n\nEther melody news\nOn the radio now\nFor you and me\nAntennae beaming sound, news and speech\n\nApparatus signal it out\n'Let it be
so'\nMarconi's words\nCommunication\nInformation\nWe're in the building of the wireless voice\nRadar, television\nStatic morse\nThe interference\nIncoherenc
e\nScrambled signals\nI want your waves\nSpectral sound\nTune in, tune out\n\nOn the radio now\nFor you and me\n\nEther melody news\nRadio for you and me\nT
he airwaves are fully free\nOscillating energy\n\nRadio for you and me\nOn the radio now\nEther melody news\nFor you and me\n\n\n\n'
```

```
['Electric broadcast',
 'The new transmission waves',
 'Turn the dial',
 'Receive the news',
 'Transistor radio',
 "We've tuned in to the ether melody",
 'The deadpan voice I want to hear',
 "It's bright and clear and full of energy",
 'Receive the tunes',
 "Music won't stop",
 'Electro pop',
 'Commercial workshop',
 'I want your time, I need your time to make a rhyme',
```

```
array([ 0.81820065, -0.84367   ,  1.8573768 ,  1.3505409 ,  0.47247204,
       -0.58283484,  0.61017805, -0.2093388 , -0.7920564 , -1.2880259 ,
        0.98495364,  1.5693108 , -1.479212  ,  0.5339909 , -4.976462  ,
        0.30785906,  1.3419511 ,  1.6382881 , -0.30673197, -1.0013002 ,
        1.7983503 ,  2.6538894 , -0.53152615, -0.25093135,  0.01624241,
       -0.84227353,  1.1582632 ,  0.3622528 , -0.8322013 ,  1.8166233 ,
        0.06263822, -1.4998399 ,  1.5758483 ,  0.6874524 ,  0.071752  ,
        0.39371368,  1.211269  , -3.0777454 ,  1.5564207 ,  1.3760288 ,
       -0.36931038, -0.8194323 , -0.4363985 , -3.5581954 , -4.0217905 ,
        2.8288076 ,  1.1315671 ,  1.4296598 , -0.67108023, -3.724654  ,
        2.5100942 ,  2.2338796 ,  1.6564642 ,  0.27263454,  2.539646  ,
        0.4593381 ,  1.566182  ,  0.3221223 ,  0.3649686 , -0.72197235,
       -0.62446946, -2.5633214 ,  0.07215827, -0.11589453, -2.3141537 ,
       -0.6950865 ,  0.5241868 , -4.6425066 , -2.9629698 ,  1.5470312 ,
        0.63070697, -1.5415114 ,  1.6130116 ,  3.3917701 ,  3.0239553 ,
       -1.0316795 ,  0.89037263,  0.48290145,  1.3781638 ,  1.4156793 ,
       -0.930899  ,  0.08161373,  1.0111905 ,  0.07994334,  1.3133733 ,
        0.16550559, -0.17067051, -0.7998781 ,  1.4522713 , -1.3283447 ,
        0.05171934, -0.42863998,  1.9529485 , -0.10906661,  2.7597723 ,
       -3.4920032 , -0.23415339, -0.34250662,  1.2611059 ,  1.3706193 ],
      dtype=float32),
```

Methodologies



Methodologies: Ensemble

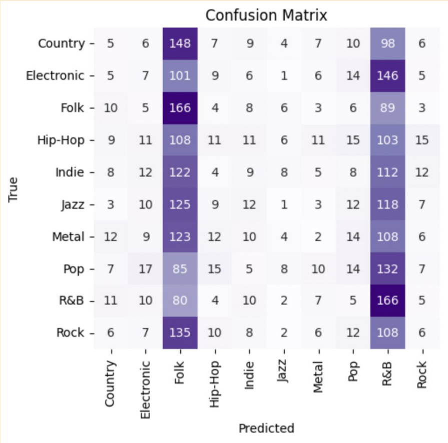
Ensemble Models

Lyric Content - Logistic Regression	Lyric Content - Bi-LSTM	Lyric Content - HAN
Best Song Content Representation + Best Per-Song Sentiment	Best Song Content Representation + Best Per-Song Sentiment ★	Best Song Content Representation + Best Per-Song Sentiment ★
Best Song Content Representation + Best Per-Lyric Sentiment ★	Best Song Content Representation + Best Per-Lyric Sentiment ★	Best Song Content Representation + Best Per-Lyric Sentiment ★

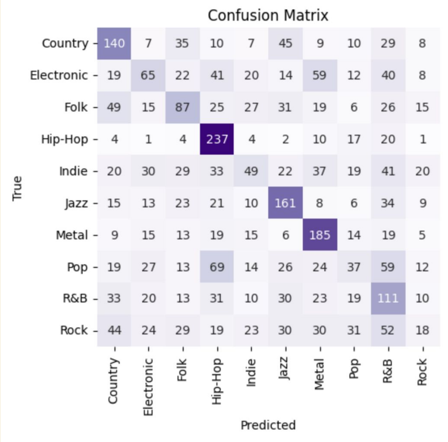
- ➢ Uses maximum weight probability to decide genre
- ➢ Even weighting for lyric content and lyric sentiment

In Progress Results: Lyric Content

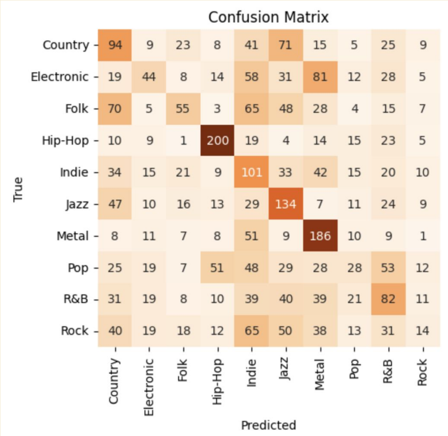
Avg of Line Embeddings, **Logistic Regression**, 12.9% val accuracy



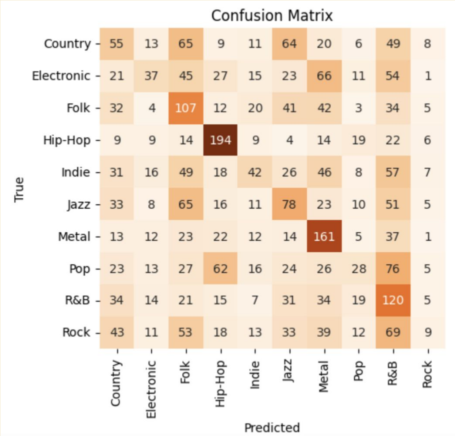
TFIDF of Line Embeddings, **Logistic Regression**, 36.33% val accuracy



Avg of Line Embeddings, **Bi-Directional LSTM**, 31.27% val accuracy

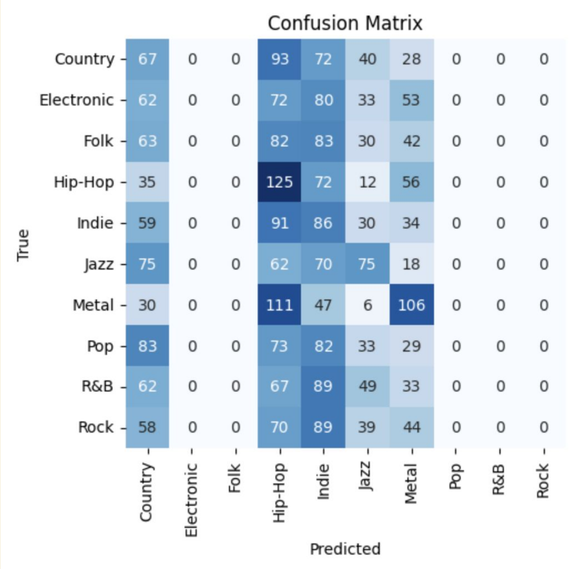


TFIDF of Line Embeddings, **Bi-Directional LSTM**, 27.7% val accuracy

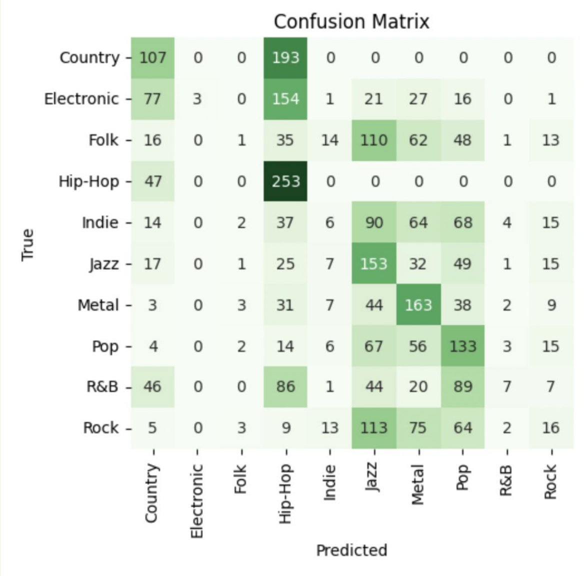


In Progress Results: Lyric Sentiment

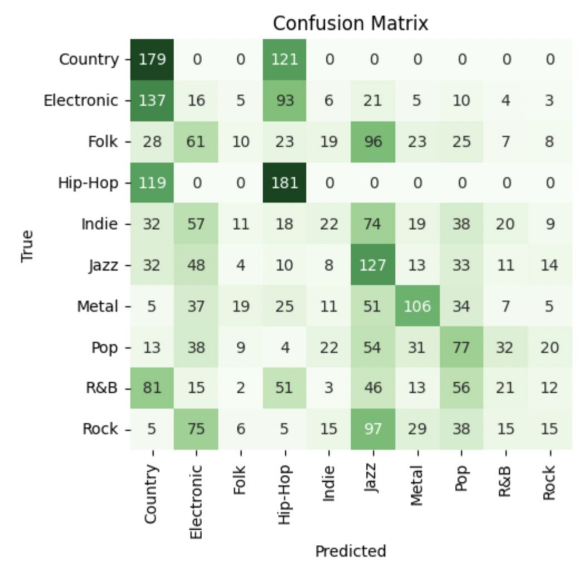
Song - Level, Decision Tree Classifier, 15.3% val accuracy



Line/Lyric - Level, Recurrent Neural Network, 28.07% val accuracy



Line/Lyric - Level, Naive Bayes Classifier, 25.13% val accuracy



Questions?

The End. Thank you!