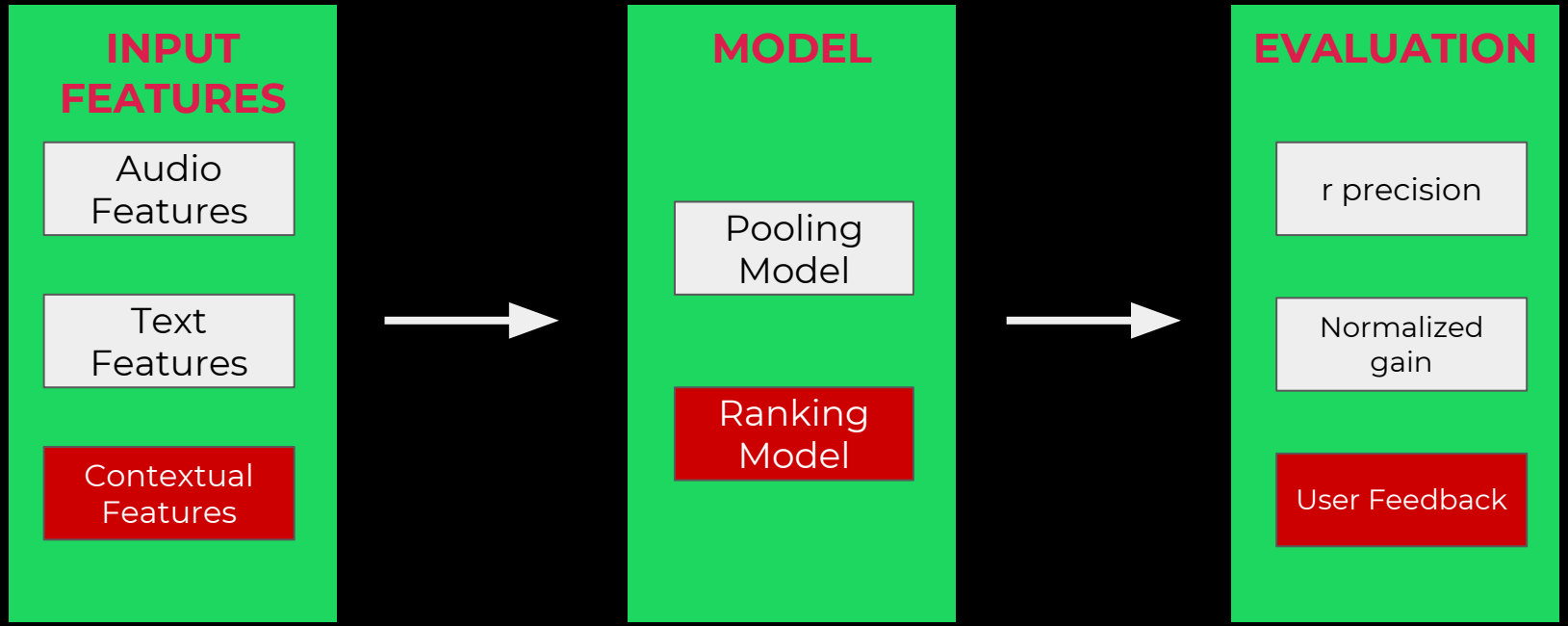
APCOMP 297r Capstone Project


# Milestone 3 presentation

Mehul Smriti Raje, Timothy Lee
Spandan Madan, Benjamin Sanchez-Lengeling

**TF:** Patrick Ohiomoba

# Next Steps, next month



# Next Steps

1. Experiments on Contextual Data

2. RECSYS pipeline

3. Smarter Ranking

4. Ground truth for Context Tasks (UX)

Full todo-list:
https://docs.google.com/document/d/1ASbP8S64QgXvXiZCmUl9KlCSquCuzyeAgNWRdLwQ4uo/edit#

# Tentative Aim:

RECSYS conference submission (April 30th deadline, May 7th)
or
RECSYS challenge submission (July)

We have evaluation metrics
We have a pooling model

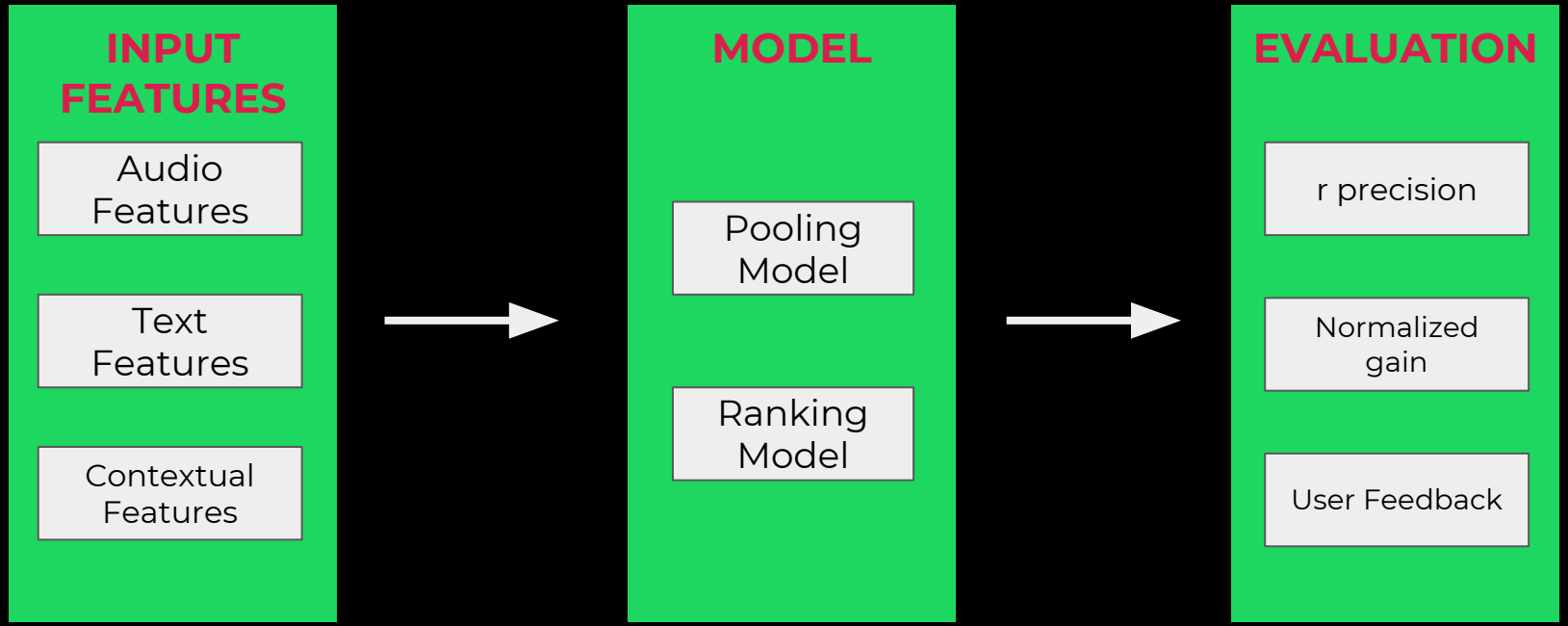
We need to improve our ranking
We want to think of intent-based completion of playlist
We need to optimize our model (embedding/sources/weights)

To build intent, we are brainstorming the UX and how to leverage the Genius/Youtube/lyrics data.

# Progress Overview:
## Generating playlists + evaluating them

1. Evaluation
2. Pooling
3. Ranking
4. UX

# Model schematic (Pipeline)



# Evaluation: experimental setup

Incomplete playlist of length k (k =5,10,25,50,100)
Held out tracks G
A set R of 500 recommend tracks

Metrics on:

- Track Level
- Artist Level
- Contextual Level (Future directions?)

# Evaluation: r-precision (good pooling)

$$\text{R-precision} = \frac{|G \cap R_{1:|G|}|}{|G|}$$

Playlist id 194

| k | r precision | NDGC |
|---|---|---|
| 1 | 0.038760 | 0.589310 |
| 5 | 0.016000 | 0.185832 |
| 10 | 0.033333 | 0.337495 |
| 25 | 0.066667 | 0.702462 |
| 100 | 0.066667 | 0.787717 |

# Evaluation: Normalized Discounted Gain

(good ranking)

$$DCG = rel_1 + \sum_{i=2}^{|R|} \frac{rel_i}{\log_2 i}$$

$$IDCG = 1 + \sum_{i=2}^{|G \cap R|} \frac{1}{\log_2 i}$$

Playlist id 194

| k | r precision | NDGC |
|---|---|---|
| 1 | 0.038760 | 0.589310 |
| 5 | 0.016000 | 0.185832 |
| 10 | 0.033333 | 0.337495 |
| 25 | 0.066667 | 0.702462 |
| 100 | 0.066667 | 0.787717 |

$$NDCG = \frac{DCG}{IDCG}$$

# Prediction Model

# Playlist Completion

## Ranking using lyrics


**Top Suggestions:**

'Why Can't We Be Friends?' by 'War'
'My House' by 'Pvris'
'Box Around The Sun' by 'MisterWives'
'Gente De Accionar' by 'Grupo Codiciado'
'The Fear' by 'Ben Howard'
'Hold Me Now' by 'Thompson Twins'
'Lump Sum' by 'Bon Iver'
'Insides Out' by 'Kid Cudi'
'Send It On' by 'D'Angelo'
'Shake' by 'MercyMe'

# Playlist Completion

## Pooling using word2vec


# Playlist Completion

## Pooling using Word2Vec


### Top Suggestions:

'Imagine - 2010 - Remaster' by 'John Lennon'
'Hotel California - Remastered' by 'Eagles'
'Dust in the Wind' by 'Kansas'
'Knockin' On Heaven's Door - Remastered' by 'Bob Dylan
'Come Together' by 'Aerosmith'
'Layla - 40th Anniversary Version / 2010 Remastered' by 'Derek & The Dominos'
'Free Bird' by 'Lynyrd Skynyrd'
'House Of The Rising Sun' by 'The Animals'
'Behind Blue Eyes' by 'The Who'
'Come Together' by 'Re Beatles'

# Ranking

Basic idea: Train new embedding, pick closest elements in neighborhood.

Embeddings:

- Doc2vec
- Joint Embedding method: Generate pairs with correct and incorrect samples, force net to
    i. Put correct songs closer to input songs than incorrect songs (first few iterations).
    ii. Finetune further to rank- Put correct samples in right order, while ensuring above doesn't change.
- Play with different input representations here: Audio/ mean word2vec/ Adjectives + noun words

# Problem: User Experience (UX)

## Evaluation on Automatic Continuation of Playlists

http://spandan-project-manager.herokuapp.com/

Contact us if you want to get song suggestions!

Thanks! Any questions?