

Recommender Systems

John Bruscella

General Assembly
DSI 113020 - March 2021



Agenda

- What is a recommender system?
- What are some types of recommender systems?
- What is Natural Language Processing (NLP)?
- How can we use NLP in a recommender system?
- Applications and future work



What is a recommender system?



Music



Books



Shopping



Travel



Food



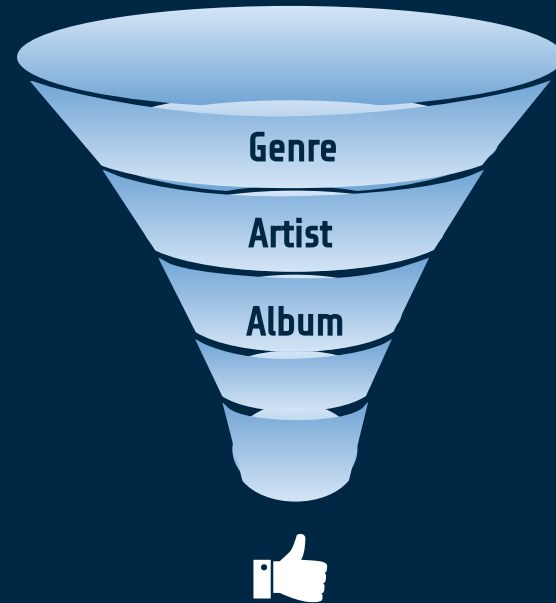
Movies

Types of Recommendation Systems

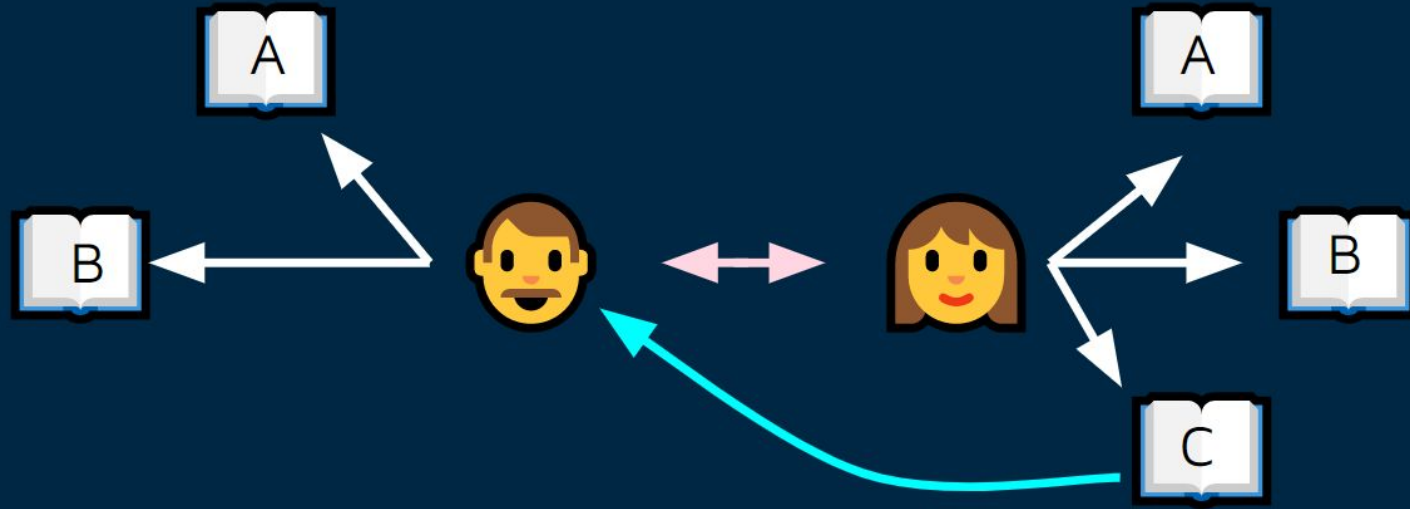
Collaborative Filtering



Content Based Filtering



Collaborative Filtering



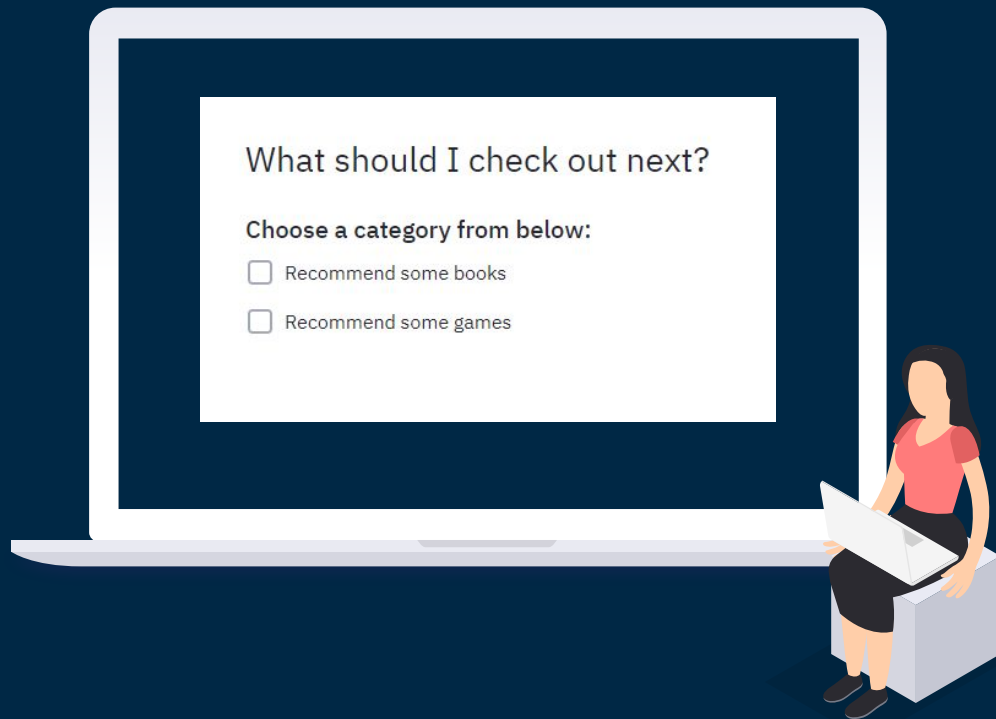
The Netflix Prize

(Collaborative Filtering)

- \$1 Million Prize
 - o 10% improvement
- Data included:
 - 100M ratings
 - 18 Thousand movies
 - 480 Thousand users
- Winning Model
 - Collaborative Filtering
 - SVD



Collaborative Filtering Demo



Collaborative Filtering - App vs. Goodreads.com

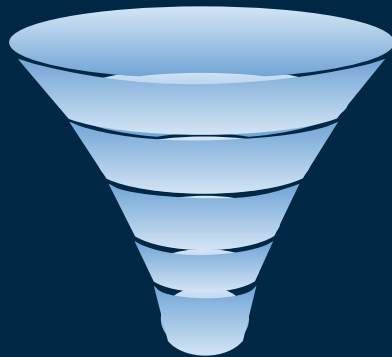
To Kill a Mockingbird

- The Great Gatsby
- The Catcher in the Rye
- Animal Farm
- Lord of the Flies

The Fellowship of the Ring

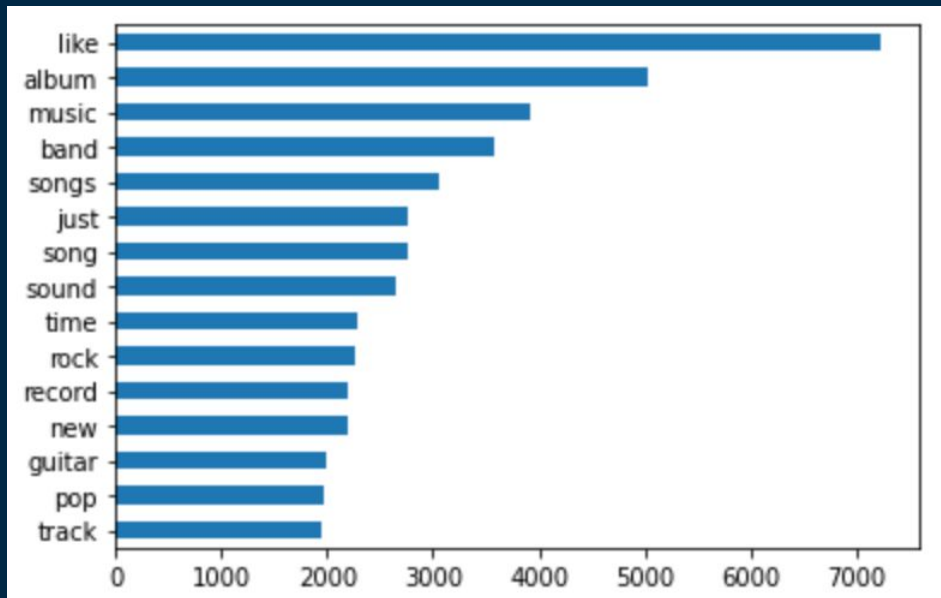
- The Two Towers
- The Hobbit

Content Based Filtering



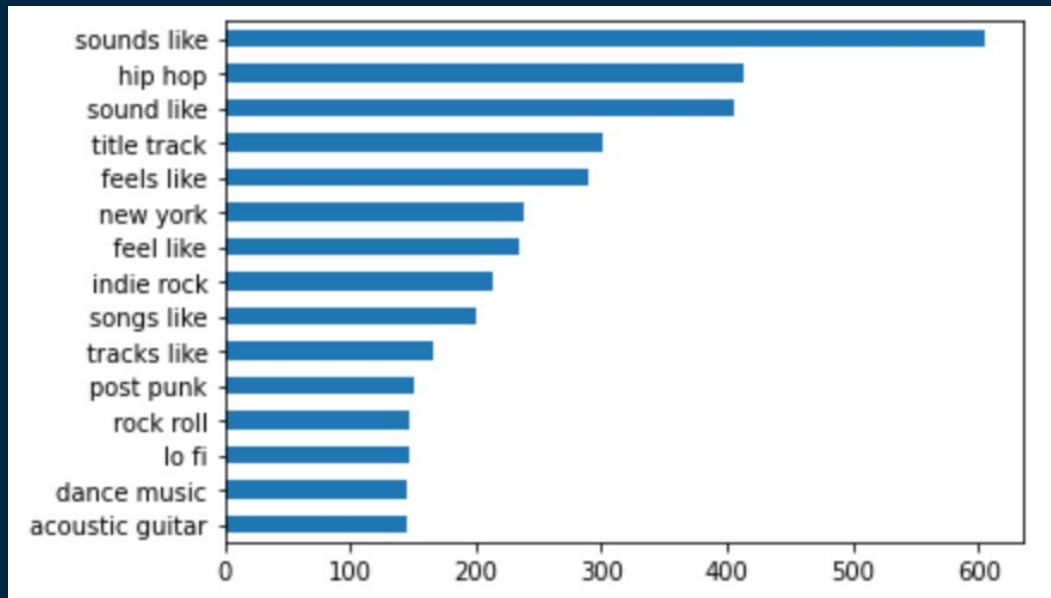
Music Review Dataset

- Pitchfork Album Reviews
 - 18,000 Reviews
 - Most frequent genre: rock
 - 800 Word Reviews
 - Word Count: 3,000 - 4,000 words



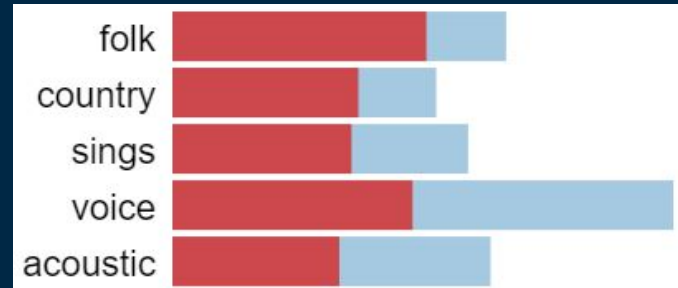
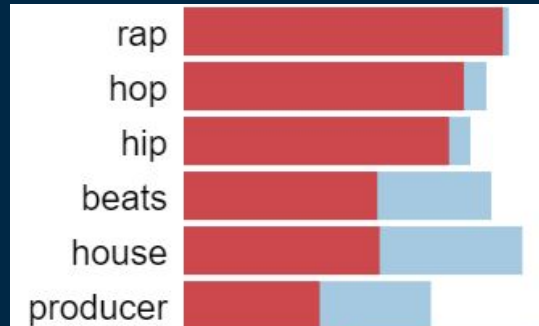
Music Review Dataset

- Pitchfork Album Reviews
 - 18,000 Reviews
 - Most frequent genre: rock
 - 800 Word Reviews
 - Word Count: 3,000 - 4,000 words



Music Review Dataset

- Pitchfork Album Reviews
 - 18,000 Reviews
 - Most frequent genre: rock
 - 800 Word Reviews
 - Word Count: 3,000 - 4,000 words
- Categories



Natural Language Processing

Stemming

Takes the Root of the word

running -> run

Lemmatizing

Base/dictionary form of the word

apples -> apple

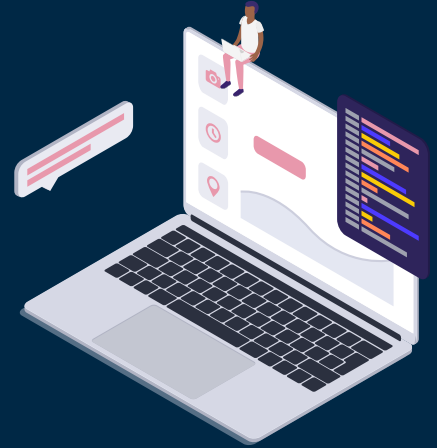
Stop Words

I, me, our, your, her, it, what, who, this



NLP Modeling

- 5000 Reviews
- Target: Review Score (1-5)
- Feature: Text Review
- Best Model
 - Logistic Regression
 - Tfidf function with stop words
 - Lemmatization
 - RMSE: 0.51
 - Balanced Accuracy Score: 0.58



Natural Language Processing

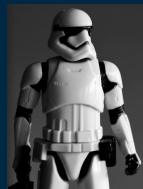
star wars	star trek	similarity
$[-1.22410014e-01,$ $-1.50970012e-01,$ $3.54151994e-01,$ $-8.94370005e-02,$ $1.56545013e-01,$ $-6.56255007e-01,$ $-1.43234998e-01,$ $1.12329997e-01, \dots]$	$[3.16754997e-01,$ $-1.58399999e-01,$ $3.91155034e-02,$ $-4.40240018e-02,$ $3.44935000e-01,$ $-5.04502654e-03,$ $-8.51249993e-02,$ $9.14998353e-04 \dots]$	0.72705

Cosine Similarities

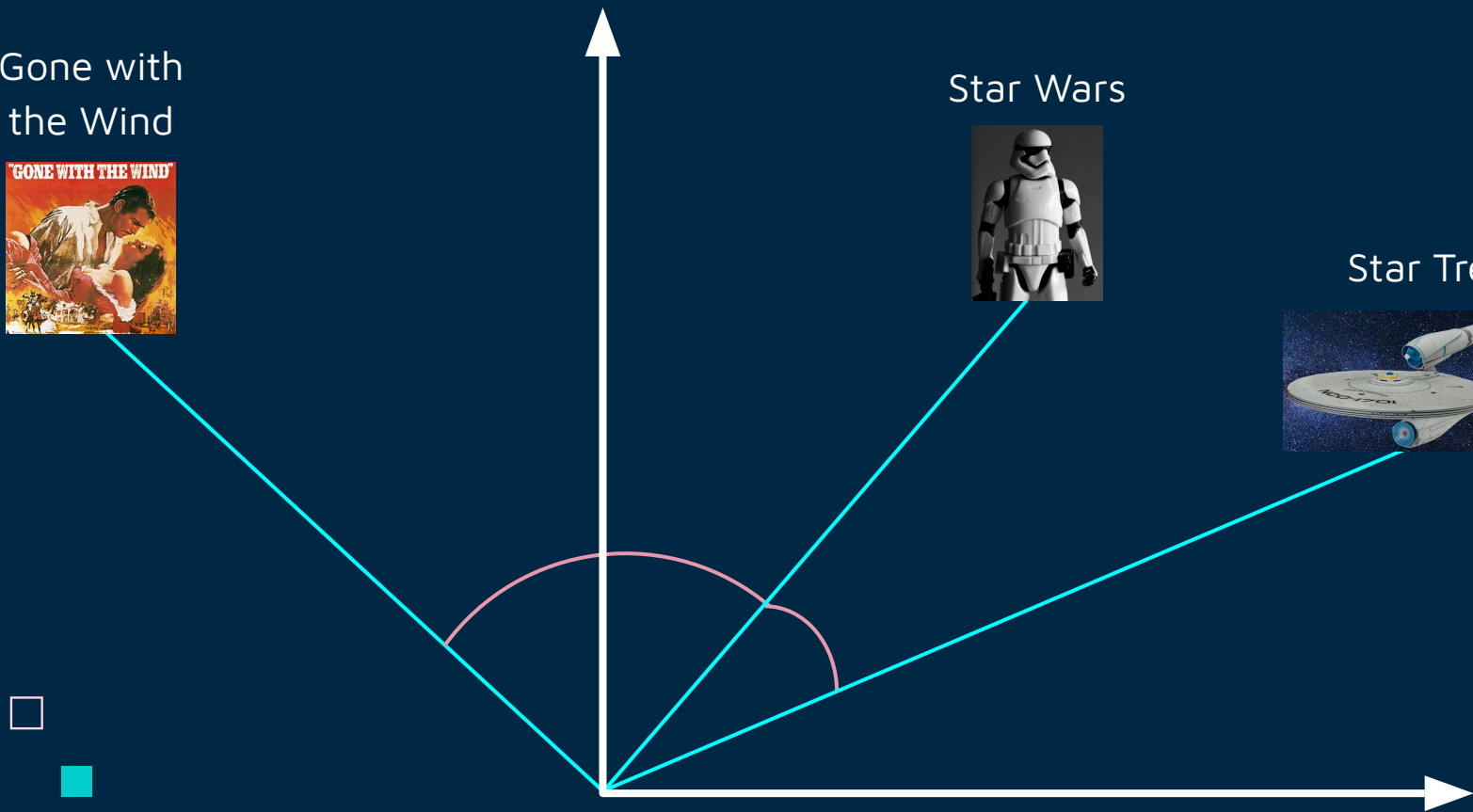
Gone with the Wind



Star Wars



Star Trek



Music Recommender Based on Language

Daft Punk, Random Access Memories

Genre: Electronic

<u>Artist</u>	<u>Album Title</u>
The Avalanches	Wildflower
Cornelius	Fantasma
Justice	Woman

Music Recommender Based on Language

Beck, Morning Phase

Genre: Rock

Artist

Album Title

Screaming Trees

Last Words: The Final Recordings

Starflyer 59

Old

Radiohead

Amnesiac: Special Collector's Edition

Music Recommender Based on Language

Bon Iver, Bon Iver

Genre: Rock

Artist

Album Title

Girls

Father, Son, Holy Ghost

Volcano Choir

Unmap

Animal Collective

Merriweather Post Pavilion

Music Recommender Based on Language

Kendrick Lamar, To Pimp a Butterfly

Genre: Rap

Added Sentiment and Polarity Score

Artist

Album Title

Pete Rock

The Surviving Elements

Birdman, Young Thug, Rich
Homie Quan

Tha Tour Part 1

Missy Elliot

Respect m.e.

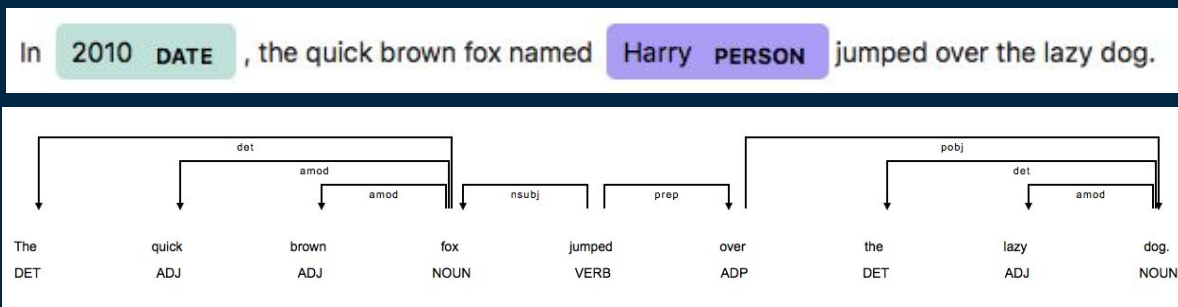
Applications & Future Work

Applications

- Company reviews to evaluate customer happiness

Future Work

- A/B testing
- Entity Recognizer
- Combine NLP with collaborative filtering
- Incorporate additional filters



Credits and Works Cited

- Music Data: Pitchfork Review dataset:
 - <https://www.kaggle.com/nolanbconaway/pitchfork-data>
- Book Data: Goodreads Book Datasets with User Rating 10M:
 - https://www.kaggle.com/bahramjannesarr/goodreads-book-datasets-10m?select=user_rating_0_to_1000.csv
- Board Game Data: BoardGameGeek Reviews
 - <https://www.kaggle.com/jvanelteren/boardgamegeek-reviews>
- Slide Template - SlidesCarnival
- Illustrations created by Sergei Tikhonov.

The background is a dark blue gradient. It is decorated with various geometric shapes: small squares in teal, pink, and orange, and thin white vertical lines of varying lengths. These elements are scattered across the slide, with a higher concentration on the left and right sides.

Thank you!

Questions?