

PODCAST RECOMMENDER SYSTEM

Eu Jin Lee

DSI 12

Instructor: Mahdi Shadkam-Farrokhni

Repo:

https://github.com/MissingNA/podcast_recommender_system

LinkedIn:

<https://www.linkedin.com/in/eeujinlee/>

PROJECT SUMMARY

Goal: Explore the world of recommender system in industry specifically in the context of podcasts.

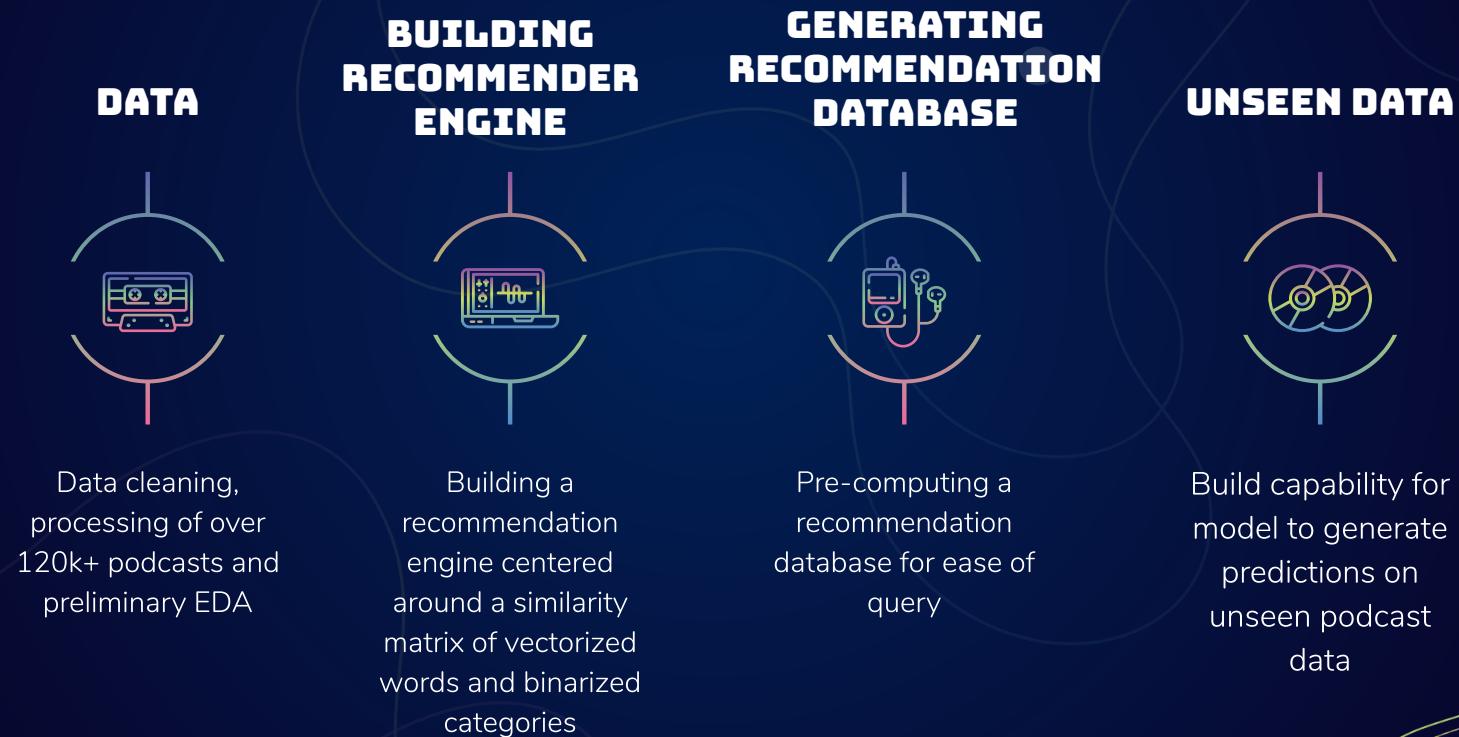
- In the pursuit of dominance within podcast industry, the need for reliable and accurate recommender systems are vital for a company's success.

Limitations: data, computing power

- While the quality and accuracy of recommendations are subjective, the general consensus stand that more data is always better. Hence, incorporating more sources of data and different types of data will definitely generate much more accurate recommendations.
- When attempting to generate recommendations, computing and running methods on large matrices requires A LOT of computing power. Scaling up becomes a challenge.

Product: An end to end (front and back end) podcast recommender application.

WORKFLOW



DATA SOURCE

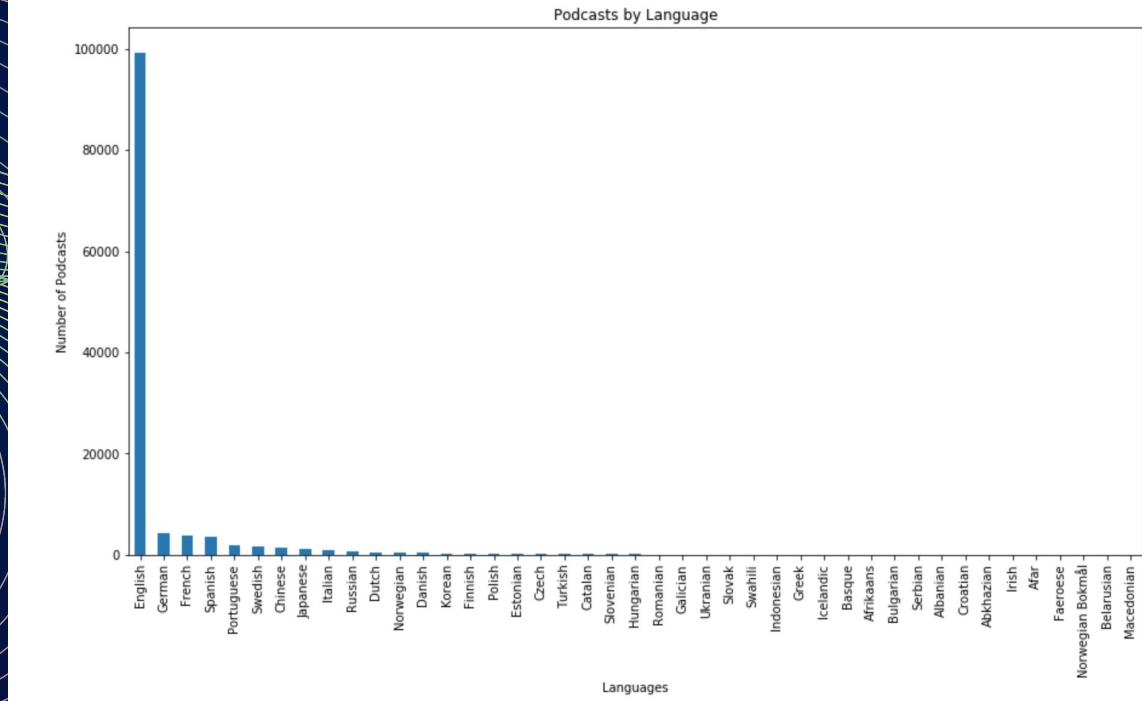


~121,000 iTunes podcasts
containing descriptions and
categories, language and id
information.

Data was downloaded from Kaggle
sourced from Listen Notes a
podcast search engine.

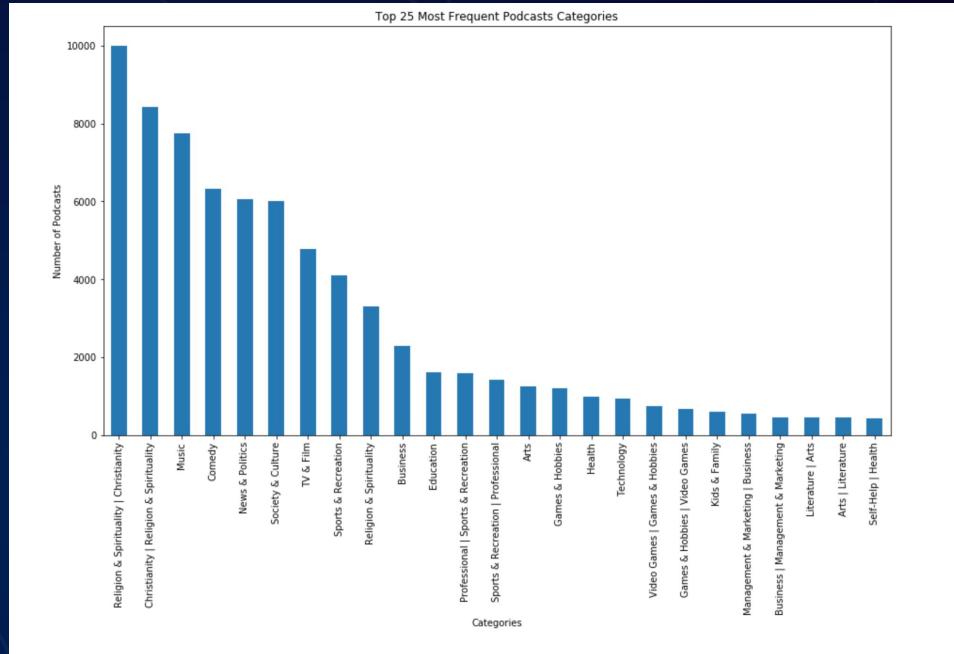
EXPLORATORY VISUALIZATION

English represents ~82% of the iTunes podcast dataset. Making it the most universal language used.



The visualization represents the top 25 most frequent podcast categories across all the languages.

Religion, Music, Comedy, New & Politics and Society & Culture claiming the top spots.



WHAT GOES INTO THE ENGINE?



VECTORIZED DESCRIPTIONS

The descriptions of all English podcasts are vectorized using CountVectorizer and filtered for the top 5000 most frequent words.



BINARIZED CATEGORIES

Many of these podcasts have multiple categories. So a categories dummy matrix of 1s and 0s.



RECOMMENDATIONS

A similarity matrix is computed using cosine similarity. The matrix ranges from -1 to 1.

1 = similar score

0 = unrelated

-1 = opposite score

EXAMPLE RECOMMENDATION RESULTS

0.6694359441465276 - Benjamin Walker's Theory of Everything - Personally connecting the dots. All of them. Benjamin Walker's Theory of Everything is a proud member of Radiotopia, from PRX. Learn more at radiotopia.fm.

0.6128240249379643 - Radio Diaries -

First-person diaries, sound portraits, and hidden chapters of history from Peabody Award-winning producer Joe Richman and the Radio Diaries team. From teenagers to octogenarians, prisoners to prison guards, bra saleswomen to lighthouse keepers. The extraordinary stories of ordinary life. Radio Diaries is a proud member of Radiotopia, from PRX. Learn more at radiotopia.fm.

0.5477203179380679 - The Heart - The Heart is an audio art project about intimacy and humanity. A proud member of Radiotopia from PRX.

0.511308302008249 - The Mortified Podcast -

The Mortified Podcast is a storytelling series where adults share the embarrassing things they created as kids—diaries, letters, lyrics & beyond—in front of total strangers. PS: It totally likes you. Mortified is a proud member of Radiotopia, from PRX. Learn more at radiotopia.fm.

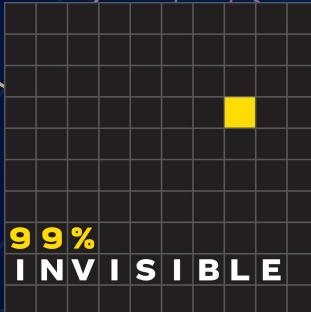
0.5031140762834015 - Strangers - From Lea Thau, Peabody award-winning producer and creator of The Moth Podcast and The Moth Radio Hour, comes her new storytelling baby, Strangers, part of KCRW's Independent Producer Project. Each episode is an empathy shot in your arm, featuring true stories about the people we meet, the connections we make, the heartbreaks we suffer, the kindnesses we encounter, and those fraught moments when we discover that WE aren't even who we thought we were. Strangers is a proud member of Radiotopia, from PRX. Learn more at radiotopia.fm.

Original - 99% Invisible -

Design is everywhere in our lives, perhaps most importantly in the places where we've just stopped noticing. 99% Invisible is a weekly exploration of the process and power of design and architecture. From award-winning producer Roman Mars. Learn more at 99percentinvisible.org. A proud member of Radiotopia, from PRX. Learn more at radiotopia.fm.

RECOMMENDATION RESULTS

Test on model was done on one of my favorite podcasts: 99% Invisible



0.28585391727227594 - Drinking Buddies Podcast - A podcast about everything from pop culture to politics to sports. Two friends that have a couple beers and discuss the past week.

0.2786600750403002 - Danger! Talks... - An open discussion podcast on subjects that I want to know more about with the people that know more about them.

0.2649420522410131 - Olympia Pop Rocks - Guire and Jemmy Joe interview local Olympians of note, and discuss pop culture, politics, local issues or whatever strikes their fancy.

0.25805375473402736 - People I Know - People I Know - the podcast that talks to interesting people so you don't have to. Each episode will see a new interesting person that I know, talk about some interesting people that they know ... not forgetting the all-important scented candle review!

0.2543629237085361 - The Raven Effect - Join professional wrestling superstar & world class idiot, Raven, along with a motley assortment of friends, enemies, know-it-alls, know nothings, professional level morons and highly functioning sociopaths, for a veritable cornucopia of stupidity covering current events, conspiracies, pop culture, politics, general muckety-muck, and any and all topics that can be properly bantered.

Original - Wait, Hold Up! -Each week, tune in as Yarel and Jessica interview people you know (or who they think you should know) who are on their grind and learn the wait, hold up moments that changed their lives. When they're not with a guest they're diving into the latest in pop culture & politics.



BEST SONG

Jupiter is the biggest planet in the Solar System

RESULTS FROM UNSEEN DATA

A friend gave me a podcast that wasn't in the dataset.

BEST VIDEO

Venus is terribly hot, even hotter than Mercury

RECOMMENDATION DATABASE SAMPLE

	ref_podcast	first_rec	second_rec	third_rec	fourth_rec	fifth_rec
0	Eat Sleep Code Podcast	Stackify Developer Things	The freeCodeCamp Podcast	Oracle Developer Podcast	Storage Developer Conference	Complete Developer Podcast
1	The Tech M&A Podcast	Technocrats (Audio)	Tech Reformation	Ado Veli Podcast	Speaking in Tech	The TechBox
2	Social Media, Mobility, Analytics, Cloud: Tech...	Two Cracked Pots	The Communicators	Lanced Podcast	Technology Wizard	The Cloudcast (.net) - Weekly Cloud Computing ...
3	Radio Leo (Video LO)	Aggregate Feed - LoadingReadyRun	Melted Chocolate Master Feed	Radio Leo (Video LO)	Radio Leo (Video HD)	Radio Leo (MP3)
4	Digital India	AMPUP Your Digital Marketing	Social Media Talks Podcast Kompass Media	The Science of Social Media	VADEMECUM BLOGERA	The EchoJunction Podcast with Adam Fraser

RECOMMENDATION DATABASE LOOKUP

ref_podcast	first_rec	second_rec	third_rec	fourth_rec	fifth_rec	
25727	99% Invisible	Strangers	The Mortified Podcast	The Heart	Radio Diaries	Benjamen Walker's Theory of Everything

MOVING FORWARD

BUILDING FLASK APP

Coding out the application using Flask.

BUILDING WEBSITE INTERFACE

Creating a front facing website interface for podcast recommender system using HTML, CSS, Javascript

HOSTING ON HEROKU

Hosting application on Heroku with AWS storage integration.



HEROKU

REFERENCES

- **Building a Podcast Recommendation Engine**
<https://www.kaggle.com/switkowski/building-a-podcast-recommendation-engine>
- **How to Clean Text for Machine Learning with Python**
<https://machinelearningmastery.com/clean-text-machine-learning-python/>
- **What is a cosine similarity matrix?** <https://medium.com/acing-ai/what-is-cosine-similarity-matrix-f0819e674ad1>
- **A Gentle Introduction to Sparse Matrices for Machine Learning**
<https://machinelearningmastery.com/sparse-matrices-for-machine-learning/>
- **How do I get indices of N maximum values in a NumPy array?**
<https://stackoverflow.com/questions/6910641/how-do-i-get-indices-of-n-maximum-values-in-a-numpy-array/38884051>
- **How to create a production-ready Recommender System**
<https://towardsdatascience.com/how-to-create-a-production-ready-recommender-system-3c932752f8ea>
- **Converting sparse matrix to dense matrix**
https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.csr_matrix.todense.html
- **20 Podcast Predictions for 2020 from Top Industry Leaders**
<https://blog.pacific-content.com/20-podcast-predictions-for-2020-from-top-industry-leaders-f4ef49e48909>
- **How to transform numpy.matrix or array to scipy sparse matrix**
<https://stackoverflow.com/questions/7922487/how-to-transform-numpy-matrix-or-array-to-scipy-sparse-matrix>

THANKS FOR LISTENING

Do you have any questions?

Eu Jin Lee

Repo:

https://github.com/MissingNA/podcast_recommender_system

LinkedIn:

<https://www.linkedin.com/in/eeujinlee/>

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.