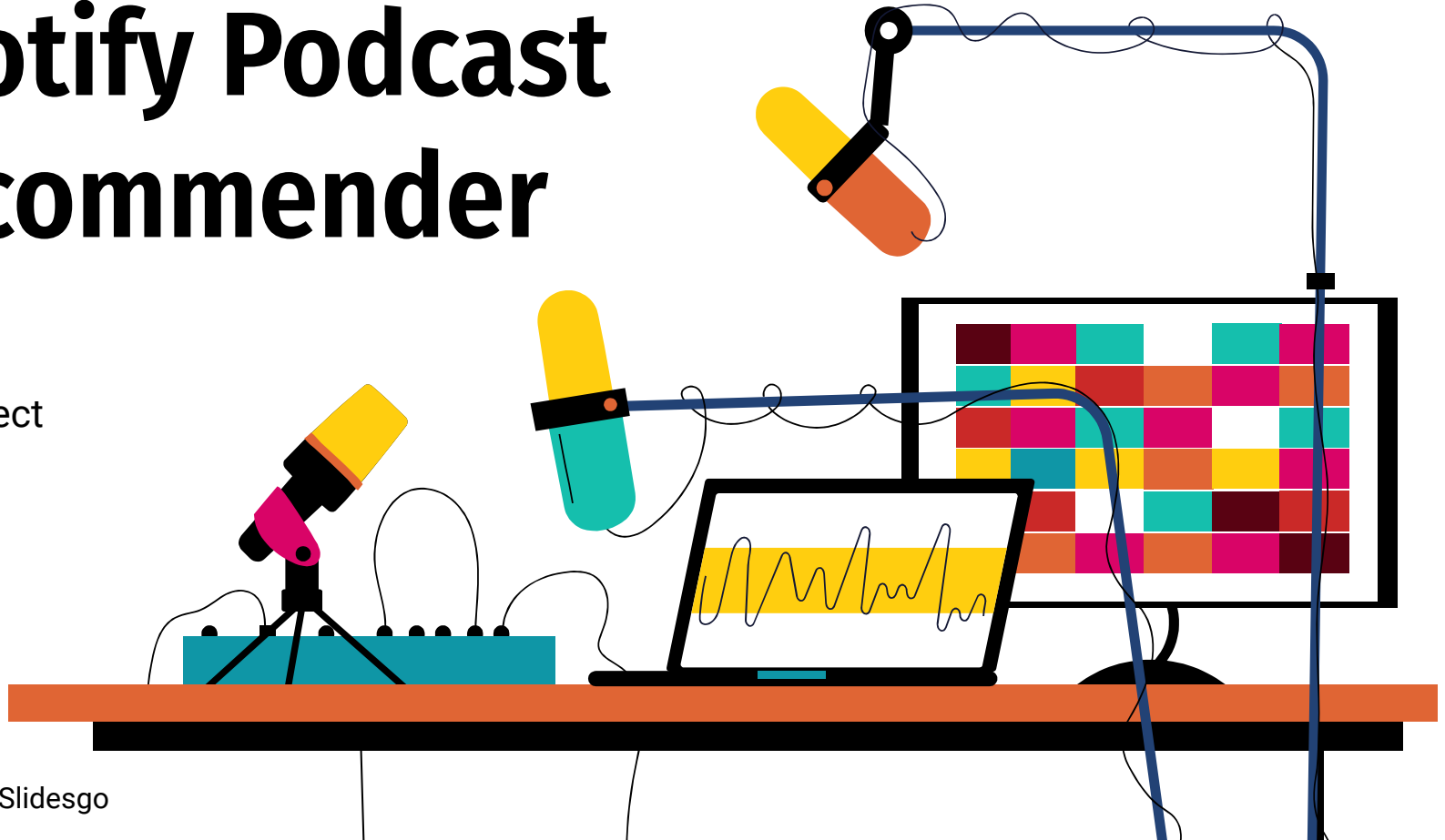# Spotify Podcast Recommender

Zoe Liao
NLP Project
2/23/22

Background

Podcasts

So many! Over 3.2 million[1]

+

Mental Health

Specific but broad: stories, advice

→

Goal

Content-based recommendation system

Process

Data
Spotify Web API: 350 shows -> 344; 20745 episodes -> 18115

Tools
NLTK, SpaCy, Scikit-learn

Topic Models
Episodes & Podcasts

Recommend
Content-based, using cosine similarity

Preprocessing/Tokenizing Details

Cleaning

- Digits
- Punctuation
- Symbols
- Websites
- non-English characters

Proper Nouns & Compounds

- Group proper nouns for Episodes
- Compound (e.g., mental health, personal development)

Lemmatize, Lowercase

- WordNet Lemmatizer
- Lowercase

Stop List

- NLTK stopwords
- Single names (for Episodes)
- Twitter, Youtube, etc.

Other

- Unigrams and bigrams for Podcasts
- Doubled nouns for Podcasts

Topic Modeling

Episodes

Vectorizer
TF-IDF: min_df = 3, max_df = 0.8

Topic Model
LSA

Topics
15 (less interpretable)

Podcasts

Vectorizer
TF-IDF: min_df = 2, max_df =0.8

Topic Model
NMF

Topics
10 (more interpretable)

Podcast Topics (NMF with TF-IDF)

Topic 1: Story/Interview
story, guest, story story, show, award, stigma, interview, professional, work, host

Topic 2: Relationship
relationship, sex, relationship relationship, life, friendship, friend, welcome, sex sex, career, host

Topic 3: Season? (less interpretable)
season, season season, topic, pain, health, science, school, mental_illness, identity, series

Topic 4: Physical Health
health, fitness, health health, wellness, fitness fitness, show, wellness wellness, nutrition, business, dr

Topic 5: Women
woman, woman woman, everything, faith, career, body, talk, host, marriage, let

Topic 6: Students? (less interpretable)
year, conversation, issue, student, school, education, college, year year, experience, mental_health mental_health

Topic 7: Sharing experience
life, life life, people, share, experience, way, journey, people people, year, challenge

Topic 8: Weight loss
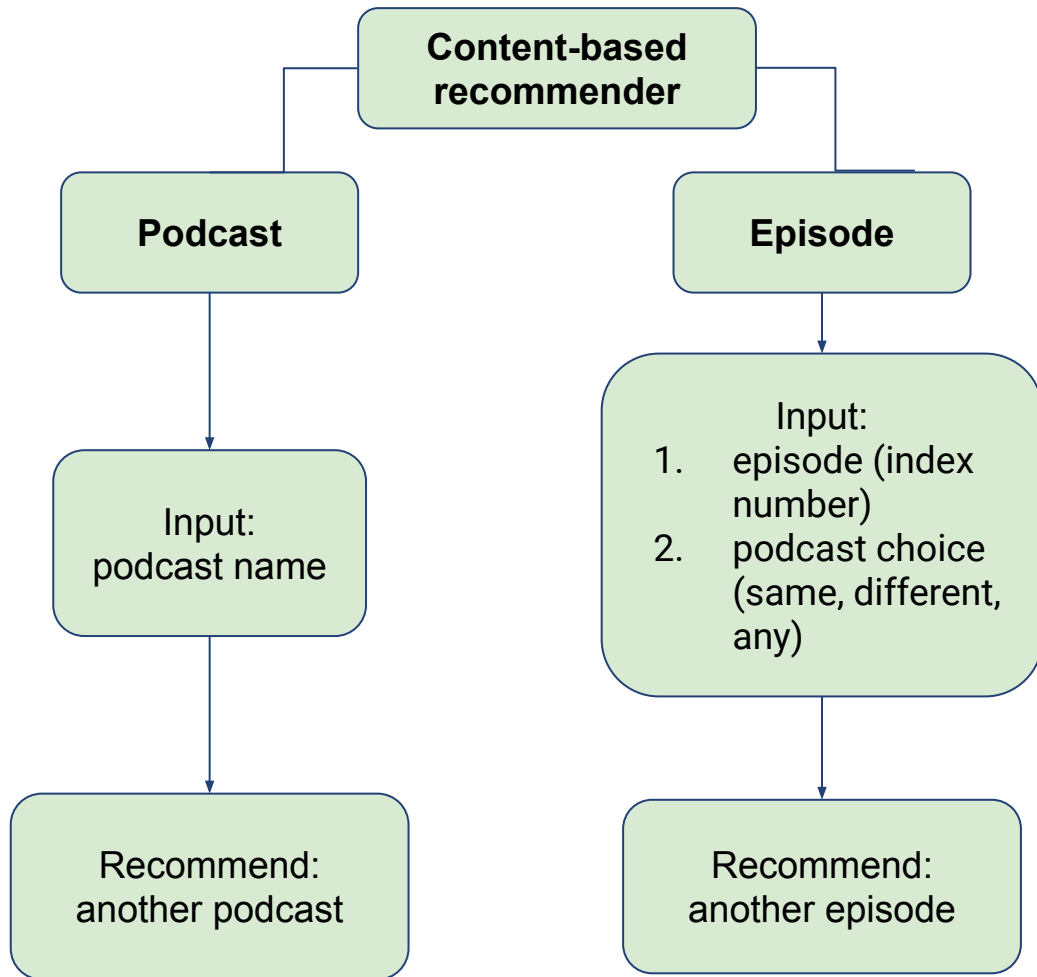loss, weight, journey, loss loss, surgery, weight loss, pound, tip, journey journey, bypass

Topic 9: Art
support, art, support support, culture, art art, people, youd, become, please, art culture

Topic 10: Therapy
im, therapy, im im, therapist, therapy therapy, name, die, der, practice, therapist therapist

Recommender

Streamlit

Content-based recommender

Podcast

Input:
podcast name

Recommend:
another podcast

Episode

Input:
1. episode (index number)
2. podcast choice (same, different, any)

Recommend:
another episode

"Good" example: Input

Episode You Chose:

Episode Name: 061. All About Friendships.

Episode Date: 2021-10-26

From podcast: Badass Confidence Coach

Description: Tim and I are talking about all things Friendship. What are the traits of a good friend? What kind of friend are you? When do you stay in it and when do you back away? Friends come and friends go. But this topic is always a relevant one. Give it a listen! Mad love! Interested in coaching and/or therapy? Reach out. www.annamarcolin.com

"Good" example: Output

Recommendation

Recommended episode: Equally Yoked Friends

From podcast: The Joi of Social Work

Date: 2021-06-06

Description: Friends! How many of us have them? How many of have equally yoked friends? In this episode we are talking about friendship - what we learned about it from the adults around us, why do we give more grace to intimate partners than friends and our opinion of having friends who hold you accountable. Let us know what you think & share your friendship stories with Joi. You can email them to thejoiofsocialwork@gmail.com. Upcoming Events: 6/26 - Social Work Appreciation Gathering, New York, NY (6/13 is the absolute last day to register) https://www.eventbrite.com/e/social-work-appreciation-gathering-tickets-154308866975?aff=eand Things To Check Out: 6/19 - More Than Academics Documentary, New York, NY https://www.eventbrite.com/e/morethanacademics-documentary-premiere-tickets-154484133201

"Bad" example: Input

Episode You Chose:

Episode Name: #6 Starting a new job in 2020

Episode Date: 2021-02-02

From podcast: Honest UX Talks

Description: In this episode, I wanted to share how did it felt to start a new job in 2020 in the product company being fully remote. ⚠️ Spoiler alert: It was definitely not easy for an offline extroverted person. Also, it's the first experimental episode alone. ➡️Pre-history and why I started a new job ➡️The choice: Product company vs Toptal ➡️Connecting with the team online ➡️Struggles of an extrovert working online ➡️Understanding a complex product ➡️What I've learned ------- Follow us on Instagram to stay tuned for the next episodes. Feel free to DM your question if you want us to discuss it. 📇 instagram.com/honest_ux_talks 👩‍💻 instagram.com/anfisign

"Bad" example: Output

Recommendation

Recommended episode: Conversation 7: How to stay mentally strong for you, your child and your family

From podcast: Leveling the Spectrum

Date: 2021-05-11

Description: Today my mom and I share our ideas on how to stay mentally strong for you, your child and your family. We start off the conversation by sharing my realization of how difficult it is for parents to have time for "self care". Self care sometimes is just caring for yourself. Trying to get parents to meditate when they have never meditated before or do other "self care" tasks outside of their normal routine can put more stress on them instead of relieving it. Sometimes all you need to focus on are the basics and care for yourself, such as brushing your teeth every morning and night, fueling your body with nutritious food, dressing yourself, etc. The goal is to maintain a healthy mind and body so that you are able to wake up most mornings and feel ready to tack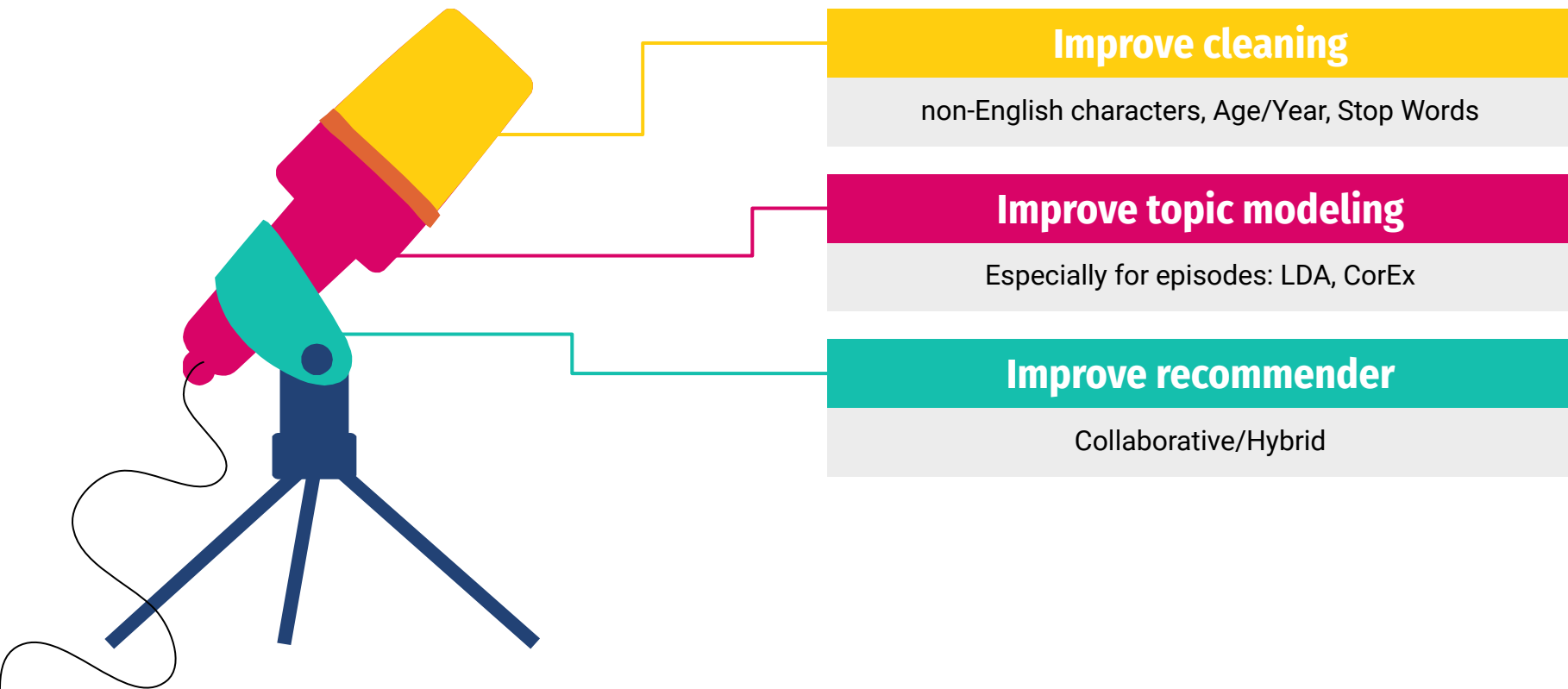le your busy day. Thank you for listening! Contact us- levelingthespectrum@gmail.com Wordpress Website- https://levelingthespectrum.wordpress.com/ Syrena's podcast "The Sesh"- https://open.spotify.com/show/60QnjzArn9p8V4NqOfVYi6?si=r1kkrfQeQoSj0iBbFfwX4A Syrena's Instagram- https://www.instagram.com/syrenamichele/?hl=en Music credit- https://www.purple-planet.com --- This episode is sponsored by · Anchor: The easiest way to make a podcast. https://anchor.fm/app

Next Steps

Improve cleaning

non-English characters, Age/Year, Stop Words

Improve topic modeling

Especially for episodes: LDA, CorEx

Improve recommender

Collaborative/Hybrid

Appendix

Episode Topics (TF-IDF with LSA)

Topic 1
date, life, created, first, story, wa, get, dating, mental, time

Topic 2
life, mental, wa, get, way, people, time, make, join, find

Topic 3
sanctuary, ssanrakkha_buddhist, guided, meditation, awareness, open, hmr, mindful, yasm, aggacitta_ssanrakkha_buddhist

Topic 4
mental, factor, delve, condition, bobby_temps, thrives, brainchild, thursday, managing, manage

Topic 5
full_ep, surprise, host, abbie_chatfield, oh, dear, the_edge, nat_penfold, brooklyn_ross, kyle_jackie

Topic 6
contain, violence, may, drug, message, voice, indian, sex, indiannoir, classifiable

Topic 7
anchor, easiest, make, way, contain, mental, violence, drug, may, mind

Topic 8
healing, card, the_tarot, the_wild_soul, soul_tarot, honoring, work, tribe, called, medicine

Topic 9
information, privacy, omnystudiocomlistener, see, supporter, become, national_gratitude_month, visit, acastcomprivacy, optout

Topic 10
supporter, become, get, let, feel, review, find, know, card, free

Topic 11
join, group, closed, get, discus, community, uncomfortable, showsubscribe, shareheres, facebookwellevatr

Topic 12
message, voice, send, life, visit, find, please, go, time, choice

Topic 13
join, message, voice, send, group, privacy, friend, closed, omnystudiocomlistener, book

Topic 14
visit, support, friend, ad, megaphonefmadchoices, choice, welcome, love, life, today

Topic 15
support, supporter, please, become, rendering_unconscious, welcome, message, voice, patreon, interview