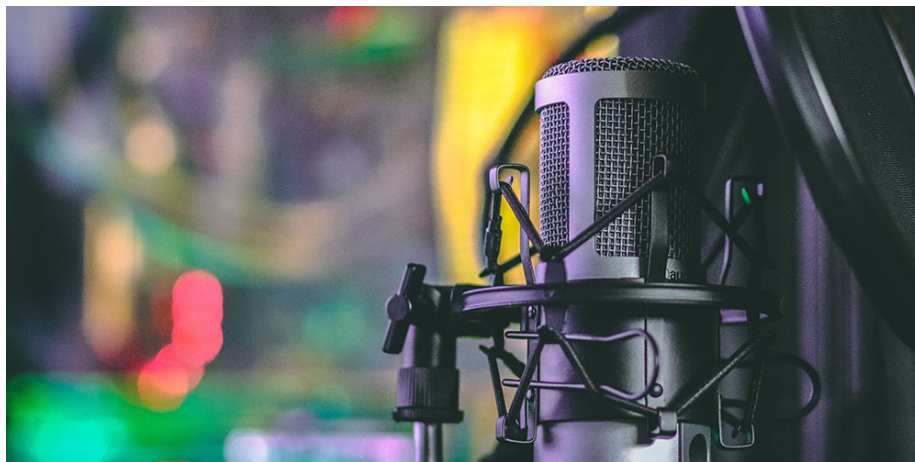


Welcome to my Content Based Podcast Recommender



This repository creates a web application to recommend offbeat podcasts based on user input. Recommendations are determined using calculated similarities between key words that the user will provide and the text description of each podcast. Let's get started!

The Data

The data was downloaded from Kaggle but originates from the Listen Notes database. The attributes used for this recommender were:

1. **uuid:** the unique id to reference the podcast
2. **title:** the title of the podcast
3. **description:** the short text that describes the content of the podcast
4. **category:** the category the podcast was assigned to
5. **language:** the language the podcast was assigned to

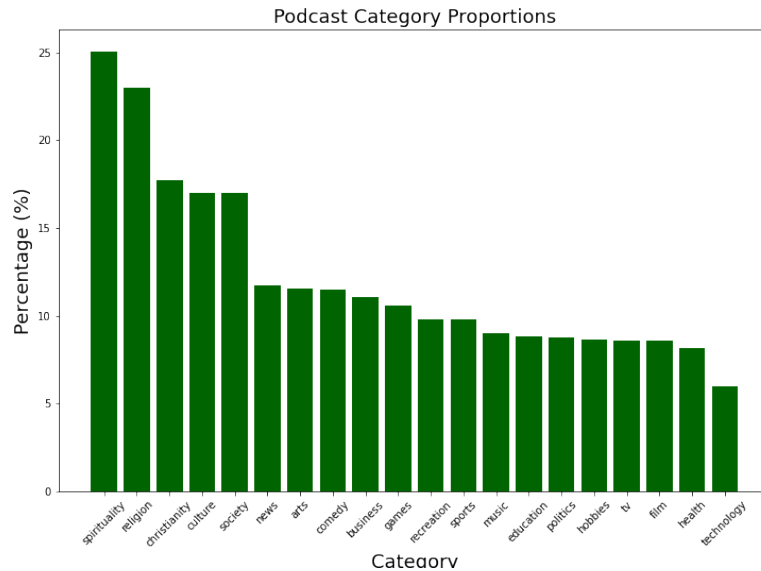
Exploration

Let's take a look at the data!

The Categories

This dataset had 83 categories each podcast could be assigned to. However, some categories were rare compared to others, so below is a bar plot of the categories that were more than 5% of the overall dataset. After further calculation, the

5 most common categories were spirituality, religion, Christianity, culture, and society.



The Recommender

To see the details of how this recommender was created, please refer to the `recommender.py` in the app folder. This is a content based recommender, which means that the recommendations are determined by the similarity in the key words and the descriptions of each podcast. Using sklearn's TFIDF vectorizer, each podcast, or row, now is the calculated weights of each word for each document (podcast). Then, utilizing cosine similarity, the similarity score of the user's key words and each podcast vector is recorded and filtered for the largest in order to return the number of recommendations the user would like.

The Web Application

To see this recommender in action, please refer to the app folder. Users will be asked to input a list of key words they want along with the number of recommendations they would like. Below is a sample output if the user were to input "bigfoot, alien, supernatural, ufo, paranormal" with 3 recommendations.

Welcome to my Content Based Podcast Recommender!



Listen Notes (2021). Retrieved from listennotes.com.

This application recommends podcasts based on similarities between the key words you want and the text descriptions. Each recommendation will come with a short description along with its cosine similarity with the key words you provided. Keep in mind that scores closer to 1 are considered very similar and scores closer to 0 are considered not similar.

Please provide a list of key words in the first box and the number of recommendations you would like in the second box. For example, first box: sports, nba, basketball, and second box: 3.
What kind of podcast are you looking for?

I would recommend

Title	Text Description	Cosine Similarity Score
REALITY ROUNDHOUSE PODCAST	Reality Roundhouse Podcast is a chat style podcast that covers topics ranging from Sports, MMA, Paranormal, Aliens, Bigfoot, UFO, Music, Movies, Pop Culture, and more.	0.59
Spirit Radio-the Paranormal Experience	Join your hosts Willy Hassell and Lynne Nickerson for a weekly radio show covering all subjects of the paranormal world with weekly guests from around the world. Ghosts, spirits, UFOs, alien encounters, bigfoot. This is the show where the paranormal becomes the normal. So join us on a journey into the unknown the unseen the unthinkable.	0.45
Scaredycast	Pop culture paranormal podcast discussing all things spooky and weird! Current happenings in the world of ghosts, UFOs, bigfoot, monsters, and more! Check out Scaredycast.com for more!	0.45

Please reach out to me if you would like to check it out! I might have the AWS EC2 off to avoid unnecessary costs.

For the Future

Some next steps are:

- Expand the data base to other languages since it only contains English podcasts
- Play with the css style of this web application