Project Title

RecoPenguingine

Project Pitch

We are creating a web app displaying a recommendation system where users can input information and select key features for personalized music & media recommendations. The users can view a variety of dashboards of their recommendations, top genres & favorites, and similar/recommended listening activity based on location/region. The web app will also allow users to view **"behind the scenes" ML model visualizations**. The final presentation is a publicly available web application and a video demonstration of the application and modeling in action.

Names & Skills

- Sophia Turnbough - **ML/Modeling**, Viz & Analysis, Python, **Project Management**
- Cody Cook - **UI Design, SDLC, Data Connection**
- German - **Python, Algorithms/SE**
- Pari - **Analytics & ML Deployment**, Dashboarding, SQL
- Abbey - Web App Customizations (HTML/CSS), UI Design, **Data Analysis & Dashboarding**

Data Ethics Checklist

#	Question	Generally	Data Breach	Example
1	Could a user sell drugs or other illegal items on your platform?	N	N	https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1072&context=sph_pubs

2	Could a user of your platform engage in sex trafficking?	N	N	https://stmuscholars.org/craigslist-and-backpage-sex-trafficking-at-your-fingertips/
3	Could a user sell class notes or cheat on their homework on your platform?	N	N	https://www.edsurge.com/news/2021-02-23-more-students-are-using-chegg-to-cheat-is-the-company-doing-enough-to-stop-it
4	Could a stalker use your project to find someone?	N	M	https://www.nytimes.com/2018/05/19/technology/phone-apps-stalking.html
5	Could your app be used to spy on or track individuals?	N	N	https://www.nytimes.com/2019/12/22/us/politics/totok-app-uae.html
6	Could your app/software access the camera or microphone and record things without users being aware?	N	N	https://gizmodo.com/these-academics-spent-the-last-year-testing-whether-you-1826961188
7	If someone uses your platform, could they be re-traumatized or have their mental health impacted in some way?	N	N	https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

8	Could your algorithm promote material that would traumatize or upset individuals?	N	N	https://www.theguardian.com/technology/2021/oct/16/tiktok-eating-disorder-thinspo-teens
9	Would your users be upset if the data you collect was given to someone else?	N	M	https://www.businessinsider.com/stolen-data-of-533-million-facebook-users-leaked-online-2021-4
10	Could a data leak potentially lead to identity theft?	N	M	https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement
11	If your site was hacked, would users of that product potentially lose their job, spouse, or family?	N	N	https://www.forbes.com/sites/zakdoffman/2019/08/23/ashley-madison-is-back-with-30-million-cheating-spouses-signed-since-the-hack/?sh=22f1ba5c3878
12	Should there be an age limitation on your product?	N	N	https://www.bbc.com/news/technology-48925623
13	Could someone use your product to find, contact, and potentially	N	N	https://www.nbcnews.com/health/aging/genetic-testing-scam-targets-seniors-rips-medicare-n1037186

	commit elder abuse?			
14	If the data on your platform was breached, could it be used to blackmail the users?	N	N	https://www.wired.com/story/parler-hack-data-public-posts-images-video/
15	Does the existance of your project imply that a particular racial group, gender, religion or other protected category is inherently bad, gross, or unwanted?	N	N	https://www.distractify.com/p/pinky-gloves-dragged
16	Could your product be used to commit hate crimes against a specific group?	N	N	https://ibmandtheholocaust.com/
17	Does the primary content of your game or algorithm focus on something considered deeply unethical?	N	N	https://www.quora.com/What-is-the-most-unethical-video-game-ever-created

18	Does your game or software contain race, gender, or other stereotypes?	N	N	https://en.wikipedia.org/wiki/List_of_controversial_video_games
19	Could users of your app scam other individuals?	N	N	https://dailyiowan.com/2021/06/21/opinion-kickstarter-scams-are-on-the-rise/
20	Is your particular algorithm biased towards predicting correctly only for one race, gender, or other group?	N	N	https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm
21	Are the users of your project, players of your game, or those being surveyed for your data aware of how their data will be used?	N	N	https://www.computerweekly.com/news/252464048/Many-search-engine-users-unaware-of-personal-data-collection

22	What are the possible misinterpretations of your results? For example - would a white supremacist or misogynist be stoked about your results if they misinterpreted it?	N	N	https://www.nature.com/articles/s41467-020-19723-8
23	Does the use or purchase of your data potentially contribute to a dangerous group or regime?	N	N	https://vertpaleo.org/svp-sends-letter-to-paleontological-community-on-myanmar-amber/
24	Could your virtual reality environment cause injury to the user?	N	N	https://bonejoint.net/blog/eight-things-you-should-know-about-virtual-reality/
25	Are your study participants or game players aware that their data will be collected and used?	Y	Y	https://www.polygon.com/features/2019/5/9/18522937/video-game-privacy-player-data-collection
26	Does your game or app contain addictive design elements	N	N	https://searchsoftwarequality.techtarget.com/tip/5-examples-of-ethical-issues-in-software-development

	without benefit to the user?			
27	Does your survey contain an aspect of compulsion or unusually large incentive, that would command users to take it even if it was to their detriment?	N	N	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4214066/
28	Could your research outcomes harm an individual or entity?	N	N	https://rutgersaaup.org/rutgers-budget-system-is-a-mess-what-will-president-holloway-do-about-it/

Data Ethics Responses

4. Could a stalker use your project to find someone?

In the event of a data breach, a stalker could potentially use the location data publicly provided by Spotify amongst other methods to determine a general location of the user.

9. Would your users be upset if the data you collect was given to someone else?

In the event of a data breach, the user could be upset their data was given to another application, but they have their privacy settings set within Spotify and are able to control data usage as they wish.

10. Could a data leak potentially lead to identity theft?

In the event of a data breach, Spotify's data could lead one to determine payment or location data in conjunction with the use of other data information sources.

25. Are your study participants or game players aware that their data will be collected and used?

The data being used is publicly available per the user's individual privacy settings on Spotify. They have the option to deny data tracking within the app and their devices.

Skills & Data Breakdown

- Spotify API connection to active streaming & location data (will only have access to those users who have allowed location data to be collected within their account privacy settings) -> using spotipy/Spotify Developer
- ETL pipeline for wrangling/cleaning
- EDA
- Feature Engineering
- Model selection + hyperparameter tuning
- Training & testing for recommendation engine
- Deployment of ML models & visualizations to web app
- Web app customizations + widget building
- Presentation & demonstration

Literature/Market Review

- Musicovery - http://b2b.musicovery.com/

    - This is an app that gives you a playlist of music to listen to depending on the mood you are feeling. It can be updated very fast and has an option to like and dislike songs. It is also an engine that you can connect through and use with their API. It can also recommend live concerts that you may be interested in and Youtube channels etc. While we would like to do something similar, we would also like to recommend popular music genres related to what other people with similar interests are listening to and hopefully get regional data for popular songs/genres within your country.

- Last.fm - https://www.last.fm/ (CBS Interactive. (2002). Play music, find songs, and discover artists. Last.fm. Retrieved January 26, 2022, from https://www.last.fm/)

    - This is an app and website that can be connected to your music streaming services. It collects your listening history then generates a report for you on what you like to listen to and statistics for that. They have an API that others can use to connect through, on top of getting user data through approval of the user's music streaming services(e.g Spotify) listening history. Even though we are trying to gather listening behaviors to create recommendations, Last.fm only collects the data but its purpose does not have specific genre/other related recommendations. This data is also displayed with stats that we plan to leave out from the user to give a more simplistic and easy-to-use system.

- Cave, N. (n.d.). Recommends music, movies, TV shows, books, authors, games, and podcasts. TasteDive. Retrieved January 26, 2022, from https://tastedive.com/

    - This is an application that is most popular as an entertainment recommendation engine for films, TV shows, music, video games, etc. This application collects the user's search history whether it's movies, TV shows, or music and the site's algorithm provides a list of similar content. Based on the liked history of a user on movies, TV series, or any other entertainment platform this recommendation engine can provide accurate information and recommendations based on user likes and dislikes. Within this application, the more a user explores or uses the site, the more the site learns about the user's preferences and makes the recommendation results better. It is similar to our project. We would like to create a recommendation system engine which helps individuals find music genres or music types based on their music listening preferences based on other listeners in the region/country "how popular their music taste is within the country".

- Cyrus. (2020, September 11). Build and deploy your own machine learning web application by Streamlit and Heroku. Medium. Retrieved January 25, 2022, from https://medium.com/analytics-vidhya/build-and-deploy-your-own-machine-learning-web-application-by-streamlit-and-heroku-d306f2d29474

    - This is a broad overview/top level article guide demonstrating how Streamlit and Heroku can be used to create a basic user-friendly web application to display

modeling and analytics concepts. For our project, we will be using some of the concepts detailed (predominantly the framework/project flow template) in the article to further explore some of the basic functionalities of the tools we will be using to connect the project to the presentation and showcase steps with the web application. We will be using different machine learning and modeling techniques and adding more features, most importantly being the recommendation engine. We will also have a different range of dashboard views of the models we used "behind the scenes", a location based recommendation feature, a custom widget integrating the user input (platform choice, user information, key feature selection options only modeling based on user valued feature importance, dashboard & custom viewing options, etc), and more detailed data engineering/coding skills exhibited. The novelty sources from user input, customization options, and adding in the geo- component, as well as the recommendation engine as a whole.

- Gibney, M. & Gnod - The Global Network of Discovery. (n.d.). Gnoosic - Discover new Music. Gnoosic. Retrieved January 26, 2022, from https://www.gnoosic.com/

    - This is an app and website that lets the user input some of their favorite bands and the app will go through multiple bands that are either of the same genre as the ones input or bands that may be fit based on the user's taste based on the bands they input. The app itself doesn't show any more than the suggestions' names and some of them recommend snippets of one of their songs, or even recommend shows for the band, if they are available, that is. Otherwise, if none of those is available to show to the user, it only gives the recommended band's name, and three buttons of "I like it", "I don't like it" or "I don't know", leaving the user to maybe go out of their way and look more directly into the given recommendation, rather than having it in the same page/app.

- Shi, J. (2021, April 28). Music Recommendation Algorithm Based on Multidimensional Time-Series Model Analysis. Music Recommendation Algorithm Based on Multidimensional Time-Series Model Analysis. Retrieved January 26, 2022, from https://www.hindawi.com/journals/complexity/2021/5579086/

    - This article studies how other platforms utilize different algorithms to make suggestions based on the user's data and their listening behavior. The article proposes a Time-based algorithm to suggest the user different music based on what they listen to "in the moment", mid-term and long-term. The model proposed

can be somewhat similar to what Pandora's recommendation system uses, "Music Genome Project", but it also tries to analyze how the user feels, as Cody mentioned, to make recommendations.

Timeline

Phase	Week Number	Task	Owner	Task Stage	Due Date
Data Collection/Connection		Spotify API connection	Sophia Grace Turnbough		
Data Collection/Connection			Abbey Rohler		
Data Collection/Connection			Paribesh Chaudhary		
Data Collection/Connection			codycook33@email.arizona.edu		
Data Collection/Connection			German Alonso Perez Arispuro		
Data Processing			Sophia Grace Turnbough		
Data Processing			Abbey Rohler		
Data Processing			Paribesh Chaudhary		
Data Processing			codycook33@email.arizona.edu		
Data Processing			German Alonso Perez Arispuro		
Data Cleaning			Sophia Grace Turnbough		

Data Cleaning	Abbey Rohler
Data Cleaning	Paribesh Chaudhary
Data Cleaning	codycook33@email.arizona.edu
Data Cleaning	German Alonso Perez Arispuro
Feature Engineering	Sophia Grace Turnbough
Feature Engineering	Abbey Rohler
Feature Engineering	Paribesh Chaudhary
Feature Engineering	codycook33@email.arizona.edu
Feature Engineering	German Alonso Perez Arispuro
Web App Customizations	Sophia Grace Turnbough
Web App Customizations	Abbey Rohler
Web App Customizations	Paribesh Chaudhary
Web App Customizations	codycook33@email.arizona.edu
Web App Customizations	German Alonso Perez Arispuro
Model Deployment	Sophia Grace Turnbough
Model Deployment	Abbey Rohler
Model Deployment	Paribesh Chaudhary

Model Deployment	codycook33@email.arizona.edu
Model Deployment	German Alonso Perez Arispuro
Validation	Sophia Grace Turnbough
Validation	Abbey Rohler
Validation	Paribesh Chaudhary
Validation	codycook33@email.arizona.edu
Validation	German Alonso Perez Arispuro
ML Models/Visualizations	Sophia Grace Turnbough
ML Models/Visualizations	Abbey Rohler
ML Models/Visualizations	Paribesh Chaudhary
ML Models/Visualizations	codycook33@email.arizona.edu
ML Models/Visualizations	German Alonso Perez Arispuro
Recommendation System	Sophia Grace Turnbough
Recommendation System	Abbey Rohler
Recommendation System	Paribesh Chaudhary
Recommendatio	codycook33@email.arizona.edu

n System			
Recommendation System		German Alonso Perez Arispuro	
Milestones & Planning → Project Approval	Final Name/Stack List	Abbey Rohler,Paribesh Chaudhary	2/2/2022
Milestones & Planning → Project Approval	Finalized Timeline/Task List	Sophia Grace Turnbough,Paribesh Chaudhary	2/2/2022
Milestones & Planning → Project Approval	Research Spotify streaming data vs. static set	codycook33@email.arizona.edu,German Alonso Perez Arispuro	2/2/2022
Milestones & Planning → Project Approval	Proposal Draft 2	Sophia Grace Turnbough,German Alonso Perez Arispuro,codycook33@email.arizona.edu,Paribesh Chaudhary,Abbey Rohler	2/2/2022
Milestones & Planning → Project Approval	Email Nick, Cristian & Will	Sophia Grace Turnbough	2/1/2022
Exploratory Data Analysis		Sophia Grace Turnbough	
Exploratory Data Analysis		Abbey Rohler	
Exploratory Data Analysis		Paribesh Chaudhary	
Exploratory Data Analysis		codycook33@email.arizona.edu	
Exploratory Data Analysis		German Alonso Perez Arispuro	

Demo + Presentation Prep	Sophia Grace Turnbough
Demo + Presentation Prep	Abbey Rohler
Demo + Presentation Prep	Paribesh Chaudhary
Demo + Presentation Prep	codycook33@email.arizona.edu
Demo + Presentation Prep	German Alonso Perez Arispuro

Final Deliverable Product Specifications

Compatible Systems:

Heroku web apps are compatible on desktop and mobile browsers.

Minimum Deliverable:

A customizable recommendation system based on user input deployed on a web app with a “Top —” dashboard, a recommendations dashboard, and 2 “behind the scenes” machine learning model views at minimum.

Expandable Deliverable:

Ideally, we will include a geo-analysis heat map of location based recommendations as well as optional viewing dashboards and/or custom features.

Data Availability:

In the event we cannot access active streaming data as we need it, we have a couple of static dataset options that we will have updated to the latest possible information per the project timeline.