MASTER OF INFORMATION AND COMMUNICATION STUDIES
Capstone Project

UNIVERSITY OF THE PHILIPPINES
OPEN UNIVERSITY

MASTER OF INFORMATION AND COMMUNICATION STUDIES

SULIT, JIMSON MONTILLA

Content-based Original Pilipino Music (OPM) recommender system centered on mood

Thesis Adviser:

Ria Mae H. Borromeo, Ph.D.
Faculty of Information and Communication Studies

29 January 2022

Permission of the classification of this academic work access is subject to the provisions of applicable laws, the provisions of the UP IPR policy and any contractual obligations:

Invention (I)	☐ Yes or ■ No
Publication (P)	☐ Yes or ■ No
Confidential (C)	☐ Yes or ■ No
Free (F)	■ Yes or ☐ No

Student's signature:

Thesis adviser signature:

University Permission Page

Content-based Original Pilipino Music (OPM) recommender system centered on mood

Jimson Sulit
November 20, 2022

Acceptance Page:

This paper prepared by JIMSON SULIT **with the title: "Content-based Original Pilipino Music (OPM) recommender system centered on mood" is hereby accepted by the Faculty of Information and Communication Studies, U.P. Open University, in partial fulfillment of the requirements for the degree Master of Information Systems.**

Ria Mae H. Borromeo, Ph.D.
NAME
Adviser

November 27, 2022
(Date)

Roberto B. Figueroa Jr.
NAME
Program Chair

December 8, 2022
(Date)

DIEGO SILANG S. MARANAN, PhD
Dean
Faculty of Information and Communication Studies

December 16, 2022

Acknowledgement

I would like to thank my adviser Dr. Ria H. Borromeo for her dedicated support and guidance. Dr. Ria continuously provided encouragement and was always willing and enthusiastic to assist in any way she could throughout the project. She was instrumental in fleshing out the ideas that resulted to the final project. I would also like to thank Pablo P. for making sure I am well fed during those sleepless nights. Finally, many thanks to all participants that took part in the testing and enabled this special project to be possible.

Dedication

To my mother, Olive, who encouraged me to fly towards my dreams and never regret my fall: always reminding me that the greatest tragedy of them all, is to never feel the burning light.

TABLE OF CONTENTS

List of Tables

List of Figures

Abstract

Recommender systems have become increasingly relevant due to the user-focused service delivery of streaming giants like Netflix and Spotify. Systems like this make it possible to offer media content that is highly similar to a user's preference and, if available, streaming history. Despite the availability of these enterprise recommender systems, recommendations are very global and general, requiring highly technical proficiency before one can use the technology to get recommendations for a very narrow genre like Original Pilipino Music (OPM).

This special project proposes a web-based recommender system using content-based filtering to recommend Original Pilipino Music (OPM) to music based on their mood, with recommendations subject to the user's input (e.g. is this song happy or sad?) that will be considered in the final recommendation. This project will also leverage audio content features available from the Spotify database for the model training part of the recommender system.

Both objective and subjective evaluations of the implemented recommender system are scoped down to the available audio features of available OPM or Filipino songs on the Spotify Database. While this project concurs with the findings of many researchers that music based solely on audio content or features does not provide very accurate recommendations, this project is a good minimum viable product (MVP) for music enthusiasts. Furthermore, the project is also proof of concept for a more robust OPM recommender system in the future, as detailed in the results and recommendations.

Keywords: music recommendation, recommender system, content-based recommendation, opm, original Pilipino music

Chapter I

THE PROBLEM DOMAIN

Statement of the Problem

Music recommender systems have become a staple in automatic music discovery through streaming platforms like Spotify, Apple Music, and Google Music, which have made music discovery a fast and easy process, given one's user data and extensive listening history. Recommender systems help users discover new music by providing recommendations in different forms but most especially through the generation of a new playlist containing songs that the system thinks the user might like.

These sophisticated enterprise-grade recommender systems from streaming giants like Spotify or Apple Music use user-generated meta-data such as play frequency and listening history, as the basis for the recommendations. In this approach, when choices are endless, music listeners and discoverers default to popular Western songs, generating a class imbalance from the frequency of metadata created as an effect of such popularity.

The problem with these metadata-based recommender systems is that they almost always do not recommend songs for which there is no data available (i.e., new songs). This also presents a barrier to people who wants to discover new OPM but does not have the means to pay for a premium streaming account that religiously generates user metadata and records listening history. Moreover, it is highly complex for these enterprise-level recommender systems to access and zero in on a recommendation for a specific market (e.g. Philippines), for a specific user. Furthermore, the user is boxed to their generated metadata and listening history, with no freedom to exercise their subjective judgments to direct the recommender system into a certain direction, given user input.

Background and Objectives of the Project

According to Rappler, Original Pilipino Music (OPM) content has garnered close to 10 billion streams on the music streaming platform Spotify. Nevertheless, this is nothing compared to the visibility and reach of globally known Western songs. When choices are endless, potential music listeners default to popular foreign songs. As part of the research for this project, foreign songs have dominated the local streaming market of the Philippines from 2017 to 2020. While this is entirely a different issue, this presents an opportunity to leverage technology to put focus on OPM and make local music discovery an easier process that could be iteratively improved along the way.

Recommending music on streaming platforms presents many challenges. With users boxed to their generated metadata and listening history, streaming giants have been working tirelessly to improve their recommender systems to take into consideration that music judgment and perception are affected by the context of the user. While not hugely impacting, characteristics like age, geography, musical knowledge and exposure are difficult to capture and fit into the equation. This makes the process of building a recommender system much more complicated than it looks.

Furthermore, there is no readily available tool for music discovery specifically for OPM. Since our focus is OPM discovery and discovery that democratizes music discovery through technology without premium subscriptions, we will rely on content-based methods. Personalized or user-centric music suggestions and recommendations typically come from algorithms that leverage collaborative filtering or content-based methods. Collaborative filtering relies on user history and previous interactions between a user and an item (i.e., a song), sometimes even considering other metrics. On the other hand, content-based methods take into account the properties or features of the items.

Given the background of the problem and the identified opportunity above, we see the need to create a highly targeted recommender system to address the existing metadata class imbalance brought by the popularity and dominance of Western songs in major streaming platforms like Spotify and Apple Music. While a multi-feature and multi-faceted approach to tackling the problem is a good recommendation for the future roadmap of this project, starting small and limiting the scope of our project makes our proposal more realistic. This project aims to focus on quantifying specified emotions (e.g., happy, sad) by users and take those into consideration to capture the subjective judgment and (additional) preference of a user to discover less popular Original Pilipino Music (OPM) songs that they could listen to.

Our approach will be to use content-based method to build our recommender system. One major issue with the existing recommenders from streaming platforms is that it is more sophisticated and it is heavily dependent on the data from similar users. In our approach, there is no need for data on other users and since we are concerned with OPM, we will be able to recommend new and unpopular items – which is the main objective of this project – put spotlight on less popular OPM songs on the web.

Significance and Scope of the Project

The proposed solution is deemed to address the following issues:

- Unavailability of music discovery tool for OPM
- Unavailability of OPM discovery tool for premium and non-premium streaming platform users (in our scope, Spotify)

- Unavailability of music recommender systems that takes in user input, not solely relying on generated metadata and listening history (which is virtually zero for new or one-time users) from major streaming platforms, a content-based recommender system that can easily be explored and used by individual users
- Unavailability of a targeted song recommender system for a specific music genre (i.e. OPM) that takes into consideration additional user input (e.g. if a sampled song is happy or sad, according to their personal judgment)

The proposed system is expected to address all these mentioned unavailability issues to a create a simple system that can be used for straightforward potential OPM discovery for individual users.

Data Scope

Our data set for the model training and testing is scoped down to approximately 1,500,000 songs from the Spotify database for the Philippine market from 2017 to 2021. These selected songs are chart-topper songs. While a huge chunk of these songs is not OPM, it is important to use them for the training as these are the tunes that are being played by Filipino users, and our goal is to cater to users who are mostly in the Philippine market. Nevertheless, for the recommendation part of the model, songs that will be used are OPM songs only.

Our approach is to use a content-based method to build our recommender system. While this approach lets us recommend new and unpopular items, one of the limitations of this approach is that evaluation is pretty subjective. However, since we will be taking in input from the user, our method works well in capturing user's subjectivity, recommending songs unique to their preference.

Documentation of Existence and Seriousness of the Problem

In the case of this project, the inspiration for this undertaking is the enterprise-level recommender systems of streaming platforms like Netflix and Spotify. Specifically, the goal is to create a tool for the issue and opportunity that we have identified – unavailability of a simple music discovery tool or recommender system specifically for OPM songs. According to published documentation, Spotify's recommender system provides suggestions for a user based on his historical interactions, the attributes of the songs/artists he listens to, and the preferences that are calculated to identify 'similar' users and cluster them together to further optimize the recommendations. One objective of the system that we propose is that while user login is required, we would not need for our users to have these historical interactions as they can input data and generate recommendations based on their answers or input.

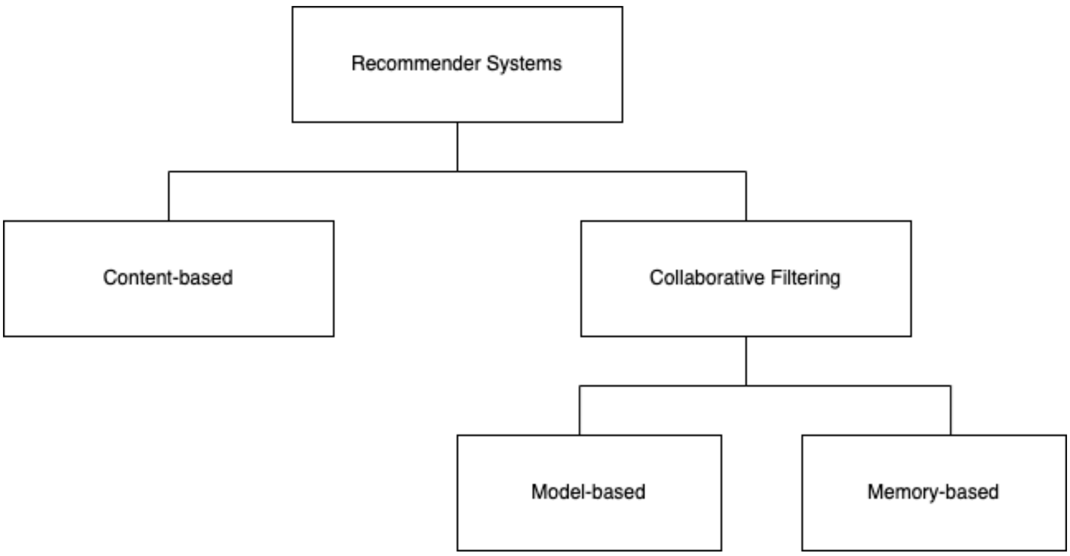

Figure 1 Recommender Systems

The main challenge in solely relying on Spotify's recommender system is the complexity of their system and the inaccessibility to add user input (i.e., a user identifying if a song is happy or sad based on personal judgment). Furthermore, the end-to-end process of giving highly relevant recommendations only works well with users who have established listening history. Although they also employ content-based models for new or budding users, there is no option on the platform to specify it by specific genre (e.g. OPM) or for users to input more data to inform the model of their preferences.

Chapter II

REVIEW OF EXISTING ALTERNATIVES

Here are some of the existing systems available on the web that inspired, and their current issues which led us to proposing this project:

Name	***Chosic*** **Music Tools**
License	Standard Copyright License
Platform (Web/Mobile/Desktop)	Web
Brief Description	*Chosic* is a web application for music discovery with multiple music tools to find music based on genre, song similarity, etc.
Basic Functionalities	Similar Song Finder Similar Artists Finder Playlist analyzer
How is this related to your system?	This web platform offers several related tools that can perform nested searches
Can you use this system instead? Why? Why not?	The platform code is not open source but the system can be used to benchmark best practices for a better user experience. Their similar song finder tool focuses solely on content-based method with no available option for users to inform the tool of their personal preferences

Table 1 Chosic Music Tools Assessment

Name	**MoodFuse**
License	MIT License (MIT)
Platform (Web/Mobile/Desktop)	Web
Brief Description	MoodFuse is a mood-based recommender system that uses a meter to recommend songs
Basic Functionalities	Genre selection dropdown Energy, danceability, and happiness meter
How is this related to your system?	The platform is similar to what we want to build in this project, a simple UI that can return a list of songs based on how a user perceived five sampled songs. While this did not explicitly used moods, energy, danceability, and happiness are music metadata that can be leveraged to further identify distinct moods.
Can you use this system instead? Why? Why not?	Open-source code but not maintained and updated as the web application is not working when it was tested as part of the research phase. Like the first one, this can be studied to inform our implementation

Table 2 MoodFuse Assessment

Name	**Mood Playlist**
License	Standard Copyright License
Platform (Web/Mobile/Desktop)	Web
Brief Description	MOOD Playlist is a playlist generator that allows you to automatically create personalized Spotify playlists that suit your mood and taste.
Basic Functionalities	Spotify Login Artist
How is this related to your system?	The tool focuses on creating playlist based on a combination of five or more artists from a user's listening history. Under the hood, I might use a similar algorithm to group OPM songs and artists and that is one similarity to the system that I want to build, aside from it creating playlist based on moods. It uses two music metadata or metrics, namely, energy and valence to group moods.
Can you use this system instead? Why? Why not?	This application is owned by Sony entertainment so this is not definitely open source. Like the first system, I can use this as a benchmark or an inspiration. I can also do further work to research related applications made by the authors and see if there is a chunk of code that may be applicable (or usable) to build the logic that I want for the system.

Table 3 Mood Playlist Assessment

While these existing systems recommend music on their own, they did not add much to the offerings of the Spotify API. The main challenge in solely relying on Spotify's recommender system is the complexity of their system and the inaccessibility to add user input (i.e., a user identifying if a song is happy or sad based on personal judgment). Furthermore, the end-to-end process of giving highly relevant recommendations only works well with users who have established listening history, and our offering is not just limited to those with established listening history. These three systems employ historical interactions and although the last two systems get input from the user, it is quite different from the implementation that we want – where user can sample songs and judge if it is, say, happy or sad according to their perception. Such additional information will be used to inform the recommendation part of the system to give users more relevant and specific offerings.

Chapter III

APPROACH TO BE TAKEN IN THIS PROJECT

Theoretical Framework

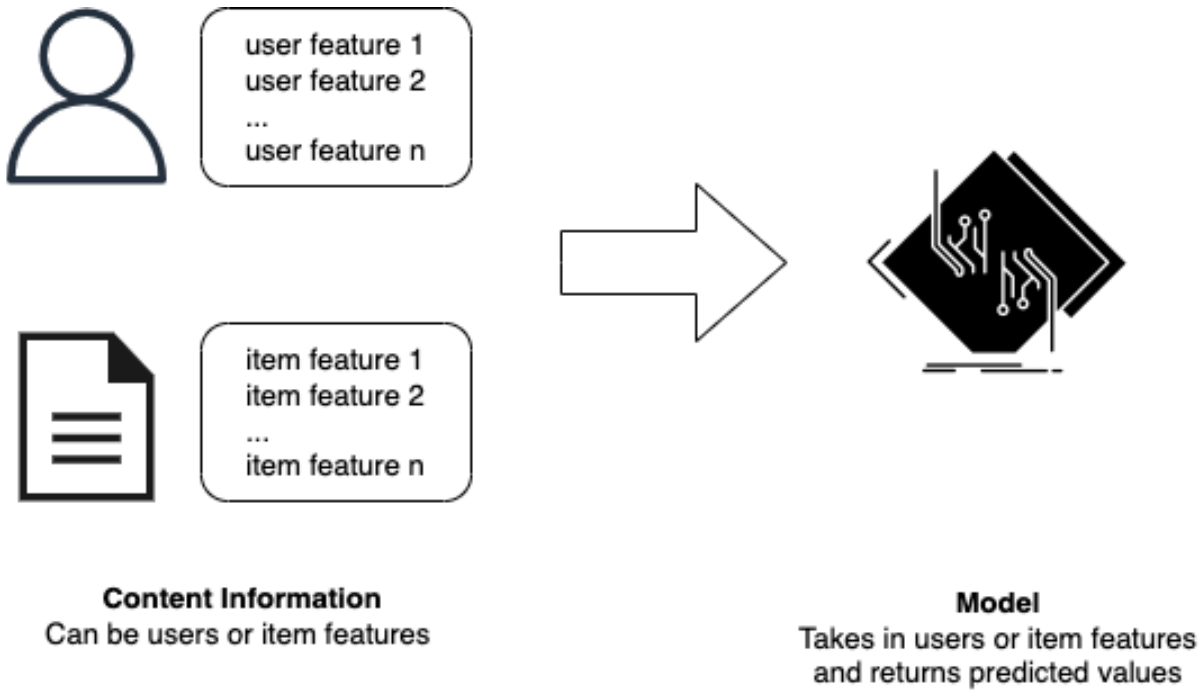

Figure 2 Content-based method framework

Content-based method

It is a huge challenge to recommend anything to new users or to recommend a new item to any users and many users or items usually have too few interactions to be efficiently handled. In the realm of recommender systems, this is called the "cold start problem". The idea of a content based method is to try to build a model based on the available features. In our case, this will be the audio file or song metadata and the user input.

Content-based methods suffer far less from this so-called “cold start problem” compared to their collaborative and hybrid method counterparts. In a content-based method, new users can be described by their characteristics. Forcing them to input features as they interact with our application also helps us shape the dimensions of the features that will be used so we can recommend more relevant items.

In this framework, we will treat the problem of recommending songs to users based on identified mood as a classification problem (e.g., happy or sad). This is where we can set a model that will be based on the user and/or item features at our disposal. Using this framework, we will also address these identified issues:

- Unavailability of music recommender systems that takes in user input, not solely relying on generated metadata and listening history (which is virtually zero for new or one-time users) from major streaming platforms, a content-based recommender system that can easily be explored and used by individual users
- Unavailability of a targeted song recommender system for a specific music genre (i.e. OPM) that takes into consideration additional user input (e.g. if a sampled song is happy or sad, according to their personal judgment)

Our goal is to create a simple system that can be used for OPM discovery using the user’s mood and build a product that can potentially springboard to future projects that can build upon our final product and improve on what we have delivered.

Rationale of the Framework

Using a content based approach will let us address the issues that we have identified earlier. Relying solely on generated metadata and listening history, and in turn collaborative filtering approaches, will make this undertaking much more complex than what it is currently scoped for. Unlike the recommender systems of giant streaming platforms, this system that we will build is meant to be used by individual users or a small group of people. Moreover, this approach is suited for building our recommender system from the ground up while taking advantage of the availability and richness of data available from the Spotify database.

Technologies you plan to consider or use

The final deliverable will be a web application where users can play around with different parameters and sample listening to the songs recommended by the system. Song recommendations will be based on the inputted values of available parameters that they chose. The following are the technologies used for the delivery of the project:

For data extraction / data preprocessing / model development and testing

- Python 3
- Jupyter Colab Notebook
- Python-enabled libraries: numpy, pandas, scikit-learn, streamlit

For modelling and feature engineering the four different mood tiers

- Scikit-learn's K-nearest neighbors' algorithm

For web design and development

- DevExpress ver. 21.2.4 for MVC Extensions
- Bootstrap ver. 5.1
- Microsoft SQL Server 2019 (database)

For pre-deployment and web deployment activities

- Heroku and Streamlit Cloud Services

Chapter IV

CHAPTER PLAN

Concept

Platform usability is the key aspect of the project and this will rely on user testing. To further improve the initial iteration of the recommender system and platform, the deliverable includes user testing and a quantitative survey asking feedback regarding the usability of the project.

Methods

Several parameters have been added for platform users to customize the recommendations of the system. These parameters are based on the original audio features from the Spotify database namely acousticness, danceability, energy, instrumentalness, valence, and tempo.

Plan for User Testing and Project Assessment

For the delivered application, testers were asked different questions regarding the usability and practicality of the delivered applications. The questions were asked to correctly evaluate the recommender system in a way that the feedback can be used to improve the existing system.

As a whole, the version 1 of the delivered system could be improved in terms of user experience. Furthermore, users asked for a more flexible tool where they themselves can tweak the different parameters available to change the results of the recommendations, according to their preference.

Usability

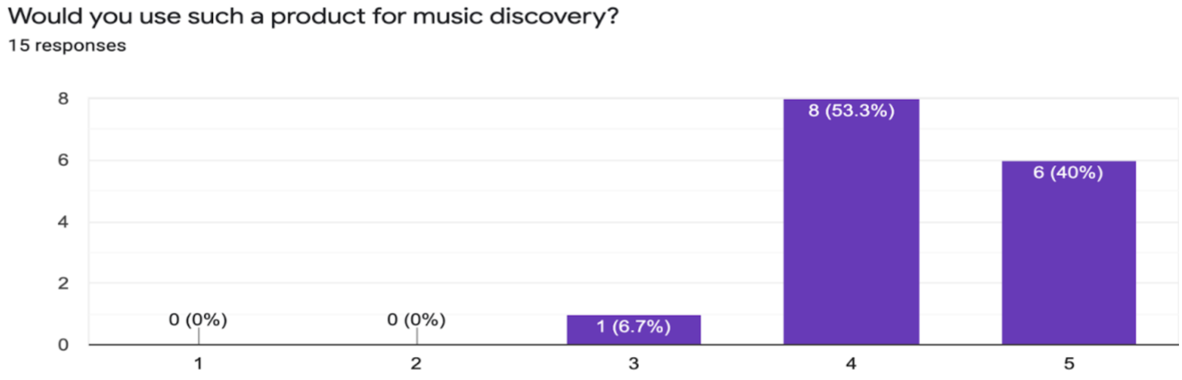

Figure 3 Usability Survey Results

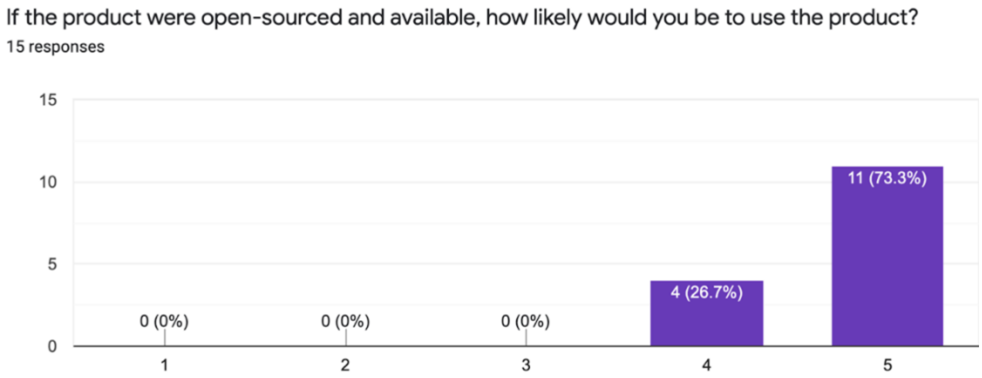

Figure 4 Usability Survey Results (continued)

Users noted their willingness and enthusiasm to use such a tool to discover new Original Pilipino Music. Furthermore, this new type discovery is placed the spotlight on the availability of quality OPM but are not mainstream. In the companion survey, users listed song and music discovery as the thing that they liked the most about the application.

In your own words, what are the things you like the most about this product?

15 responses

discovering OPM songs that are not mainstream

listening for free to new songs that might not be available in other platforms

knowing about audio features and how it affects a song

discovering indie pinoy hip hop

being able to discover new music

playing around the tool and listening to new music

many new filipino songs to listen to

listening to old and new OPM songs

I like learning about the different dimensions of a song and how it affects one's preference for music

Figure 5 Product likeability

Final Requests

As part of the user testing feedback, testers were also asked how they would want to improve the product. We can see from the user feedback the huge interest in giving them the freedom to play around or customize the major parameters without them getting too technical with the different audio features and how it can change a song recommendation.

In your own words, what are the things that you would most like to improve in this new product?

15 responses

- more customization
- more indie or songs from small bands
- more functionalities more song
- more song choices or categories and filters
- simpler functions
- the audio names should be simplified for a normal user
- put more functionality so user will use this as a regular tool
- saving my session and seeing history of recommendations
- A login optionality to see my recommendations and to notifications for newly released songs

Figure 6 Product feedback

The final Graphical User Interface (GUI) was improved and designed according to the request of alpha testers. Certain issues were encountered with the original plan of letting users sample and label five songs where they will either tag it as happy or sad. This original approach was problematic for a few reasons. First, the necessary API calls to Spotify's database require an enterprise account. Certain resources provided workarounds to delay API calls to adhere to the limits of a Spotify developer account but nothing worked enough such that the tool would be usable to the users. Although the backend work for this original approach is complex, users also felt that the tool is too simplistic for them to use and requested for more freedom to customize the results of the recommender system. Hence, we changed the approach in relation to the user feedback and inputs of the adviser to improve the system while staying true to the original proposal for the special project.

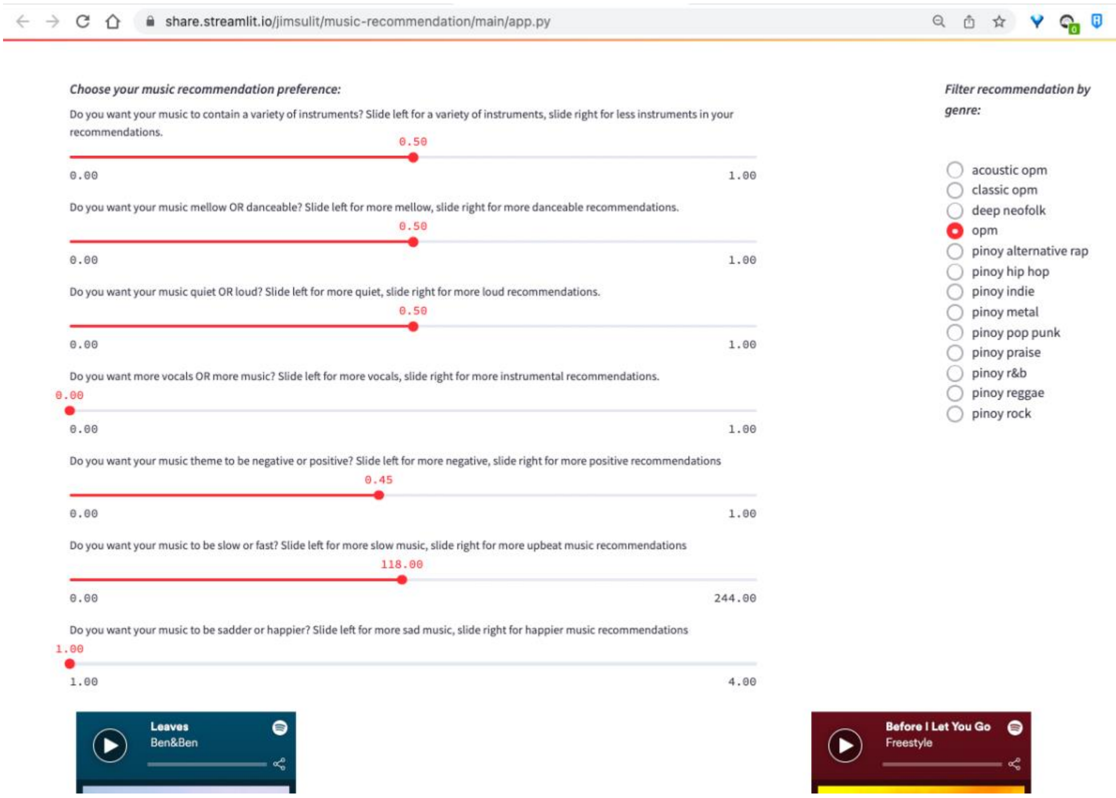

Figure 7 Final recommender system front-end / GUI

Main Parameters of the Recommender System

Several parameters have been added for users to customize the recommendations of the system. These parameters are based on the original audio features from the Spotify database namely *acousticness, danceability, energy, instrumentalness, valence, and tempo.* These audio features make up the first six parameters. For our main parameter, tweaking the song to *less (sadder) or more (happier)*, the range of values were feature engineered from the audio features *energy* and *valence,* we are calling it *mood label value* for this project*.* The different tiers or value groups were classified and approximated to make sure that we have a balanced song data bank.

	artist_name_x	track_name	energy	valence	mood_label
0	SUD	Sila	0.376	0.280	2.0
1	Silent Sanctuary	Pasensya Ka Na	0.262	0.265	2.0
2	Up Dharma Down	Tadhana	0.322	0.511	3.0
3	Moira Dela Torre	Malaya (Camp Sawi Original Motion Picture Soun...	0.125	0.108	1.0
4	Up Dharma Down	Sigurado	0.669	0.963	4.0
...	...	...	...	...	...
557	Kiyo	Ikaw lang	0.499	0.501	3.0
558	Skusta Clee	Karma	0.567	0.439	3.0
559	M Zhayt	Para Paraan	0.491	0.600	4.0
560	mrld	Ligaya	0.533	0.343	2.0
561	Zack Tabudlo	Pano	0.457	0.415	3.0

Figure 8 Original data frame including the engineered *mood label values*

Mood Label Values Table				
1.0	2.0	2.5	3.0	4.0
Saddest	Sadder	Neutral (Middle ground)	Happier	Happiest

Table 4 Mood label values table

Security Testing

No web and user experience impacting issues were found in the web application and cloud service. Moreover, since user login is removed as a functionality, no critical user information would be subject to tampering or defacement as the web application continues to go live.

Test	Status
Checking for website accessibility	**PASSED**
Checking for missing HTTP header - Strict-Transport-Security	**PASSED**
Checking for missing HTTP header - Content Security Policy	**PASSED**
Checking for missing HTTP header - X-Frame-Options	**PASSED**
Checking for missing HTTP header - X-XSS-Protection	**PASSED**
Checking for missing HTTP header - X-Content-Type-Options	**PASSED**
Checking for missing HTTP header - Referrer	**PASSED**
Checking for client access policies	**PASSED**
Checking for use of untrusted certificates	**PASSED**

Checking for enabled HTTP debug methods	**PASSED**
Checking for secure communication	**PASSED**
Checking for directory listing	**PASSED**
Checking for domain too loose set for cookies	**PASSED**
Checking for HttpOnly flag of cookie	**PASSED**
Checking for Secure flag of cookie	**PASSED**
No malware detected by scan	**PASSED**
No injected spam detected	**PASSED**
No defacements detected	**PASSED**
No internal server errors detected	**PASSED**

Table 5 Security Testing Results

Chapter V

RESULTS AND DISCUSSION

Original Approach: Cosine Similarity

The main motivation to use cosine similarity is because Spotify uses this concept to an extent to power their enterprise-level recommender system. The goal is to create a proof of concept or a local level version of their recommender system focused on Original Pilipino Music (OPM). Essentially, cosine similarity is a way for a system to understand how similar two things are.

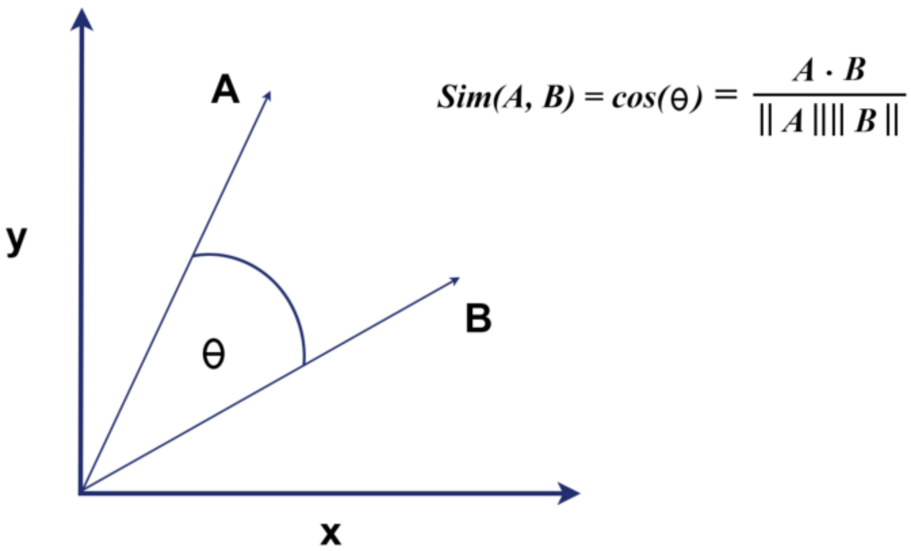

Figure 9 Cosine similarity formula

Figure 9 illustrates the concept of cosine similarity. It follows that we calculate the distance between the points (in our case, this could be the song or the chosen audio features of the song). The theory behind it is that the smaller the distance, the more similar they are. The complexity of the work comes from manually standardizing and *vectorizing* the selected audio features for the project. Furthermore, the operation of comparing and looping songs on a one-to-one basis proved to be computationally verbose and inefficient. There must be a solution to deliver this efficiently and elegantly but that might take even more time to deliver the system and will surely result in missing the deadline by a long stretch.

Failure in debugging surrounding the issues behind the manual standardization and vectorization of audio features for cosine similarity analysis caused the longest project timeline delay. The original idea is to implement this logic from scratch to have a good baseline for future work. Originally, the system will take input from the user (asking if the displayed song is happy or not). This logic fits well with cosine similarity. Since manual standardization proved to be more complex given the time at hand, it was ultimately decided to look for an alternative and easier approach to deliver the recommender system.

This is a classic trade-off encountered in the software development life cycle: a choice between a more complex system adhering to original plans or a simpler system that is stripped down of proposed complexity but is enough to satisfy the essence of the project. Hence, we looked for an alternative approach to deliver the system and taking advantage of the inputs from the user feedback survey.

Alternative Approach: K-Nearest Neighbors (k-NN) + rule-based mood classification

Since this is essentially a data classification problem, looking through a variety of distance-based algorithms is the best place to start to look for an alternative solution. The k-nearest neighbor classification (k-NN) is one of the most popular distance-based algorithms. This classification method is based on measuring the distances between the test sample and the training samples to determine the final classification output. Traditionally, a k-NN classifier works naturally with numerical data.

The k-NN algorithm, when implemented in music mood classification, looks at similar songs and assumes that they belong to the same category because they seem to be near to each other. There are other techniques that are also used to classify songs but historically, k-NN yields better results especially in multi-class/multi-feature processing for classification.

In this project, we use the k-NN algorithm to group songs according to the following features: *acousticness, danceability, energy, instrumentalness, valence, and tempo.* This way, the recommender system can suggest songs that are very similar to each other. As for the second part of the main recommender system logic, we solved for the product of the *energy* and *valence* feature values and standardized them across our song bank. From here, we created a rule-based tiers/groups. Below illustrates the resulting engineered feature.

Mood Label Values Table				
1.0	2.0	2.5	3.0	4.0
Saddest	Sadder	Neutral (Middle ground)	Happier	Happiest

Table 6 Mood label values table

The rule-based mood classification is an important component of the project to make sure that while we have taken a different direction from what was proposed for IS 295A, we still maintain the essence of the project – a recommender system centered on defined mood levels/values. We used the product of *valence* and *energy* as these two features are closely correlated. Although we could use *danceability*, we want to avoid wrongly labeling upbeat sad songs.

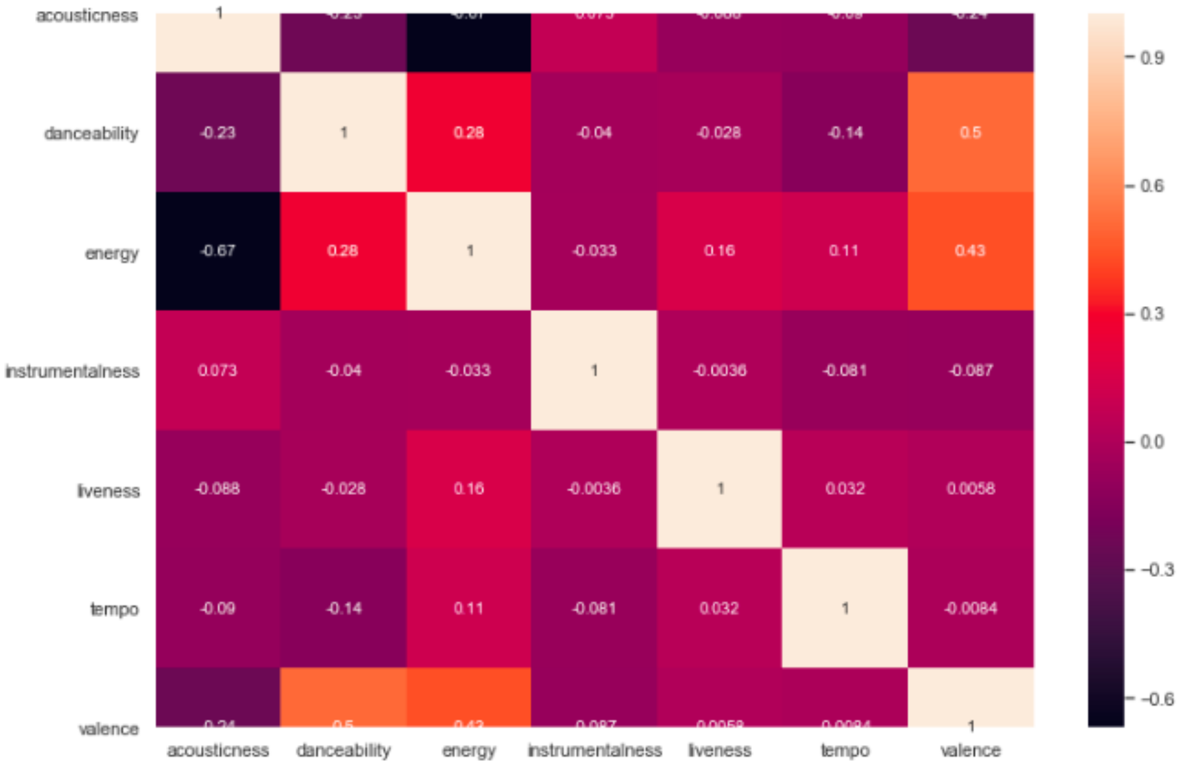

Figure 10 Audio features correlation heat map

User Feedback – Final Version of the Recommender System

Going back to our users who tested the v1, after applying the workaround and taking into consideration the inputs, they have unanimously validated that such tool would be essential for OPM song discovery, with most of them mainly discovering small artists in the *Indie* and *Hip-Hop* music spaces. They also provided insights regarding the mood label values slider on the web application. While they have noticed the differences of the level of 'sadness' or 'happiness' of the recommended songs when they move the slider, most of them remarked that such nuances might not be noticeable for a normal user, one who might not necessarily be categorized as an OPM or music enthusiast.

Limitations of the Recommender System

As pointed out by the users during our testing discussions for the final version of the recommender system, the recommendations from tweaking the mood label values could still be improved. I have detailed the following limitations for the current implementation:

- The data set contains songs from Spotify's Top 200 from 2015 to 2021. As such, our main song bank is limited to approximately 500 OPM songs.
- We are limited to the constraint of songs that appeared to the annual Top 200 songs on Spotify. Once we add new songs to the existing song bank, we have to define new constraints and requirements to justify the addition of new songs. Furthermore, we would also have to deep dive to another analysis and see if we need additional steps to standardize, normalize, and vectorize the audio features of the new songs or we will retain the current implementation.
- A user pointed out that we could train a model to explicitly classify and label songs as either sad or happy. The missing link to this approach is we need to have a bigger data set and ensure that we have an almost balanced dataset to assure that the model would not be biased to certain audio features or genres. This could also be a potential future work idea.

Continuing the discussion on the final version, as an exercise to potentially improve the project for future work, I have also asked the users what features they would like to see for them to pay for this kind of service. Or if paying is out of question, what would make them regular users of the application. In both cases, the stand out answer is inputting either the name of the artist or song, and returning similar artists or songs. This includes a dashboard of the artists that they follow and potential releases and the option to integrate to their favorite music streaming provider including the option to check their listening history. In conclusion, these are all potential future work.

Exploratory Data Analysis Insights

Our data set include OPM chart-toppers in Spotify from 2015 to 2021. Looking at the distribution of OPM songs below, most of the chart-toppers were from the years 2017-2021. Although it is not seen in the data visualization below, the oldest song in our data set is the song *Panalangin* by the APO Hiking Society. The years 2017-2021 were good years for OPM on Spotify. It is sensible that most of these songs are from these years as that is when Spotify became more famous in the Philippines. It is worthy to note that while years 2017-2021, Filipino listeners also listened to old songs from the 1980s and 1990s, the oldest being from the year 1980.

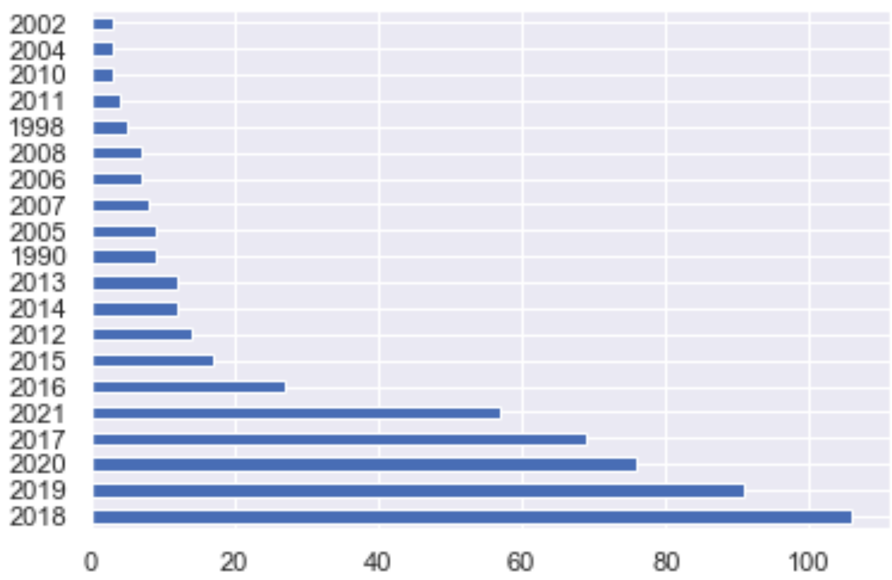

Figure 11 OPM Chart-toppers by year

OPM Artist Distribution

Although 2017-2021 were good years for OPM on Spotify, foreign artists still make up 80% of the charts. This could be another future exploratory work to get to the bottom of why this is the status quo.

```
# data from 2015 to 2021
print("Percentage of OPM artists in Spotify Top 200:", str(df_ph.shape[0]/df_artists.shape[0] * 100) + "%")
```

```
Percentage of OPM artists in Spotify Top 200: 20.162224797219004%
```

Figure 12 % of OPM artists in Spotify Top 200

Percentage of OPM artists in Spotify Top 200: ***20.16%***

Popularity and Danceability

An interesting relationship that we have seen during the exploration is that most songs that are popular are highly danceable. Popularity was a classification used by Spotify and is included as part of the data that can be used by developers. According to Spotify, “popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played a lot now will have a higher popularity than songs that were played a lot in the past.

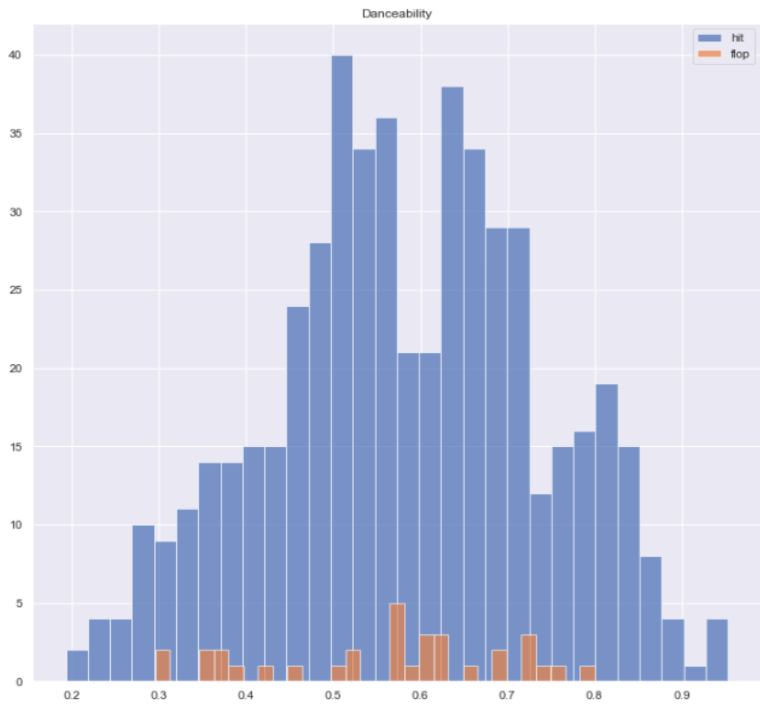

Figure 13 Song popularity and danceability relationship

Audio Feature Trends

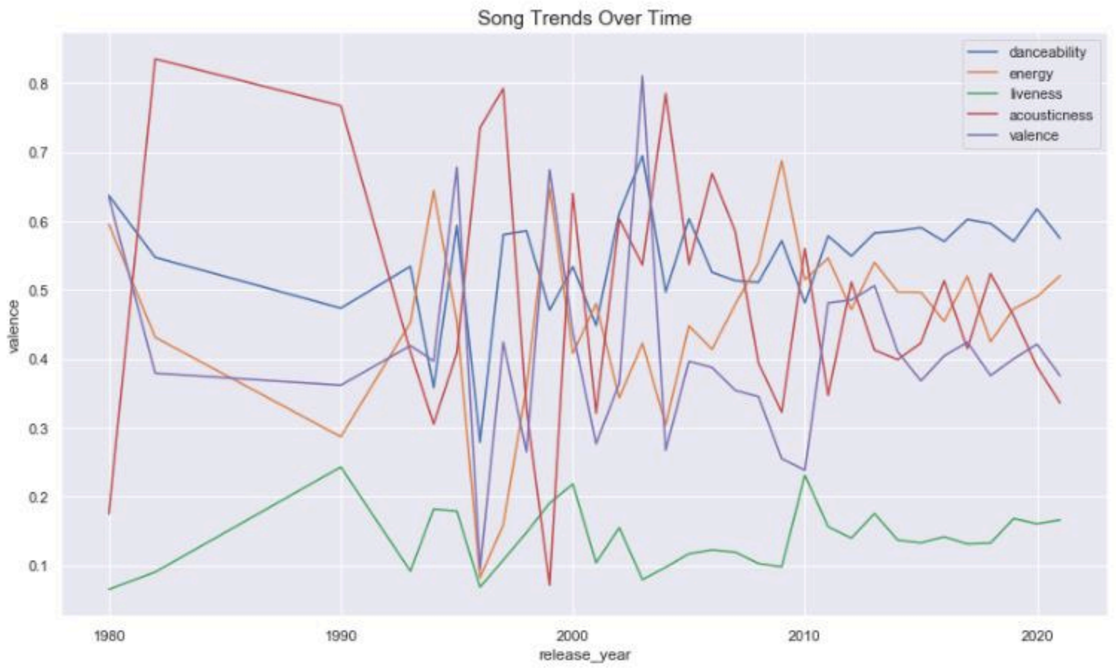

Figure 14 OPM Song audio features average trend over time

The audio features of songs over the years change but the most varied and prominent fluctuations happen with energy, danceability, and acousticness features. This could mean that, while not strictly observed, OPM has gone through so many musical evolutions since 1980. Therefore, for future work, this feature evolution could be taken into consideration in improving the results of an OPM recommender system.

Development Challenges

The original direction for this project is to implement an algorithm from scratch using cosine similarity to preprocess and classify the songs. For the algorithm input, users will label and sample five songs – they will be asked if they think a song is either happy or sad. Developmental issues were encountered, particularly in the algorithm that was originally developed. While it has worked fine on a local machine, limitations of the required dependencies when it was deployed to online web services like *Heroku* became a source of multiple problems. The development phase needed two additional weeks’ worth of work consumed just to troubleshoot this deployment dependency issues. Given the time constraint, resources ultimately led to using another cloud service called *Streamlit Cloud Services.*

As stated above, the original idea is to implement a classification algorithm from scratch employing cosine similarity. The biggest blocker to this approach is the complexity that surrounds and the time it takes to manually standardize and vectorize the features that will be used for mood classification. In the end, the original idea provided so many potentials but would take a longer time, realistically speaking. For the completion of this project, the clear decision is to simplify the project and implement an existing library or algorithm to execute the classification.

Chapter VI

CONCLUSIONS

In terms of usability for version 1, users have said that the system might be helpful but not an absolute necessity. In our studies, we have learned that an effective management information system (MIS) is a user-machine interface system for providing information to support the organizational operations, management, analysis, and decision-making functions. While we have taken a different direction from the original proposal, an effective MIS is a system that is operational, usable, and satisfies the user requirements.

The strict adherence to the original proposal and requirements can also be negated by the need to deliver an operational system to the stakeholders. This is a testament to the fact that iterative feedback from experts and users is crucial to delivering a product that will satisfy all stakeholders, regardless of how much the complexity of the product was reduced. Moreover, we have learned that product complexity can be deprioritized in favor of delivering a simpler but working system that still captures the essence of the requirements. The tradeoff between giving up complexity and delivering a working product will always be an important consideration. In the context of the project, the latter should be favored.

Undertaking this special project also highlights the importance of listening to end-users and experts that can help us brainstorm simpler but more intuitive iterations of the system. From a developer perspective, complex engineering would translate to better user feedback. This is not always the case, as evidenced by the development issues that were encountered. Soliciting detailed user feedback and brainstorming with experienced professionals and experts would always yield a more reliable final product.

In terms of the OPM-related insights that we have gathered, OPM artists making up 20% of the Top 200 is a sobering fact that something must be done for Filipinos to appreciate more OPM. Further work to democratize OPM music discovery is imperative. Moreover, OPM artists could take notes from foreign chart-topping songs to see what they can do to make OPM songs reach more Filipino listeners. These audio features can provide further insights on what minor tweaks they can employ without giving up their unique artistry and musicality. Furthermore, aspiring OPM artists can emulate what was done by OPM chart-toppers from 2017 to 2021 to replicate their streaming success. For example, more danceable tunes seem to translate to more chart-topping songs.

Chapter VII

RECOMMENDATIONS

The original approach of implementing the algorithm from scratch which gives better granularity and accuracy in terms of classifying the songs can be a future work for this project and this can be integrated to the deployed version to better improve the classifications and recommendations. Furthermore, as pointed out by a user, training a model to explicitly classify and label songs to different emotions can also be a future direction to take. With a bigger balanced data set, this can be undertaken, from here, other emotions can be added and can be computationally classified given the newly identified groups and thresholds.

As an OPM music discovery tool, we could also add the functionality of like a *search engine* for similar artists and songs. This might be a herculean project to undertake but that would be valuable for music enthusiasts. Moreover, with this idea, a dashboard integrating to a user's preferred streaming application can also be added together with a dashboard that intuitively displays personal listening analytics.

Another potential future direction is to focus and double down on one specific OPM genre and generate insights for small indie artists. The 20% composition of OPM artists in the chart-toppers in the Philippines is a compelling reason to take the valuable insights discovered from this project and further deep dive on what existing artists or bands can do to reach more listeners, as evidenced by the trends and characteristics of the songs that usually becomes a hit with Filipino users. This could be another future exploratory work to get to the bottom of this discovered status quo.

REFERENCES

A. Printed Book/E-book

1. Printed book with one author

N.-H. Liu, “Comparison of content-based music recommendation using different distance estimation methods,” Applied Intelligence, vol. 38, no. 2, pp. 160–174, 2012.

B. Serials

1. Journal Article with Multiple Authors

P. Cano, M. Koppenberger, and N. Wack, “Content-based music audio recommendation,” Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA '05, 2005.

S. Sharma, P. Fulzele and I. Sreedevi, "Novel hybrid model for music genre classification based on support vector machine," 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2018, pp. 395-400, doi: 10.1109/ISCAIE.2018.8405505.

E. N. Tamatjita and A. W. Mahastama, "Comparison of music genre classification using Nearest Centroid Classifier and k-Nearest Neighbours," 2016 International Conference on Information Management and Technology (ICIMTech), 2016, pp. 118-123, doi: 10.1109/ICIMTech.2016.7930314.

X. Wang and Y. Wang, “Improving content-based and hybrid music recommendation using Deep Learning,” Proceedings of the 22nd ACM international conference on Multimedia, 2014.

C. Website

1. Website

D. R. Chaudhuri, S. A. Khan, and S. Firdos, “Content-based music recommendation system,” *Medium*, 17-May-2020. [Online]. Available: https://medium.com/@dibyendu19034/content-based-music-recommendation-system-74f30bccc239. [Accessed: 30-Jan-2022].

S. Das, “Beginners Guide to learn about content based recommender engine,” *Analytics Vidhya*, 24-Sep-2015. [Online]. Available: https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/. [Accessed: 31-Jan-2022].

M. Hilsdorf, “Build your first mood-based music recommendation system in python,” *Medium*, 18-Oct-2021. [Online]. Available: https://towardsdatascience.com/build-your-first-mood-based-music-recommendation-system-in-python-26a427308d96. [Accessed: 30-Jan-2022].

C. Hu, “Why Spotify's music recommendations always seem so spot on,” *Popular Science*, 03-Dec-2021. [Online]. Available: https://www.popsci.com/technology/spotify-audio-recommendation-research/. [Accessed: 30-Jan-2022].

“Content-based filtering Recommendation Systems,” Google. [Online]. Available: https://developers.google.com/machine-learning/recommendation/content-based/basics. [Accessed: 31-Jan-2022].

Appendices

APPENDIX A

Deliverables and Milestones