MASTER OF TECHNOLOGY
PROJECT REPORT

S-Finder
Intelligent Music Recommender System

Team members

Xu Jiayu
Cai Zheng
Tang Enze

Contents

1 EXECUTIVE SUMMARY

Music has been a cultural product since AD 1. This form is a way to convey and express feelings. As time goes on, more and more outstanding musicians have created one classic and popular song after another. After being sung and rewritten from generation to generation, the song itself can move people's hearts or express higher spiritual significance. Music can inspire people, give them hope, make them believe in something better, or keep them away for a while. When you need it, it will be your good friend, when you need it, you will find it. The proverb "Put on the music, pause the memory, stop the pain, put the joy back" is the easiest thing to do. Music is a means of connecting all cultures. Different cultures have music, and we can easily see how different cultures are combined.

It is because of modern people's strong desire and need for music that many music apps related to it have emerged. In the age of the information explosion, huge network resources provide a great convenience for our daily life. Whether it's pop, rock, folk, electronica, jazz, absolute music, rap, metal, world music, New Age, classical, indie, ambient music, and so on. According to the school. But, sometimes, if you have more choices, then it's hard to find your favorite song in a big pile. Many music apps recognize this problem and offer smart personalized advice to help them find their hobbies and preferences. However, the current recommendation system on the music market is not perfect, and the types of songs obtained by users are often single, which will produce the effect of "cocoon" over time.

Our project team wanted to solve the recommendation flexibility problem in NetEase Cloud Music platforms. Our intelligent recommendation system can conveniently help users to choose a new singer and his or her songs to enjoy. Only providing users the NetEase Cloud Music URL of their favorite playlist, the recommendation system will provide users with some best choices according to their existing music in a real and short time.

We used crawl skills to scrape a large amount of music in the NetEase

Cloud Music platform to build a dataset. After gaining a large amount of information from some dimensions of music, we selected several features, such as ‘singer’, ‘category’, ‘release time’, ‘lyrics emotion’ and so on. Firstly, we decreased the dimensions of features by PCA, and then we collected user information and found the most suitable music by Item-KNN algorithm which will be recommended to users. Besides, we designed a fresh and NetEase Cloud Music-type-like interface to make users a better search experience.

Our team has a clear work distribution and feels very happy to work together to create such an intelligent system. And definitely we hope our project can be put into commercial use and be used and receive feedback from more users. We can also upgrade the database in real time and may set a database for users in the future if more and more users are willing to use our commendation system.

2 PROBLEM DESCRIPTION

A large number of music categories not only meet users' needs for different kinds of songs but also have problems choosing the most suitable music they would like to listen to and enjoy with their friends, hard to clearly find out singers and music according to their hobbies and interests. Many music platforms have realized this and implemented intelligent recommendations for helping users discover their own preferences. It is so important and self-evident to have a strong and satisfying song recommendation module in a music platform.

Most music platform always recommends their music that is sold at a high price or some music that is popular in the recent period. Some other suitable music is ignored by users and users, and that music is probably forgotten in the corner. There is no perfect recommendation system that only recommends singers and music according to users' current favorite music inventory. It is difficult to obtain the optional music effectively and efficiently.

2.1 PROJECT OBJECTIVE

Our music intelligent recommendation system enables users to get new singers and music ideas quickly considering many aspects of the music's label, lyric emotion, number of users comments and evaluation. Also, our system will reduce the time users consider which music to start to a large degree. More convenience and flexibility are our original intention.

We hope our project can be added to those existing music platforms and companies as a complement to the recommendation system. In today's market, our system can meet the following requirements:

1. Attract new users

Former recommendation systems will collect music that users have listened to and according to music to give new advice. But when an absolutely new user just starts his NetEase Cloud Music, he has no data in his inventory, so obtaining the users' music preference resources is not enough and the prediction is low. Then our system will send the user to the

NetEase Cloud Music store to let users choose some music they are interested in. Our team will consider letting our system receive users' feedback and recalibrate the recommendations system has given.

2. Expand the scope of users

Our free and flexible recommendation system can mobilize the creativity of users to make their own preference music list, and satisfy users with different music preferences. It can also attract other platforms' users to NetEase Cloud Music, so the system can even create business value.

3. Excellent experience

The most advantage of our system is that it can shorten the waiting time and give the most suitable and perfect related music to the users. A richer music experience can be created with our system in the NetEase Cloud Music platform.

3 KNOWLEDGE MODELING

3.1 GENERAL DESCRIPTION

As we came up with this idea of building an intelligent recommendation system, we planned to apply Item-KNN models into our system to provide recommendation work.

Briefly, our system-building process consists of 3 different parts. And each part applied one model we have learned in our class and courses.

We will explain how we applied 3 models to the 3 corresponding parts of our project.

In this project, we used the thesaurus of the Dalian University of Technology to analyze the song data and obtained the emotional data of the singer, which was used as the input of the project model. Keywords are extracted from different lyrics, and the corresponding emotion levels of different words are extracted as features, which can be divided into NA, NN, NP and so on. And then you get the vector. So, you can get the distance to the target. The output of the model is the singer closest to the emotion.

3.2 PREPROCESS

For the preprocess section, MinMaxScaler and the PCA method are mainly used to process data.

1. MinMaxScaler method

The method scales all features to the value between 0 and 1. The method is aimed to prevent the effect of feature values.

In the MinMaxScaler section, all feature values will be scaled by the following function:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (3-1)$$

The X represents the previous data, and X_{min}, X_{max} represent the minimum and maximum data of all.

2. PCA method

In the PCA section, the PCA method will be applied to extract the data's features and reduce the data's dimension. The data will finally be represented by several principal components. To determine the number of principal components, Scree Plot, which shows the relation between eigenvalue and number of principal components, is applied in the program. The result is shown in the diagram below.

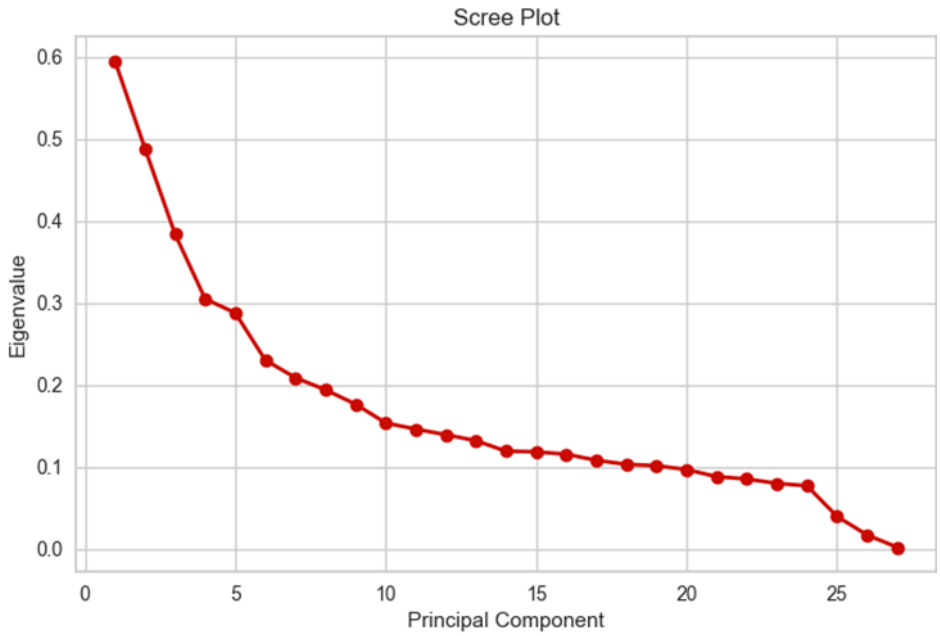


Figure 3-1 Scree Plot

The final result determines the number of principal components by choosing the elbow point.

3.3 ITEM-KNN

In the machine learning section, we will perform the Item-KNN method to give a list of recommended music that is similar to the user's playlist.

1. Input

To measure the preference of the user, the model first calculates the sum of each feature of all songs that are listed in the input user library. Then a MinMaxScaler similar to the previous step is performed to scale the vector, which forms our user-vector

2. Item-KNN

Item-KNN method is performed between the user-vector and database to find 10 singers that is closest to the user preference. To measure the distance between the user-vector and each point in the data-base, the model use the standard Euclidean distance.

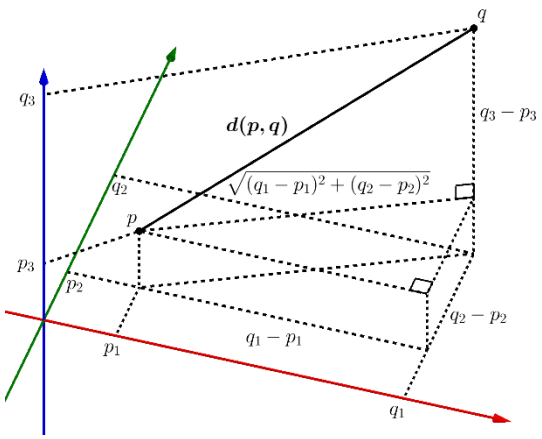


Figure 3-2 Euclidean distance

The model first calculates the distance between the user input and all other points in the data-base. The formula is as follows:

$$distance(u,v) = \sqrt{\sum_{i=1}^{21}(x_i - y_i)^2} \qquad (3-2)$$

$$u = (x_1, x_2 \dots x_{21}), \qquad v = (y_1, y_2 \dots y_{21})$$

In the formula, u represents the input vector, and v represents a vector in the database. $x_1, x_2 \dots x_{21}$ represent 21 feature value of input vector x, and $y_1, y_2 \dots y_{21}$ represents feature values of one vector in the database.

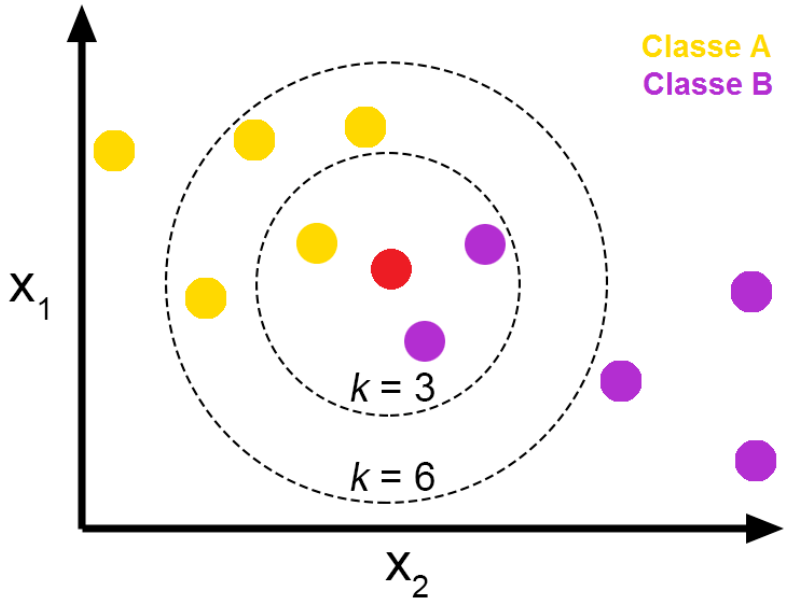


Figure 3-3 Item-KNN method

Then the model will sort all points in order of distance from low to high, and pink up 10 points with the smallest distance to be the final recommended results.

4 SOLUTION OUTLINE

4.1 SYSTEM ARCHITECTURE

The system architecture diagram illustrates how the application in the front-end interfaces with the python service, data processing, database and recommendation engine back-end work together. Below I will describe how our subsystems cooperate with each other from a typical user scenario.

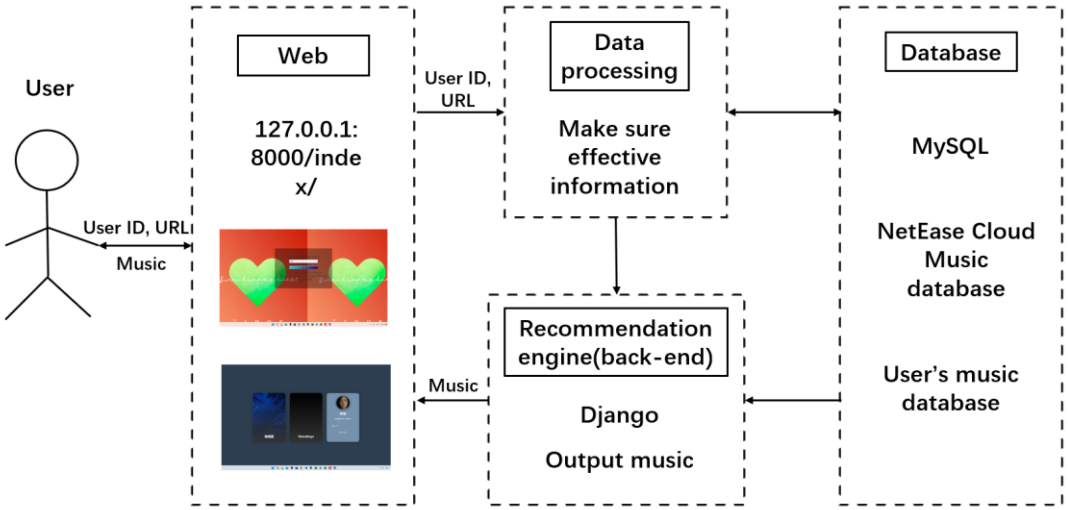


Figure 4-1 System architecture

The typical user usage scenario is as follows. The user is tired of listening to a few music all the time and wants to change to another similar music that may be suitable to him. The user opens our website and logs in to his NetEase Cloud Music website and pastes his page link into our recommendation page. Besides, he must give himself a name (user ID). His user ID and his link URL must be never registered compared with the database that stores the information of all registered users. That is how data processing works, making sure effective login information.

After this part, both the data processing part and database part send information to the recommendation engine (back-end), all music data and the user's music data. And the recommendation system will find the most similar and suitable music by the KNN model.

Considering the diversity of recommended content, we ask that the first

three of them come from the recommendation engine, and the last seven are randomly selected from the same style of music. Then, 3 pieces of music will be sent to the user, this 3 music are the most suitable music for users to play through the analysis of the recommendation system. Music that exists in the user's music library will be selected and will not be sent to the user again.

4.2 PROJECT SCOPE

1. The music library

The recommendation intelligent system depends highly on the richness of the library. At present, our data comes from crawling the music data on the NetEase Cloud Music store including 140,000 songs. They covered most of the songs existing today, except for songs that are listened to by very few people and with very few comments.

2. The recommendation engine

Based on the data we obtained, we selected many features (emotion categories such as NN, NE, NA …) to construct our similarity matrix. And they decrease the dimension of these features by PCA, and It is easy and convenient to expand or reduce features (There may be new types and features of music in the future).

4.3 SYSTEM FEATURES

From the user's perspective, the most useful and convenient function is researching songs that may be suitable to users based on the playlist of users.

1. Recommend music based on the user's playlist

All user’s need is to log in to their NetEase Cloud Music library interface, input their interface URL to our recommendation platform, and once again a name of their own, they can instantly get the singers and music recommended by the system.

2. Analyze the singer’s style using lyrics in their songs

In the final word ontology, emotions are divided into 7 categories and

21 categories. The purpose of constructing this resource is to provide a convenient and reliable auxiliary means for sentiment analysis and propensity analysis of Chinese texts in the field of sentiment computing. Chinese sentiment lexical ontology can be used to solve the problem of multi-class sentiment classification, and it can also be used to solve the problem of general propensity analysis. This analysis of lyric emotion is a major feature of this system.

4.4 CORE IMPLEMENT

Here is the detailed system map, Now I will go through some

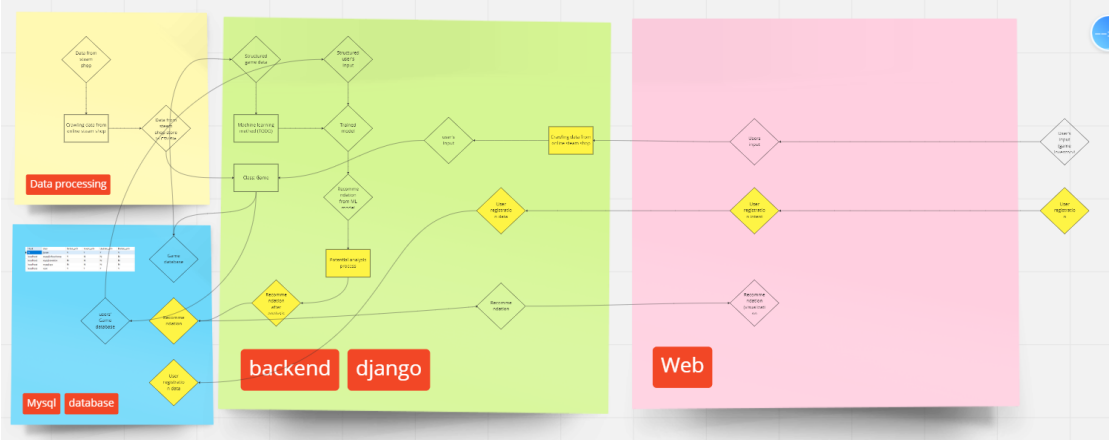


Figure 4-2 System map

1. Data Retrieving – Crawling method

Crawling 140,000 songs in NetEase Cloud Music. Information: song name, singer name, lyric. Save in JSON file. Raw data waiting to be analyzed. raw data waiting to be analyzed. Crawling the playlist using the link provided by the user. Will mention this later in the use case.

2. Data Processing – Emotion dictionary & Jieba

Jieba: Python package.

- Chinese automatic tokenizer, segment words in sentences and extract keywords

- Emotion dictionary:

- Provided by the Dalian University of Technology.

- 21 emotion tags are distinguished.

- For high-frequency words in lyrics, give them emotional labels and intensity

3. Data Processing – lyric analysis & singer emotional label

For each song, use Jieba to extract key information in the lyrics, and add emotional tags to each song against the emotional thesaurus. Add the emotional tags of all songs of a singer and average them to get the singer's emotional tag, which is also the input of the machine learning method.

For the playlist provided by the user, the emotional tag of each song is also calculated, and the average value becomes the emotional tag of the playlist

4. Recommendation Results – KNN

Perform the KNN method to take all singer emotional tags and the user's playlist emotional tags as input. All sentiment labels are rendered in a Euclidean space. find 10 singers' emotional tag that is closest to the playlist emotional tag

4.5 SYSTEM DEVELOPMENT & IMPLEMENTATION USING TOOLKITS

1. MySQL

We made the database using MySQL, storing all NetEase Cloud Music data and all user information data and the information of the user who is querying the music. Some data was stored in the database and the other data is sent to the recommendation engine and other parts.

2. Django

Django was applied when we built the front-end and the web. In turn have connections with users and databases and recommendation engines. Django's architecture is more concise and easier to work with other components in the project. Django allows us to configure routes at will, and can directly render the front end with parameters carried by the back end. After editing the class, you can use the shortcuts in Django to automatically complete the database construction. In addition to being proficient in using the Django framework.

3. The recommendation engine

Based on the data we obtained, we selected many features (Single singer or not, release time, RPG…) to construct our similarity matrix. And the decreased dimension of these features by PCA, and It is easy and convenient to expand or reduce features (There may be new types and features of music in the future).

4. Data crawling package

In this project, our team used packages such as requests, beautifulsoup and XPath to solve various problems when crawling data. Technically, when using python to crawl web pages, it is generally to crawl the HTML file of the web page, and then interpret it and find the required resources. In this process, various problems may be encountered, for example:

1. Required elements are hidden by the HTML file.
2. The website detects that the browser accesses the interface multiple times and rejects the browser's access
3. Website requires authentication to access

For the first problem. We can locate hidden elements' positions by using the XPath package. Entering the path to the target element class, the package can reveal the element.

For the second problem. Using the Selenium package to simulate normal users using browsers. Its access parameters are exactly the same as those of normal users using browsers, and the access behavior is relatively more like normal users, and it is not easy to be hit by anti-crawler policies.

For the third problem. Set cookies to simulate an authenticated user, access the website using the cookies set and the program can pass the authentication.

4.6 BUSINESS CASE/ MARKET RESEARCH

Users are always the guarantee of all commercial value and the core source of income, so we should first do a good job in our recommendation system to attract more users to use our system. Faster update solutions and more scientific and diverse recommendations are the foundation of our

foothold in the market. It's a virtuous circle, and more and more platforms will cooperate with us.

In fact, the commercial value that recommendation systems can profit from mainly comes from three aspects: advertising, e-commerce and value-added services (such as membership, etc.). Recommendation techniques can be used to do better, accelerate the process of realization, and generate more revenue.

1. Advertisements

So-called advertisements, that is, on the products on the Internet advertising, by putting the AD exposure to the user or the user to click on ads to earn profits, makes the third party (i.e., advertisers) money (if not self-built advertising platform, will flow outsourced advertising alliance, yield is divided into advertising alliance), known as the wool in pigs. Of course, there are many kinds of advertising, including display advertising, brand advertising, effect advertising, information flow advertising, and so on. Our project, a music recommendation system, was originally designed as a music recommendation for NetEase Cloud Music users because we think there should be a recommendation function for such a music platform. It can be cooperated with the NetEase Cloud Music platform or even other music platforms. It can generate a link or a separate area interface on the music platform, and the blank areas of the interface can be used as advertising spots. The probability of users clicking will be greater, which can bring a better conversion effect than the market. By quantifying these indicators, the commercial value generated by the recommendation system in advertising exposure can be reflected.

2. E-commerce

Here, the commodities in e-commerce refer to generalized goods, including virtual goods (such as online novels, online courses, etc.) in addition to physical goods. The e-commerce of the platform is realized, and our project platform makes money through the merchants who enter. Through the efficient distribution of commodities, the recommendation system recommends commodities to users who are willing to buy them,

promotes the sale of commodities, and thus obtains more shares from merchants.

3. Value-added services

Value-added service mainly refers to "characteristic service", our project can guarantee the basic service at the same time, beyond the conventional, personalized service. For example, the introduction of membership, premium packages and other value-added services.

The recommendation system can accurately recommend the member programs that he or she is interested into non-member users, so as to enhance the exposure of member programs in non-member users and promote non-member users to buy members. The personalized recommendation that may be set in the future can better explore the deep interest needs of users, accurately reach the interest points of users, and thus bring more accurate member-specific recommendations. In addition, personalized recommendations can also allow more obscure but high-quality recommendations to get exposure. The better the recommended experience, the easier it is for users to find what they want to see on the platform, the higher the natural engagement, and the easier it is to keep paying.

4.7 FUTURE IMPROVEMENTS

1. The update of the music library

In the future, more and more music will emerge in the market and the data and information about the new music are needed to be upgraded in our raw database in case we can recommend the users the most popular and brand-new music in real time. A timed task is needed for us to update the music library regularly.

At present, the way we treated new users is that we jump the page to the NetEase Cloud Music store interface. In the future, we can try to recommend the use of some music that is popular at that time or recommend most to other users, or we can use a random algorithm to randomly recommend different kinds of music.

2. Feedback system

We may make a feedback system into our recommendation system. Collecting the feedback from the user, the system will upgrade and recognize the music which is recommended to users more accurately and meet the users' taste.

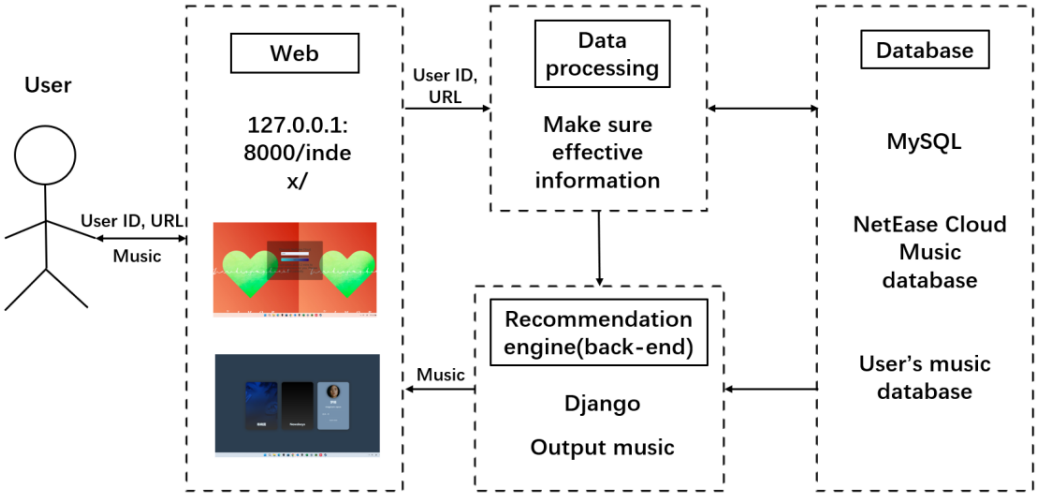
3. Improve algorithm efficiency

Currently, we support two matching methods: top-n matching and informed search matching. The default is informed search. But if the number of recommendations increases, the time complexity of the algorithm will increase exponentially. In the future, we can specify rules according to the number of recommendations we need, and combine these two strategies to formulate new rules to recommend.

Appendix 1: Project Proposal

PROJECT PROPOSAL

Date of proposal: 02/10/2022
Project Title:
Group ID (As Enrolled in LumiNUS Class Groups): Practice Project Group 4 **Group Members (name, Student ID):** **Xu Jiayu A0261690A** **Tang Enze A0261984R** **Cai Zheng A0261940H**
Sponsor/Client: *(Company Name, Address and Contact Name, Email, if any)*
Background/Aims/Objectives: Music has been a cultural product since AD 1. This form is a way to convey and express feelings. As time goes on, more and more outstanding musicians have created one classic and popular song after another. After being sung and rewritten from generation to generation, the song itself can move people's hearts or express higher spiritual significance. Music can inspire people, give them hope, make them believe in something better, or keep them away for a while. When you need it, it will be your good friend, when you need it, you will find it. The proverb "Put on the music, pause the memory, stop the pain, put the joy back" is the easiest thing to do. Music is a means of connecting all cultures. Different cultures have music, and we can easily see how different cultures are combined.
Project Descriptions:



We aim to deliver a system which recommends music to users, the system will follow the process:

1. Using python retrieve music data from NetEase Cloud Music.
2. Preprocessing the data and storing the data locally.
3. Training the model using the formatted data.
4. Processing user input to extract the feature
5. Obtain knowledge based on the feature.
6. Process the user's input with the module. Output the expected recommendation

We came up with 2 processes for the project. The main difference of these 2 processes is, adapt machine learning when processing data, or adapt machine learning in decision-making, and the team will determine the method that works for them in the process of exploration

process 1:
Original data -> Model

User input -> Model + search engine -> Classified output

process 2:
Original data -> Label the data using Model (KNN) -> labeled data

User input indicates the requirement -> Search engine -> Data which satisfied the requirement

To achieve our desired goal. We need to develop these terms:

1. The outlook of the Web (If we decide to use the web). Using HTML CSS JavaScript.
2. The backend of the system, Android studio (If we decide to use an Android APP). Java or python (Depend on whether we embedded machine learning into our decision process or not)

3. Original data from NetEase Cloud Music. We plan to crawl data from NetEase Cloud Music Shop.
4. Implementation of Machine learning, we haven't decided on the specific method to be used, but we are inclined to use unsupervised methods.

Optional Action:

1. Deployment of the system. We will consider this after the product is done.
2. Database to store user's info. We will consider this after the main function is done.
3. User-base recommendation. We will consider this after the main function is done.

Appendix 2: Mapped System functionalities against knowledge, techniques and skills of modular courses: MR, RS, CGS

1. Data Preprocess

For the preprocessing part, the model first applied MinMaxScaler to scale the data in the range of [0,1], then used the PCA method to extract the features and reduce the dimension of the data. To determine the number of principal components in the PCA method, a scree plot is applied to discover the relationship between eigenvalues and the number of principal components.

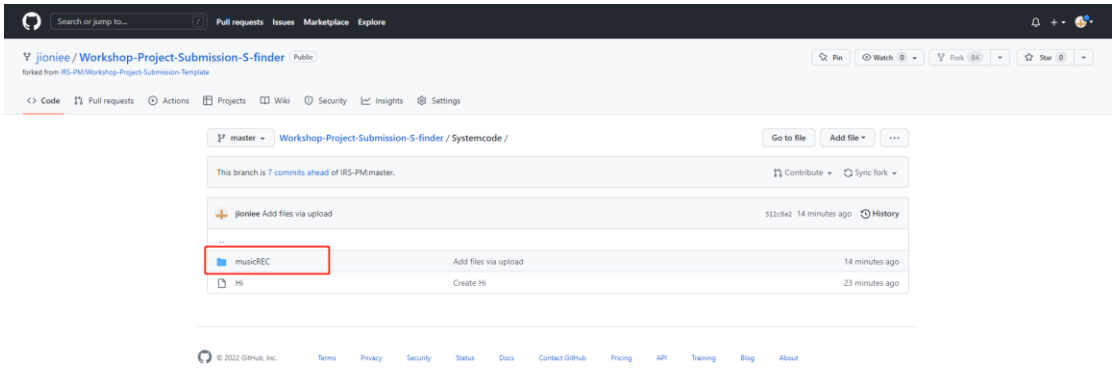
2. Machine Learning Method

In the machine learning part, the program constructed a model with Item-KNN, which measures the standard Euclidean distance between the input user playlist and all other singers in the database. The algorithm finally predicts and recommends a top-10 music list that is most similar to the input user playlist.

Appendix 3: Installation and User Guide

Download the code from the GitHub repository

https://github.com/jioniee/Workshop-Project-Submission-S-finder/tree/master/Systemcode



git clone https://github.com/jioniee/Workshop-Project-Submission-S-finder.git

Download python 3.10

https://www.python.org/downloads/

Download pycharm

https://www.jetbrains.com/pycharm/

Download Mysql

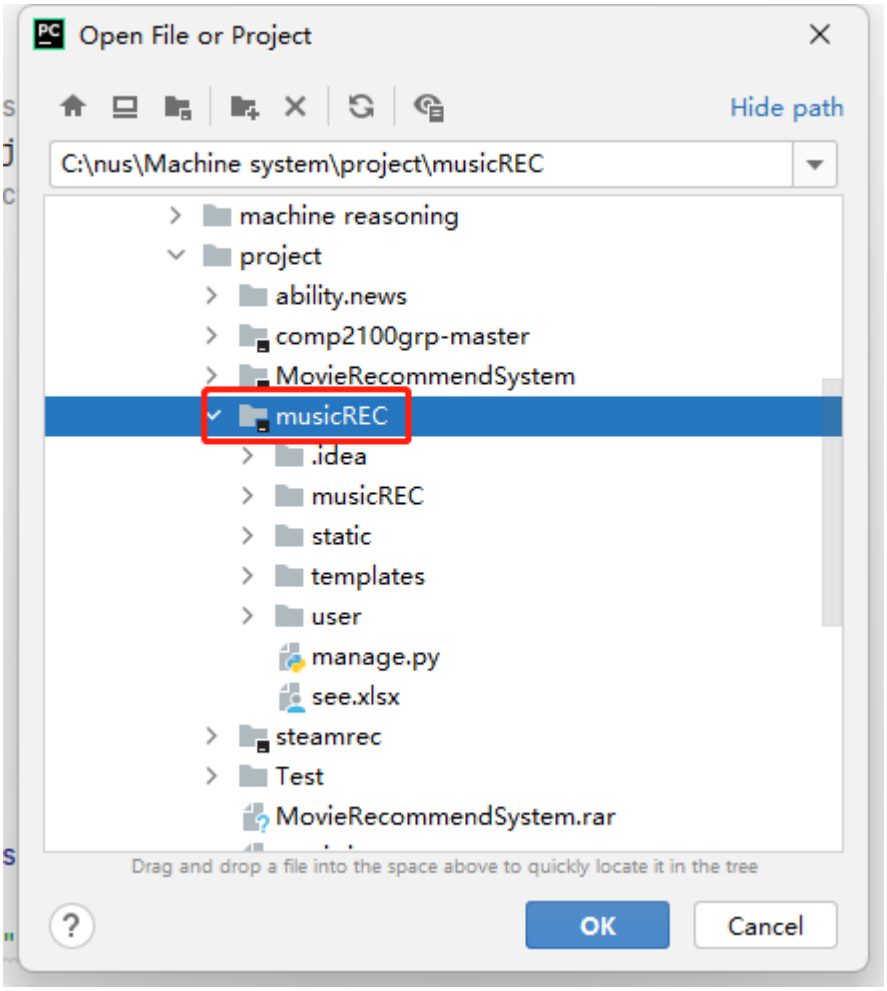
https://www.jetbrains.com/pycharm/

Download navicat (Recommended)

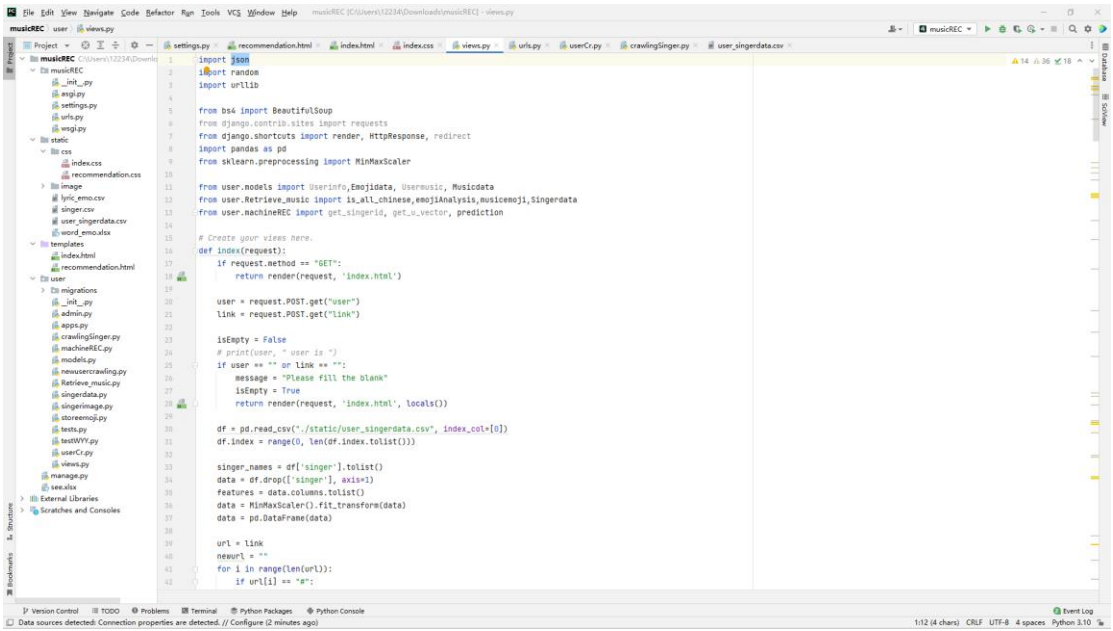
https://navicat.com/en/

Navicat is not compulsory, but Navicat can improve the process of managing the database. Navicat provides a 14-days free trial

Open project using pycharm



The page should look like this:

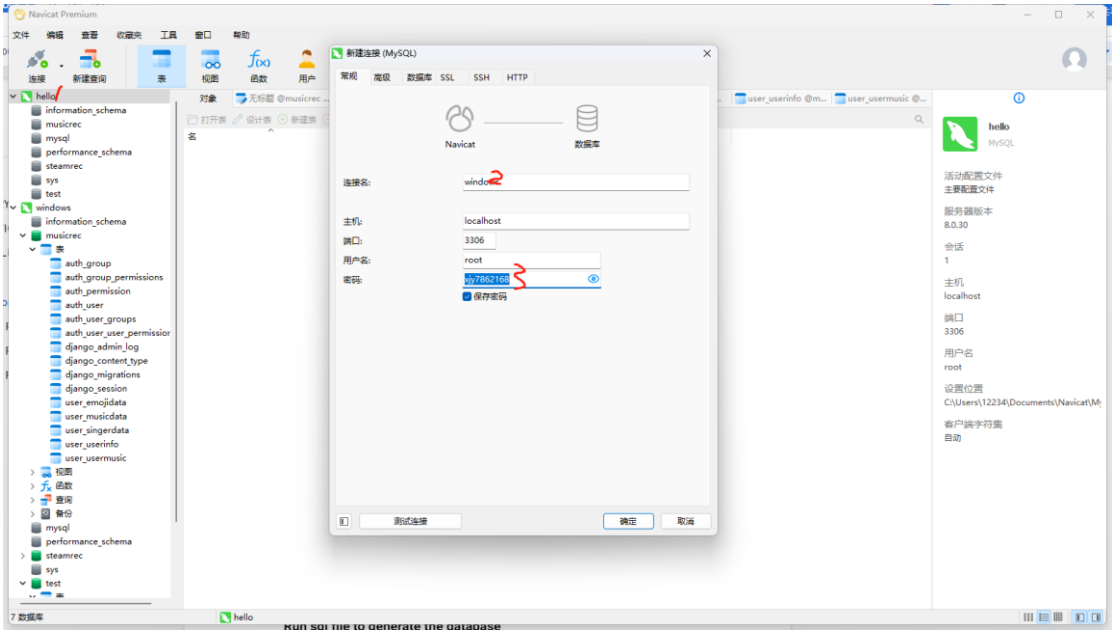


Starting MySQL

Configure according to your system environment：

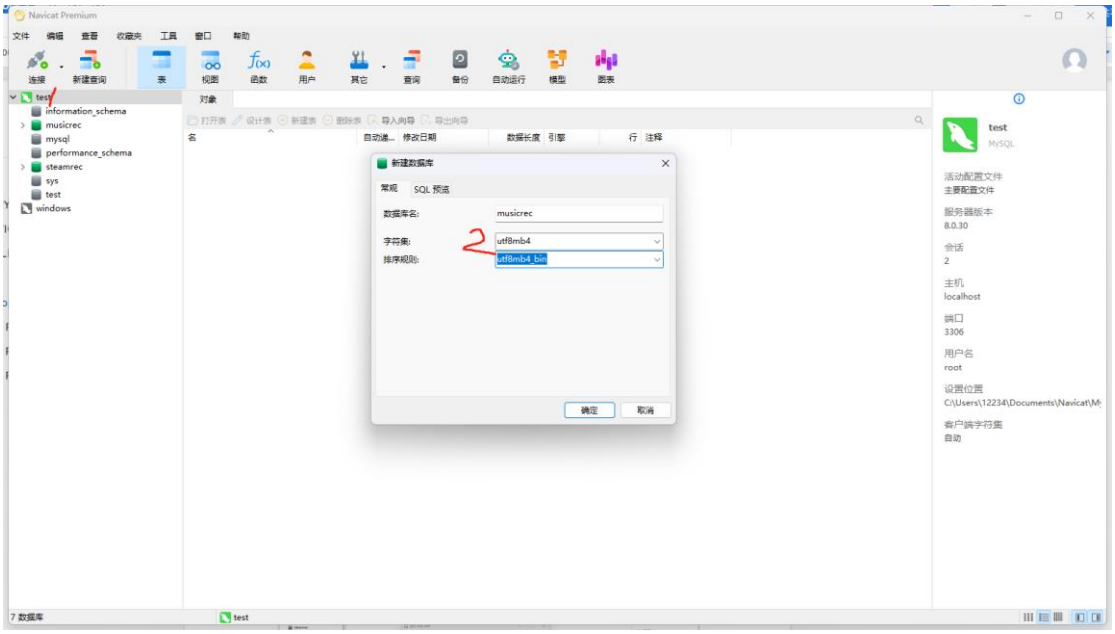
https://dev.mysql.com/doc/mysql-getting-started/en/

Create connection and database using Navicat



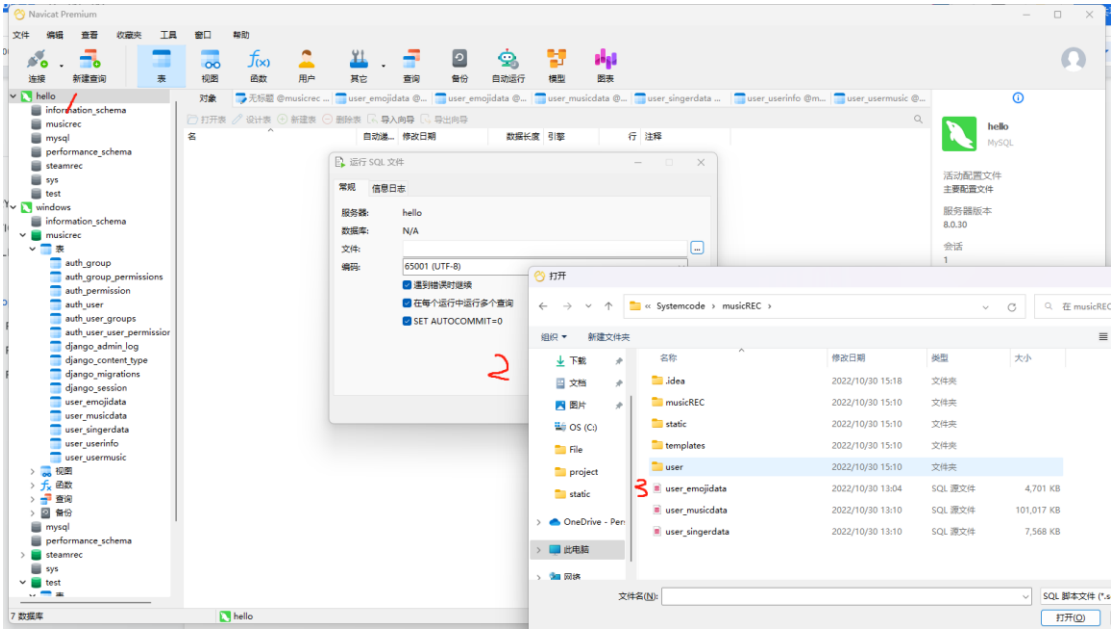
1. hit connection
2. choose MySQL
3. fill in the connection and password as shown in the image.

Create a database in Navicat



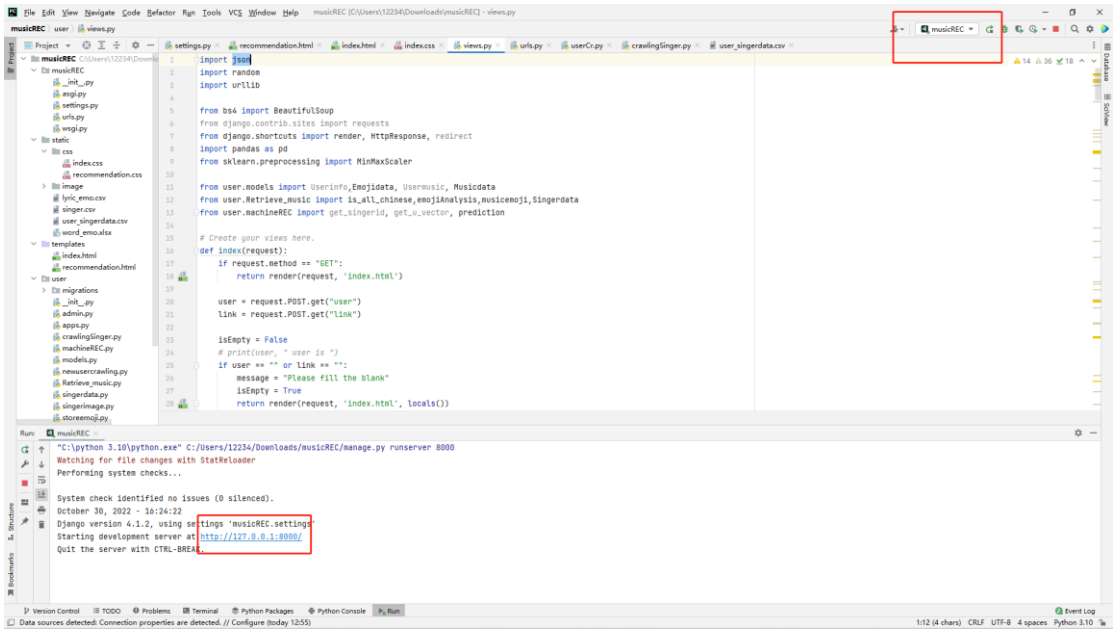
1. **right click on the connection we just created.**
2. **Input the database name and config as shown in the picture**

Run SQL file to generate the data



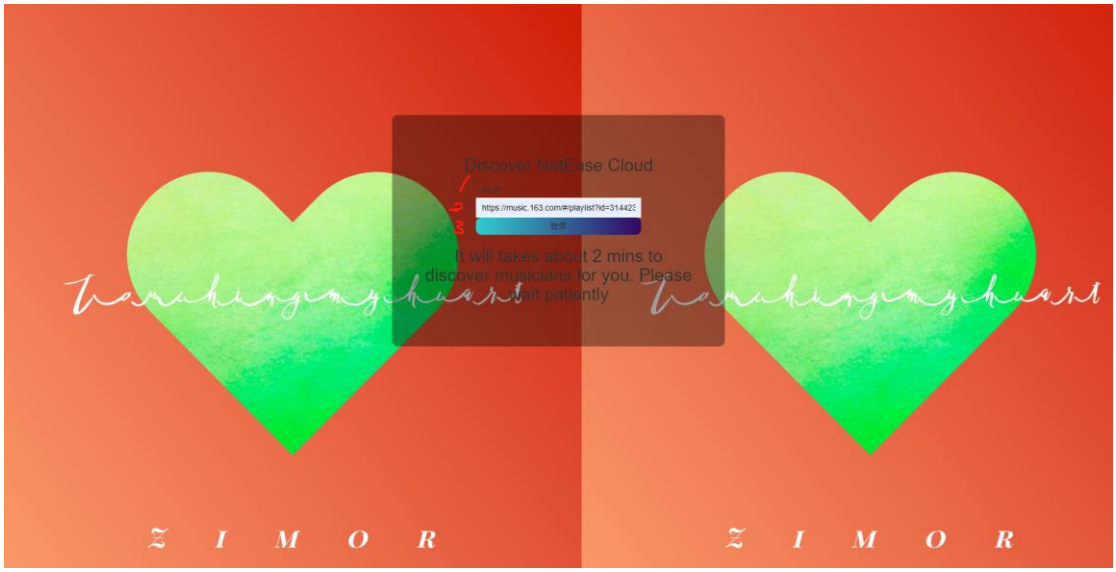
1. right click the database just created, choose run SQL file
2. click “upload”
3. select 3 SQL files in the root directory separately.

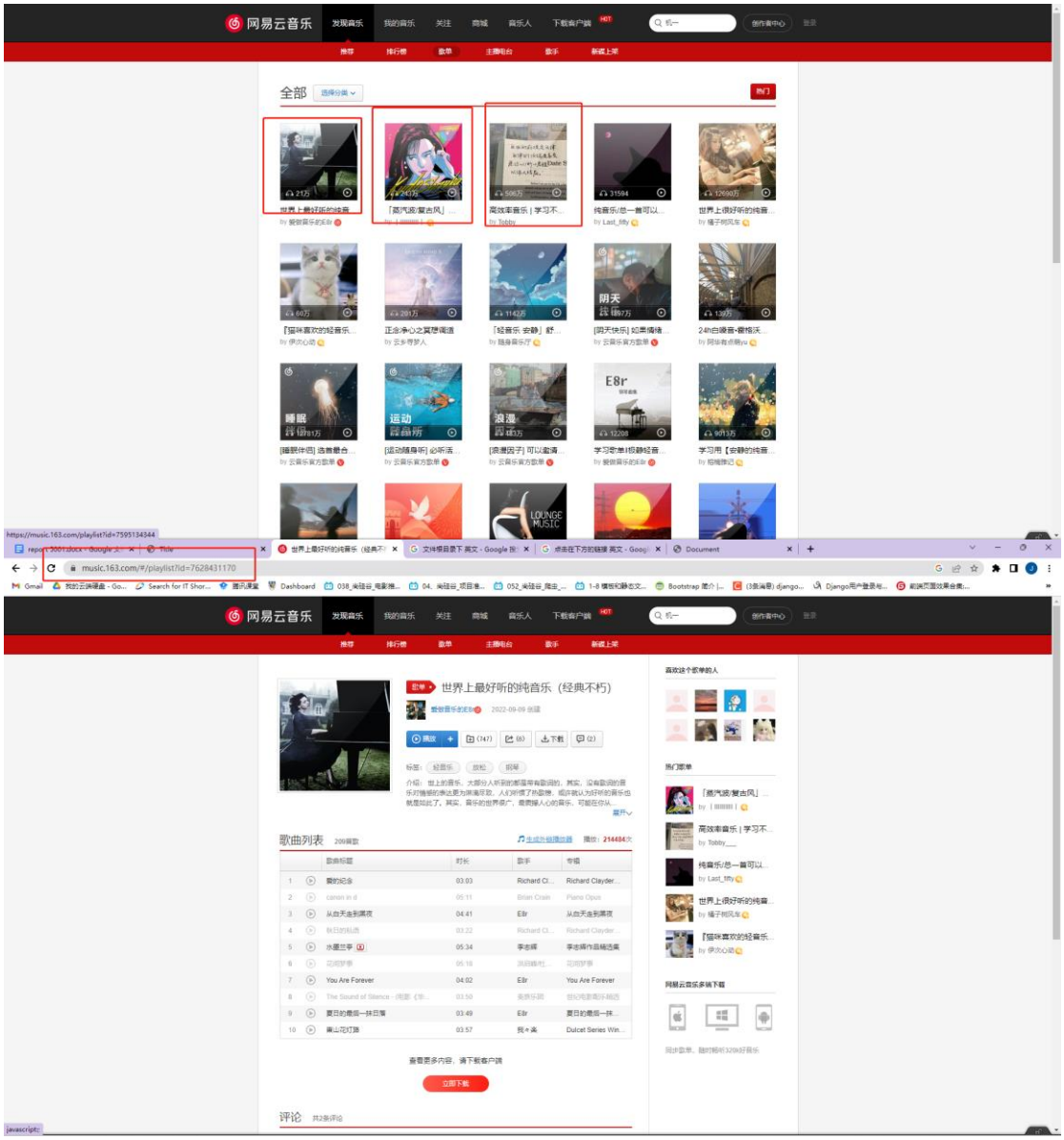
Run the program in pycharm



Run the program. After the system is ready. Click on the link below

Using the program:

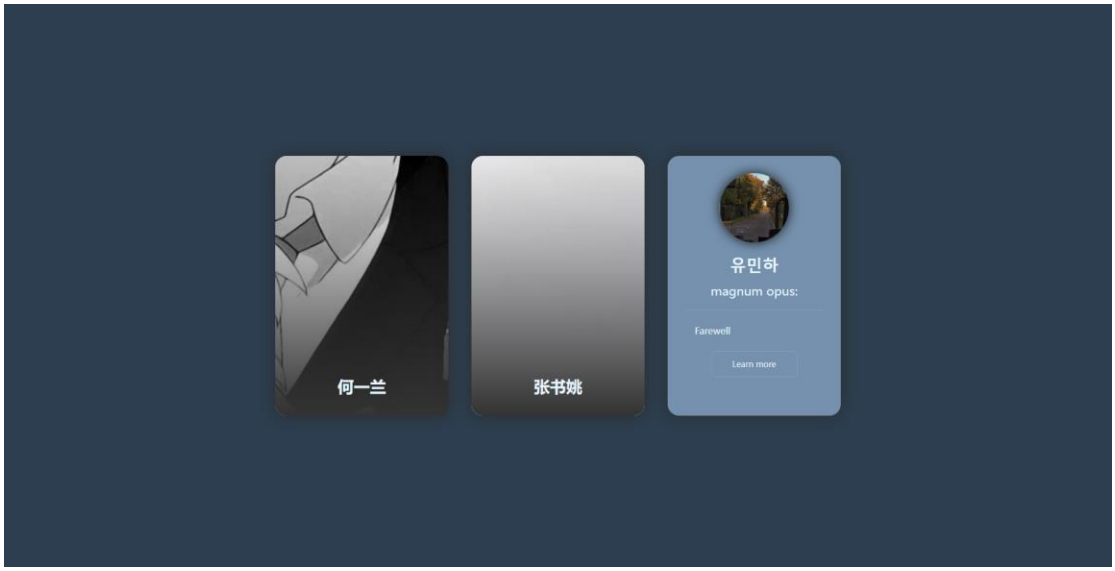




Using the program:

1. Fill in your name.
2. open NetEase Cloud Music with your browser, link: https://music.163.com/
3. Click on any playlist
4. Entering the page, copy the link and paste it to the website.
5. Click the button

Recommendation:



The system will recommend 3 musicians to you based on the playlist you provide.

Appendix 4: Individual Report – Xu Jiayu

Xu Jiayu – e0983084

My contribution to the team

In the team, I play the role of the leader, planning meetings and managing the progress of the team. Skills are based on the experience of each member of the team. I arranged for them to do the right job. This increases the productivity of the team. Even though it was the first project work for our team, we still completed the task with good quality and achieved the expected goals. In development, I am responsible for the front end and back end of the system. For the front end of the project, I borrowed some templates and made improvements to make the display of the page match the theme of the project. The backend of the project is a huge process, which includes custom types, definitions and connection to the database to the backend. Provide data for the project (by writing and running crawler functions), assign routes and let the backend render the frontend pages.

What I learned from this project

This is the first time I'm responsible for the development of a web page alone, and I've tried many new techniques, such as using Django in python as a development framework. Before this, I only worked on projects using java as a programming language, and I found that compared to spring boot, Django's architecture is more concise and easier to work with other components in the project. Django allows us to configure routes at will, and can directly render the front end with parameters carried by the back end. After editing the class, you can use the shortcuts in Django to automatically complete the database construction. In addition to being proficient in using the Django framework. I also had a lot of fun writing python scraping functions. I have no experience writing crawler functions before. In this project, I used packages such as requests, beautifulsoup and XPath to solve various problems when crawling data. Technically, when using python to crawl web pages, it is generally to crawl the HTML file of

the web page, and then interpret it and find the required resources. In this process, various problems may be encountered, for example:

1. Required elements are hidden by the HTML file.

2. The website detects that the browser accesses the interface multiple times and rejects the browser's access

3. Website requires authentication to access

For the first problem. We can locate hidden elements' positions by using the XPath package. Entering the path to the target element class, the package can reveal the element.

For the second problem. Using the Selenium package to simulate normal users using browsers. Its access parameters are exactly the same as those of normal users using browsers, and the access behavior is relatively more like normal users, and it is not easy to be hit by anti-crawler policies.

For the third problem. Set cookies to simulate an authenticated user, access the website using the cookies set and the program can pass the authentication.

What needs to be proved

We have more features planned, but many are too late to complete. This was due to my miscalculation of the difficulty of the project, but considering this was the first project experience for most of our team members, it was not a bad result. The team didn't devote enough time to testing and iterating on the project, and in practice, it was more important to ensure that the basic functionality was implemented than adding additional functionality. We should allow time to test the project to ensure the robustness of the implemented functionality

How can I apply the knowledge and skills

I can apply my project management skills to future project development. As a project manager, I am able to analyze everyone's areas of expertise and assign tasks to them flexibly. At the same time, monitor the progress of the project. If there are unexpected situations in the project, use your own communication and management skills to properly solve these

problems. As a software engineer, I can apply the project development knowledge learned in this project to my work. I can skillfully use the backend of the Django development system to quickly build basic components for the project, such as routing configuration, data configuration, and define classes. As a data analyst, I can apply my experience writing crawlers to this project. Use different python packages to solve various problems during data crawling. Broaden the channels of data that can be analyzed.

Appendix 5: Individual Report – Cai Zheng

Cai Zheng – e0983334

My contribution to the team

I am mainly responsible for the data preprocessing and machine learning part of our program.

1. Data preprocessing

For the data processing part, I applied some data processing skills that learned from the module, such as the MinMaxScaler method, which scales all data into the range between 0 and 1, and the PCA method, which helps to extract the features and reduce the dimension of the data.

2. Machine Learning Algorithm

For the machine learning part, I wrote the function to transform the input user playlist into the vector that machine learning needs. And I built the Item-KNN model from the given music database and used it to predict a recommendation list that is most similar to the input user playlist.

3. Typesetting work

I used my word knowledge to help with some typesetting work.

What I learned from this project

During the project, I learned much about project development skills in a team as well as techniques to build machine learning models.

1. Teamwork

First of all, the project taught me how to work in a team. I am mainly responsible for designing the machine learning part, and I have to communicate with the people who integrate my model into the whole system very much. I need to tell them the input, output and main usage of my function, and I learned much about communication skills and teamwork during the process.

2. Implementing machine learning algorithms

Since I'm responsible for designing the machine learning part, I applied a lot of data processing skills that learned from the module such as data

normalization and PCA method to process the data and got a better result, and I also built the Item-KNN model and successfully predicted reasonable result with it. During the process, I got a better understanding of the machine-learning skills that we learned in class.

3. Mastering machine learning kits in python

During the project, I got familiar with many machine learning kits in python such as scikit-learn, pandas, NumPy and so on. They offered me many conveniences when I try to build the machine learning model. I searched and learned many useful functions and applied them in my project. I believe they can still help me in future projects.

What needs to be improved

Due to the limitation of time and resources, there are many things that are yet to be improved in our system.

1. No enough features from the song database.

It's a pity that some quantified features, such as the length of the song, rate, and release time are too hard to get and are thus not included in our database.

2. Lack of user database

There was no enough time for us to get a database of users, so we failed to use a collaborative filtering method based on user similarity though the method had been written, and there's no quantitative evaluation method to evaluate the performance of our system since it requires massive user data in training and testing set.

3. Long prediction time

In the previous plan, we decide to apply a clustering method to cluster similar songs together first and narrow down the range of KNN. However, the performance of clustering was so poor that we had to cut it down. If there's more time for us to find a better way of clustering, the prediction time could have been greatly reduced.

These cities will continue to inspire me to do better in future projects.

How can I apply the knowledge and skills

Machine learning is a powerful tool in the history of humans. The machine learning method learned in the module can be applied to solve many real-world problems. For instance, the project reveals a method to recommend items that fit people's preferences, and it could help people discover things they like and create more profits for companies. The precious team-working experience will help me work better in teams, communicating with the people responsible for other parts such as the front end and back end better. And the python skills I developed will also continue to help me in future project development work.

Appendix 6: Individual Report – Tang Enze

Tang Enze A0261984R

Personal contribution

The experience of completing this project with Xu Jiayu and Cai Zheng has been a lot for me. We spent a lot of time with these project members from setting the theme to the end and putting our ideas into practice bit by bit. This made me feel a sense of accomplishment. My project members are also very good at communicating and helping each other. Discuss and solve problems with each other, it is an honor to work with such a team.

In actual development, I was mainly responsible for every presentation and the report and videos, providing my ideas and opinions that can attribute to the final project, such as original project ideas, every feature, algorithm model and frameworks we used. Besides, I learned how to crawl data from the web by myself and applied the skill to crawling music data as our project's raw data. Finally, I was also mainly involved in the style of the user interface and the testing and adjustment of the recommendation system in various environments.

What is learned?

During this project, I really learned a lot. To be honest, because I changed my major from energy and power engineering to intelligent systems, it is really a tough process. This project is the first complete project about computer knowledge in my study life. I will elaborate on the following points.

1. Project

This is the first project that I've been involved in that was quite complete. I've made the connections between the pieces and know how every part works together to implement the final function. There are Users, web, data processing, database and recommend the engine (backend).

I think I learned a lot from this project, not only the detailed models and algorithms but also the whole process of front-end and back-end work

together. I know the whole process of a project: direction, framework, thinking and implementation. I know how to crawl data and know that we need to give features and decrease feature dimension becoming a low-dimension vector finally, know how we can apply the model we have learned in class to actual problems. I also learned many frameworks, such as Django, to make front-end web and how to connect front-end to back-end, and how to make our page display look better. More importantly, I learned how a developer's approach and ideas to solve the problem and so on. I think these things are valuable experience accumulation for me at the beginning of my computer career.

2. Teamwork

Because it is project cooperation, So I discussed with Xu Jiayu and Cai Zheng online in zoom lots of times to develop plans and processes, refine models, assign work, etc. Luckily, my teammates are both responsible for the project and are good at recommendation algorithms to determine the front-end and back-end API, which made the development efficiency higher. So, I always ask them for relevant knowledge, understand the role of each part in the whole project, and find relevant learning materials and videos to learn and improve relevant parts. When I have a problem in the development process, the project members will share their experience and knowledge to help me, so the efficiency in the development process is very high, and this experience also helps me learn how to communicate with the team. Effective teamwork gets twice the result with half the effort.

How can you apply the knowledge and the skill gained?

First of all, I gained the skills and experience to crawl data from a website. I think I can apply this skill in many projects because a lot of current work requires crawling data to build data sets.

Secondly, I have also understood the process of back-end development and how the front and back end interact, which is an essential skill and experience for my future work. Combined with the call to the database in the future, I believe I can apply this knowledge in many development projects. The application of different models in different project contexts

will cover the vast scope of machine learning and deep learning.

Third, I learned how to cooperate and communicate with the team. Regular meeting summary is very necessary, the document can help us to arrange work processes, clearly see everyone's division of labor, and also can be summarized to sort out what we have done all the work of some links, data set position, etc. Details can be recorded, convenient to everyone in view when necessary. In teamwork, I developed good code habits. To facilitate others to read, the code should have certain norms. Communication and exchange between people often play a decisive role, for the efficiency of work and the maintenance of human relations are extremely important.