

State-Of-The-Art Music Recommendation

Katharina Stengg

University of Klagenfurt

Seminar Paper: Seminar in Artificial Intelligence

`katharina.stengg@aau.at`

Abstract. Nowadays tons of music is available to almost everyone almost all the time. Additionally new songs are released and immediately available on a daily basis. This highlights the need of a good song recommendation technique to prevent users from information overload. There exists a variety of methods for this task, some only from a computer scientists' view but also a lot of interdisciplinary work, especially when it comes to include sociological or psychological factors. This seminar paper aims to present some of the most recent datasets and methods for music recommendation using either traditional or more specific approaches. Another core aspect of this seminar paper is the critical discussion and evaluation with the goal to answer the main hypothesis on whether we already achieve sufficient results using state-of-the-art methods.

Key words: recommender systems, music recommendation, artificial intelligence, deep learning, lstm, content-based, collaborative, knowledge-based, hybrid, datasets, methods, evaluation

1 Introduction

According to Jannach *et al.* [10] the overall goal of recommender systems is to build systems that help users in making decisions.

Recommendations should be applicable to the broad mass of users but also suggest on such a personal level that it outperforms field experts.

Such recommendation systems aim to relieve users from the information overload they need to handle when making a decision, without the actual need to consult an expert or even pay for the service.

In general literature differs between the following four main traditional categories of recommendation systems:

- Collaborative Filtering
- Content-based Filtering
- Knowledge-based Recommendation
- Hybrid Recommendation

Collaborative Filtering Collaborative Filtering (CF) is based on other user's opinions and ratings.

This recommendation method is similar to people's behaviour in real world. If we like the camera we bought, the friends we tell about might also consider to buy it. If we don't like it, they probably also won't buy it. Collaborative filtering is mainly used to recommend an item, predict the rating for an item or provide constrained suggestions. [11]

Content-based Filtering In comparison to collaborative filtering this approach doesn't rely on user's ratings but on items with similar ratings. A user profile containing the user's interest is used. [12]

Knowledge-based Recommendation Recommendations are made based on available recommendations and user's demands. This technique does not suffer from the so-called cold-start problem.[12]

Hybrid Recommendation Each of the categories mentioned above comes with a variety of advantages and disadvantages.

Hybrid recommendation systems are combinations of other recommendation approaches to eliminate certain limitations and improve the overall performance. There is a variety of possibilities to combine methods. Switching between recommendations, combining the scores to an overall score, having various recommendations as output and more. [12]

Due to the field of recommendation systems being a very broad field, this seminar paper aims to focus on the core task of music recommendation.

Since there has been music, there were in some way music recommendations happening.

With the rise of online streaming providers the recommendation tasks switched from personal recommendations of a salesman in a CD-shop to an app or website providing recommendations to users. Other than that the core task didn't change. A music recommendation system has to introduce the customer or user to a new, available song, artists or even playlist which also matches their taste and therefore produces high user satisfaction.

Those recommendations not only need to satisfy the user's need but also affect artists, trends and charts. The more people listen to a song, the higher it is voted in the charts and the more income the artist receives. This leads to the concern of biased recommendation systems impacting not only the user but also other people and the music industry itself.

Later in this work two publications dealing with gender and personality bias in recommendation systems will be discussed.

Music recommendation is a highly competitive field in the corporate world. Due to limited access to company internal algorithms the focus of this work is to

examine publications in the research domain.

The main hypothesis of this seminar paper is:

State-of-the-art music recommendation: Are we there yet?

This question indicates that there is a lot of new development in the field of music recommendation but we don't have an answer to the really important question of whether we already achieve sufficient results in real-world usage scenarios such as when streaming music on a telephone.

To answer this question this seminar paper aims to analyze various recent research approaches for the task of music recommendation. Additionally recent publications evolving around various problems and concerns of music recommendation are considered.

Nowadays various disciplines work on this topic such as *Computer Science*, *Human Computer Interaction* (HCI), *Psychology*, *Sociology*, *Information Processing*, *Cloud Computing* and more. Using this knowledge we can say that the core topic of recommendation systems is broadly spread among various fields and therefore a lot of different aspects need to be taken into account to get high qualitative recommendations.

An additional difficulty of this domain is the human and personal factor. It can be really hard to generalize because of different users and their personality, mood, intention and other factors. Recent work did a lot of research into this direction.

To summarize the following chapters aim to provide an overview of recent developments for the task of music recommendation followed by a critical discussion and evaluation as well as conclusion.

2 State-of-the-art

This section provides an overview of recent publications in research for the core task of music recommendation.

2.1 Datasets for Music Recommendation

One core aspect for achieving good predictions in general is a suitable dataset. Many of the methods mentioned in this seminar paper to some extent rely on publicly available datasets. This subsection aims to provide an overview on the latest released datasets for music recommendation.

In the year 2022 Moscati *et al.* proposed a multi-modal large scale dataset suitable for the task of music recommendation. The dataset contains 109 269 songs with each having 26 features. It is called *Music4All-Onion* because the structure of the dataset is based on an onion with various different layers. Inner layers

tend to be less semantic meaningful while outer layers have a strong semantic meaning.

The inner most layer contains acoustic features and is called *Audio-layer*. This layer is followed by the *EMD-layer* which contains processed features evolving around the lyrics of a song. *EGC-layer* is the name of the middle layer which provides information on the corresponding genres. The next layer is called *UGC-layer* and contains data concerning the tags of the tracks. *DC-layer* is the outermost layer which consists of various video features.

The authors did some benchmarking experiments and stated that they achieved promising results.[15]

Grigorev *et al.* stated that only little available datasets include aspects recently proven to be of importance such as situational or emotional data. For collecting the dataset they used a method split up into three main parts. The first part was a lab study where they tried to collect user's preferences using the *Million Song Dataset* (MSD). As the second part of the study the authors tried to collect situational data and feedback. The feedback was received via smart wristbands and an app which provides emotional feedback and music recommendations. The final step included the usage of a situational recommender built with the data gained from the previous study's part.

In comparison to other available datasets *SiTunes* includes traditional recommendation data, psychological and physiological data as well as explicit and implicit feedback in both a lab and also a real-world usage scenario. [16]

Schedl *et al.* presented a novel large-scale dataset consisting of over 2 billion listening events and 120 000 users. To generate the dataset they used an available API to crawl data from the platform Last.fm¹. After some postprocessing they additionally included Spotify URI's to make the audio easily accessible from the dataset. Other core feature dimensions present are contextual and content data as well as demographic information such as age, gender or country of the users.

The dataset cannot only be used for music recommendation but also for various forms of music classification and analysis to for example gain insights in various unfairness produced by different recommendation systems. This broad usage variety is one of the main advantages of this new dataset.

Additionally *LFM-2b* is not only a large-scale dataset but also consists of a wide timeframe with entries ranging from the year 2005 to the year 2020. The authors stated that they also plan further extensions to make the dataset even more universally usable.[17]

¹ <https://www.last.fm/de/>

2.2 Music Recommendation Methods

This subsection aims to provide an overview on methods recently developed or published for improving the field of music recommendation.

CAME, content and context-aware music embedding for recommendation, is a novel approach proposed by Wang *et al.*. They stated that traditional methods often have limitations due to missing additional information aspects. Therefore they aim to combine content and contextual information for a personalized music recommendation system. Content information is for example metadata, tags or lyrics of songs. Contextual information consists of the user's behaviour such as the session sequence when the user listens to tracks or sequences of music played. The combination of those two information types is done via a heterogeneous information network (HIN). Their approach *CAME* then uses deep learning techniques such as CNN and attention mechanisms. Evaluation was done qualitatively and quantitatively on a real-world dataset crawled from Xiami music². The authors concluded that their approach outperforms the baseline approaches. *CAME* differs from other state-of-the-art recommendation systems because it also incorporates dynamic music features, heterogeneous information and various characteristics of music. [1]

Moscato *et al.* proposed a novel approach where they included user's moods, emotions and personality traits for the task of music recommendation. According to the authors recent publications revealed evidence for the importance of such aspects in terms of a good recommendation. Their approach extends a content-based recommendation method with features of user's mood and personality.

The authors used a dataset from Deezer³ as well as two sources for user's personality informations. They were able to show that the newly introduced factors are of great importance for good music recommendations. [7]

In 2020 Hansen *et al.* tried to tackle challenges such as dealing with short songs or user's listening to the same songs multiple times. For this they used a dataset from Spotify⁴ and a neural network architecture.

Their main hypothesis was that contextual information such as the device or a timestamp as well as sequential user behaviour are important dimensions for music recommendation systems. They were able to conclude that both contextual and sequential data are of high value for the core task of music recommendation. [8]

In their work Niyazov *et al.* proposed a novel approach for content-based music recommendation. They aimed to recommend suitable music to users derived

² <https://www.mi.com/global/>

³ <https://www.deezer.com/de/offers>

⁴ <https://open.spotify.com/intl-de>

from the acoustic similarity of songs. This means that they use the acoustic signal of a song transformed into a suitable representation.

To find the best approach they performed two separate experiments and afterwards compared the quantitative results.

The first experiment is based on acoustic feature analysis. They extract acoustic features from songs, transform them into a vector representation and use 10 Nearest Neighbors with Euclidean distance to get the 10 best recommendations per song.

The second experiment is using artificial neural networks. The networks are trained using spectrogram representations of the acoustic signals. Using the model a vector representation for each song is predicted. Similar to the first approach using 10-Nearest Neighbors the best 10 recommendations per song are gathered. Both approaches got compared with random recommendations, with and without a genre limitation, using precision, recall, F1-score and a version of the F1-score with a threshold of 0.4. This means that all recommendations below this value are considered as irrelevant. The value was introduced by the authors based on their subjective evaluation during the process of development.

Comparing the quantitative measurements the authors were able to conclude that both approaches outperformed the random recommendation with the second approach using Neural Networks achieving the best results.

Additionally the authors investigated genre-specific clustering via visualizing vector spaces on a plane. Due to genre-specific clustering being closely related to the acoustic sound of songs this again highlighted their findings that approaches using ANN achieve better results. [3]

La Gatta *et al.* proposed a novel approach for music recommendation using hypergraph embedding. They split their work into three main parts. First they investigated how to model and store the relevant data in a hypergraph-based structure. Afterwards they focussed on generating hypergraph embeddings from their dataset. The third part of their work was the actual recommendation task where they used graph-based machine learning to generate the top-k track recommendations for each user.

One core challenge of their work was the modelling of scenarios such as a song being part of an album, users listening to songs, et cetera. This was done using a hypergraph-based structure which has vertices and edges but also metadata information.

For evaluation La Gatta *et al.* compared their novel approach with other state-of-the-art music recommenders such as *CAME* [1]. They were able to surpass them at some quality metrics. [2]

In the year 2022 the authors of [6] proposed a novel approach called *DNBMR* (Deep neural based music recommendation) to prevent the cold start problem many other systems face when they rely mostly on historical song data of users. They incorporated music attributes as well as music temporally liked by user's. With a combination of *LSTM* (long short term memory) and neural networks their approach beat two other baseline recommendation systems. Evaluation

was done using a dataset from a large Asian streaming platform provided on the platform *Kaggle* in form of a competition and the metrics *Accuracy*, *Recall*, *F-score*, *AUC* (area under curve), *MAP* (mean average precision) and *user coverage*. This approach benefits from the fact that as soon as a user listens to a song their taste is grasped.

In their work Bontempelli *et al.* tried to improve the Flow algorithm of the music provider Deezer. This algorithm is basically a playlist recommender system as it provides its user's with new songs in the form of a personalized playlist on Deezer. A major disadvantage of this recommender tool was that it didn't include user's emotions and therefore users did not like their recommended songs. Therefore main goal of this work was to include an emotional aspect into the recommendations. This was done via the introduction of six mood categories namely *Chill*, *Focus*, *Melancholy*, *Motivation*, *Party* and *You and Me*. Songs were manually annotated with moods by experts to provide a foundation of labeled data. The general approach is a collaborative filtering method. Additionally they included findings from an audio content analysis. The authors successfully released the improved version named *Flow Moods* in 2021 as a mobile version and in 2022 it was also included in the desktop Deezer application. Results were very promising and could be used to analyze the categorized listening behaviour of Deezer users. [20]

In 2023 a novel approach for music recommendation based on user's emotion was published. This work incorporated the field of human computer interaction as the user was sitting in front of a computer and the detection of the emotions was done live.

The authors tried to predict the user's emotion based on the facial expression defined by various facial characteristics using a neural network algorithm.

Depending on the emotion a suitable *Spotify*⁵ playlist was displayed. To prevent users from being unsatisfied with the song they now are listening to, the surface of the application still provided the option for an alternative playlist and therefore implicit also mood selection. For evaluation they used *Accuracy*, *Loss*, *F1-score*, *Precision* and *Recall*. In comparison to other approaches the focus here was more on the emotion recognition rather than the actual playlist recommendation.[5]

Table 1 summarizes various limitations of current approaches in the field of music recommendation.

2.3 User Experience

This subsection summarizes a publication evaluating the experience of users when being confronted with corporate recommendation systems as for example the in-house tools from various popular music providers.

⁵ <https://open.spotify.com/intl-de>

Limitations	
Limited data	Approaches such as [3] use only a limited amount of data to train their models. Jannach <i>et al.</i> [21] mentioned that sometimes smaller subsets of the data are used due to limited resources.
Cold-start problem	Some works try to tackle the cold-start problem. It basically means that there is too little data present for accurate predictions. La Gatta <i>et al.</i> [2] state, that collaborative filtering methods suffer from this problem.
Reproducibility	Works such as [9] showed that one limitation of current approaches in this field is their reproducibility.
Unsuitable methodologies	In their work [9] Cremonesi <i>et al.</i> discussed various methodological issues visible in current approaches. This ranges from unsuitable experiments to a lack of research questions.
Inference times	There is little information to be found on actual recommendation times. Users should receive fast recommendations.

Table 1. Summary of limitations of current approaches for music recommendation

In their work [14] the authors aim to gain knowledge about user’s experiences on two popular streaming platforms namely *Spotify*⁶ and *Deezer*⁷.

They used a list of desired criterias from another work and investigated to which extent the platforms meet those. Following aspects were considered: *user activity, satisfaction, feedback, cold-start, cognitive load, learning, and personality and preferences*.

After performing the experiment with 10 participants the authors analyzed the results. Both platforms suggestion systems did not meet the desired criterias. The provider Deezer surpassed Spotify in terms of evaluated data.

In addition this work aims to explain the need for adapting recommendation systems for other goals than just achieving a high accuracy score. They were able to show that the inclusion of various other views increases user’s joy when using a platform for playing music.

⁶ <https://open.spotify.com/intl-de>

⁷ <https://www.deezer.com/de/offers>

2.4 Evaluation of Methods for Music Recommendation

This subsection includes recent work evaluating state-of-the-art music recommendation methods in terms of their quality and aiming to find out whether any unfairness is present in the predicted suggestions.

In 2020 Melchiorre *et al.* proposed their work where they analyzed the personality bias of common algorithms in the area of music recommendation. They stated that other work already proved that there is for example a bias on item or user level in multimedia recommendations. Additionally there exists work proving a correlation between music preferences and personality traits.

For their experiment they created a novel dataset combining user's personality traits and music consumption. To gather this data they used Twitter microblog posts with certain hashtags as well as the music metadata dataset MusicBrainz. Using the *IBM Personality Insight API* the authors were able to integrate the user's personality traits according to the *OCEAN* model of psychology. The authors compared *Recall@K* and *NDCG@K* (normalized discounted cumulative gain), with *K* being the length of the recommendation list, of three popular music recommenders namely *SLIM*, *EASE* and *Mult-VAE*. They were able to prove that low scores on personality traits such as *openness*, *extraversion* and *conscientiousness* lead to higher performance while low scores for *neuroticism* and *agreeableness* lead to low performance according to the chosen metrics. Finally they stated some limitations of their work as well as planned extensions. [4]

Ferraro *et al.* published a work where they analyzed the gender imbalance in song recommendation systems. A core driver of their research were interviews with musicians where the musicians wanted female artists to be recommended more frequently than it currently happens. To analyze whether there is actually a gender bias, the authors incorporated gender information into a dataset and performed recommendations using a collaborative-filtering method named *ALS*. They were able to find that the ranking of results includes a bias where the tendency first shows music from male artists. To improve the situation they suggested a re-ranking method. In the future they want to test their re-ranking method, which is basically a feedback loop, on datasets more similar to real-world usage. [13]

3 Critical discussion and evaluation

This section aims to provide a critical view and evaluation of current methods in the field of music recommendation.

One critical core aspect is the usage of incomplete data. Not every novel approach uses a suitable and state-of-the-art dataset. Some even pick a small percentage of the data available for better computational efficiency, this doesn't model real-world scenarios where we have tons of data and users and complex relationships inbetween. Another aspect is the access to data.

For example in one work [4] the authors crawled Twitter data to gain insights on the personality traits and the music usage of users. One of their main limitations is that users only post what they want others to know. There probably is a huge bias in that data as the real, deep, personal information is missing.

A large percentage of work in the field of recommender systems doesn't perform proper evaluation methods. Approaches such as [3] are considered decent just through analyzing various performance metrics on a test dataset but not through real usage tests. This leads to the problem that many research approaches in the field of recommender systems are not applicable for real-world scenarios. In their work Cremonesi and Jannach evaluated recent work in the field of recommender systems for their reproducibility, evaluation methodologies, baselines and research questions to answer the questions whether they actually improve the state-of-the-art. If they found evidence that there is no real progress, they tried to find reasons for that and provided recommendations for researchers but also reviewers and chairs. They suggested to move from a strong reliance on accuracy and well-explored problems to more practice related work. [9]

In their paper Jannach *et al.* tried to find out why deep learning approaches constantly outperform state-of-the-art baselines in new publications but don't win against more traditional solutions in competitions held at conferences. One potential reason is the computational expensiveness of deep learning approaches in combination with large datasets. As an alternative to long waiting times people tend to use different approaches or use a smaller subset of the dataset which could result in a different performance. Datasets used in such competitions tend to have a high data sparsity which could lead to a bad performance of deep learning approaches due to the risk of overfitting. Additionally the presence of a lot of meta data could result in a confused model and deep learning models tend to bypass side information. The authors highlight that researcher should not solely rely on deep learning methods but also consider different approaches such as traditional machine learning, depending on the use case.[21]

Summarizing we can say that every approach has its own limitations. However this should not prevent researchers from using proper and reproducible techniques. Many approaches rely on old or not perfectly suitable datasets because it results in huge effort to improve or create new datasets. Others use incomplete methodologies or evaluation practices for various reasons.

One has to note that this seminar paper only contains publicly available music recommendation systems. Outside of the research environment recommender systems is a very competitive field. Large companies such as the music provider service *Spotify*⁸ have a desperate need of good music recommendation to satisfy their user's needs and to stand out from their competitors. A large share

⁸ <https://open.spotify.com/intl-de>

of their employees probably does similar work but is following a confidentiality agreement due to their contract.

3.1 Ethical Evaluation of Music Recommendation

This subsection highlights the view on responsibility and ethics as well as effects on the society when performing music recommendation and recommendations in general.

In 2016 Paraschakis published a publication where one part is about several ethical challenges which increase the difficulty of developing recommendation systems. The paper doesn't focus on music recommendation but more on recommender systems in general and their usage in e-commerce. The author stated that the ethical part is often neglected but has various important dimensions to consider.

Collecting data Users are not always aware of the fact that part of their actions are tracked and later used for example for the generation of a dataset.

User profiling Some algorithms strongly rely on user profiles which could publish private information of users or be used for attacks such as a profile injection.

Public datasets As a developer of a new recommendation system large publicly available datasets help a lot. They consist of personas personal information and behaviour which makes them on the other hand morally questionable. Recent incidents show that gaining personal information on the people behind such a datasets content is possible. If something like that happens it could lead to severe consequences.

Filering of suggestions Many recommendation systems don't filter their recommendations. This means that they don't consider the fact that some recommendations could be triggering or inappropriate for certain user groups. Additionally often there is no balance between benefit for the user and benefit for the company. According to the author incidents such as recommendations with a tendency of higher priced suggestions happened and are morally questionable.

Algorithmic opacity Usually a user only sees the recommendation but no additional information. This could lead to frustration or other feelings of the user.

Bias and behaviour manipulation Other work already highlighted the presence of biases and unfairness where recommender systems provided different results to users depending on various personality traits et cetera. How do developers ensure that such unfairness does not happen? An example of e-commerce recommendations where one user group gets suggestions of higher priced items than the other again highlights the problems arising with such biases.

A/B Testing A/B Testing is often used by companies to test their latest changes or developments without changing the whole system for all users. For this kind of test half of the users experience the new version and the rest the old version of the recommendation system. This helps companies to on the one hand gain information on user's interaction with and reaction concerning the latest changes but also performance information. Users don't know that they are currently part of an experiment and obviously did not agree on any kind of participation. Some users therefore use the real system while others serve as a kind of test objects which is again morally questionable if it for example effects decisions including money et cetera.[18]

2021 Levina wanted to gain insights into possible societal and ethical problems arising when developing and using a recommender system based on machine learning. For this the author designed a fictional recommender system in the form of a food delivery service which aims to recommend food choices based on user information. This fictional recommender system was then used to identify societal and ethical concerns using analysis techniques such as a data process-oriented approach. All concerns were summarized in a list with the following points standing out:

Data access There is a possible situation where various stakeholders could gain access to user's data without the consent of the people. This should be prevented at all cost with for example an explicit permission systems accepted or declined by the user.

Data state and life cycle The user is not aware of the current state of his data or the life cycle, in other words the period of time the data will be available and used. For this the author suggests developers to include a section for frequently asked questions (FAQ) somewhere on their site or in their app. Additionally there should be an automated deletion operation for data when it's at the end of the life cycle.

Data storage In addition to the previous keypoint users also are not aware of where and how their data is stored and whether it is also used for any additional purpose. According to the author responsible developers and stakeholders should internally establish guidelines and a concept which is transparently told to users. This ensures trust and calmness.

Societal concerns Various societal concerns were mentioned. The effect of the app or program on society in general needs to be taken into consideration. It is recommended to skip public test phases and instead perform test phases in a supervised setting, for example in a lab study. Another recommendation by the author is to not perform A/B Testing as the user is not aware of being part of the test. Instead it is suggested to first to a supervised lab testing phase or provide a beta version with information about the changes and usage. Changes should be rolled out to all users at once. The final important aspect this work highlights is the limited traceability of the effects of the app and the recommendation system on the society. For this surveys, traffis analysis and restaurant contact is recommended.[19]

4 Conclusions

Recommender systems in general is a very large and broadly spread research topic with a huge amount of publications.

With recent development in the field of artificial intelligence, especially neural networks, a lot of new approaches and ideas were published.

As also [9] stated, not every novel approach is improving the state-of-the-art. This can occur to various reasons such as the usage of unsuitable datasets or poor evaluation not considering the different important aspects of a decent recommendation system.

The core focus of this seminar paper is to summarize current developments in the research field of music recommendation. Additionally a focus is on concerns and the critical discussion of the proposed methods. The main hypothesis has a special focus on the applicability in the real world to for example get song recommendation when using a music streaming provider.

There is a lot of work aiming to include new topics, aspects and relevant information to further improve the task of music recommendation. Often unfortunately a lot of limitations occur and methods are not really applicable for real-world usage or even not reproducible.

Unfortunately many recommendation systems suffer from producin biased suggestions. Studies showed a variety of bias ranging from a gender bias to a bias depending on the time or the device used to listen to songs. This bias not only affects the user's which consume the recommended music but also artists and their income as well as the music industry in general.

To conclude we can say that there is a tendency to use novel methods such as deep learning in comparison to the usage of traditional approaches.

Additionally there is a lot of research evolving around the incorporation of information and benefits from other disciplines such as for example the soziological or psychological effects to improve state-of-the-art music recommendation.

To answer the hypothesis: Unfortunately, **we are not there yet**.

The field of music recommendation is very broad and a lot of different aspects highlighted in this seminar paper need to be taken into account when aiming for a decent music recommender system. But overall most of the recent publications point into the right direction.

References

1. Wang, Dongjing and Zhang, Xin and Yu, Dongjin and Xu, Guandong and Deng, Shuiguang: Came: Content-and context-aware music embedding for recommendation. In: IEEE Transactions on Neural Networks and Learning Systems, pp. 1375–1388 (2020)
2. La Gatta, Valerio and Moscato, Vincenzo and Pennone, Mirko and Postiglione, Marco and Sperli, Giancarlo: Music recommendation via hypergraph embedding. In: IEEE transactions on neural networks and learning systems (2022)
3. Niyazov, Aldiyar and Mikhailova, Elena and Egorova, Olga: Content-based music recommendation system. In: 2021 29th Conference of Open Innovations Association (FRUCT), pp. 274–279 (2021)
4. Melchiorre, Alessandro B and Zangerle, Eva and Schedl, Markus: Personality bias of music recommendation algorithms. In: Proceedings of the 14th ACM Conference on Recommender Systems, pp. 533–538 (2020)
5. Priyanka, V Tejaswini and Reddy, Y Reshma and Vajja, Dharani and Ramesh, G and Gomathy, S: A Novel Emotion based Music Recommendation System using CNN. In: 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 592–596 (2023)
6. Singh, Jagendra and Sajid, Mohammad and Yadav, Chandra Shekhar and Singh, Shashank Sheshar and Saini, Manthan: A novel deep neural-based music recommendation method considering user and song data. In: 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1–7 (2022)
7. Moscato, Vincenzo and Picariello, Antonio and Sperli, Giancarlo: An emotional recommender system for music. In: IEEE Intelligent Systems, vol. 36, pp. 57–68 (2020)
8. Hansen, Casper and Hansen, Christian and Maystre, Lucas and Mehrotra, Rishabh and Brost, Brian and Tomasi, Federico and Lalmas, Mounia: Contextual and sequential user embeddings for large-scale music recommendation. In: Proceedings of the 14th ACM Conference on Recommender Systems, pp. 53–62 (2020)
9. Cremonesi, Paolo and Jannach, Dietmar: Progress in recommender systems research: Crisis? What crisis?. In: AI Magazine, vol. 42, number 3, pp. 43–54 (2021)
10. Jannach, Dietmar and Zanker, Markus and Felfernig, Alexander and Friedrich, Gerhard: Recommender systems: an introduction. In: Cambridge University Press (2010)
11. Schafer, J Ben and Frankowski, Dan and Herlocker, Jon and Sen, Shilad: Collaborative filtering recommender systems. In: The adaptive web: methods and strategies of web personalization, pp. 291–324 (2007)
12. Burke, Robin: Hybrid recommender systems: Survey and experiments. In: User modeling and user-adapted interaction, vol. 12, pp. 331–370. Springer (2002)
13. Ferraro, Andres and Serra, Xavier and Bauer, Christine: Break the loop: Gender imbalance in music recommenders. In: Proceedings of the 2021 conference on human information interaction and retrieval, pp. 249–254 (2021)

14. de Assunção, Willian Garcias and Zaina, Luciana Aparecida Martinez: Evaluating user experience in music discovery on deezer and spotify. In: Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems, pp. 1–1 (2022)
15. Moscati, Marta and Parada-Cabaleiro, Emilia and Deldjoo, Yashar and Zangerle, Eva and Schedl, Markus: Music4All-Onion–A Large-Scale Multi-faceted Content-Centric Music Recommendation Dataset. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 4339–4343 (2022)
16. Grigorev, Vadim and Li, Jiayu and Ma, Weizhi and He, Zhiyu and Zhang, Min and Liu, Yiqun and Yan, Ming and Zhang, Ji: SiTunes: A Situational Music Recommendation Dataset with Physiological and Psychological Signals. In: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval, pp. 417–421 (2024)
17. Schedl, Markus and Brandl, Stefan and Lesota, Oleg and Parada-Cabaleiro, Emilia and Penz, Davd and Rekabsaz, Navid: LFM-2b: A dataset of enriched music listening events for recommender systems research and fairness analysis. In: Proceedings of the 2022 Conference on Human Information Interaction and Retrieval, pp. 337–341 (2022)
18. Paraschakis, Dimitris: Recommender systems from an industrial and ethical perspective. In: Proceedings of the 10th ACM conference on recommender systems, pp. 463–466 (2016)
19. Levina, Olga: Towards Implementation of Ethical Issues into the Recommender Systems Design. In: ICCGI 2021, The Sixteenth International Multi-Conference on Computing in the Global Information Technology, pp. 6–11 (2021)
20. Bontempelli, Théo and Chapus, Benjamin and Rigaud, François and Morlon, Mathieu and Lorant, Marin and Salha-Galvan, Guillaume: Flow moods: Recommending music by moods on deezer. In: Proceedings of the 16th ACM Conference on Recommender Systems, pp. 452–455(2022)
21. Jannach, Dietmar and de Souza P. Moreira, Gabriel and Oldridge, Even: Why are deep learning models not consistently winning recommender systems competitions yet? A position paper. In: Proceedings of the Recommender Systems Challenge 2020, pp. 44–49 (2020)