Journal of New Music Research

ISSN: 0929-8215 (Print) 1744-5027 (Online) Journal homepage: www.tandfonline.com/journals/nnmr20

Hit song science: a comprehensive survey and research directions

Danilo B. Seufitelli, Gabriel P. Oliveira, Mariana O. Silva, Clarisse Scofield & Mirella M. Moro

To cite this article: Danilo B. Seufitelli, Gabriel P. Oliveira, Mariana O. Silva, Clarisse Scofield & Mirella M. Moro (2023) Hit song science: a comprehensive survey and research directions, Journal of New Music Research, 52:1, 41-72, DOI: 10.1080/09298215.2023.2282999

To link to this article: https://doi.org/10.1080/09298215.2023.2282999

Published online: 20 Nov 2023.

JOURNAL OF NEW MUSIC RESEARCH
2023, VOL. 52, NO. 1, 41–72
https://doi.org/10.1080/09298215.2023.2282999

Routledge
Taylor & Francis Group

Hit song science: a comprehensive survey and research directions

Danilo B. Seufitelli, Gabriel P. Oliveira, Mariana O. Silva, Clarisse Scofield and Mirella M. Moro

Computer Science Department, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

ABSTRACT

Hit Song Science (HSS) is an emerging topic that aims to unveil the success dynamics within the music industry. Considering the growth of the area, we provide a comprehensive study with a complete review of the main topics of this interdisciplinary field from a computer science perspective. We also define a generic workflow for HSS, introduce taxonomies for success measures and musical features, and categorize the main current learning algorithms. Overall, this survey may serve as a starting point for future research on HSS, as it emerges as a promising field that benefits both the academy and the music industry.

ARTICLE HISTORY

Received 25 October 2021
Accepted 7 November 2023

KEYWORDS

Hit song science; music information retrieval; hit song prediction; music computing; music-related features

1. Introduction

Music is not only one of the most important forms of cultural expression but also entails one of the most dynamic worldwide industries. Over recent decades, there has been a significant change in how people consume music, from physical records to streaming services. Furthermore, there is a wide variety of streaming formats, including premium signature, ad-supported services (such as YouTube, Vevo, and free-account Spotify), and streaming radio services (such as Pandora and SiriusXM). Combined, streaming revenues accounted for 83% of total revenues in the US music industry by the end of 2021.[1]

Indeed, music streaming services have become the main source of revenue within the global music market, reaching a total of US$ 16.9 billion by the end of 2021.[2] For the seventh consecutive year, such growth is mainly driven by fans' engagement on paid streaming services. In 2021, the total accounts for such services rose to 523 million, with an associated revenue increase of 24.3%. As a result, the landscape of the music industry has become more complex, encouraging artists to reinvent strategies to maintain their presence in the market and expand it to reach new audiences.

As an example of reinvention, 'Baby Shark', a famous nursery rhyme, appeared on charts for the first time in January 2019, after a video by South Korean educational brand Pinkfong sparked a viral dance challenge. Such a YouTube video clip helped the song reach No. 32 on the Billboard Hot 100 for 20 weeks, with 395 million streamings in the first half of 2019. In 2021, on-demand audio streaming reached a new single-year high of 988.1 billion streams in the United States, with the increasingly influential TikTok helping to accelerate music trends as younger consumers discover songs both new and classic through popular memes and dance challenges.[3] In this context, predicting whether a song will become a hit remains a valid concern of the music industry, which constantly seeks to increase its revenue while dealing with different audiences.

Sucha prediction (hit or not) has significant importance for different groups. For the music industry CEOs, it may assist in maximizing future success by helping them choose whom to invest. Furthermore, by investing correctly in potential artists/songs and their distribution, music providers may increase physical and digital album sales and improve on-demand audio streaming services. Artists may also profit by identifying the most suitable songs to lead an album to stardom. For music consumers, it may help to decide if an album is worth buying because it may potentially contain three to five hits instead of being a *one-hit-wonder* album. In addition, consumers play an essential role in the music market as they are the musical universe 'feedbackers'. In other words, musical success is measured based on consumer response and updated based on their evolving preferences. Essentially, this ability is the main drive

CONTACT Mirella M. Moro mirella@dcc.ufmg.br Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Belo Horizonte, MG CEP 31270-901, Brazil

[1] 2021 Year-End Music Industry Revenue Report (RIAA): http://www.riaa.com/reports
[2] IFPI Global Music Report 2022: http://gmr.ifpi.org/
[3] MRC Year-end Report: https://mrcdatareports.com/wp-content/uploads/2022/01/MRC_YEAREND_2021_US_FNL.pdf

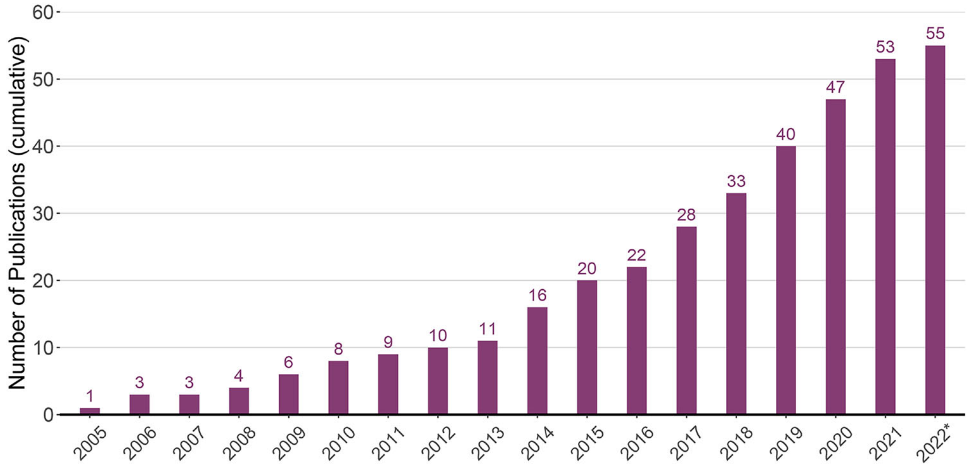


Figure 1. Hit Song Science publications (cumulative), 2005–2022* (search performed in September 2022; i.e. before the year ended).

behind the research field referred to as Hit Song Science (HSS), which Pachet (2011) defines as 'an emerging field of science that aims at predicting the success of songs before they are released on the market.' Overall, HSS focuses on predicting whether a song will be a hit before its distribution using automated approaches, such as machine learning algorithms. Therefore, we consider this definition of HSS throughout this survey.

The study of HSS has two different perspectives: the prediction viewpoint and the scientific one. The former is justified by HSS involving variable issues such as a complex combination of micro-emotions (also related to personal history), music specifics, and other elements that escape our understanding, and how all of them may determine the success of a song. Conversely, HSS as a science is a bold view against the assumption that such variables hinder prediction, and then it relies on features that would give a convenient result. The discussion of HSS being a wish or science raises many issues for the discipline. As pinpointed by T. Li et al. (2011), HSS should also rely on 'the psychology of music listening and the effects of repeated exposure, the paradoxical nature of the Western media broadcasting system, radios in particular,[4] and the social influence human beings exert and receive from each other'.

Although true, such arguments have not stopped companies and startups from creating and selling their solutions (e.g., Polyphonic HMI, MixCloud, and Hit Songs Deconstructed), nor have they stopped the research efforts from the scientific perspective. In particular, HSS is considered a specific task within Music Information Retrieval (MIR), a more comprehensive field that aims to extract relevant information from music content (Pachet, 2011). As MIR is a well-established research field, a few survey articles cover its main topics, including compressed audio files (Zampoglou & Malamos, 2014), systems using microphone input (Maršık et al., 2015), MIR systems (H. Li et al., 2017), and content-based MIR (Murthy & Koolagudi, 2018). Regarding MIR tasks, some are more studied than others with exclusive surveys. For example, Sturm (2012) and Corrêa and Rodrigues (2016) focus on music genre recognition and classification, whose goal is to determine the genre of a song from a set of musical features.

Overall, Hit Song Science has been a constant research topic since 2005, as well illustrated in Figure 1. The publications considered in such a figure (and the remaining of the article) were obtained mainly from DBLP[5] until September 30, 2022, by using the following keywords: *hit song science*, *hit song prediction*, *musical success*, and *music popularity*. Such keywords were considered individually in our search. The complete list of the 55 papers considered for this survey is listed in the Appendix. Besides such an increasing interest, to the best of our knowledge, this is the first survey article exclusively focussed on HSS. Here, we use a bottom-up approach to investigate similarities among the selected articles regarding their methods, features, and algorithms. Specifically, we model such similarities to define a generic workflow on Hit Song Prediction, whose steps are individually presented and discussed in specific sections later on.

Our goal is to answer several open questions related to this field, including *what are the main research problems of HSS? How is musical success defined? What are the most used learning techniques for solving HSS problems?*

[4] This paradox highlights the mutual influence between a radio station and the public, where not only does the station shape public opinion, but the public also has the power to impact what gets broadcasted.

[5] DBLP Digital Library: http://dblp.uni-trier.de/

What are the challenges and research opportunities for this field? We answer such questions throughout our bottom-up strategy by reviewing the literature on HSS to describe its main research problems, frequently used data sources, musical success definitions, features, and learning methods. Furthermore, we propose novel taxonomies for (i) success measures, (ii) features, and (iii) learning methods used in this field to consolidate the existing knowledge in HSS and guide future research on this topic. Finally, we compare the reviewed articles on such characteristics, and our analyses reveal that there is not a 'feat'. for an ideal hit song prediction model, as its performance depends on subjective decisions made in the analysis process. Hence, this survey bridges computer science and music-related fields, motivating upcoming research on Hit Song Science.

The remainder of this article is organized into six sections. First, in Section 2, we present an overview of the HSS field and propose the generic workflow for its prediction problem. The following sections reflect the main steps of such a workflow: (i) most common data sources (Section 3); (ii) success perspectives (Section 4); (iii) features (Section 5); and (iv) learning methods (Section 6). Further, in Section 7, we point out important research directions to guide future work on HSS. Finally, we present a general discussion and concluding remarks in Section 8.

2. An overview on hit song science

In this section, we define Hit Song Science and present an initial review to investigate the similarities between the main studies in this field (Section 2.1). We use this review as a bottom-up strategy to define a generic workflow that guides the next sections of this survey (Section 2.2).

2.1. Hit song science

Hit Song Science (HSS) is a multidisciplinary field where computer science meets conventional music-related topics, such as music theory, sociology of music, and cultural markets. It involves acquiring and analyzing music data with different modalities from distinct data sources. Such analysis resorts to Information Retrieval, Machine Learning, Data Mining, and other advanced technologies to connect music and science. In this process, the main objective is to detect, or rather predict, whether a given song will become a chart-topping hit. Mainly, it boils down to distinguishing hits from non-hits, where a hit is usually a song featured at the top of music charts.

HSS emerges as a field of predictive studies to better understand the relation between the inherent characteristics of songs and their popularity. The premise of HSS is that popular songs are similar to the features that make them appealing to most people. Such attributes could then be explored through Machine Learning techniques to predict whether a song will top or appear in the popularity charts. Predicting the popularity of musical tracks provides huge benefits for all parties involved in the global music industry, as it improves revenues by focusing on potential hits. Moreover, predicting hits from social music media (e.g., Spotify and Deezer) can help improve revenues even further through advertising and publicity.

Despite being a recurring topic within Music Information Retrieval (MIR), the concept of HSS was first introduced by the Polyphonic HMI,[6] in 2003. This artificial intelligence company developed a machine learning software called Hit Song Science, which uses mathematical algorithms and statistical techniques to predict the success of a song in the current market. According to Polyphonic, its software could anticipate the success of artists such as Norah Jones, Jennifer Lopez, and Robbie Williams. Such a revolutionary tool allowed scientists and music enthusiasts to break down millions of past hit songs into their mathematical features. Despite such an unprecedented effort, it took time after the emergence of this tool for the issue to gain significant attention as a research direction, as informed in Figure 2.

Early works in this area focus mainly on song-related features and introduce different types of approaches, including boosting classifiers (Dhanaraj & Logan, 2005), experimental studies (Salganik et al., 2006), adaptive algorithms (Chon et al., 2006) and Support Vector Machines (Dhanaraj & Logan, 2005; Pachet & Roy, 2008). As more techniques for learning from data became available and widely used, recently proposed approaches have become more advanced and elaborated. Specifically, more robust methods have been applied not only to analyze bigger and more complex data but also to learn additional insights (Herremans et al., 2014; Martín-Gutiérrez et al., 2020; Yang et al., 2017; Zangerle et al., 2019). Moreover, with the popularization of social networks, information about music consumers' tastes becomes available and easy to explore (Abel et al., 2010; Araujo et al., 2017; Bischoff et al., 2009; Y. Kim et al., 2014; Koenigstein et al., 2009; Shulman et al., 2016; Yu et al., 2019). As a result, many studies have also relied on social information as hit song predictors (Interiano et al., 2018; Ren et al., 2016; Silva & Moro, 2019; Silva et al., 2022, november 7–11, 2019).

Although hits have been around for decades, the research on Hit Song Science is still recent and has been increasing every year (see Figure 1). Indeed, efforts in

[6] Polyphonic HMI, (January, 2020), http://bit.ly/polyphonic-hmi

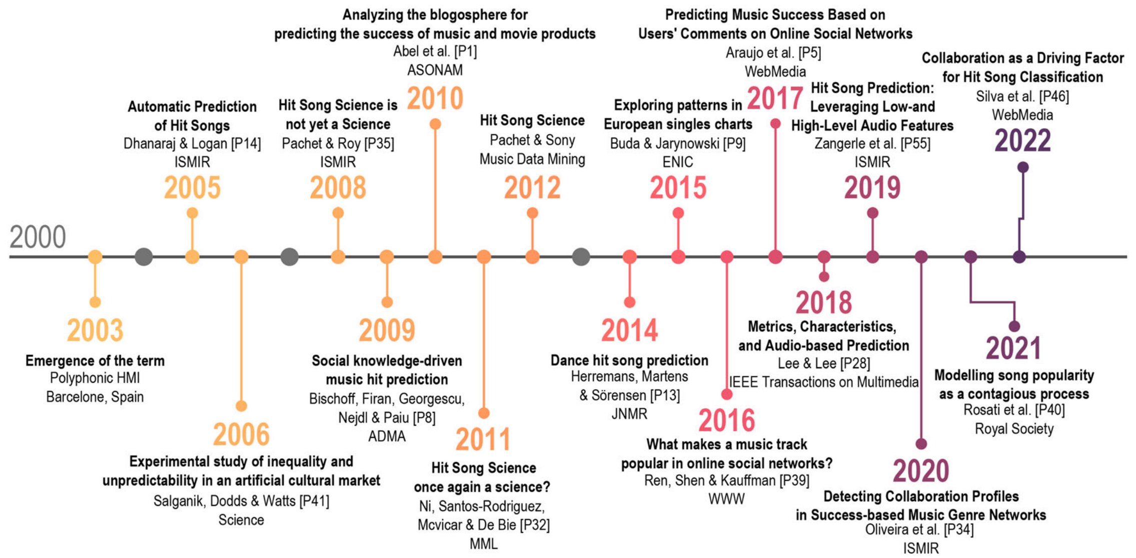


Figure 2. Hit Song Science research timeline in a nutshell.

such an area are growing, with researchers using different perspectives and proposing new approaches to study the problem of predicting hit songs (as illustrated in Figure 2). Despite the diversity of viewpoints, our research has identified certain similarities in the studies' methodologies through a preliminary review. Therefore, we propose a generalization of such workflows in the next section.

2.2. Generic workflow for hit song prediction

Despite the varied alternatives, most existing studies tackle HSS as a *Hit Song Prediction* problem, in which classification and regression are both chosen as the most common solution tasks. For classification, given a set of songs' features, the task is to classify a song as a hit or a non-hit. Specifically, classification algorithms receive a set of different songs' features (acoustic, metadata, lyrics, etc.) as input to execute their internal processing respecting their particularities (which differ each technique) and then predict a class (label) for each song instance (i.e., *hit* or *non-hit*). Note the focus is on defining a hit song by *discrete values*. On the other hand, regression solutions aim to predict a continuous outcome (y) variable based on the value of one or multiple predictor variables (x). Moreover, classification approaches lose relevant information about song popularity due to binary conversions.

Such tasks fall under the umbrella of supervised machine learning, which is designed to learn based on a labeled dataset. However, other types of learning have also been used to predict musical success, as well as to find patterns and potential factors that influence songs' popularity. In general, such approaches include clustering algorithms (Silva et al., 2019), statistical analysis (Araujo et al., 2017; Shin & Park, 2018), prediction (Dewan & Ramaprasad, 2014; Silva & Moro, 2019), and social networks analysis (Buda & Jarynowski, 2015; Silva et al., 2019).

Overall, most machine learning solutions follow a predefined pipeline. Therefore, we propose a generic, simple workflow for the *Hit Song Prediction* problem as one contribution of this survey, as shown in Figure 3. From one or more selected data sources, defining a proper success

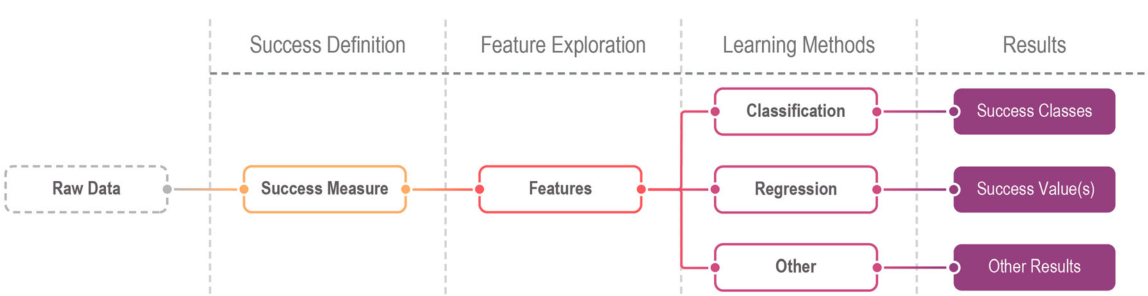


Figure 3. Generic workflow for the Hit Song Prediction problem.

measure to evaluate the prediction model is required (Success Definition). Moreover, such data sources guide which features will be considered as input (Feature Exploration) of the chosen Learning Methods. Then, the output (Results) may be in the form of success classes and values, among others. Such a workflow is reflected in the structure of this article, where a specific section covers each step in the prediction process.

3. Music data acquisition

Following the premise of Hit Song Science (HSS), the first step of most approaches is gathering data regarding song characteristics and success. However, these concepts can be seen in many facets, as the definitions are open to different visions. For instance, data about a given song can be acoustic and/or lyric-based, while its popularity may be measured considering its position within a chart or its sales revenue. Besides, information concerning consumers' behavior may be aggregated to HSS analyses to enhance the results. Therefore, data from multiple sources are necessary and useful to build better models for analyzing and predicting musical success. In this section, we describe and classify the main and most commonly used data sources into four categories according to their purpose: popularity, acoustic characteristics, lyrics, and social behavior.

3.1. Common data sources

Regarding song popularity, research studies usually consider information such as position in charts to determine whether a song is a hit or not. The US-based magazine Billboard[7] is the most consolidated data source, providing many different types of rankings since the 1940s. The Hot 100 is the most commonly used, as it is a weekly list of the 100 most popular songs (regardless of music genre or style) in the US, considering data from radio airplay, sales, and streaming activity (Askin & Mauskapf, 2017; Bischoff et al., 2009; S. T. Kim & Oh, 2021; Y. Kim et al., 2014; Koenigstein et al., 2009; Lee & Lee, 2018; Middlebrook & Sheik, 2019; Nunes & Ordanini, 2014; Raza & Nanath, 2020; Silva & Moro, 2019; Silva et al., 2022, november 7–11, 2019; Tsiara & Tjortjis, 2020; Vötter et al., 2021, 2022). Billboard also aggregates the weekly rankings in a Year-End Hot 100 Chart, which is used in some studies within HSS (Singhi & Brown, 2014, 2015). Still, some studies consider other specific Billboard charts in their analyses. For example, Chon et al. (2006) focus on one specific genre by using the Top Jazz Chart, which is based only on the albums' sales. Also, Lee and Lee (2015) obtain data from The Rock Songs Chart, a weekly list of the 50 most popular rock songs. Such authors believe solutions may produce cleaner results and better insights when focussing on specific genres.

As the world becomes more connected, local engagement shapes the global music environment. In such a way, some studies consider charts from outside the US in their analyses and predictions. The United Kingdom is the second most considered market, having its charts published by the Official Charts Company[8] (OCC) (Herremans et al., 2014; Interiano et al., 2018; Ni et al., 2011). Besides using British charts, Fan and Casey (2013) also collect Chinese hit songs for comparison purposes. Moreover, there are studies considering other European countries (e.g., France, Belgium, and Germany) (Buda & Jarynowski, 2015; Herremans & Bergmans, 2017) and Asian markets such as South Korea (Shin & Park, 2018) and Indonesia (Febirautami et al., 2018). Other popularity approaches use YouTube views and likes (Chiru & Popescu, 2017) and sales data provided by platforms such as Amazon (Abel et al., 2010) and Nielsen SoundScan (Askin & Mauskapf, 2017; Berns & Moore, 2012; Dewan & Ramaprasad, 2014).

Now, changing the subject to features, acoustic characteristics of a song are important tools for describing its structure. Besides being better discussed in Section 5, they have been largely used since early HSS research studies, such as Dhanaraj and Logan (2005), which use in-house databases as their data source. With the evolution of the Music Information Retrieval (MIR) field, new sources take place in HSS, such as EchoNest API, with more than a trillion data points on over 34 million songs in its database (Herremans et al., 2014). Several studies use this API for extracting features such as tempo, time signature, song duration, and loudness (Askin & Mauskapf, 2017; Fan & Casey, 2013; Herremans & Bergmans, 2017; Herremans et al., 2014; Ni et al., 2011; Singhi & Brown, 2015). Nonetheless, with the expansion of music streaming services and the acquisition of EchoNest by Spotify in 2014, its Developer API[9] is now the main source of acoustic features, thus being used by most recent studies (Al-Beitawi et al., 2020; Araujo et al., 2019, 2020, 2017; Febirautami et al., 2018; Gao, 2021; Kamal et al., 2021; Kaneria et al., 2021; S. T. Kim & Oh, 2021; Martín-Gutiérrez et al., 2020; Matsumoto et al., 2020; Middlebrook & Sheik, 2019; Oliveira et al., 2020; Raza & Nanath, 2020; Silva & Moro, 2019; Silva et al., 2022, november 7–11, 2019). Nonetheless, there are still other sources used, such as the Million

[7] Billboard Charts: http://www.billboard.com/charts

[8] Official Charts Company: http://www.officialcharts.com/charts

[9] Spotify Developer API: http://developer.spotify.com/documentation/web-api

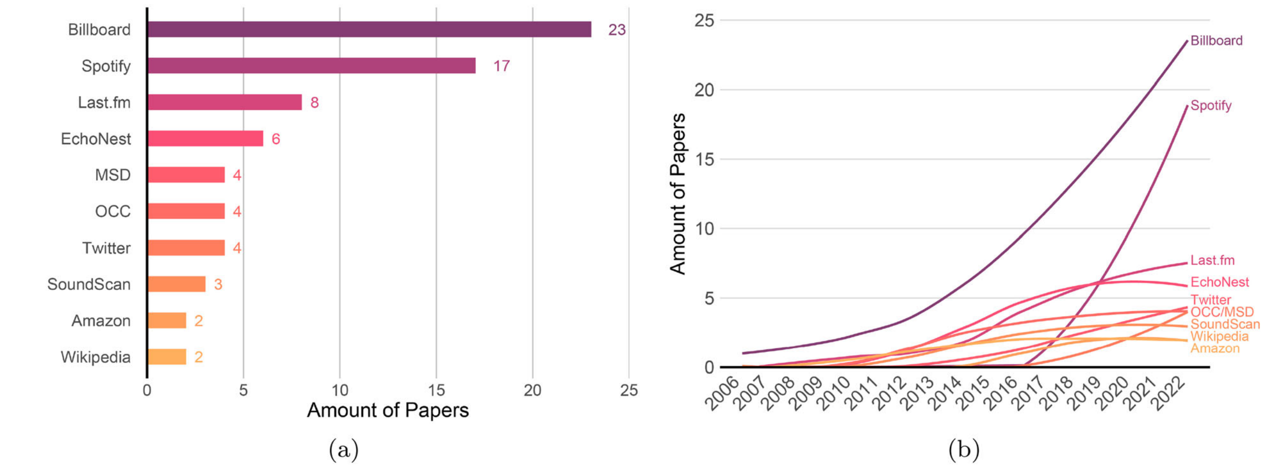


Figure 4. Most commonly used data sources in Hit Song Science (a) and their evolution over the years (b).

Song Dataset[10] (MSD) (Rajyashree et al., 2018; Zangerle et al., 2019) and AcousticBrainz[11] (Interiano et al., 2018; Vötter et al., 2021, 2022).

Data frequently used within HSS papers also include song lyrics, mainly for generating features related to rhyme and text. From the early years of HSS, there has been no consensus on the best and most reliable source for song lyrics, and each study considers a different source. For example, websites such as Astraweb Lyric Search (Dhanaraj & Logan, 2005), MetroLyrics (Singhi & Brown, 2015), MusicSongLyrics (Ren et al., 2016), and LyricsMania (Chiru & Popescu, 2017) are used as lyrics sources by several authors. However, more recent studies, such as Martín-Gutiérrez et al. (2020) and Kamal et al. (2021), use Genius, which has an exclusive API for developers to collect data in a simple and fast way, with no need to use web crawlers or HTML pages.

Finally, a different kind of data source has recently emerged: social media, which changes how people share their opinions and impacts several areas, including the music industry. Therefore, consumer behavior plays a key role in musical success analysis, and online platforms such as Last.fm[12] are largely used to collect listener-based data and features (Bischoff et al., 2009; Dewan & Ramaprasad, 2014; Herremans & Bergmans, 2017; Ren & Kauffman, 2017; Ren et al., 2016; Shulman et al., 2016; Vötter et al., 2021, 2022). Moreover, blogging platforms are also important sources of information about people's feelings about a given song, album, or artist. For example, Abel et al. (2010) use Spinn3r[13] to collect more than 100 million blog posts in the music domain. More recent studies collect data from social networks such as Twitter, Instagram, and Facebook to analyze users' behavior related to a new musical release (Araujo et al., 2017; Cosimato et al., 2019; Y. Kim et al., 2014; Tsiara & Tjortjis, 2020).

3.2. Overall considerations

Songs are complex and dynamic objects that can be analyzed in different ways, and Hit Song Science (HSS) has emerged as a field where studies try to use many of these facets in their models. Therefore, a convenient approach is to collect information about songs through multiple sources to complement audio-based musical success prediction. The most frequently used data sources in HSS are presented in Figure 4(a). Note that chart organizations, such as Billboard (US) and Official Charts Company (UK), are the primary sources of hit song lists, which are the basis of most studies. Only after this step do authors collect song features (e.g., acoustic, lyrics, and social). Nonetheless, this data source may create a regional bias, as local markets worldwide behave differently by recognizing specific artists and music genres. As the world becomes more globalized, these markets constitute a driving force within the music industry.

Furthermore, Figure 4(b) presents the evolution of data sources in HSS over the years. Over the last few decades, the world has seen a significant change in how people consume music, from physical records to streaming services. This change is also reflected in the research within HSS, as Spotify became the second most used data source in 2020. In addition to acoustic features, this platform offers relevant information, such as global and regional charts (due to its presence in more than 70

[10] Million Song Dataset: http://millionsongdataset.com/
[11] AcousticBrainz: http://acousticbrainz.org/
[12] Last.fm API: http://www.last.fm/api/
[13] Spinn3r: http://docs.spinn3r.com/

countries) and user behavior. Therefore, streaming platforms are becoming a powerful data source for research in HSS and MIR fields.

4. Measuring success

Besides acquiring data (previous section), HSS research also defines a success metric fundamental for predicting success. People usually associate musical success with fame, richness, and power. Such a connection might seem reasonable, but defining and measuring the success of a song can be an abstract process. Understanding music popularity remains a topic of great interest for music-related industries and researchers within the MIR community. Notably, success measures are often summarised in music top charts and used further to understand a song's current and long-term rankings. Also, the success of a song can be exploited as the target variable of prediction models.

Different measures of musical success in the Hit Song Science literature set the criteria required for a song to be perceived as successful. Such definitions evolve across time and culture, making it difficult to propose a unified and objective definition. There is a clear need for a proper classification to unify such interpretations and enable a fair comparison among them. Therefore, we propose a hierarchical taxonomy in Figure 5 that characterizes success measures from different perspectives. The classification is based on an overview of the existing literature and includes three categories: Top-Charts (Section 4.1), Economy (Section 4.2), and Engagement (Section 4.3). Table 1 maps selected works across all three perspectives and measures of success, including hit and non-hit definitions, when available.

4.1. Top-Charts perspective

Musical success has been measured by relying on top-charts provided by radio stations, trade magazines, regional markets, and streaming platforms. A top-chart is the numerical ranking of songs based mainly on retail

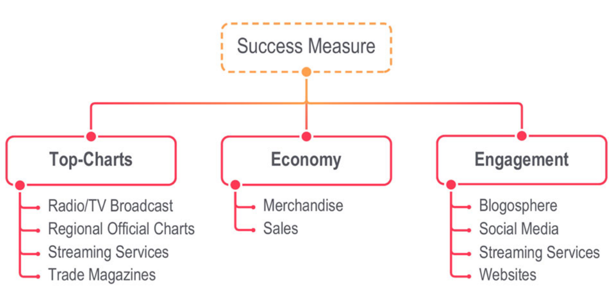


Figure 5. Proposed hierarchical taxonomy for music success measures from three perspectives.

sales (physical and digital), radio and television plays, and online streaming. Moreover, since the early 2000s, many tools have started to track the vast activity taking place online, from streaming to social network actions. As a result, new revenue sources have emerged–such as streaming platforms, digital downloads, and live shows online–updating the definition of a successful song. Still, music charts have become an increasingly reliable musical success metric for both the industry and artists.

From a historical perspective, trade magazines, such as Billboard Magazine, have effectively provided the main history of successful songs. Its first chart was published in 1936, *Hit Parade*, which was a term used to rank popular songs based on data from manual reports filled out by radio stations and stores. A variety of song charts followed, which eventually were consolidated into the *Billboard Hot 100*[14] (best-selling singles) and *Billboard 200*[15] (best-selling albums). All Billboard charts currently combine record sales, radio airplay, digital downloads, and streaming activity. Therefore, there is an undeniable appeal for artists and record labels to be able to predict the path of their songs along with the Billboard charts–artists want to compose hit songs, and labels want to invest in more popular artists.

Even with the advent of changes in the music world (namely, the introduction of digital distribution mechanisms), Billboard remains the most visible chart in the music industry. Indeed, most studies that define success from a Top-Charts perspective consider Billboard charts to be ground truth (about 48%). Each chart summarises popularity statistics that reflect record sales and airplay data of any given week, as well as stores song, artist, and album metadata. Ultimately, such charts are the most common and suitable solution adopted in the HSS literature to assess the quality of predictions.

Specifically, approaches modeling hit prediction as a classification task[16] usually define a song as successful if it is featured in any weekly Billboard chart at least once (Middlebrook & Sheik, 2019; Silva et al., 2022, november 7–11; Singhi & Brown, 2014, 2015; Vötter et al., 2021, 2022). However, in such cases, the definition of *non-hits* or *flops* is far more challenging. With no data available on less popular songs (i.e., official 'flops charts'), there is no consensus on this concept. To address such a challenge, Singhi and Brown (2014) and Silva et al. (2022, november 7–11) consider *flops* as the non-top-charted songs by all singers who have hit songs on the Billboard Year-End Hot 100 chart. A similar solution

[14] Billboard Hot 100: http://www.billboard.com/charts/hot-100
[15] Billboard 200: http://www.billboard.com/charts/billboard-200
[16] Classify songs into non-hits and hits.

Table 1. Success perspectives and measures used in Hit Song Science.

Perspective	Success Measure	Hit definition	Non-hit definition	Reference Alias
Top-Charts	Chart Appearance	Charted songs	Non-charted songs	Dhanaraj and Logan (2005), Interiano et al. (2018), Middlebrook and Sheik (2019), and Zangerle et al. (2019)
Top-Charts	Chart Appearance	Charted songs	*Hit-artists*' non-charted songs	Dewan and Ramaprasad (2014), Singhi and Brown (2015), Vötter et al. (2021), Vötter et al. (2022), and Silva et al. (2022, november 7–11)
Top-Charts	Chart Appearance	Spotify Charts Top 50	Spotify Charts Viral 50	Araujo et al. (2020)
Top-Charts	Chart Appearance	Weekly Top 200 Spotify Chart	N/A	Oliveira et al. (2020)
Top-Charts	Chart Appearance	Top 100 Trending Spotify Chart	N/A	Al-Beitawi et al. (2020)
Top-Charts	Chart Position	Top 1–x | $x \in \{1, 3, 5, 10, 20, 30, 40, 50\}$	Top y-100 | $y \in \{2, 4, 6, 11, 21, 31, 41, 51\}$	Bischoff et al. (2009)
Top-Charts	Chart Position	Top 1–x | $x \in \{5, 10, 20, 30, 40, 50, 100\}$	N/A	Koenigstein et al. (2009)
Top-Charts	Chart Position	Top 1–5	Top 30–40	Ni et al. (2011) and Singhi and Brown (2014)
Top-Charts	Chart Position	Top 1–x | $x \in \{5, 20\}$	Bottom 1–x | $x \in \{5, 20\}$	Fan and Casey (2013)
Top-Charts	Chart Position	Top 1–x | $x \in \{10, 20\}$	Bottom 1–x | $x \in \{10, 20\}$	Herremans et al. (2014)
Top-Charts	Chart Position	Top 1–1	Top 2–100	Nunes and Ordanini (2014)
Top-Charts	Chart Position	Top 1–15	N/A	Buda and Jarynowski (2015)
Top-Charts	Chart Position	Top 1–20	Under chart 1–20	Herremans and Bergmans (2017)
Top-Charts	Chart Position	Top 1–10 | Top 11–100 | Top 101–200	Non-charted albums	Cosimato et al. (2019)
Top-Charts	Chart Position	Top 1–20	Bottom 10	Raza and Nanath (2020)
Top-Charts	Chart Position	Top 1–10	Top 11–100	S. T. Kim and Oh (2021)
Top-Charts	Rank Score$^{\mathbf{R}}$	N/A	N/A	Abel et al. (2010), Y. Kim et al. (2014), Lee and Lee (2015), Lee and Lee (2018), and Tsiara and Tjortjis (2020)
Top-Charts	Chart History	N/A	N/A	Pachet and Roy (2008), Shin and Park (2018), Araujo et al. (2019), and Silva and Moro (2019)
Top-Charts	Peak position	N/A	N/A	Chon et al. (2006), Frieler et al. (2015), and Askin and Mauskapf (2017)
Top-Charts	Weeks on chart	N/A	N/A	Chon et al. (2006), Frieler et al. (2015), Askin and Mauskapf (2017), and Ren and Kauffman (2017)
Top-Charts	Time2TopRank	N/A	N/A	Ren and Kauffman (2017)
Economy	Billboard Units	N/A	N/A	Araujo et al. (2017)
Economy	Record Sales	N/A	N/A	Abel et al. (2010) and Berns and Moore (2012)
Engagement	Listening Logs	N/A	N/A	Ren et al. (2016) and Shulman et al. (2016)
Engagement	Artists' Number of Followers	N/A	N/A	Silva et al. (2019)
Engagement	Number of Views	N/A	N/A	Chiru and Popescu (2017)
Engagement	Number of Likes	N/A	N/A	Chiru and Popescu (2017)
Engagement	Number of Plays	N/A	N/A	Yu et al. (2019)
Engagement	Market Share of Downloads$^{\mathbf{M}}$	N/A	N/A	Salganik et al. (2006)
Engagement	Hit Score$^{\mathbf{H}}$	N/A	N/A	Yang et al. (2017) and Vötter et al. (2021)
Engagement	Spotify Popularity Score	N/A	N/A	Araujo et al. (2017), Martín-Gutiérrez et al. (2020), Matsumoto et al. (2020), Silva et al. (2019), Kaneria et al. (2021), Kamal et al. (2021), and Gao (2021)
Engagement	Number of Streams	$num\_streams \geq 2M$	$num\_streams < 2M$	Febirautami et al. (2018)

$^{\mathbf{R}}$**Rank Score**: $rank\_score(i) = max\_rank - rank(i) + 1$ $^{\mathbf{M}}$**Market Share of Downloads**: $m_i = d_i / \sum_{k=1}^{S} d_k$
$^{\mathbf{H}}$**Hit Score**: $hit\_score(i) = log(playcount(i)) \times log(num\_users(i))$

is proposed by Singhi and Brown (2015), who experimentally evaluated four different definitions of *flops*, ranging from a broad to a very narrow perspective. Alternatively, Middlebrook and Sheik (2019) merge a Billboard dataset into a set of Spotify collected songs, thus setting the tracks that never appeared on the chart as non-hits.

Another common strategy in classification models is assuming a track is popular if it exceeds a certain *popularity threshold* (Martín-Gutiérrez et al., 2020). Most studies train the classifiers on several rank ranges (Bischoff et al., 2009; Cosimato et al., 2019; Herremans et al., 2014; Y. Kim et al., 2014; Koenigstein et al., 2009; Rajyashree et al., 2018; Raza & Nanath, 2020; Silva et al., 2022, november 7–11; Tsiara & Tjortjis, 2020), considering hit songs as those tracks that have reached a predefined peak chart position. For non-hits, in general, randomly select about the same number of hit songs from the set of music tracks with Billboard positions greater than the threshold established. For instance, when the rank 1–10 is the interval of a hit, non-hit songs are randomly selected from the 11–100 rank positions. The median value of the ranking has also been explored as a popularity boundary. Alternatively, Lee and Lee (2015, 2018) formulate the prediction problem as binary classification, i.e., high vs. low values of multiple popularity metrics extracted from Billboard charts, where the median value of each pattern sets the boundary of the two classes (*hit* and *non-hit*).

Some researchers have also used the charts' statistics as success measures, including the highest position on the chart in any week of a year (peak position) and the number of weeks the track has been or was on the chart (weeks) (Askin & Mauskapf, 2017; Chon et al., 2006). Such an approach is more common when a regression (or ranking) task is considered, where the proposed model learns to predict numeric scores. In contrast, most previous works use the Billboard rank score as a continuous output variable (Y. Kim et al., 2014; Koenigstein et al., 2009; Silva & Moro, 2019; Zangerle et al., 2019). On the other hand, Chon et al. (2006) aimed to predict an album's future based on two concepts: *lifecycle* and *lifespan*. The *lifecycle* of an album is a trajectory of its weekly positions in a chart, while *lifespan* is defined to be how long the lifecycle of an album is in terms of the number of weeks. Following a different path, Nunes and Ordanini (2014) assess relative success by comparing Billboard's Hot 100 songs that reached the #1 position.

Music charts based on physical sales, digital downloads and streaming activity are not exclusively present in the United States. Since 1952, the United Kingdom has released the UK Singles Chart listing the top-selling singles in the country by the New Musical Express magazine.[17] Nowadays, the chart is called the Official Singles Chart,[18] which is compiled by the Official Charts Company (OCC), listing the top-selling singles in the UK. As a reliable alternative source, the Official Singles Chart has also been explored from different perspectives, including considering different rank ranges (Fan & Casey, 2013; Herremans et al., 2014; Ni et al., 2011) and simply defining songs' success as making it' into the charts (Interiano et al., 2018). Finally, other sources of music charts featured in the literature review include Spotify Charts (Araujo et al., 2020; Oliveira et al., 2020), the Pandora effort (Pachet, 2011; Pachet & Roy, 2008), MixRadio charts (Rosati et al., 2021), regional charts (Buda & Jarynowski, 2015; Fan & Casey, 2013; Herremans & Bergmans, 2017; Shin & Park, 2018; Yoo & Kim, 2010), radio broadcast (Dhanaraj & Logan, 2005), and streaming/website charts (Al-Beitawi et al., 2020; Araujo et al., 2019; Ren & Kauffman, 2017).

4.2. Economy perspective

Economy indicators are also useful as a conventional and quantitative measure of musical success. Such economy gauges can be quantified regarding profits, revenues, or dividends. Fisher et al. (2010) state that totals revenues can be broken down into two main sources: *performance fees* and *recorded music*, whether in physical or digital format. However, along with the digital revolution, various revenue earnings have emerged. The International Federation of the Phonographic Industry (IFPI),[19] a non-profit members' organization, provides an annual global recorded music report based on five segments: *physical*, *digital* (excluding streaming), *streaming*, *synchronization revenues* (revenue from the use of music in advertising, film, gaming, and TV) and *performance rights* (use of recorded music by broadcasters and public venues).

Other sources of information on recorded music sales include Amazon Sales Rank (Abel et al., 2010; Dewan & Ramaprasad, 2014), Nielsen SoundScan (Berns & Moore, 2012; Dewan & Ramaprasad, 2014), and Billboard units (Araujo et al., 2017). The Amazon Best Sellers Rank (BSR),[20] also known as the 'Amazon Sales Rank', is a score that Amazon assigns to a specific product based on historical sales data. It has been a popular tool over the years, as Amazon is one of the largest online CD retailers. Abel et al. (2010) measure the music sales performance by using such a rank, aiming to capture a product's popularity compared with others in its category.

Nielsen SoundScan[21] has also been a source for album sales data as an information system that tracks music sales and music video products throughout the United States and Canada. Nielsen's sales data is compiled weekly from over 39,000 global retail outlets and has been a trusted and vital resource for companies aiming to have a complete picture of music sales for over two decades. Moreover, Nielsen's data serves as the primary sales source for the Billboard music charts, making it the largest source of sales records in the music industry. Most studies use such data sources as post-factum popularity information over songs (Berns & Moore, 2012) and albums (Dewan & Ramaprasad, 2014). As an alternative, Billboard album-equivalent units can also be collected as a data source from an economy perspective. Specifically, the album equivalent is a measurement unit that defines the consumption of music that equals the purchase of one album copy. This consumption includes streaming and song downloads plus traditional album sales. Araujo et al. (2017) manually collected the weekly Billboard Top 10 record chart to gather sales units during the album's release date.

[17] New Musical Express magazine: http://www.nme.com/
[18] Official Singles Chart: http://www.officialcharts.com/charts/singles-chart
[19] International Federation of the Phonographic Industry (IFPI): http://www.ifpi.org
[20] The Amazon Best Sellers Rank: http://www.amazon.com/Best-Sellers/zgbs
[21] Nielsen SoundScan: http://www.nielsen.com

4.3. Engagement perspective

Music has always contained a social dimension, whether shaped through the engagement of artists or listeners. Such an aspect affects how people consume and engage with music and can impact musical performance success. In particular, digital applications have become a powerful tool when discussing and measuring success by offering ways to share information about music–and to share the music itself. Different social media services (like Facebook, YouTube, and Twitter) are designed to attract audiences and encourage them to discover new artists, share recommendations and consume music. Moreover, most music streaming platforms, such as Last.fm and Spotify, integrate music listening and social interactions into a single service. Therefore, social engagement metrics can be a valuable tool for measuring music success.

There are several metrics over which social media and streaming platforms can assess success, including views (Chiru & Popescu, 2017) and likes (Chiru & Popescu, 2017; Shulman et al., 2016) on social media; or digital downloads (Salganik et al., 2006), ratings (Rajyashree et al., 2018), number of streams (Araujo et al., 2017; Febirautami et al., 2018; Gao, 2021; Kamal et al., 2021; Kaneria et al., 2021; Martín-Gutiérrez et al., 2020; Matsumoto et al., 2020; Silva et al., 2019), and play counts (Dewan & Ramaprasad, 2014; Ren et al., 2016; Shulman et al., 2016; Vötter et al., 2021, 2022; Yang et al., 2017; Yu et al., 2019) on streaming/digital platforms. As depicted in Figure 4, the most commonly used digital application in HSS research is Spotify, followed by The Echo Nest platform. The former is the most popular global audio streaming subscription service today, which manages and shares over 50 million tracks.[22] The latter is a music intelligence and data platform for developers and media companies, acquired by Spotify in March 2014.

Then, in March 2016, The Echo Nest API was shut down, and developers were encouraged to move to the Spotify API instead. The Echo Nest API provided a buzz-measuring score, called *hotttnesss*, derived from mentions on the web, mentions on music blogs, music reviews, play counts, and others. A similar score is available on Spotify Web API, called *popularity*, which is based on the total number of plays compared to other tracks and how recent those plays are. Although Spotify uses several algorithms to determine popularity, the more a song is played, the higher its score. The play count can be comparable to digital downloads within the scenario of streaming services. In other words, the popularity rating can translate valuable and helpful information about musical success. Most studies have used such scores as the label or response variable in learning models for predicting popularity (Araujo et al., 2017; Febirautami et al., 2018; Gao, 2021; Kamal et al., 2021; Kaneria et al., 2021; Martín-Gutiérrez et al., 2020). Differently, Silva et al. (2019) consider not only the popularity score but also the number of followers to define a successful artist.

Authors have also explored play counts using different platforms as a success measure. Examples include the music streaming service KKBOX,[23] which mainly targets the music market of Southeast Asia. It features over 50 million legal tracks and is currently available in Taiwan, Hong Kong, Japan, Singapore, and Malaysia, with over 10 million users. Interested in distinguishing hits and non-hits, Yang et al. (2017) consider playing counts of a song from the KKBOX streaming service to define song success. Likewise, Yu et al. (2019) use the number of plays of one artist's songs as the target variable. However, in this case, the dataset was collected from a competition organized by the Alibaba Group[24] in China.

Besides streaming platforms, a popular music data service is the website Last.fm.[25] Founded in the United Kingdom in 2002, it has been used as a music recommender system, where a detailed profile of each user's musical taste is available. As a result, it is a valuable source of information about the listeners' behaviour. Dewan and Ramaprasad (2014) and Ren et al. (2016) characterize song success based on Last.fm listeners, which are highly correlated with Amazon Sales rank. Similarly, Shulman et al. (2016) define song success based on the number of people who have listened to or loved a song, almost as a sort of consumption measure. Another famous website is YouTube,[26] which offers a wide variety of user-generated and corporate media videos and encourages user engagement within the platform. Metrics like the number of views and like votes can be a valid measure of popularity (Chiru & Popescu, 2017).

4.4. Overall considerations

In summary, the success of a song can be described from distinct perspectives, including how many times a song has reached a top-chart, has been played on video clips, has been consumed via streaming, and so on. Essentially, existing measures of success fall into three categories: *Top-Charts*, *Economy*, and *Engagement*, representing the different points of view. Researchers traditionally seek to explain the phenomena of artistic success based on accumulated wealth, prestige, and notoriety. However, with

[22] Spotify Company Info: http://newsroom.spotify.com/company-info/

[23] KKBOX: http://www.kkbox.com
[24] Alibaba Group: http://www.alibabagroup.com
[25] Last.fm: http://www.last.fm
[26] YouTube: http://www.youtube.com/

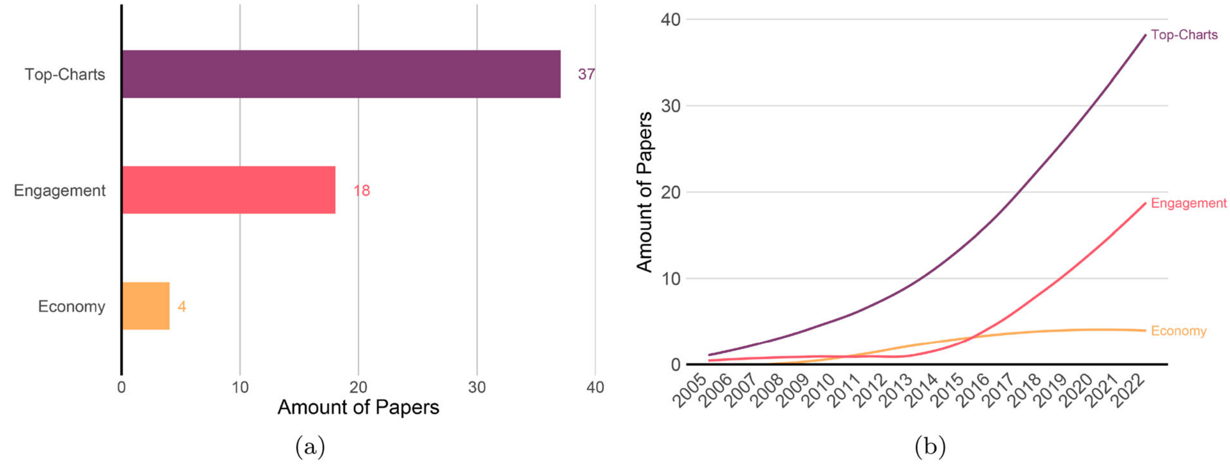


Figure 6. Most common success perspectives in Hit Song Science (a) and their evolution over the years (b).

the recent changes in the music industry, there are new markets and possibilities to measure artists' attractiveness and fans' engagement.

Even after such changes, the *Top-Charts* perspective remains the most accepted and, perhaps, the most reliable, as shown in Figure 6(a). With almost 63% (37 of 59) of the considered studies, the music charts became the primary source of tools measuring music success, alongside a booming music industry (Figure 6(b)). On the other hand, with the introduction of digital distribution mechanisms, social media and streaming platforms have created new ways of measuring artistic success, mainly about listeners' engagement. This phenomenon becomes clear in Figure 6(b), with the curve of the Engagement perspective increasing in the 21st century.

Such results may indicate the definition of success in HSS studies depends mainly on the data available. Likewise, easy access to data on the web can be the main reason behind the vast majority of studies using the definition of success from top-charts, social media, and streaming platforms. Although the *Economy* perspective is supposed to be a strong indication of artistic success, album sales data availability is limited. Information such as profits, revenues, or dividends is generally not publicly available. The data sources on album sales are generally in the form of rankings without explicitly disclosing the sales figures. Therefore, to correctly measure artistic success, a complete and easily accessible source of data encompassing all three perspectives is fundamental.

5. Features

The success of a song is usually associated with a collection of factors related to the musical scenario. For example, recent studies show that characteristics including high *danceability* and low *instrumentalness* increase the popularity of songs. In other words, such songs tend to be more exciting, as a danceable music structure puts the audience in a good mood (Al-Beitawi et al., 2020). Besides explicit features, recent research also considers the strength of artist collaborations in producing hit songs, expanding research on musical success to another level (Silva et al., 2022, november 7–11, 2019). Such perspectives use a large set of features (e.g., danceability, instrumentalness, etc.) in the HSS context, which are the basis of hit song prediction models. However, there is no unique set of features for a successful model.

In this section, we propose and describe a novel taxonomy for the most frequently used features in HSS, presented in Figure 7. Here, we can divide such descriptors into two main groups according to their relation

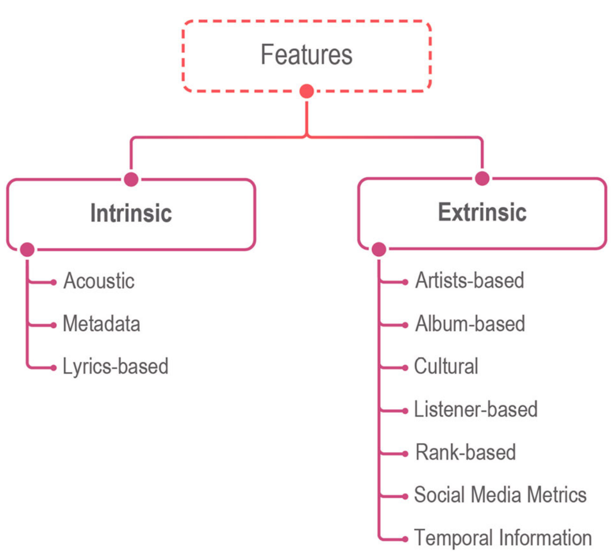


Figure 7. Proposed hierarchical taxonomy for musical features. Intrinsic features are directly extracted from the song, while extrinsic ones are related to other objects and agents that influence the success of a song.

with the song itself, which is the main object of this field. First, the *Intrinsic* features (Section 5.1) are those directly extracted from the audio (i.e. acoustic fingerprints, lyrics) and information such as genre, duration, and the number of artists. The second group includes all features related to *Extrinsic* agents or objects that may influence directly or indirectly the musical success, i.e. artist popularity, album sales, and the number of streams for the considered song (Section 5.2).

5.1. Intrinsic features

Here, we present the intrinsic features of music. Section 5.1.1 presents the Acoustic Features, which rely solely on musical data extracted from the audio (including different aspects of audio properties). Section 5.1.2 presents the Metadata features, which include information such as title, duration, genres as well the participant singers. Finally, we discuss features based on song lyrics in Section 5.1.3.

5.1.1. Acoustic features

Despite different definitions of music, we can certainly assume music is an art form whose medium is sound. Moreover, it is composed of numerous core elements, including pitch (melody and harmony), rhythm, dynamics, and the qualities of timbre and texture. However, due to the countless existing styles of music, some of these elements can be emphasized, diminished, or omitted.

There are different taxonomies for categorizing elements of music. One type of taxonomy divides musical descriptors into three dimensions: timbre, pitch, and rhythm (Scaringella et al., 2006). Alternatively, hierarchical taxonomies go beyond such features and consider other categories such as symbolic- and audio-based features (Corrêa & Rodrigues, 2016; Fu et al., 2011). However, many works use acoustic features that can be easily collected from sources such as Spotify or EchoNest. Hence, Table 2 summarises the most used acoustic features followed by a brief description (mainly extracted from Spotify[27]) and the works that use such features.

In the face of such diversity of elements, several works differ in the set of features used to classify hit songs. For instance, Dhanaraj and Logan (2005) extract features from each song describing its main sounds. In particular, the authors characterize sounds using Mel-frequency Cepstrum Coefficients (MFCC), features focussing on music timbrel aspects. Using a different source, Pachet and Roy (2008) consider a set of 49 audio features taken from MPEG-7 audio standard, including spectral characteristics (Spectral Centroid, Kurtosis, and Skewness, HFC, Mel Frequency Cepstrum Coefficients), temporal (ZCR, Inter-Quartile-Range), and harmonic (Chroma) features. Such features were selected given their generality, i.e. they do not contain specific musical information or musically ad hoc algorithms. Lee and Lee (2018) follow a similar path by taking 82 MPEG-7 features from each considered sound, besides the complexity features (Chroma, Rhythm, Timbre, and Arousal) and MFCC features.

Other works extract their musical features from The Echo Nest platform as song descriptors, including *tempo, time signature, song duration, loudness* (Ni et al., 2011), *mode, danceability, energy, key* (Herremans et al., 2014), *liveness*, and *speechiness* (Askin & Mauskapf, 2017; Fan & Casey, 2013; Herremans & Bergmans, 2017). Overall, such features match those used to represent a song globally. In addition to the set of Echo Nest features, Ni et al. (2011) compute detailed summaries of the songs such as the *Coefficient of Variance of Loudness* and *Harmonic Simplicity*. Likewise, Herremans et al. (2014) include two additional features to incorporate a temporal aspect: *Timbre* and *Beatdiff* . In contrast, Singhi and Brown (2015) consider not only *danceability*, *loudness*, *energy, mode, tempo*, but also *mean, median*, and *standard deviation* of *timbre, pitch*, and *beat duration* vectors.

Although researchers continue to use such musical attributes, the data source has shifted. All acoustic features are now available in the Spotify API since March 2014, when Spotify acquired the Echo Nest. The main features used in recent works are *acousticness, danceability, energy, instrumentalness, key, liveness, duration, mode, speechiness, tempo, time signature,* and *valence* (Al-Beitawi et al., 2020; Araujo et al., 2020; Febirautami et al., 2018; Gao, 2021; Kamal et al., 2021; Kaneria et al., 2021; S. T. Kim & Oh, 2021; Matsumoto et al., 2020; Middlebrook & Sheik, 2019; Raza & Nanath, 2020; Silva et al., 2022, november 7–11; Vötter et al., 2021, 2022). In addition to using high-level features from the Spotify API, other works include low-level features such as MFCCs (Martín-Gutiérrez et al., 2020), the *Tonnetz* (Harte et al., 2006), the *Chromagram* (vector of twelve elements indicating how much energy is released by each class tone) (Shepard, 1964), the *Octave-based Spectral Contrast* (considers the spectral peak, the spectral valley and its difference in each sub-band) (Dan-Ning et al., 2002), the *Spectral Centroid* (frequency indicator for the energy where the spectrum is centered) (Tzanetakis & Cook, 2002), *Spectral Bandwidth* (Martín-Gutiérrez et al., 2020) and *Zero Crossing Rate* (ZCR) (Araujo et al., 2019).

In a different perspective, Ren et al. (2016) extract only *tempo* (fast, moderate, slow) and *melody* (pitch, rhythm)

[27] Spotify API Reference: http://developer.spotify.com/

Table 2. The most used acoustic features, their brief explanation and corresponding works.

Feature – Description	Reference Alias
acousticness – A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.	Askin and Mauskapf (2017), Febirautami et al. (2018), Middlebrook and Sheik (2019), Araujo et al. (2020), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
danceability – Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is the least danceable and 1.0 is the most danceable.	Fan and Casey (2013), Herremans et al. (2014), Askin and Mauskapf (2017), Interiano et al. (2018), Febirautami et al. (2018), Zangerle et al. (2019), Middlebrook and Sheik (2019), Araujo et al. (2020), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
energy – Energy varies from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud and noisy; e.g. death metal has high energy, and a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.	Fan and Casey (2013), Herremans et al. (2014), Febirautami et al. (2018), Middlebrook and Sheik (2019), Martín-Gutiérrez et al. (2020), Araujo et al. (2020), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
instrumentalness – This feature predicts whether a track contains no vocals. 'Ooh' and 'aah' sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly 'vocal'. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.	Ren and Kauffman (2017), Interiano et al. (2018), Zangerle et al. (2019), Middlebrook and Sheik (2019), Araujo et al. (2020), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
key – The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation. E.g. $0 = C$, $1 = C\sharp/D\flat$, $2 = D$, and so on. If no key is detected, it is -1.	Fan and Casey (2013), Herremans et al. (2014), Askin and Mauskapf (2017), Zangerle et al. (2019), Middlebrook and Sheik (2019), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
liveness – Liveness detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.	Fan and Casey (2013), Middlebrook and Sheik (2019), Araujo et al. (2020), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
loudness – The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing the relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.	Ni et al. (2011), Fan and Casey (2013), Herremans et al. (2014), Zangerle et al. (2019), Middlebrook and Sheik (2019), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
mode – Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.	Fan and Casey (2013), Y. Kim et al. (2014), Herremans et al. (2014), Ren et al. (2016), Askin and Mauskapf (2017), Middlebrook and Sheik (2019), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
pitch – Pitch is a subject feature, which is determined (perceived) by what the ear judges to be the most fundamental frequency of the sound.	Frieler et al. (2015), Ren et al. (2016), Rajyashree et al. (2018), and Zangerle et al. (2019)
rhythm – As a recurring pattern of tension and release in music, it describes how certain patterns occur and recur in the music and is related to the 'danceability' of the music. Beat and tempo (beat-per-minute, BPM) are two important cues that describe the rhythmic content of the music which have been utilised in music classification.	Lee and Lee (2015), Ren et al. (2016), Lee and Lee (2018), Interiano et al. (2018), Rajyashree et al. (2018), Zangerle et al. (2019), and Middlebrook and Sheik (2019)
song duration – The duration of a track in seconds as precisely computed by the audio decoder.	Ni et al. (2011), Fan and Casey (2013), Herremans et al. (2014), Frieler et al. (2015), Ren and Kauffman (2017), Araujo et al. (2019), Middlebrook and Sheik (2019), Matsumoto et al. (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
speechiness – Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audiobook, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.	Fan and Casey (2013), Middlebrook and Sheik (2019), Araujo et al. (2020), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
tempo – This is the overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.	Ni et al. (2011), Fan and Casey (2013), Ren et al. (2016), Middlebrook and Sheik (2019), Matsumoto et al. (2020), Raza and Nanath (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
timbre – Timbre is the quality of a musical note or sound that distinguishes types of musical instruments, or voices. Timbre vectors are best used in comparison with each other.	Herremans et al. (2014), Lee and Lee (2015), Lee and Lee (2018), Interiano et al. (2018), and Martín-Gutiérrez et al. (2020)
time signature – It is an estimated overall time signature of a track. The time signature (meter) is a notational convention to specify the number of bits per bar (or measure).	Ni et al. (2011), Fan and Casey (2013), Herremans et al. (2014), Middlebrook and Sheik (2019), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)
valence – With values from 0.0 to 1.0, this metric describes the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).	Askin and Mauskapf (2017), Middlebrook and Sheik (2019), Araujo et al. (2020), Matsumoto et al. (2020), S. T. Kim and Oh (2021), Kaneria et al. (2021), Kamal et al. (2021), Gao (2021), and Silva et al. (2022, november 7–11)

as acoustic features. Lee and Lee (2015) extract features from audio signals in viewpoints of musical aspects, namely *chroma, rhythm*, and *timbre*. In another approach, Ren and Kauffman (2017) describe the musical construct vector (MCV), with Theme, Mood, Instrumental, and Genre, which reflect how acoustic content is perceived. Such high-level semantics are extracted from lower-level musical features, such as timbre, rhythm, and tempo, using machine-based methods.

Using yet other data sources, Interiano et al. (2018) gather several features, including *timbre, tonality, danceability, voice, gender* (male/female), and *mood* from AcousticBrainz. Finally, Zangerle et al. (2019) extract high and low-level features using Essentia's pre-compiled extractors, which provide a variety of spectral, time-domain, rhythm, and tonal descriptors. It provides around 40 basic features (e.g., MFCCs, dissonance or silence rate), 11 rhythm features (e.g., beats per minute or onset-rate), and 13 tonal features (e.g., key or harmonic pitch class profiles) that serve as low-level input for our task. The high-level features include musical genre, mood, timbre, vocals/voice, or danceability.

As an alternative, some researchers opt to use tools to extract features from audio files. A popular example is the *librosa* python package (McFee et al., 2015) for music and audio analysis that includes features as *tempo* (beats every moment) (Araujo et al., 2019), MFCC (impersonates a few sections of the human discourse generation and discourse discernment), and the *Consonant Element* (the symphonic segment inside a sound flag) (Rajyashree et al., 2018), or even the *Spectral Centroid*, the *Spectral Flatness*, and *Zero Crossings*. Another useful tool for audio feature extraction is the MeloSpySuite (Frieler et al., 2013) software, which includes a set of stand-alone *commandline* tools for extracting numerical or textual characteristics of melodies. Chiru and Popescu (2017) apply the Discrete Fourier Transform (DFT) signal in the WAV file to take the sound intensity chart in the frequency function. Since the magnitude represents the sound intensity, they extract for every second of every song the frequency of the highest magnitude, i.e. the largest weight of all sound frequencies that are heard in a second.

5.1.2. Metadata

In the music context, metadata is descriptive information about a song. It is often used for discovery and identification and is, therefore, one of the core elements in MIR research. This type of feature usually includes basic information such as title, author, genres, and so on Abel et al. (2010), Frieler et al. (2015), Askin and Mauskapf (2017), Ren and Kauffman (2017), Shin and Park (2018), Interiano et al. (2018), Zangerle et al. (2019), Middlebrook and Sheik (2019), Oliveira et al. (2020), Kamal et al. (2021), Rosati et al. (2021), Gao (2021), Vötter et al. (2021), Silva et al. (2022, november 7–11), and Vötter et al. (2022). However, some descriptive information about the song may often not be directly related to the audio signal itself. For example, artist location, artist familiarity, artist hotness, song hotness (Herremans et al., 2014), lists of song tags (Bischoff et al., 2009), song type (Solo, Group, or Collaboration) (Shin & Park, 2018), explicity (whether a track has explicit content) (Araujo et al., 2020; Middlebrook & Sheik, 2019), and available markets for release (Martín-Gutiérrez et al., 2020) may summarise musical aspects as well.

Although metadata is traditionally used to provide digital identification, in the HSS context, a useful purpose is to help find relevant information and discover musical resources. Even though most meta-information is generally discarded when building a prediction model, some researchers have been seeking to assess hit song predictors through metadata. Pachet and Roy (2008) use a music and metadata database provided by the HiFind Company. The HiFind metadata are grouped in 16 categories, representing specific dimensions of music including style, genre, musical setup, main instrument, country, situation, mood, character and language. By using a different database, Nunes and Ordanini (2014) also explore the instrumentation. Specifically, the authors consider the number of instrument types audible for each song as the principal independent variable.

5.1.3. Lyrics-based features

Lyrics form an integral musical component and can help solve complex MIR tasks. The lyrics of a song contain specific emotional content and have more power to change mood than audio features alone (Singhi & Brown, 2015). However, lyrics are considerably ignored in the overall MIR research compared to acoustic features, although lyric-based features remain a useful predictor widely used in the Hit Song Prediction task. In particular, lyrics have been considered a significant component of what makes a song a hit. For example, Dhanaraj and Logan (2005) extract descriptive features based on the semantic content of songs from lyrics using a Probabilistic Latent Semantic Analysis (PLSA) method (Hofmann, 1999), which is an effective way to assess the similarity between songs based on lyrics (Logan et al., 2004). In particular, each song is converted to a vector representing the likelihood that a song is about a pre-learned topic.

Further studies explore other lyrics-based features. Singhi and Brown (2014) propose a novel hit song detection model using lyric features alone. Specifically, the authors consider a complete set of 24 rhyme and syllable features of the Rhyme Analyzer (Hirjee & Brown, 2010), including syllables per line, rhymes per line, and links per line. The CMU Pronunciation Dictionary (Elovitz et al., 1976) was additionally used to transcribe plain lyrics of songs to a sequence of phonemes with indicated stress, resulting in seven new metric features. The authors found the quality of hit prediction improves as lyric length increases. Moreover, Singhi and Brown (2015) also show rhyme, meter and lyrics matter to hit detection, and complexity is related to being a hit.

Along with the advent of AI and machine learning, researchers have explored more advanced and powerful techniques to extract lyric-based features. Ren et al. (2016) apply Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to learn five topic distributions from lyrics. Likewise, Ren and Kauffman (2017) use LDA to build a topic model to learn the semantic themes from a dataset of 4410 tracks. To such a task, they complement the acoustic content and give the artist's meaning behind the music. The results found around 65% of the tracks were about 'love' and 'life'. In a bag-of-words fashion, Chiru and Popescu (2017) extract words from lyrics along with their frequencies of appearance. The authors claim that lyrics are the most useful features in identifying whether a song will be successful or not.

Finally, to assess multimodal learning, Martín-Gutiérrez et al. (2020) consider a collection of features evoked from different modalities, including text, audio, and meta-data. Regarding the text modality, a set of descriptors is extracted from the corpus of the song lyrics. Using NLP techniques, a stylometric analysis results in the following features: the total number of sentences, the average number of words per sentence, the total number of words, the average number of syllables per word, a sentence similarity coefficient, and a vocabulary wealth coefficient. Sentiment analysis on song lyrics has also been applied to predict the nature of hit songs (Raza & Nanath, 2020). Recently, Kamal et al. (2021) have performed a sentiment analysis to obtain a value between -1 to $+1$, representing the polarity of the lyrics (-1 corresponds to negative, 0 to neutral, and $+1$ to positive lyrics). They observed most popular songs have a neutral sentiment, and there are more popular songs with positive sentiments than negative sentiments.

5.2. Extrinsic features

In this section, we present the extrinsic perspectives of the songs, which model the musical ecosystem by incorporating social media, market data, and so on. In summary, Section 5.2.1 goes over information about artists; Section 5.2.2 overviews the features extracted from albums; Section 5.2.3 brings up discussions about cultural aspects of songs; Section 5.2.4 regards the listeners as an important resource in HSS; Section 5.2.5 overviews features based on chart rankings; Section 5.2.6 reveals the importance of social networks to HSS; and finally, Section 5.2.7 discusses the temporal features of musical success.

5.2.1. Artist-based features

Some features are not directly related to the music structure, such as those based on artists. Among the characteristics analyzed by several works (Matsumoto et al., 2020; Middlebrook & Sheik, 2019; Nunes & Ordanini, 2014; Oliveira et al., 2020; Ren & Kauffman, 2017; Ren et al., 2016; Silva et al., 2022, november 7–11), we highlight the following: (i) basic information (e.g., the artist's title/id) (Middlebrook & Sheik, 2019); (ii) demographic data (e.g., age, race, gender and nationality) (Nunes & Ordanini, 2014); (iii) the type of artist (e.g., whether solo, group or collaboration) (Shin & Park, 2018); and (iv) the artists' awards are associated with big record labels (Ren et al., 2016).

Other works use artist popularity to help predict hit songs (Interiano et al., 2018). However, the set of popularity features varies among studies. One possibility is to consider information extracted from platforms such as Last.fm, which usually provides a set of five tags from each artist in the database previously labeled by the platform users (Ren & Kauffman, 2017; Shulman et al., 2016). Other features can be included, such as the number of tags assigned to an artist, the number of listeners, the best position ever achieved in the Billboard charts (Bischoff et al., 2009), and the number of followers of the artist or a popularity score (Martín-Gutiérrez et al., 2020).

Regarding artists' popularity, other studies apply network science metrics to a success-based artist network (Matsumoto et al., 2020; Oliveira et al., 2020). For example, Silva et al. (2022, november 7–11) and Silva et al. (2019) consider well-known metrics dependent on the node (Clustering Coefficient, Eigenvector, Degree, and Weighted Degree) and related to the whole graph (Closeness, Eccentricity, and Betweenness). To deepen such analyses, Silva and Moro (2019) also consider the popularity over time of the artists from the *rank_score*[28] extracted from Billboard charts. From a multi-perspective approach, Silva et al. (2022, november 7–11) explore features within three musical context factors: *song*, *album* and *artist*, where the latter captures artists' collaboration profile quantitatively and qualitatively, as well as the number of genres and albums the artist owns.

[28] The *rank_score* is the inverted rating on a success chart.

Beyond such charts, Ren and Kauffman (2017) measure artist reputation and leverage information on the news on the Grammy, American, and Billboard awards. They also consider relevant artist record labels because major labels have more resources to produce and promote high-quality tracks. Similar work by Askin and Mauskapf (2017) adds a dummy variable if a song was released on a major record label. However, they also include a set of dummy variables in each model to account for the number of songs an artist had previously placed on the charts. These previous song count dummies capture artists' relative visibility or popularity at the moment of a song's release. Finally, they also construct a variable called 'multiple memberships' to account for artists who have released songs under different names or band compositions.

5.2.2. Album-based features

Similar to artist-based features, albums also have useful descriptive information. Indeed, additional information related not only to the social environment but also the marketing strategies may be considered. For instance, in the music industry, a music release should be an essential component of any music promotion strategy (Martín-Gutiérrez et al., 2020). Depending on the release format, a song can reach the top of the charts more quickly. While albums and EPs are suitable for attracting attention and building a fanbase, singles promote the album and keep fans engaged.

Listeners today have particular behaviors regarding music consumption, so artists need to adapt their strategies to be successful. Hence, to fit such novel listening habits, labels opt to release a steady stream of singles to generate momentum and enthusiasm. In other words, the album type can act as a significant predictor. Middlebrook and Sheik (2019) consider, as predictors, album-based information extracted from Spotify, in addition to track, artist, and audio features. In particular, the authors explore both *album_type* (album, single, or compilation) and *album_release_date* (the date the album was first release) features. Likewise, Silva et al. (2022, november 7–11) consider the *album_type* as predictor, as well as the albums' total number of tracks.

Also, based on the marketing strategies of the artists' careers, Bischoff et al. (2009) rely on a list of assumptions for their music hit prediction algorithm. One hypothesis is that previous album of the same artist directly influence the songs' future success. Thus, an album's popularity is measured by the highest position reached on Billboard, named as *peak position*. Moreover, the authors only include the top-5 albums that reached positions in the charts since some artists have many previously released albums.

In HSS studies, the focus is usually on determining whether one can predict the success of a song. However, some researchers have engaged in predicting albums' success (Araujo et al., 2017; Chon et al., 2006; Dewan & Ramaprasad, 2014). Going further, Chon et al. (2006) try to predict an album's future. In particular, the authors investigate not only how long an album will stay on a top-chart (i.e. the album's lifespan) but also whether an album's position can be predicted. To do so, they develop an algorithm to calculate Euclidean distances between the first weeks' sales history of a new album and the same number of weeks from the average life-cycle patterns and determine the expected lifespan with the minimum distance.

5.2.3. Cultural features

From a cultural perspective, we can identify a specific feature used in HSS. Specifically, Buda and Jarynowski (2015) investigate distances between European countries in a Cartesian/Euclidean two-dimensional way. Although the most typical representation is geographical distances, other matrices of distances are possible. For instance, the cultural map of Europe was built by political scientists based on the World Values Survey. Hence, two dominant dimensions were chosen (explaining 70% of variations between countries): traditional values versus secular-rational values on the vertical y-axis and survival versus self-expression values on the horizontal x-axis.

Although no other study uses cultural features in its models, recent research specialized in cuts of data for selecting different markets. For example, Fan and Casey (2013) find significant differences in hit song prediction by comparing Chinese and British charts. In addition, Oliveira et al. (2020) use data from Spotify to build genre networks from eight countries besides the global scenario. Thus, it is possible to identify the importance of local aspects (mapped into regional genres) in defining the hits for each market.

5.2.4. Listener-based features

The listeners who are music lovers can contribute a lot to an artist's success, and they usually make up the artist's fan base. In this category of features, the works seek to use the listeners' characteristics in the hit song prediction algorithms in several ways. For example, Pachet and Roy (2008) use a set of 632 labels on the artists' popularity, and listeners manually interpret such labels. On the other hand, Yang et al. (2017) collect a set of Taiwanese listener records over one year in collaboration with KKBOX Inc. The final dataset gathered the play counts of 30 K users for around 125 K songs.

A trend-line proposed by Shulman et al. (2016) captures information about the early adopters of songs that become a hit. The features gathered are their popularity, seniority, or activity level, which might be proxies for their influence. The authors split such features into two sub-categories: roots, which are features of the first users to adopt a song; and researchers, the features averaged over other early adopters. As for network density, the similarity between early adopters works as follows: high similarity implies a niche market or one that people with similar interests are likely to adopt, while low similarity might indicate an item that could appeal to a wide variety of people.

Similarly, Herremans and Bergmans (2017) work with a dataset of 854,060 records, which includes time, date, user, song, and artist. Such a dataset contains a row for each song each week (instance) and a column (feature) for each user indicating the play count to a particular song. They also enriched the dataset with a predictive feature set to 1 in case a potential song becomes a hit in the future. Furthermore, Berns and Moore (2012) define a dataset using individual classifications according to the listeners' genre preferences. Each survey participant was given a list of six musical genres to rank from 1 (likes the most) to 6 (likes the least). Then, the final dataset consists of the top three genres for each participant.

5.2.5. Rank-based features

Dealing with hit song predictions requires reinventing the creation and use of features. Several works use features based on the musical rankings provided by entities that measure the popularity of the songs (Tsiara & Tjortjis, 2020). For example, Bischoff et al. (2009) extract some implicit features for both artists and tracks by combining their top-reached position on Billboard and their HITS scores–computed by applying the HITS algorithm on a graph using artists, tracks, and tags as nodes. In a more straightforward approach, Koenigstein et al. (2009) add the information of a song's debut rating on Billboard as an input to the proposed algorithm. Such a song's debut rating on Billboard is a valid attribute that may have additional explanatory information for the algorithm.

In the same direction, some works consider that the ranking of each song depends on both streaming and download volumes (Yoo & Kim, 2010); others add the period that the song remained on the Billboard chart after its first entry (Y. Kim et al., 2014; Middlebrook & Sheik, 2019). Lee and Lee (2015) consider the chart's first performance as ranking scores for the first and second weeks. The initial chart performance is significant in predicting popularity from two perspectives. First, it is a part of the entire popularity pattern over time and conveys partial information about long-term popularity patterns. Second, the popularity of a song's early stage influences future popularity by increasing the song's visibility.

Ren and Kauffman (2017) analyze the top-rank position before release; in other words, if a song gets in Top 50/100/150 before the new track is released. They also consider the first top-chart rank when a track first reaches a top-chart ranking on Last.fm. Shin and Park (2018) include features such as the song's inaugural rank on the chart, peak rank, time to reach the peak position after appearing on the chart, and the time to exit from the chart after reaching the peak.

5.2.6. Social media features

Features based on social networks can enrich the task of hit song prediction because the artist's fan bases share music content. Hence, a wide range of possibilities in defining features for prediction is available. For instance, in the work of Salganik et al. (2006), the participants made decisions about which songs to listen to; and while listening to a song, they were invited to rate it from one to five stars (the interval from 'I hate It' to 'I love it'). Then, to understand the social influence, the participants could download the song and see how many times each song was previously downloaded. Thus, the participants received a relatively weak signal about the social influence, which they were free to use or ignore, in addition to their own musical preferences.

Besides being the leading cause of music piracy, peer-to-peer (P2P) networks are also used to sample music before purchasing it. In such a direction, Koenigstein et al. (2009) investigate the relations between music file sharing and sales using P2P music information. They compare file-sharing songs to their popularity on Billboard charts and show a solid correlation (0.88–0.89). Finally, they show how this correlation can improve algorithms to predict a song's success on Billboard.

From a distinct social perspective, Abel et al. (2010) characterize the music albums on a given day by extracting the following features from the blogosphere: number of posts that include the music album title; number of posts in which the artist name appears; and number of posts where the artist name and album title appear. Likewise, Y. Kim et al. (2014) collect tweets with music-related keywords and calculate the song and artist popularities as several tweets associated with a song and artist, respectively.

Dewan and Ramaprasad (2014) study the interplay between blog buzz, radio play, and music sales at the album and song levels. They gather weekly data on the volume of song-level blog buzz from Google Blog Search, song-level unit sales, and radio play measured by the number of 'spins' from Nielsen SoundScan. The blog buzz is measured by the number of blogs that mention the artist and song names in a given week. They find the relationship between song buzz and sales is stronger for niche music than mainstream music and less popular songs inside albums. We can relate this buzz to the commonly called 'early adopters'. Shulman et al. (2016) defend that most of the predictive power comes from looking at how quickly songs reach their first early adopters. Hence, they study the feature structures of the network around early adopters, splitting them into two sub-categories: ego network features to relate the early adopters to their local networks, and subgraph features that hold only connections between the early adopters.

Ren et al. (2016) understand that social context positively impacted the data sample analysed, improving music content as a predictor for a music track's popularity on the web. Thereby, for predicting such song popularity online, they consider the number of user comments in the first five weeks after the track was released as a feature. Similarly, Araujo et al. (2017) collect messages from Twitter referring 30 days before the release of the examined albums. They choose such a time window to keep up with the growth in listeners' expectations as the launch date approaches. Also using data from Twitter, (Tsiara & Tjortjis, 2020) performed sentiment analysis on collected tweets regarding a song and its artist to predict the charts in the future.

With another social network data source, Ren and Kauffman (2017) consider comments about music on Last.fm after its release date (EarlyStageComments) in a time *t*. If the first few weeks of comments were sufficient to predict a track's popularity duration, the number of weeks was appropriate for *t* value. Lastly, Cosimato et al. (2019) use the following social media features: number of reached people on social media (fans); interest in specific albums measured by how much the album's author is a trend on the web; and collective opinion about the album or the singer on social media. Besides, they consider that a famous singer could influence the success of a new artist by collaborating on one or more of his songs.

5.2.7. Temporal information

The last classification proposed for extrinsic features is temporal information. Such characteristics deal with the speed with which music reaches success (after initial single release) (Buda & Jarynowski, 2015), or even how long it remains at the top. Another view is to understand the behavior of the songs before they are included in the charts, the time during their fame (already in the charts), until the moment when it stops appearing on the charts (Shulman et al., 2016). Some works analyze temporal distance to the release date; that is, the number of days when a music album will be or was released (Abel et al., 2010).

Although non-temporal features contribute to hit song prediction, the temporal ones are relevant because they describe the evolution of musical success and allow to identify trends in such time-related data. Such features are also valuable as they make the prediction more realistic by increasing the accuracy of the models (Shulman et al., 2016). However, despite their great relevance, temporal data are challenging because the main data sources do not provide them, impacting the number of studies considering such features.

5.3. Overall considerations

This section presented the main features used in works that predict a hit song. We also proposed a taxonomy to better organise the types of existing musical features divided into *Intrinsic* and *Extrinsic* features (Figure 7). The first group contains the intrinsic characteristics (genotypes) of the audio of the songs, as well as the related information, whether to describe the song or its lyrics. The second one is the group of features that describe the extrinsic characteristics associated with the songs (phenotypes). Such a group of features comprises data obtained from artists, albums, cultural aspects (closely linked to the genre), information from listeners, data from charts, and social and temporal networks.

Figure 8 presents the most used features and their evolution over time. Not surprisingly, the *Intrinsic* features are the most used predictors within HSS, as they are the most natural way to perceive music. In fact, they were the first set of features used by early HSS studies (Dhanaraj & Logan, 2005; Ni et al., 2011; Pachet & Roy, 2008) (Figure 8(b)). Social media gained notoriety in the music scene with the popularisation of online platforms in the 2010s. Hence, *Extrinsic* features assessing social interactions were included in the prediction models, based on the hypothesis that the popularity of a song may be directly related to how much people talk about it online (Cosimato et al., 2019). As depicted by Figure 8(a), social media metrics are in the top three most used features and rank #1 within the *Extrinsic* ones.

Overall, music is a multimodal concept that can be described in several aspects. For instance, it can be translated into audio signals, encapsulated into the lyrics and descriptive metadata, or even social web content. Nevertheless, most current research focuses on an unimodal or bimodal approach (e.g., acoustic + lyrics, acoustic + social, etc.). Although recent studies that consider such scenarios have good results, some aspects are not properly adopted, such as cultural, temporal, or album-based factors. Therefore, we believe that a more holistic view could better express prior knowledge of the data

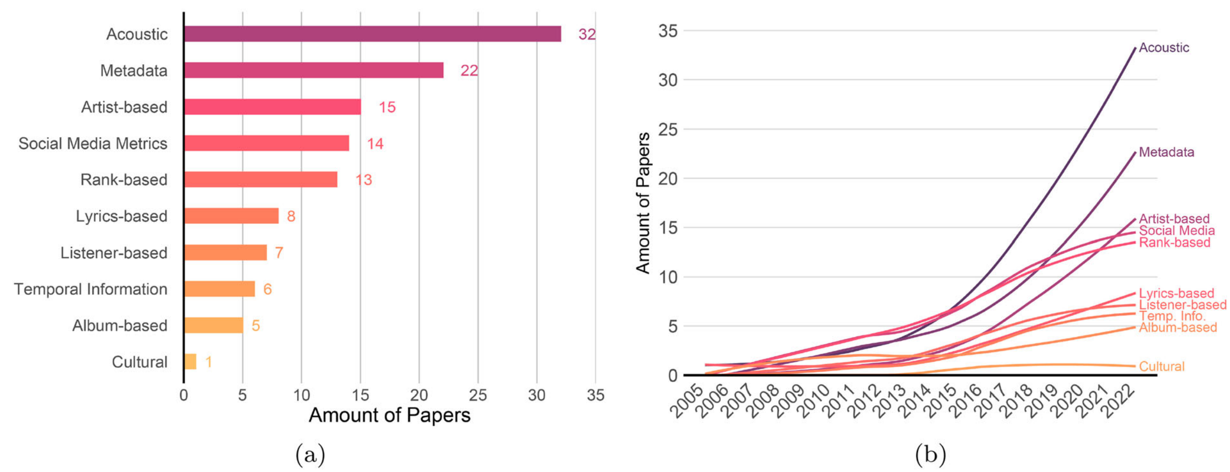


Figure 8. Most commonly used musical features as predictors (a) and their evolution over the years (b).

structure and fully exploit the valuable information that is encoded in it.

6. Learning methods

One of the main steps of Hit Song Science is to discover the set of predictors that contribute to the success of a song. In general, most works use machine learning approaches for such a task. In this section, we summarise the principal techniques of machine learning used for hit song prediction. We cover the main classification algorithms in Section 6.1, highlight the principal algorithms of regression in Section 6.2, and go over other approaches in Section 6.3.

6.1. Classification

Several studies tackle hit song prediction as a binary classification task. Given a set of songs' features, the task is to classify it as a hit or a non-hit. Table 3 summarises the selected works that apply Classification algorithms to Hit Song Science. Such a table includes the adopted success perspective, list of features, number of songs, classifiers utilized, and the accuracy of the best algorithm, which is in bold. Several algorithms address the binary classification task; the most studied in hit song prediction are as follows.

6.1.1. Random forest

Random Forest classifiers are flexible and easy-to-use machine learning algorithms that frequently produce excellent results, even without adjusting hyperparameters. Random forest creates a set of decision trees (a forest) from a randomly selected subset of the training set. In a decision tree, the internal nodes represent a test on a feature, the branches represent an exit from the test, and the leaves represent a vote to a class label (i.e. hit or non-hit). Next, the algorithm aggregates the votes from different decision trees to elect the final class of the test object. In other words, an object is assigned to a class that has the most votes from all trees.

In HSS, many works use the random forest algorithm to predict the success of a song (Araujo et al., 2019; Cosimato et al., 2019; Frieler et al., 2015; Gao, 2021; Kamal et al., 2021; Kaneria et al., 2021; Y. Kim et al., 2014; Rajyashree et al., 2018; Raza & Nanath, 2020; Ren & Kauffman, 2017; Ren et al., 2016; Shulman et al., 2016; Silva et al., 2022, november 7–11; Vötter et al., 2021, 2022). We highlight Interiano et al. (2018) who add the 'superstar' variable (i.e. if the artist had appeared on top charts) based on music acoustic features. Such a variable quantifies the contribution of purely musical characteristics to the song's success and suggests the time scale of fashion dynamics in popular music. Similarly, Middlebrook and Sheik (2019) test four classification models (Logistic Regression, Neural Network, Random Forest, and Support Vector Machine) on a dataset with approximately 1.8 million hit and non-hit songs. Overall, the best model is the random forest, which predicts Billboard song success with 97% accuracy in Cosimato et al. (2019).

6.1.2. Naive Bayes

Naive Bayes classifiers are supervised learning algorithms that apply Bayes' theorem with the naive assumption of conditional independence between every pair of features. In other words, such a classifier estimates the probability of a hit or non-hit holding on the assumption of musical features being conditionally independent. Due to the independence assumption, the class-conditional probability for every feature combination does not need to be calculated; only the conditional probability of each

Table 3. Main features and machine learning methods used in classification approaches for Hit Song Science.

Year	Ref.	Success	Features	# Songs	Classifier		Acc.
2005	Dhanaraj and Logan (2005)	Top-Charts	Acoustic, Lyrics-based	1700	**SVM**, Boosting Classifiers		0.69 [A]
2008	Pachet and Roy (2008)	Top-Charts	Acoustic, Metadata, Listener-based	32,000	**SVM**		0.41 [B]
2009	Bischoff et al. (2009)	Top-Charts	Metadata, Rank-based, Artist-based, Album-based, Social Media Metrics	317,058	SVM, Naive Bayes, **Bayesian Networks**, Decision Trees		0.75
	Koenigstein et al. (2009)	Top-Charts	Rank-based, Social Media Metrics	N/A	**Decision Trees**		0.89
2010	Abel et al. (2010)	Economy	Metadata, Temporal Info., Social Media Metrics	N/A	Naive Bayes, RBF Neural Network, SVM, **Decision Table**, One Rule, BFTree, LAD Tree, Simple Cart		0.50
2011	Ni et al. (2011)	Top-Charts	Acoustic	5947	**Perceptron**		0.58
2014	Herremans et al. (2014)	Top-Charts	Acoustic, Metadata, Artist-based	3452	Decision Tree, **RIPPER**, Naive Bayes, Logistic Regression, **SVM**		0.85
	Y. Kim et al. (2014)	Top-Charts	Rank-based, Artist-based, Social Media Metrics	178	**Random Forest**		0.84
	Nunes and Ordanini (2014)	Top-Charts	Metadata	2399	**Logistic Regression**		0.65
	Singhi and Brown (2014)	Top-Charts	Lyrics-based	6815	**Bayesian Network**		0.86
2015	Frieler et al. (2015)	Top-Charts	Acoustic, Metadata	266	**Random Forest**		0.52
	Lee and Lee (2015)	Top-Charts	Acoustic, Rank-based	867	**MLP**		0.71
	Singhi and Brown (2015)	Top-Charts	Acoustic, Lyrics-based	6815	**SVM**, Bayesian Networks		0.69 [A]
2016	Ren et al. (2016)	Engagement	Acoustic, Lyrics-based, Artist-based, Social Media Metrics	1961	Decision Tree, SVM, **Random Forest**, Bagging	>	0.7
	Shulman et al. (2016)	Engagement	Temporal Info., Social Media Metrics, Listener-based	5.8M	**Logistic Regression**, Random Forest, SVM		0.81
2017	Herremans and Bergmans (2017)	Top-Charts	Acoustic, Temporal Info., Listener-based	982	RIPPER, **Logistic Regression**, SVM, Naive Bayes		0.79 [A]
2017	Ren and Kauffman (2017)	Top-Charts	Acoustic, Metadata, Lyrics-based, Rank-based, Artist-based, Social Media Metrics	3881	SVM, Bagging, **Random Forest**		0.80
2018	Febirautami et al. (2018)	Engagement	Acoustic	233	**Decision Tree**		0.73
	Interiano et al. (2018)	Top-Charts	Acoustic, Metadata	500,000	**Random Forest**		0.86
	Lee and Lee (2018)	Top-Charts	Acoustic	16,686	**SVM**		0.70
	Rajyashree et al. (2018)	Engagement	Acoustic	8000+	Logistic Regression, SVM, Naive Bayes, Random Forest, Neural Network		N/A*
2019	Araujo et al. (2019)	Top-Charts	Acoustic	N/A	Ada Boost, Random Forests, Bernoulli, Gaussian Naive Bayes, **SVM**		0.89
	Cosimato et al. (2019)	Top-Charts	Social Media Metrics	N/A	**Random Forest**, SVM, MLP		0.97
	Middlebrook and Sheik (2019)	Top-Charts	Acoustic, Metadata, Rank-based, Temporal Info., Artist-based, Album-based	1.8M	Logistic Regression, Neural Network, **Random Forest**, SVM		0.89
2020	Araujo et al. (2020)	Top-Charts	Acoustic, Metadata	N/A	**SVM**, Gaussian Naive Bayes, kNN, Logistic Regression	>	0.8 [A]
	Martín-Gutiérrez et al. (2020)	Engagement	Acoustic, Metadata, Lyrics-based, Artist-based	101,939	**Deep Neural Network**		0.83
	Matsumoto et al. (2020)	Engagement	Acoustic, Artist-based	N/A	**SVM**		0.81
	Raza and Nanath (2020)	Top-Charts	Acoustic, Lyrics-based	647	**Logistic Regression**, Decision Trees, Random Forest, Naive Bayes		0.52
	Tsiara and Tjortjis (2020)	Top-Charts	Rank-based, Social Media Metrics	N/A	Decision Table, Filtered Classifier, **Logistic Model Tree (LMT)**, Logistic Regression		0.96
2021	S. T. Kim and Oh (2021)	Top-Charts	Acoustic	6209	**Logistic Regression**		0.68 [A]
	Kaneria et al. (2021)	Top-Charts	Acoustic	37,236	Logistic Regression, Decision Tree, **Random Forest**, Naive Bayes, KNN, XGBoost		0.91
	Kamal et al. (2021)	Engagement	Acoustic, Metadata, Lyrics-based	18,000+	**Random Forest**, SVM, Decision Tree, K-Nearest Neighbours, Logistic Regression and Naïve Bayes		0.85
	Gao (2021)	Engagement	Metadata, Acoustic	130,663	Multi-variate Linear Regression, Logistic Regression, Decision Tree, Random Forest, Boosting Tree, **Neural Networks**		0.83
	Vötter et al. (2021)	Engagement	Acoustic, Metadata	73,482	Logistic Regression, Random Forest, **SVM**, W&D Neural Networks, FCN		0.71
2022	Silva et al. (2022, november 7–11)	Top-Charts	Acoustic, Metadata, Album-based, Artists-based	911,027	Random Forest, Gradient Boosting, Support Vector Machines (**SVC** and NuSVC), MLP		0.82 [A]
	Vötter et al. (2022)	Top-Charts	Acoustic, Metadata	73,482	Logistic Regression, Random Forest, **SVM**, Deep Feedforward Neural Network, KNN, CNN		0.71

Ref.: Reference alias **Acc.**: Accuracy * Final results are not presented.
[A] **AUC**: Area Under the Receiver Operating Characteristic (ROC) Curve [B] **min-f1**: minimum F-Measure

feature x given Y has to be measured. It is a practical advantage since a reasonable estimate of the probability can be obtained without a very large training set (Abel et al., 2010; Araujo et al., 2019, 2020; Bischoff et al., 2009; Herremans & Bergmans, 2017; Herremans et al., 2014; Kamal et al., 2021; Kaneria et al., 2021; Rajyashree et al., 2018; Raza & Nanath, 2020). Although such an independence assumption is a weak assumption in practice, several studies prove that naive Bayes competes fairly with more sophisticated classifiers, having similar or slightly inferior performance than other approaches. For example, Herremans and Bergmans (2017) not only focus on audio features to predict a hit song but also includes social media listening behaviors to identify early adopters. They use a large dataset of social listening behavior from Last.fm. The first better result of AUC was the logistic regression (0.79) facing Naive Bayes to the second one (0.70), while SVM performs poorly (0.50). Still, naive Bayes is especially resistant to isolated noise points and robust to irrelevant attributes.

6.1.3. Bayesian networks

Bayesian networks are probabilistic graphical models that use Bayesian inference for probability computations, composed of random variables represented as nodes and their conditional dependencies represented as directed edges. The joint probability of the variables outlined in the directed and acyclic graph can be estimated as the product of the individual probabilities of each variable through the edges, conditioned on the node's parent variables. In other words, Bayesian networks are graphs that express how the occurrence of certain variables depends on another state. With satisfactory results, the hit song prediction strategy is used by Bischoff et al. (2009), Singhi and Brown (2014), and Singhi and Brown (2015). For instance, Bischoff et al. (2009) reach a value of 0.883 for the AUC measure, 0.788 precision, and 0.858 recall for hits, while the overall accuracy is 81.31%.

6.1.4. Support vector machine

SVM is a computer science concept for supervised learning methods that analyze data and recognize patterns. The standard SVM takes a dataset as input and, for each given data record, predicts which of two possible classes it is part of. SVM is a non-probabilistic binary linear classifier based on the theory of statistical learning (Cortes & Vapnik, 1995). In summary, SVM finds a line of separation (called as hyperplane) between data from the hit and non-hit classes. Such a line seeks to maximize the distance between the closest points concerning each class.

Many HSS approaches use SVM as one of the main algorithms for hit song prediction (Abel et al., 2010; Araujo et al., 2019, 2020; Bischoff et al., 2009; Dhanaraj & Logan, 2005; Herremans & Bergmans, 2017; Herremans et al., 2014; Kamal et al., 2021; Lee & Lee, 2018; Matsumoto et al., 2020; Middlebrook & Sheik, 2019; Pachet & Roy, 2008; Rajyashree et al., 2018; Ren & Kauffman, 2017; Ren et al., 2016; Shulman et al., 2016; Silva et al., 2022, november 7–11; Singhi & Brown, 2015; Vötter et al., 2021, 2022). The first exploration within the hit song science domain is by Dhanaraj and Logan (2005). The authors use acoustic characteristics and lyrics to build a Support Vector Machine, which is then tested over a small dataset with promising results. On the other hand, Bischoff et al. (2009) trained (with Billboard as ground truth) and tested the classifier on the total set of instances (both hits and non-hits), corresponding to each of the hit class ranges, using social media data (from Last.fm) as features. Also, Singhi and Brown (2015) use weighted-cost SVMs (in LIBSVM), which assign different misclassification costs to instances depending on the class they belong to. Also, using the LIBSVM package, Lee and Lee (2018) suggest three different experiments investigating the features' popularity prediction performance. They train SVMs with the extracted features to perform binary classification of each popularity metric. The boundary of the two classes is set to the median value of each popularity metric in the training data set. The radial basis function (RBF) is used as the kernel function of the SVMs. Despite SVM overall achieving satisfactory results, the best performance of the SVM classifier achieved an accuracy of 89% in work by Araujo et al. (2019) when predicting the song's popularity two months in advance.

6.1.5. Decision tree

Decision trees are non-parametric supervised machine learning methods widely used in classification tasks. In a decision tree, a decision is made by walking from the root node to the leaf node. Although decision trees are conceptually simple, they are relevant predictors, and their complexity is logarithmic in the prediction stage. However, decision trees have some problems that can degrade their predictive power. A tree grown to its maximum depth can overfit the training set and may degrade its predictive power to new data. Pruning the decision tree may mitigate such a problem. In addition, they are unstable models (high variance) because minor variations in training data can result in entirely different trees. Training several different trees and aggregating their predictions can avoid high variance.

In hit song science, decision trees are used for predicting hits as well (Bischoff et al., 2009; Febirautami et al., 2018; Gao, 2021; Herremans et al., 2014; Kamal et al., 2021; Kaneria et al., 2021; Koenigstein et al., 2009;

Raza & Nanath, 2020; Ren et al., 2016). One of the most popular algorithm implementation is the C4.5 Algorithm (Witten et al., 2011). According to Quinlan (1993), C4.5 Algorithm uses a divide-and-conquer approach to building trees recursively. Also, Decision Trees can be considered one of the easiest models to understand classification due to the linguistic nature (Martens et al., 2008). As mentioned by Herremans et al. (2014), the tree data structure consists of decision nodes and leaves. The leaves specify the class value (in this case, hit or non-hit), and the nodes specify a test of one of the features. When a path from the node to a leaf is followed based on the feature values of a particular song, a predictive rule can be derived (Ruggieri, 2002). As the best result for the Decision Tree classifier, Koenigstein et al. (2009) achieves 89% of accuracy by predicting a song's success on the Billboard in advance, using Peer-to-Peer information.

6.1.6. Neural network

In computer science, neural networks are computational models inspired by the central nervous system (as the brain). They are capable of performing machine learning, recognizing hidden patterns and correlations in raw data, grouping and classifying them, and – over time – continually learning and improving. Neural networks are generally presented as systems of linked neurons, which can compute input values, simulating the behavior of biological neural networks. A single neuron is a component that calculates the weighted sum of several inputs, applies a function, and forwards the results. Each neuron receives signals from input variables and passes on a weighted and treated version of that signal. In parallel, such neurons form a hidden layer of the neural network. The output of each neuron is a variable in the input of another hidden layer. Such hidden layers may be stacked, then producing a deep neural network.

Regarding research on the hit song science, few works use neural networks (Abel et al., 2010; Gao, 2021; Martín-Gutiérrez et al., 2020; Middlebrook & Sheik, 2019; Rajyashree et al., 2018; Vötter et al., 2021, 2022). For example, Rajyashree et al. (2018) define three layers besides an input: two hidden and one output layers. The recent outbreak of Deep Learning models has changed the paradigm in pattern recognition and classification tasks in HSS. Recent research Martín-Gutiérrez et al. (2020) propose different experiments for predicting the popularity of a song and describes a final end architecture composed of two main stages: a deep encoder and a deep neural network based on the model 1-A, which they selected as the best model since it provides good performance and a low computational cost. They achieve the best performance with an accuracy of 83.46% by using Neural Networks. Similarly, Gao (2021) shows that most songs can be found with about 83% accuracy according to audio features and artists' history profiles using Neural Networks.

6.1.7. Multi-layer perceptron

MLP is an example of a neural network and can be considered a logistic regression classifier, where the input is first transformed using a learned non-linear transformation. An MLP consists of at least three layers of nodes: input, at least one hidden layer, and output. Except for the input nodes, each node is a neuron that uses a non-linear activation function. However, a single hidden layer is sufficient to set MLPs as a universal approximator.

The research proposed by Lee and Lee (2015) uses MLP with one hidden layer as a classification method for predicting hit songs. They tried various numbers of hidden neurons, and as a result, the number of hidden neurons was set to 15, which produced good results. Their classifiers are trained to perform binary classification (high vs. low) of the six popularity patterns defined in the paper. MLP is also one of the classifiers used by Cosimato et al. (2019) to predict the music album rank in the Billboard 200 Chart, alongside Random Forest and SVM. However, in this case, the Random Forest model is the one that performed better, with an accuracy of 97%, while MLP achieved 86% of accuracy. Both studies are incomparable, considering a distinct set of features and success metrics.

6.1.8. Logistic regression

Logistic regression is a supervised classification algorithm aiming to produce a model that predicts values taken by a categorical variable, often binary, from a series of continuous and/or explanatory variables. Logistic regression is distinguished from linear regression because the response variable is categorical. Logistic regression becomes a classification technique only when a threshold is defined, and the setting of such a threshold is a crucial aspect of Logistic regression and is dependent on the classification problem itself. As a prediction method for categorical variables, logistic regression is comparable to the supervised techniques proposed in automatic learning (decision trees, neural networks, etc.) or even predictive discriminant analysis in exploratory statistics.

Several studies use the logistic regression algorithm applied to hit song science (Araujo et al., 2020; Herremans & Bergmans, 2017; Herremans et al., 2014; S. T. Kim & Oh, 2021; Middlebrook & Sheik, 2019; Nunes & Ordanini, 2014; Rajyashree et al., 2018; Raza & Nanath, 2020; Shulman et al., 2016; Tsiara & Tjortjis, 2020; Vötter et al., 2021). An interesting example is by Nunes and Ordanini (2014), which uses logistic regression to document how the absolute number of distinct musical instrument types perceptible in a song affects its chances of being hit song. The results suggest that songs that do not follow conventional instrumentation and, instead, include an atypically low or high number of instruments have greater chances of becoming a hit. However, Shulman et al. (2016) maintain the best performance with about 81% accuracy with logistic regression. Most of the predictive power comes from looking at how quickly items reach their early adopters.

6.2. Regression

This section highlights prediction solutions whose main objective is to return a continuous outcome (y) variable based on the value of one or multiple predictor variables (x). Table 4 summarises the selected works that apply Regression algorithms to Hit Song Science. It includes the adopted success perspective, list of features, number of songs, regressors utilized, and the score of the best algorithm, which is in bold. Hence, Regression is also chosen for predicting hits, and the most applied algorithms in hit song prediction are as follows.

6.2.1. Linear regression

Linear regression is one of the simplest methods of Machine Learning. It is composed of an equation to estimate the expected value of a target variable y, given the values of a set of input variables X. Such regression is called linear because it considers that the relationship of response to variables is a linear function of some parameters. Linear regression models are often adjusted using the least squares approach, but they can also be adjusted by minimizing the lack of fit in some other standard (with fewer absolute deviations from the regression) or minimizing a penalty version of the minimum squares.

In the context of hit song science, the features (as discussed in Section 5) compose the set of variables X, and the expected value y is whether the song is a hit or not. Considering the Regression task, Linear Regression is the main algorithm used to predict hit songs (Abel et al., 2010; Araujo et al., 2017; Fan & Casey, 2013; Y. Kim et al., 2014; Yang et al., 2017). We highlight the study of Yang et al. (2017), which computes the audio signals to compose the feature vectors as the input of a single-layer neural network model; in other words, in effectively linear regression.

6.2.2. Support vector regression

SVR is an extension of the SVM method (initially developed for class prediction) with the functionality of numerical regression. It determines the hyperplane that separates the instances of the target attribute by analyzing the distance between the instances positioned at the boundaries of the classes (Vapnik, 1998). The family of methods derived from SVM uses kernel functions to produce mathematical transformations in the data, expanding the dimensionality of the representation in order to make them linearly separable. SVR is flexible as

Table 4. Main features and machine learning methods used in regression approaches for Hit Song Science.

Year	Ref.	Success	Features	# Songs	Regressor	Score
2009	Koenigstein et al. (2009)	Top-Charts	Rank-based, Social Media Metrics	N/A	**Decision Trees**	10.1 AE
2010	Abel et al. (2010)	Economy	Metadata, Temporal Info., Social Media Metrics	N/A	Linear Regression, **SMO Regression**, Bagging REPTree	73.28 MAE
2013	Fan and Casey (2013)	Top-Charts	Acoustic	752	Linear Regression, **SVM**	0.39 ER
2014	Y. Kim et al. (2014)	Top-Charts	Rank-based, Artist-based, Social Media Metrics	178	Linear Regression, **SVR**	0.75 R^2
2017	Araujo et al. (2017)	Top-Charts, Engagement	Album-based, Social Media Metrics	N/A	**Linear Regression**	0.96 R^2
	Yang et al. (2017)	Engagement	Social Media Metrics, Listener-based	20,000	Linear Regression, Neural Networks	N/A*
	Ren and Kauffman (2017)	Top-Charts	Acoustic, Metadata, Lyrics-based, Rank-based, Artist-based, Social Media Metrics	3881	SVM, Bagging, **Random Forest**	0.73 C
	Askin and Mauskapf (2017)	Top-Charts	Acoustic, Metadata, Artist-based	27,000	**OLS Regression**, Binomial Regression	0.43 R^2
	Chiru and Popescu (2017)	Engagement	Acoustic, Lyrics-based	6000+	**SVM**, kNN	0.16 ER
2019	Zangerle et al. (2019)	Top-Charts	Acoustic, Metadata	95,067	**Neural Networks**	43.84 MAE
	Yu et al. (2019)	Engagement	Metadata, Rank-based, Social Media Metrics	10,842	**SVM**, Neural Networks	90.6 RMSE
2020	Martín-Gutiérrez et al. (2020)	Engagement	Acoustic, Metadata, Lyrics-based, Artist-based	101,939	**Deep Neural Network**	0.09 MAE
	Matsumoto et al. (2020)	Engagement	Acoustic, Artist-based	N/A	**SVM**	6.61 MAE
	Tsiara and Tjortjis (2020)	Top-Charts	Rank-based, Social Media Metrics	N/A	**SVR**, Random Forest, Bagging	4.05 MAE
2021	Vötter et al. (2021)	Engagement	Acoustic, Metadata	73,482	**Linear Regression**, Random Forest, SVM, Convolutional Neural Network	0.47 C
2022	Vötter et al. (2022)	Engagement	Acoustic, Metadata	73,482	**Linear Regression**, Random Forest, SVM, Deep Feedforward Neural Network, KNN, CNN	0.46 C

AE: Absolute Error **MAE**: Mean Absolute Error **ER**: Error Rate **R^2**: Coefficient of Determination **C**: Correlation **RMSE**: Root Mean Squared Error
* Although hit song prediction is addressed as a regression problem, the models are evaluated with ranking metrics.

it allows the setup of how much error is acceptable in the model and will find an appropriate line (or hyperplane in higher dimensions) to fit the data. The objective function of SVR is to minimize the coefficients (specifically, the loss function coefficient vector), not the squared error. SVR has a significant advantage: its computational complexity does not depend on the dimensionality of the input space (Awad & Khanna, 2015). In addition, it has excellent generalization capabilities with high forecasting accuracy.

There are several works using SVR on hit song science (Chiru & Popescu, 2017; Fan & Casey, 2013; Y. Kim et al., 2014; Matsumoto et al., 2020; Ren & Kauffman, 2017; Tsiara & Tjortjis, 2020; Yu et al., 2019). For example, Y. Kim et al. (2014) use SVR for predicting the popularity of the music based on Billboard rank. They use the radial basis function (RBF) for the kernel of the SVR model. SVR achieves a considerably high performance compared with other regression models (linear regression and linear regression quadratic). It is interesting to notice that SVR solves different variations of hit song prediction. For example, Ren and Kauffman (2017) use different combinations of Machine Learning methods to predict popularity duration, including SVR. They measure music popularity by using a Music Construct Vector with musical variables.

6.3. Others

There is still no consensus on the best algorithm for predicting hit songs in HSS. Many researchers look for alternatives not limited to classification and regression tasks. Table 5 shows such solutions. In total, we identify 12 different approaches among the HSS studies, as follows.

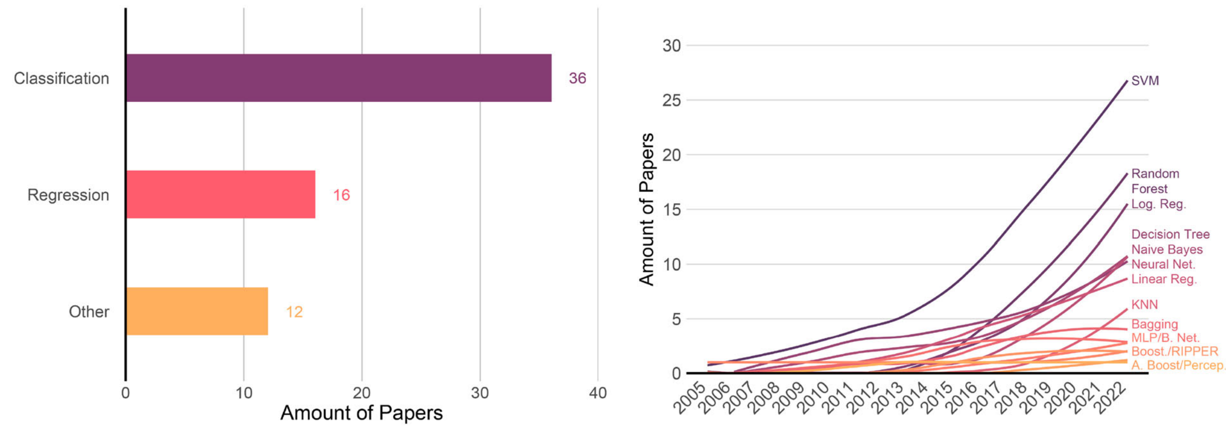
Experimental Study – Salganik et al. (2006) measure the success of a song based on the market share of downloads of such songs. **Time Series Prediction** – Chon et al. (2006) consider the life cycle of an album trajectory as the weekly positions from the first week to the last week on the Top Jazz chart. **Rankings** – Yoo and Kim (2010) measure the song's hotness by considering the high ranking songs that are given by download volumes, streaming volumes, and adjusted total volumes. **Neural Prediction** – With an unconventional approach, Berns and Moore (2012) use functional magnetic resonance imaging (fMRI) to measure the brain responses of a small group of teenagers when listening to songs by unknown artists and, as a measure of popularity, sales (SoundScan) of these songs were totaled over three years. **Statistical Analysis** – Some studies use as features the number of listeners, high level of popularity of artists who have a high number of followers, and the ranking of the song/album/artist to describe the popularity over time (Dewan & Ramaprasad, 2014; Shin & Park, 2018; Silva & Moro, 2019). **Social Networks** – There are also studies that consider user engagement to measure the popularity of songs (Buda & Jarynowski, 2015; Silva et al., 2019). **Cluster Analysis** – Al-Beitawi et al. (2020) perform a cluster analysis on the Top 100 Trending Spotify Song, considering ten musical features. **Epidemiological Analysis** Rosati et al. (2021) use the ability of the standard susceptible–infectious–recovered (SIR) epidemic model to fit the download time series of popular songs, and they conclude that such social processes underlying song popularity are similar to those that drive infectious disease transmission.

6.4. Overall considerations

The search for an ideal machine learning algorithm that predicts hit songs motivates researchers and the music industry to test various possibilities and combinations. There are algorithms with accuracy around 95%; still, in general, such works have limited scope and data, and they represent distinct facets that contribute to achieving musical success. Indeed, capturing the features necessary for prediction is hard. From another perspective,

Table 5. Main features and methods used in other approaches for Hit Song Science.

Year	Ref.	Success	Features	# Songs	Method
2006	Salganik et al. (2006)	Engagement	Social Media Metrics	48	Experimental Study
	Chon et al. (2006)	Top-Charts	Album-based	N/A	Time Series Prediction
2010	Yoo and Kim (2010)	Top-Charts	Rank-based	N/A	Ranking
2012	Berns and Moore (2012)	Economy	Listener-based	120	Neural Prediction
2014	Dewan and Ramaprasad (2014)	Engagement	Rank-based, Social Media Metrics	1000	Statistical Prediction
2015	Buda and Jarynowski (2015)	Top-Charts	Rank-based, Temporal info., Listener-based, Cultural		Social Network Analysis
2018	Shin and Park (2018)	Top-Charts	Metadata, Rank-based, Temporal info., Artist-based	7560	Ranking and Statistical Analysis
2019	Silva et al. (2019)	Engagement	Artist-based	2144	Statistical Analysis & Social Network Analysis
	Silva and Moro (2019)	Top-Charts	Rank-based, Artist-based	7185	Statistical Prediction
2020	Al-Beitawi et al. (2020)	Top-Charts	Acoustic	100	Cluster Analysis
	Oliveira et al. (2020)	Top-Charts	Artist-based	13,380	Statistical Analysis & Social Network Analysis
2021	Rosati et al. (2021)	Top-Charts	Metadata, Artist-based	950	Statistical & Epidemiological Analysis



(a) (b)

Figure 9. Most common learning algorithms in Hit Song Science (a) and their evolution over the years (b).

companies must deal with constant changes in the music market that generate new algorithm input features. For example, one of the first significant changes was the evolution of music consumption from physical media (CDs, DVDs) to digital media (streaming, downloads). More recently, the market is still trying to deal with trends in social networks (e.g.,TikTok, Instagram, and Twitter) that produce 'successes' by making them an intersection for creative expression and playful sociality (Abidin, 2020).

Consequently, we cannot say there is *one ideal* algorithm for HSS or hit song prediction. Nevertheless, people from the music industry (e.g., producers) who have access to large amounts of data (which are mostly not available to academia) can use the insights highlighted here to select or propose algorithms that work for their market. After all, in a billion-dollar industry, any percentage improvement in the bottom line can mean several million dollars in return. Nonetheless, by the universe of selected works in this survey, we note more significant favoritism towards classification algorithms, as Figure 9(a) shows. Among classification algorithms, we report a preference for Support Vector Machine and Random Forest (Figure 9(b)). Such algorithms are easy to implement and present satisfactory results. On the other hand, relatively fewer studies consider the prediction of hit songs as a regression task, and the most used algorithms are classic Linear Regression and Support Vector Regression. Other approaches use data from social networks, experimental studies, rankings, statistical analysis, clustering, time series prediction, and neural prediction.

7. Research directions

Based on our overview of the techniques and tasks relevant to Hit Song Science, we now identify and discuss new directions for research that require further investigation. These topics are not the only open research problems within HSS, but they are key factors that may shed light on the science of what makes a song successful.

Dealing with multiple sources. Data integration is one of the main issues in many Computer Science research fields. In Hit Song Science, this topic is becoming more relevant and necessary, as there is no unique data source for all necessary features and data. For instance, to the best of our knowledge, no data source provides both acoustic and lyrics-based features. Furthermore, the lack of a unique and universal identifier for each music makes integrating several data sources very challenging. Besides, information such as the musical genre(s) of a given song is not standardized in all data sources, mainly due to the blurred line existent between music styles that are close to each other.

Regional markets' diversity. Most studies on Hit Song Science use data from the American market (e.g., Billboard Hot 100 Chart and Amazon Sales Data). This may be because the United States is the biggest music market in the world, which may facilitate the acquisition and use of such data. Research studies that consider music markets other than the USA focus mainly on European countries, such as the United Kingdom. However, many other relevant markets with distinct characteristics and behaviours require an individual analysis of success. For example, South Korea, China, and Brazil are among the top 10 music markets in the world[29], with a vibrant music scene and popular regional genres. Such genres have become popular in the global scenario as they connect with other well-established music genres (e.g., collaborations involving pop, k-pop, and Latin genres such as reggaeton). Therefore, as local engagement

[29] IFPI Global Music Report: http://gmr.ifpi.org/

shapes the global environment, future work on HSS must consider the regional aspect, thus ensuring that music culture within such countries is accounted for.

Lack of standardized success metrics. Defining a song's popularity is still challenging, and each research study in HSS uses specific success metrics. In Section 4, we discussed and proposed a taxonomy for such metrics, but as there is no standard, researchers are unable to perform a fair comparison between their work and the existing literature on the subject. Hence, finding a way to generalize success properly would support future work on HSS to capture popularity definitions more accurately. Moreover, it would expose their findings to a commonly understood metric, allowing a complete evaluation by comparing performance with current work (as presented here).

Importance of social aspects. The ever-growing popularisation of social networks in the last two decades has deeply changed the music industry. The propagation of songs on such platforms is fundamental to their success, as the viral phenomenon of songs on social media may lead a newly released one to stardom or even lead back a great hit from the past to the top of the charts. Since marketing dramatically impacts the future success of songs, it is increasingly important to consider the latest social platforms and features, which could also give strong indications of a song's hit potential. Although such features have been used in previous research, novel approaches to HSS must combine audio and social data to enhance hit prediction efficiency.

Framework for predicting and recommending hit songs. One of the main goals of Hit Song Science is to predict whether a given song will become a hit or not. Consequently, extending this goal into a recommendation task in which hit songs will be returned based on the listener's musical preferences is reasonable. Although several studies assess such issues, there is still no support tool for running such models and then standardizing assessment. For example, a framework on hit song prediction/recommendation should be able to receive a given success measure as input, a set of predefined features, and an algorithm to run the machine learning task (e.g., classification and/or regression). In fact, this would represent a significant advance in the field of HSS, with benefits to both the academy and the music industry, as it allows scientists to better understand the success phenomenon and record label CEOs to invest in selected songs and artists properly.

8. Concluding remarks

In this survey, instead of emphasizing what worked and what did not, we aim to present which approaches are the most used within HSS, as well as the main findings on the subject. To do so, we follow a generic workflow (see Figure 3) of four phases, which model this survey's structure. First, we described the most commonly

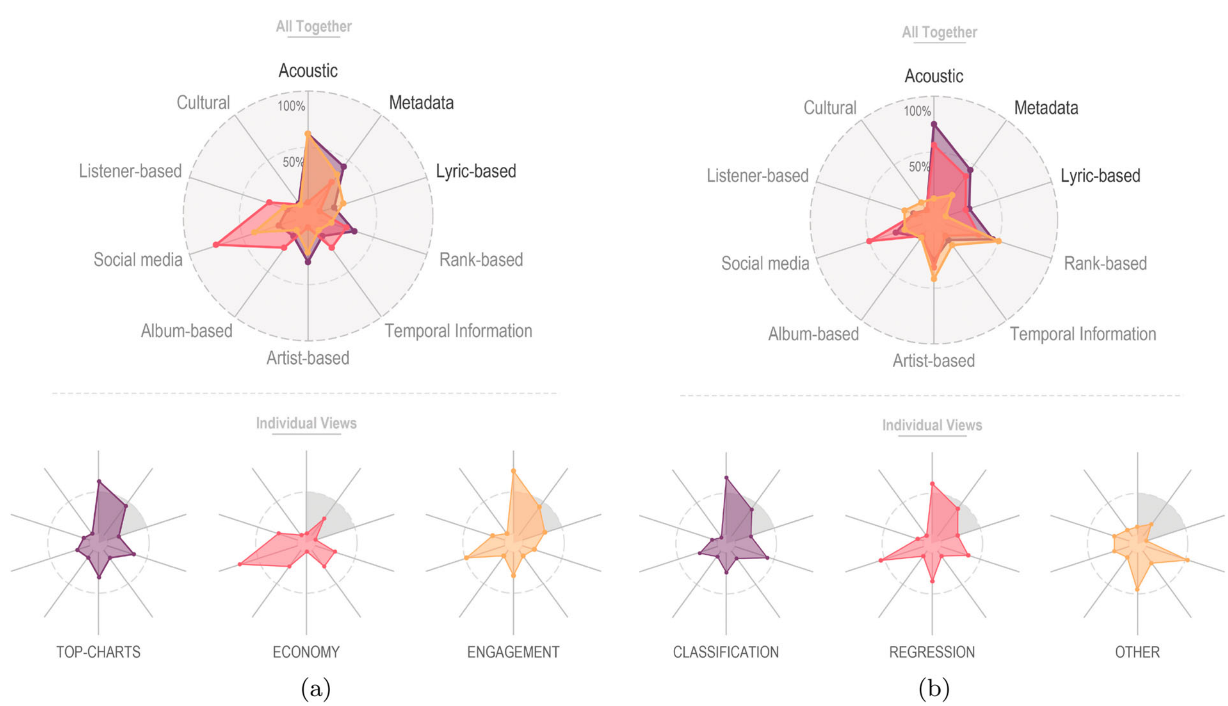


Figure 10. Comparison of HSS papers associating the features with the adopted success measures (a) and learning methods (b). The darker radial comprises the Intrinsic features, and the lighter the Extrinsic ones. (a) Success measures and (b) Machine Learning tasks.

used data sources and discussed their applications. Second, we proposed a hierarchical taxonomy to classify the different success measures existent in the literature into three perspectives: *Top-Charts, Economy,* and *Engagement*. Next, we assessed the most frequently used features in hit song prediction models. Finally, we summarised the main learning methods (i.e.classification, regression, and others) to predict whether a song will be a hit or not.

Despite following such a workflow, not all works use the same combinations of success measures, musical features, and learning methods. Figure 10 compares the choice of types of features used with (a) the success perspectives and (b) the machine learning tasks. For success, most works considering the *Top-Charts* perspective use *Intrinsic* features (mostly acoustic). However, extrinsic ones such as *Rank-based* are also present, which is expected given the success measure choice. Both *Economy* and *Engagement* include different types of features directly related to their nature. For instance, social media, listener-based, and rank-based features are strongly present in the *Economy* view, as they are fundamental to building up marketing strategies to boost sales. In contrast, works within the *Engagement* perspective rely mainly on information about music consumption and social engagement (i.e., social media).

Regarding machine learning tasks, acoustic features are the most used in classification and regression approaches. Such features are usually represented in numerical variables, making their processing easier and further statistical analyses easier since many learning algorithms require this format as input. Furthermore, compared to the classification task, studies that tackle the hit song prediction as a regression problem also consider social media metrics. Finally, the extrinsic features are considerably used by other types of solutions, given the more flexible nature of such diverse techniques (e.g., clustering, statistical and social network analysis). Again, our objective is not to detail what works and what does not, as this is very particular concerning each target. However, we reaffirm our desire to exhaustively list what the research community has produced from the beginning of the HSS area to the present day.

In conclusion, HSS emerges as a multidisciplinary field within Music Information Retrieval (MIR), aiming to predict the success of a song before its release. Its interdisciplinary nature promotes not only benefits to the MIR community but also to the music industry as a whole. Moreover, such prediction studies may help music industry CEOs to maximize expected success by properly investing in selected songs/artists. We also identified key open research issues within HSS, revealing a broad scope for brand new improvements and advancements in such a field. Therefore, we believe that this article sheds light on the science behind musical success, serving as a base material for future research on Hit Song Science.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

References

Abel, F., Diaz-Aviles, E., Henze, N., Krause, D., & Siehndel, P. (2010). Analyzing the blogosphere for predicting the success of music and movie products. In *International conference on advances in social networks analysis and mining, ASONAM* (pp. 276–280). IEEE Computer Society. https://doi.org/10.1109/ASONAM.2010.50

Abidin, C. (2020). Mapping internet celebrity on TikTok: Exploring attention economies and visibility labours. *Cultural Science Journal*, *12*(1), 77–103. https://doi.org/10.5334/csci.140

Al-Beitawi, Z., Salehan, M., & Zhang, S. (2020). Cluster analysis of musical attributes for top trending songs. In *Hawaii international conference on system sciences* (pp. 1–7). ScholarSpace. https://doi.org/10.24251/HICSS.2020.017

Araujo, C. V. S., M. A. P. de Cristo, & Giusti, R. (2019). Predicting music popularity using music charts. In *IEEE international conference on machine learning and applications* (pp. 859–864). IEEE. https://doi.org/10.1109/ICMLA.2019.00149

Araujo, C. V. S., de Cristo, M. A. P., & Giusti, R. (2020). A model for predicting music popularity on streaming platforms. *RITA*, *27*(4), 108–117. https://doi.org/10.22456/2175-2745.107021

Araujo, C. V. S., Neto, R. M., Nakamura, F. G., & Nakamura, E. F. (2017). Predicting music success based on users' comments on online social networks. In *Brazilian symposium on multimedia and the web* (pp. 149–156). ACM. https://doi.org/10.1145/3126858.3126885

Askin, N., & Mauskapf, M. (2017). What makes popular culture popular? Product features and optimal differentiation in music. *American Sociological Review*, *82*(5), 910–944. https://doi.org/10.1177/0003122417728662

Awad, M., & Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Apress Media.

Berns, G. S., & Moore, S. E. (2012). A neural predictor of cultural popularity. *Journal of Consumer Psychology*, *22*(1), 154–160. https://doi.org/10.1016/j.jcps.2011.05.001

Bischoff, K., Firan, C. S., Georgescu, M., Nejdl, W., & Paiu, R. (2009). Social knowledge-driven music hit prediction. In *International conference advanced data mining and applications* (pp. 43–54). Springer. https://doi.org/10.1007/978-3-642-03348-3_8

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. http://jmlr.org/papers/v3/blei03a.html

Buda, A., & Jarynowski, A. (2015). Exploring patterns in european singles charts. In *European network intelligence conference* (pp. 135–139). IEEE Computer Society. https://doi.org/10.1109/ENIC.2015.27

Chiru, C., & Popescu, O.-G. (2017). Automatically determining the popularity of a song. In *Rough sets – international joint conference* (pp. 392–406). Springer. https://doi.org/10.1007/978-3-319-60837-2_33

Chon, S. H., Slaney, M., & Berger, J. (2006). Predicting success from music sales data: A statistical and adaptive approach. In *ACM workshop on audio and music computing multimedia, AMCMM* (pp. 83–88). Association for Computing Machinery. https://doi.org/10.1145/1178723.1178736

Corrêa, D. C., & Rodrigues, F. A. (2016). A survey on symbolic data-based music genre classification. *Expert Systems with Applications*, *60*, 190–210. https://doi.org/10.1016/j.eswa.2016.04.008

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Cosimato, A., De Prisco, R., Guarino, A., Malandrino, D., Lettieri, N., Sorrentino, G., & Zaccagnino, R. (2019). The conundrum of success in music: Playing it or talking about it?. *IEEE Access*, *7*, 123289–123298. https://doi.org/10.1109/ACCESS.2019.2937743

Dan-Ning, J., Lu, L., Zhang, H.-J., Tao, J.-H., & Cai, L.-H. (2002). Music type classification by spectral contrast feature. In *Proceedings. IEEE international conference on multimedia and expo* (pp. 113–116). IEEE Computer Society. https://doi.org/10.1109/ICME.2002.1035731

Dewan, S., & Ramaprasad, J. (2014). Social media, traditional media, and music sales. *MIS Quarterly*, *38*(1), 101–121. https://doi.org/10.25300/MISQ/2014/38.1.05

Dhanaraj, R., & Logan, B. (2005). Automatic prediction of hit songs. In *International conference on music information retrieval* (pp. 488–491). ISMIR. http://ismir2005.ismir.net/proceedings/2024.pdf

Elovitz, H. S., Johnson, R. W., McHugh, A., & Shore, J. E. (1976). Automatic translation of English text to phonetics by means of letter-to-sound rules (Technical Report). Naval Research Laboratory.

Fan, J., & Casey, M. (2013). Study of Chinese and UK hit songs prediction. In *Proceedings of international symposium on computer music multidisciplinary research* (pp. 640–652). The Laboratory of Mechanics and Acoustics. https://cmmr2013.prism.cnrs.fr/Docs/CMMR2013Proceedings.pdf

Febirautami, L. R., Surjandari, I., & Laoh, E. (2018). Determining characteristics of popular local songs in Indonesia's music market. In *International conference on information science and control engineering* (pp. 197–201). IEEE. https://doi.org/10.1109/ICISCE.2018.00050

Fisher, C., Pearson, M. M., Goolsby, J. R., & Onken, M. H. (2010). Developing measurements of success for performing musical groups. *Journal of Services Marketing*, *24*(4), 325–334. https://doi.org/10.1108/08876041011053033

Frieler, K., Abeßer, J., Zaddach, W.-G., & Pfleiderer, M. (2013). Introducing the Jazzomat project and the Melo (S) py library. In *International workshop on folk music analysis* (pp. 76–78). Meertens Institute and Department of Information and Computing Sciences, Utrecht University. https://dspace.library.uu.nl/handle/1874/276246

Frieler, K., Jakubowski, K., & Müllensiefen, D. (2015). Is it the song and not the singer? Hit song prediction using structural features of melodies. *Jahrbuch Musikpsychologie*, *25*, 41–54. https://pada.psycharchives.org/bitstream/595eff5e-2e9c-4e6a-a0ab-8019d1a64b2b

Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, *13*(2), 303–319. https://doi.org/10.1109/TMM.2010.2098858

Gao, A. (2021). Catching the earworm: Understanding streaming music popularity using machine learning models. In *E3S web conf.* EDP Sciences (Vol. 253, p. 03024). https://doi.org/10.1051/e3sconf/202125303024

Harte, C., Sandler, M., & Gasser, M. (2006). Detecting harmonic change in musical audio. In *ACM workshop on audio and music computing multimedia* (pp. 21–26). Association for Computing Machinery. https://doi.org/10.1145/1178723.1178727

Herremans, D., & Bergmans, T. (2017). Hit song prediction based on early adopter data and audio features. In *International society for music information retrieval conference, ISMIR – late breaking demo*. ISMIR.

Herremans, D., Martens, D., & Sörensen, K. (2014). Dance hit song prediction. *Journal of New Music Research*, *43*(3), 291–302. https://doi.org/10.1080/09298215.2014.881888

Hirjee, H., & Brown, D. G. (2010). Rhyme analyzer: An analysis tool for rap lyrics. In *International society for music information retrieval conference*. ISMIR.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Annual international ACM SIGIR conference on research and development in information retrieval* (pp. 50–57). Association for Computing Machinery. https://doi.org/10.1145/312624.312649

Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., & Komarova, N. L. (2018). Musical trends and predictability of success in contemporary songs in and out of the top charts. *Royal Society Open Science*, *5*(5), 171274. https://doi.org/10.1098/rsos.171274

Kamal, J., Priya, P., Anala, M. R., & Smitha, G. R. (2021). A classification based approach to the prediction of song popularity. In *2021 international conference on innovative computing, intelligent communication and smart electrical systems (ICSES)* (pp. 1–5). IEEE.

Kaneria, A. V., Rao, A. B., Aithal, S. G., & Pai, S. N. (2021). Prediction of song popularity using machine learning concepts. In S. K. V. & K. G. Rao (Eds.), *Smart sensors measurements and instrumentation* (pp. 35–48). Springer Singapore.

Kim, S. T., & Oh, J. H. (2021). Music intelligence: Granular data and prediction of top ten hit songs. *Decision Support Systems*, *145*, Article 113535. https://www.sciencedirect.com/science/article/pii/S0167923621000452

Kim, Y., Suh, B., & Lee, K. (2014). # nowplaying the future billboard: Mining music listening behaviors of Twitter users for hit song prediction. In *International workshop on social media retrieval and analysis* (pp. 51–56). ACM. https://doi.org/10.1145/2632188.2632206

Koenigstein, N., Shavitt, Y., & Zilberman, N. (2009). Predicting billboard success using data-mining in P2P networks. In *IEEE international symposium on multimedia* (pp. 465–470). IEEE.

Lee, J., & Lee, J.-S. (2015). Predicting music popularity patterns based on musical complexity and early stage popularity. In *Workshop on speech, language & audio in multimedia* (pp. 3–6). ACM. https://doi.org/10.1145/2802558.2814645

Lee, J., & Lee, J.-S. (2018). Music popularity: Metrics, characteristics, and audio-based prediction. *IEEE Transactions on Multimedia*, *20*(11), 3173–3182. https://doi.org/10.1109/TMM.2018.2820903

Li, H., Tang, Z., Fei, X., Chao, K.-M., Yang, M., & He, C. (2017). A survey of audio MIR systems, symbolic MIR systems and a music definition language demo-system. In *IEEE international conference on e-business engineering*, (pp. 275–281). IEEE Computer Society. https://doi.org/10.1109/ICEBE.2017.51

Li, T., Ogihara, M., & Tzanetakis, G. (2011). *Music data mining* (1st ed.). CRC Press, Inc.

Logan, B., Kositsky, A., & Moreno, P. J. (2004). Semantic analysis of song lyrics. In *IEEE international conference on multimedia and expo (ICME)* (pp. 827–830). IEEE Computer Society. https://doi.org/10.1109/ICME.2004.1394328

Maršík, L., Pokorný, J., & Ilcık, M. (2015). A survey on music retrieval systems using microphone input. In *Annual international workshop on databases, texts, specifications and objects, CEUR workshop proceedings* (pp. 131–140). CEUR-WS.org. http://ceur-ws.org/Vol-1343/poster7.pdf

Martens, D., Baesens, B. B., & Van Gestel, T. (2008). Decompositional rule extraction from support vector machines by active learning. *IEEE Transactions on Knowledge and Data Engineering*, *21*(2), 178–191. https://doi.org/10.1109/TKDE.2008.131

Martín-Gutiérrez, D., Peñaloza, G. H., Belmonte-Hernández, A., & García, F. Á. (2020). A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access*, *8*, 39361–39374. https://doi.org/10.1109/ACCESS.2020.2976033

Matsumoto, Y., Harakawa, R., Ogawa, T., & Haseyama, M. (2020). Context-aware network analysis of music streaming services for popularity estimation of artists. *IEEE Access*, *8*, 48673–48685. https://doi.org/10.1109/ACCESS.2020.2978281

McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *Python in science conference* (pp. 18–25). SciPy Organizers.

Middlebrook, K., & Sheik, K. (2019). Song hit prediction: Predicting billboard hits using spotify data. *CoRR* abs/1908.08609. http://arxiv.org/abs/1908.08609

Murthy, Y. V. S., & Koolagudi, S. G. (2018). Content-based music information retrieval (CB-MIR) and its applications toward the music Industry: A review. *ACM Computing Surveys*, *51*(3), 45:1–45:46. https://doi.org/10.1145/3177849

Ni, Y., Santos-Rodriguez, R., Mcvicar, M., & De Bie, T. (2011). Hit song science once again a science? In *International workshop on machine learning and music* (pp. 355–360). ACM.

Nunes, J. C., & Ordanini, A. (2014). I like the way it sounds: The influence of instrumentation on a pop song's place in the charts. *Musicae Scientiae*, *18*, 392–409. https://doi.org/10.1177/1029864914548528

Oliveira, G. P., Silva, M. O., Seufitelli, D. B., Lacerda, A., & Moro, M. M. (2020). Detecting collaboration profiles in success-based music genre networks. In *International society for music information retrieval conference* (pp. 726–732). ISMIR.

Pachet, F. (2011). Hit song science. In G. T. T. Li & M. Ogihara (Eds.), *Music data mining* (Chap. 10, pp. 305–326). CRC Press.

Pachet, F., & Roy, P. (2008). Hit song science is not yet a science. In *International conference on music information retrieval, ISMIR* (pp. 355–360). ISMIR.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann, Elsevier.

Rajyashree, R., Anand, A., Soni, Y., & Mahajan, H. (2018). Predicting hit music using MIDI features and machine learning. In *International conference on communication and electronics systems* (pp. 94–98). IEEE. https://doi.org/10.1109/CESYS.2018.8724001

Raza, A. H., & Nanath, K. (2020). Predicting a hit song with machine learning: Is there an apriori secret formula? In *International conference on data science, artificial intelligence, and business analytics (DATABIA)*. online (pp. 111–116). IEEE. https://doi.org/10.1109/DATABIA50434.2020.9190613

Ren, J., & Kauffman, R. J. (2017). Understanding music track popularity in a social network. In *European conference on information systems* (pp. 374–388). AIS. http://aisel.aisnet.org/ecis2017_rp/25

Ren, J., Shen, J., & Kauffman, R. J. (2016). What makes a music track popular in online social networks? In *International conference companion on World Wide Web* (pp. 95–96). https://doi.org/10.1145/2872518.2889402

Rosati, D. P., Woolhouse, M. H., Bolker, B. M., & Earn, D. J. D. (2021). Modelling song popularity as a contagious process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *477*(2253), Article 20210457. https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2021.0457

Ruggieri, S. (2002). Efficient C4. 5 [classification algorithm]. *IEEE Transactions on Knowledge and Data Engineering*, *14*(2), 438–444. https://doi.org/10.1109/69.991727

Salganik, M. J., Sheridan Dodds, P., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science (New York, N.Y.)*, *311*(5762), 854–856. https://doi.org/10.1126/science.1121066

Scaringella, N., Zoia, G., & Mlynek, D. (2006). Automatic genre classification of music content: A survey. *IEEE Signal Processing Magazine*, *23*(2), 133–141. https://doi.org/10.1109/MSP.2006.1598089

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, *36*(12), 2346–2353. https://doi.org/10.1121/1.1919362

Shin, S., & Park, J. (2018). On-chart success dynamics of popular songs. *Advances in Complex Systems*, *21*(3–4), 1850008. https://doi.org/10.1142/S021952591850008X

Shulman, B., Sharma, A., & Cosley., D. (2016). Predictability of popularity: Gaps between prediction and understanding. In *International conference on web and social media* (pp. 348–357). AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13129

Silva, M. O., & Moro., M. M. (2019). Causality analysis between collaboration profiles and musical success. In *Brazilian symposium on multimedia and the web* (pp. 369–376). ACM. https://doi.org/10.1145/3323503.3349549

Silva, M. O., Oliveira, G. P., Seufitelli, D. B., Lacerda, A., & Moro, M. M. (2022, November 7–11). Collaboration as a driving factor for hit song classification. In *Proceedings of the 28th Brazillian symposium on multimedia and the web, webmedia 2022,* Association for Computing Machinery (pp. 66–74).

Silva, M. O., Rocha, L. M., & Moro., M. M. (2019). Collaboration profiles and their impact on musical success. In *ACM/SIGAPP symposium on applied computing* (pp. 2070–2077). ACM. https://doi.org/10.1145/3297280.3297483

Singhi, A., & Brown., D. G. (2014). Hit song detection using lyric features alone. In *International society for music information retrieval conference (ISMIR): Late-breaking demo*. ISMIR.

Singhi, A., & Brown., D. G. (2015). Can song lyrics predict hits. In *International symposium on computer music multidisciplinary research* (pp. 457–471). The Laboratory of Mechanics and Acoustics. https://cmmr2019.prism.cnrs.fr/Docs/proceedingsCMMR2015.pdf

Sturm, B. L. (2012). A survey of evaluation in music genre recognition. In *Adaptive multimedia retrieval* (pp. 29–66). Springer. https://doi.org/10.1007/978-3-319-12093-5_2

Tsiara, E., & Tjortjis, C. (2020). Using twitter to predict chart position for songs. In *IFIP artificial intelligence applications and innovations* (pp. 62–72). Springer. https://doi.org/10.1007/978-3-030-49161-1_6

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293–302. https://doi.org/10.1109/TSA.2002.800560

Vapnik, V. (1998). The support vector method of function estimation. In J. A. K. Suykens & J. Vandewalle (Eds.), *Nonlinear modeling* (pp. 55–85). Springer. https://doi.org/10.1007/978-1-4615-5703-6_3

Vötter, M., Mayerl, M., Specht, G., & Zangerle., E. (2021, November 29–December 1). Novel datasets for evaluating song popularity prediction tasks. In *IEEE international symposium on multimedia, ISM 2021, Virtual event*, 166–173. IEEE.

Vötter, M., Mayerl, M., Specht, G., & Zangerle, E. (2022). HSP datasets: Insights on song popularity prediction. *International Journal of Semantic Computing*, 1–23. https://www.worldscientific.com/doi/abs/10.1142/S1793351X22400104

Witten, I. H., Frank, E., & Hall., M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann, Elsevier.

Yang, L.-C., Chou, S.-Y., Liu, J.-Y., Yang, Y.-H., & Chen, Y.-A. (2017). Revisiting the problem of audio-based hit song prediction using convolutional neural networks. In *International conference on acoustics, speech and signal processing* (pp. 621–625). IEEE. https://doi.org/10.1109/ICASSP.2017.7952230

Yoo, B., & Kim, K. (2010). Online music ranking service: Ranking mechanism based on popularity and slot effect. In *Pacific Asia conference on information systems* (pp. 615–626). AISeL. http://aisel.aisnet.org/pacis2010/31

Yu, H., Li, Y., Zhang, S., & Liang., C. (2019). Popularity prediction for artists based on user songs dataset. In *Proceedings of international conference on computing and artificial intelligence* (pp. 17–24). ACM. https://doi.org/10.1145/3330482.3330493

Zampoglou, M., & Malamos, A. G. (2014). Music information retrieval in compressed audio files: A survey. *New Review of Hypermedia Multimedia*, *20*(3), 189–206. https://doi.org/10.1080/13614568.2014.889223

Zangerle, E., Vötter, M., Huber, R., & Yang., Y.-H. (2019). Hit song prediction: Leveraging low- and high-level audio features. In *International society for music information retrieval conference* (pp. 319–326). ISMIR.

Appendix. List of considered papers

Table A1. List of the 55 papers considered in this survey and their corresponding Data Sources, Success Perspectives, Considered Features, and the machine learning task used for Hit Song Science.

Year	Ref.	Data Sources	Success	Features	ML Task
2005	Dhanaraj and Logan (2005)	Oz Net Music Chart, In-house database, Astraweb Lyrics	Top-Charts	Acoustic, Lyrics-based	Classification
2006	Chon et al. (2006)	Billboard	Top-Charts	Album-based	Other
	Salganik et al. (2006)	purevolume	Engagement	Social Media Metrics	Other
2008	Pachet and Roy (2008)	HiFind Database	Top-Charts	Acoustic, Metadata, Listener-based	Classification
2009	Bischoff et al. (2009)	Last.fm, Billboard	Top-Charts	Metadata, Rank-based, Artist-based, Album-based, Social Media Metrics	Classification
	Koenigstein et al. (2009)	Gnutella, Billboard	Top-Charts	Rank-based, Social Media Metrics	Classification, Regression
2010	Abel et al. (2010)	Spinn3r, Amazon	Economy	Metadata, Temporal Info., Social Media Metrics	Classification, Regression
	Yoo and Kim (2010)	N/A	Top-Charts	Rank-based	Other
2011	Ni et al. (2011)	Official Charts Company, EchoNest	Top-Charts	Acoustic	Classification
2012	Berns and Moore (2012)	MySpace, Nielsen SoundScan	Economy	Listener-based	Other
2013	Fan and Casey (2013)	Official Charts Company, ZhongGuoGeQuPaiHangBang (China), EchoNest	Top-Charts	Acoustic	Regression
2014	Dewan and Ramaprasad (2014)	Nielsen SoundScan, Google Blog Search, Last.fm, Amazon, Allmusic.com	Economy, Engagement	Rank-based, Social Media Metrics	Other
	Herremans et al. (2014)	Official Charts Company, Billboard, EchoNest	Top-Charts	Acoustic, Metadata, Artist-based	Classification
	Y. Kim et al. (2014)	Twitter, Billboard	Top-Charts	Rank-based, Artist-based, Social Media Metrics	Classification, Regression
	Nunes and Ordanini (2014)	Billboard	Top-Charts	Metadata	Classification
	Singhi and Brown (2014)	Billboard	Top-Charts	Lyrics-based	Classification
2015	Buda and Jarynowski (2015)	European Music Papers, radio, TV, Internet	Top-Charts	Rank-based, Temporal info., Listener-based, Cultural	Other
	Frieler et al. (2015)	Earwormery Database, Polyhex UK, Geerdes Database	Top-Charts	Acoustic, Metadata	Classification
	Lee and Lee (2015)	Billboard	Top-Charts	Acoustic, Rank-based	Classification
	Singhi and Brown (2015)	EchoNest, Billboard, Metro Lyrics	Top-Charts	Acoustic, Lyrics-based	Classification
2016	Ren et al. (2016)	Last.fm, Wikipedia, 7digital, Google Lyrics, MusicSong Lyrics	Engagement	Acoustic, Lyrics-based, Artist-based, Social Media Metrics	Classification
	Shulman et al. (2016)	Last.fm	Engagement	Temporal Info., Social Media Metrics, Listener-based	Classification
2017	Araujo et al. (2017)	Twitter, Spotify, Billboard	Economy, Engagement	Album-based, Social Media Metrics	Regression
	Askin and Mauskapf (2017)	Billboard, Discogs, Echo Nest, SoundScan	Top-Charts	Acoustic, Metadata, Artist-based	Regression
	Chiru and Popescu (2017)	YouTube	Engagement	Acoustic, Lyrics-based	Regression
	Herremans and Bergmans (2017)	The Ultrapop 50, Last.fm, EchoNest	Top-Charts	Acoustic, Temporal Info., Listener-based	Classification
	Ren and Kauffman (2017)	Last.fm	Top-Charts	Acoustic, Metadata, Lyrics-based, Rank-based, Artist-based, Social Media Metrics	Classification, Regression
	Yang et al. (2017)	KKBOX	Engagement	Social Media Metrics, Listener-based	Regression
2018	Febirautami et al. (2018)	Spotify	Engagement	Acoustic	Classification
	Interiano et al. (2018)	Official Charts Company, MusicBrainz, AcousticBrainz	Top-Charts	Acoustic, Metadata	Classification
	Lee and Lee (2018)	Billboard		Acoustic	Classification
	Rajyashree et al. (2018)	Million Song Dataset	Engagement	Acoustic	Classification
	Shin and Park (2018)	Gaon Music Charts	Top-Charts	Metadata, Rank-based, Temporal info., Artist-based	Other
2019	Araujo et al. (2019)	Spotify	Top-Charts	Acoustic	Classification
	Cosimato et al. (2019)	Billboard, iTunes, Spotify Twitter, Instagram, Facebook, YouTube, Newspapers	Top-Charts	Social Media Metrics	Classification
	Middlebrook and Sheik (2019)	Spotify, Billboard	Top-Charts	Acoustic, Metadata, Rank-based, Temporal Info., Artist-based, Album-based	Classification
	Silva and Moro (2019)	MusicOSet	Top-Charts	Rank-based, Artist-based	Other
	Silva et al. (2019)	MusicOSet	Engagement	Artist-based	Other
	Yu et al. (2019)	N/A	Engagement	Metadata, Rank-based, Social MediaMetrics	Regression
	Zangerle et al. (2019)	Million Song Dataset, Billboard	Top-Charts	Acoustic, Metadata	Regression
2020	Al-Beitawi et al. (2020)	Spotify	Top-Charts	Acoustic	Other
	Araujo et al. (2020)	Spotify	Top-Charts	Acoustic, Metadata	Classification
	Martín-Gutiérrez et al. (2020)	SpotGenTrack	Engagement	Acoustic, Metadata, Lyrics-based, Artist-based	Classification, Regression

(*continued*).

Table A1. Continued.

Year	Ref.	Data Sources	Success	Features	ML Task
	Matsumoto et al. (2020)	Spotify	Engagement	Acoustic, Artist-based	Classification, Regression
	Oliveira et al. (2020)	Spotify	Top-Charts	Artist-based	Other
	Raza and Nanath (2020)	Billboard	Top-Charts	Acoustic, Lyrics-based	Classification
	Tsiara and Tjortjis (2020)	Twitter, Billboard	Top-Charts	Rank-based, Social Media Metrics	Classification, Regression
2021	S. T. Kim and Oh (2021)	Spotify, Billboard	Top-Charts	Acoustic	Classification
	Kaneria et al. (2021)	Spotify	Engagement	Acoustic	Classification
	Kamal et al. (2021)	Spotify	Engagement	Acoustic, Metadata, Lyrics-based	Classification
	Rosati et al. (2021)	MixRadio	Top-Charts	Metadata, Artist-based	Other
	Gao (2021)	Spotify	Engagement	Acoustic, Metadata	Classification
	Vötter et al. (2021)	AcousticBrainz, Billboard, Million Song Dataset, Last.fm	Top-Charts, Engagement	Acoustic, Metadata	Classification, Regression
2022	Silva et al. (2022, november 7–11)	MusicOSet	Top-Charts	Acoustic, Metadata, Album-based, Artists-based	Classification
	Vötter et al. (2022)	AcousticBrainz, Billboard, Million Song Dataset, Last.fm	Top-Charts, Engagement	Acoustic, Metadata	Classification, Regression