

# Settlement and Distribution Structure Proposal of Music Royalties in Korea and Its AI-based Applications

[Youngmin Kim](#)\*, Dong Hwan Kim, Sunho Park, YongHwa Kim, Jisoo Hong, Sunghee Hong, Jinsoo Jeong, Byounghyo Lee, Hyeonchan Oh

Posted Date: 8 August 2023

doi: 10.20944/preprints202308.0575.v1

Keywords: music sound source; copyright information; log data; artificial intelligence; database; settlement; distribution



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Settlement and Distribution Structure Proposal of Music Royalties in Korea and Its AI-Based Applications

Youngmin Kim <sup>1,\*</sup>, Dong Hwan Kim <sup>2</sup>, Sunho Park <sup>3</sup>, YongHwa Kim <sup>1</sup>, Jisoo Hong <sup>1</sup>,  
Sunghee Hong <sup>1</sup>, Jinsoo Jeong <sup>1</sup>, Byounghyo Lee <sup>1</sup> and Hyeonchan Oh <sup>1</sup>

<sup>1</sup> Korea Electronics Technology Institute, Seoul 03924, Republic of Korea; rainmaker@keti.re.kr

<sup>2</sup> CATBELL COMPANY, Seoul 06526, Republic of Korea; stan@catbellcompany.com

<sup>3</sup> CHOROCK SOFT, Seoul 06123, Republic of Korea; doctorf@chorock.co.kr

\* Correspondence: rainmaker@keti.re.kr; Tel.: +82-2-6388-6695

**Abstract:** Digital music is one of the most important things in market due to the music royalties' distribution in Korea. As the music market is transformed to the digital music market such as download the music and streaming, the distribution of music royalties starting from online service provider (OSP) has come to be a highly important part for music rights holders. One of the most important issues in current music royalties' distribution in Korea is the unfair way of music royalties due to indiscriminate repeat streaming of digital music. To prevent this, music log usage data from several OSPs was collected in day-based system, however, there was a limit to identify detailed information on the use of music in current state. This paper analyzed the structural problems and limitations of settlement of music royalties, and provides a structure in which there can be transparent settlement and distribution between users and rights holders as one of the institutional measures. We'll also propose various AI-based applications using music log usage data. The proposed system will hopefully be used for public purposes.

**Keywords:** music sound source; copyright information; log data; artificial intelligence; database; settlement; distribution

## 1. Introduction

Due to the development of streaming-based web services and mobile devices, the modern music industry is becoming digital. Unlike the past where the music charts were ranked based on offline sales of cassette tapes and CDs, now the number of downloads, streams, and media appearances are becoming more important factors [1–3]. Compared to the analogue sales method based on offline sales, the digital method was expected to provide a clear and efficient alternative to music royalties, but in fact, it is difficult to assume that the settlement or distribution is transparent and efficient in Korea due to a few reasons.

First, as the system changed from analogue to digital, the music industry also made attempts to digitize existing music (UCI issuance). However, each online service provider (OSP), which is one of the elements that dominate the Korean music industry, has a different UCI issued and used, so it was not easy to unify the UCIs.

Second, there are no institutional methods for transparent settlement. For transparent settlement of music, it is necessary to accurately collect how much the music log data have been downloaded and streamed. The data was collected accurately for individual OSPs because OSP provided music service; however, they were reluctant to disclose information such as the number of subscribers and information. After several incidents and lawsuits related with music royalties, the Ministry of Culture, Sports and Tourism, Korea Copyright Commission, and Korea Music Content Association discussed guidelines for the log collection system in 2013 [4]. With the help of discussion, each OSP is to provide log information to Korea Copyright Commission every day. However the log collection system needs to be improved because of unclear management.

Third, each Association could make unsettled dividends because of unclear sales information. In order to completely secure log collection data, additional costs are incurred, which may cause unnecessary disputes. In fact, transparent disclosure of data increases the workload of data generators and collectors such as trust service providers or OSPs, and there is a lack of manpower capable of managing errors for data quality.

This study analyzes and suggests the followings for solving these problems and protecting music users, rights holders, and public interests: 1) Analyzing the structural problems and limitations of calculating and settling music usage, 2) Providing a structure for transparent settlement and distribution between users and rights holders, 3) Supposing several ways based on artificial intelligence technology to utilize the music log data for public purposes, 4) Providing alternatives to regain transparent copyrights and trust in the music industry by presenting legislation plans.

## 2. Collection of Music Usage Information and Analysis of Structural Limitations

Copyright trust service providers have the authority for settlement and distribution since they manage the rights delegated by copyright holders, music record producers, or performers. For example, when they do settlement and distribution of music revenues to copyright holders, the copyright trust service provider sums up the music sales volume for a certain period of time (e.g. three months) and provide the data in the form of a report to each copyright holder. In this case, it is difficult to figure out the settlement details because the settlement reports do not include accurate information and there were number of complex products from OSPs in the reports. In other words, the OSPs calculate the music sales for each user ID (not each music song) and then settle and distribute the music sales and profits to the copyright holders (in this case copyright trust service providers). So, there were evitable errors between music sales profit from user ID and royalties from the song(s).

To solve this issue, the Ministry of Culture, Sports and Tourism, Korea Copyright Commission, and Korea Music Content Association discussed guidelines for log collection, and it was decided in 2013 that each OSP is to provide log information to Korea Copyright Commission every day. However, the log collection system was based on MOU, there is not a decent way to obtain log information or prohibit suspension of transmission. At the moment, they agreed that the log collection system gave information to the music ranking chart website for statistical use, and the settlement and distribution of music royalties was performed by separated systems of Associations. Table 1 is an example of the log item standard field.

**Table 1.** An example of the log item standard field.

Item	Example	Remarks
UCI code	uci:i500-12345678-1234567890-1	-
Granted code	S1234	From UCI
Sales Year/Month/Day/Time	20121001-13:05:30	Sales YYMMDD-HH:MM:SS
Service site	http://www.melon.com	-
Song ID	0258576	Self-categorization
Song	Back in time	Self-categorization
Album code	2031	-
Album name	The moon embracing the sun	-
Artist name	Lyn	-
Service code	D-50	Standardization code from UCI
Distributor name	Recording Industry Association of Korea	-
Distributor song code	A101629692	-
Release day	20120215	Release YYYYMMDD
Producer name	Pan Entertainment	Agency or Producer name
ICN code	ICN.201=1000077661-A101629692	-
ISRC code	KR-A01-99-00137	-

As shown in Table 1, there were five essential elements (UCI code, Song ID (from OSP), Service code (Download/Streaming), Sell time, Asp name) for settlement and distribution of music royalties. According to the regulations on the collection of music royalties of Korea Music Copyright Associations, transmission royalties defines as the higher of following usage formulas using the information [4].

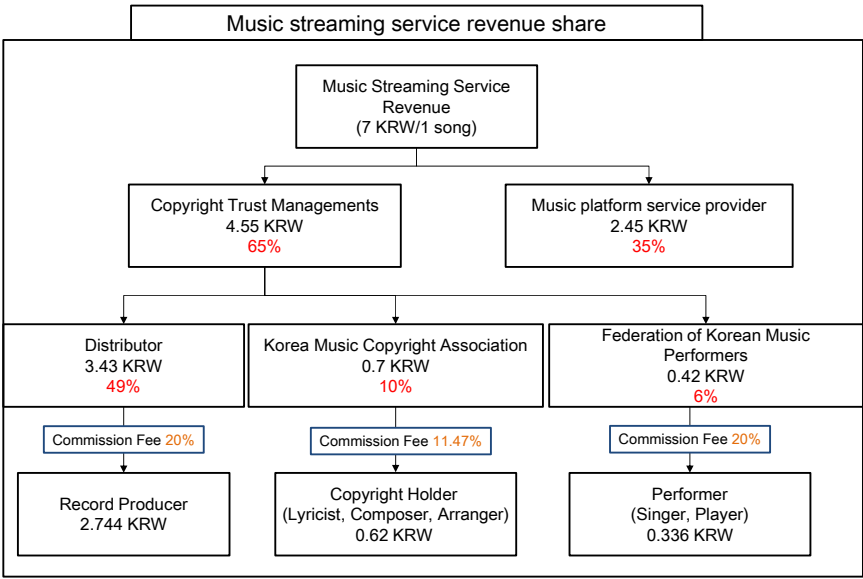
$$\begin{aligned} \text{Royalties} &= (0.7 \text{ KRW (rate per song)}) \times (\text{number of uses}) \times (\text{share or } 700 \\ &\quad \text{KRW per month (rate per subscriber)}) \times (\text{number of subscribers}) \times \\ &\quad (\text{management ratio}) \text{ or} \\ \text{Royalties} &= (\text{Sales}) \times (10.5\% \text{ (music royalty rate)}) \times \text{management ratio} \end{aligned} \quad (1)$$

(For Korea Music Copyright Association, the management ratio is 96%)

Considering this, it seems as if the system could be operating plausibly for settlement since music usage log data was collected by MOU and the transmission royalties are derived according to the formula (1). However, the above system for accurate settlement and distribution of royalties has a few deficiencies due to the several reasons.

First of all, there was an inherent problem with the log collection system. The music usage log information, which is transmitted from major OSPs in Korea on a daily basis based on MOU without institutional support by law, was originally used for statistical use. Therefore, for the issue of settlement, major OSPs delivered the transmission royalties to the each Association based on the above formula and announced sales data. Each Association distributed the royalties based on the portion of music right holders. However, the transmitted music usage log data from OSPs was song-based sales data; the information collected from the log collection system is to show how many times certain music is streamed or downloaded at a certain time, not information on which music the users listened to, and how many times they did. Therefore, it is highly difficult to confirm whether the data was accurate since actual sales data was done by monthly payment per user ID. Moreover, only domestic major OSPs were participating in MOU while global operators such as Google, Apple, and Spotify, etc. are not participating.

The second problem was originated from the profit distribution structure that is focused on big production companies or OSPs. Figure 1 shows an example of the profit distribution structure of streaming music. When the streaming music royalty per song was 7 KRW, the royalty of the copyright holders (lyricist, composer, and arranger) will be 0.62 KRW, while the OSP and the Korea Music Copyright Association will receive 2.45 KRW and 0.7 KRW, respectively [5]. In other words, a substantial amount of the revenue from music sales is likely to be given to major OSPs or famous music record producers (intermediate distributors). Therefore, individual music copyright holders rely on the trust of the copyright trust service providers, and it makes the transmission royalties to be manipulated [6].



**Figure 1.** An example of the music streaming service revenue share.

The third problem may arise from the Personal Information Protection Act. When individual right holder requested the detail report of transmission royalty of music to the Association which plays role in having authority of settlement and distribution of royalty from individual right holder, the report was asked by the Association to major OSPs. However, there were number of complex product music sales in OSPs, so individual right holder could not expect an accurate detailed report. The music sales information, number of subscribers, and subscriber details are key elements that are related to major OSP’s profits, so they may not provide such data accurately. Currently, music usage information was not disclosed to the general public and researchers, and it was not disclosed in detail to even the copyright holders.

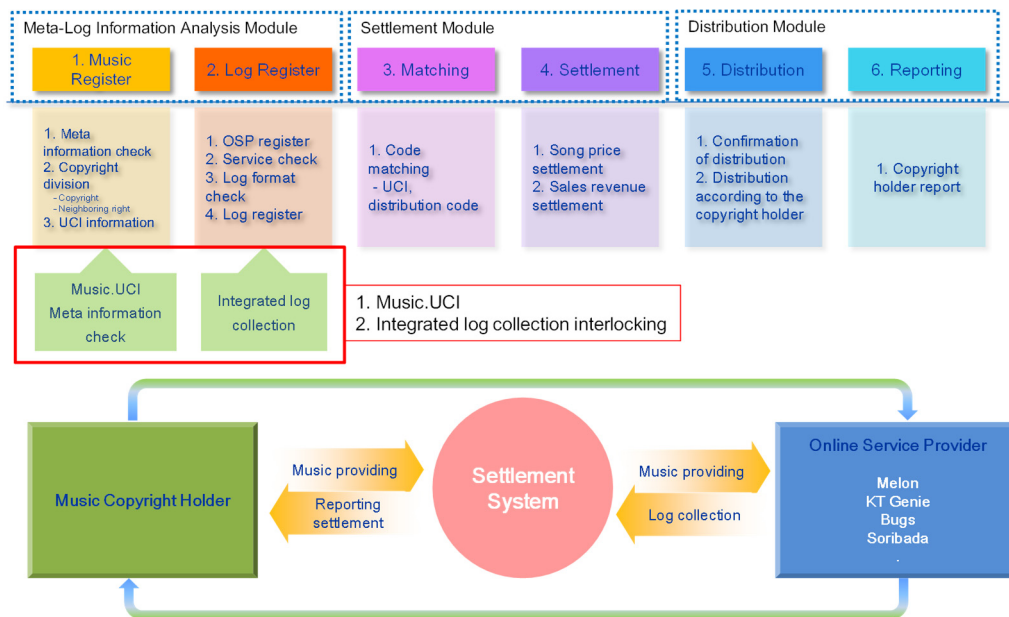
The aforementioned problems could be solved by forcing amendment of law, which is to calculate sales by obtaining both subscriber information of OSPs and PG information of OSPs. However, the information is each OSP’s business key information that is related to their profits, so many different interests were entangled in this information.

**3. Suggesting a Settlement-Distribution Structure for Music Usage**

This study suggests the following alternative structure using the existing log collection system to settle and distribute the royalties for music usage. Because the music rights holders have the right to access the music usage status and information of the rights holders, so this structure will be one of the efficient ways to perform transparent settlement and distribution of royalties.

Figure 2 shows the proposed settlement-distribution structure based on meta-log information analysis. This structure consists of three modules, meta-log information analysis module, settlement module, and distribution module. Those modules could be successfully constructed by spontaneous request of music right holders.





**Figure 2.** Proposed settlement-distribution structure based on meta-log information analysis.

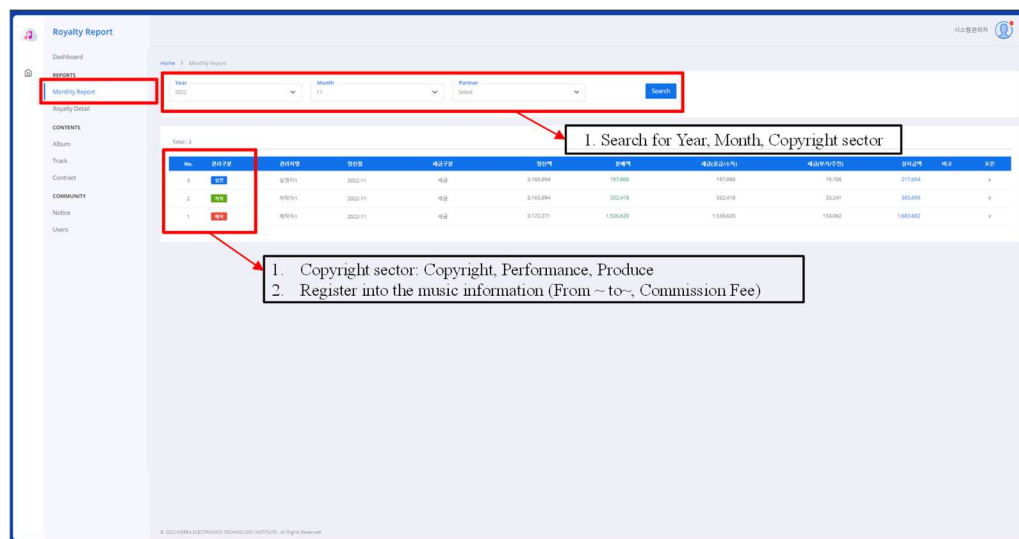
First, meta-log information analysis module performs the copyrighted music registration and log information registration. Only the music copyright holder who has an authority of music usage status and information could register the copyrighted music itself for settlement and distribution. Once the music was registered, the basic metadata could be obtained from Music UCI data of the Recording Industry Association of Korea. The right of a single music source can be divided into copyright and neighboring rights (production rights and performance rights). The current log collection system gathered sales month, OSP classification, and a song code according to the OSP, service code according to the OSP, association type, UCI, and sales volume data. For a settlement, it is necessary that the rate per song and reported settled sum for separated OSP (rate per song multiplied by the number of music usage) will be needed. It is highly complicated issues because it deeply involved stakeholders (e.g., OSPs); however, the music copyright holder could request this information for their rights. Once the issues among the aforementioned stakeholders are solved, the proposed structure could automatically retrieve the music usage log data from the log collection system as long as the music right holders register their copyrighted music.

The second step consists of a match module and a settlement module. A match module performs a matching process between the copyrighted music by registering and music usage log data from log collection system. The settlement process could be done by rate of song or share of total sales of song according to the formula (1). The method of share of total sales of song is higher than that of the rate of song in most cases, and the information could be requested by copyright holders because of their rights, while the information from major OSPs was confidential. Recently several OSPs (e.g., NAVER VIVE, etc.) adopted the settlement method of paying for music usage so that the fee goes directly to the music right holders.

The thirds step consists of a distribution of royalties and a report system. When the settlement for distribution is finalized, the distribution ratio will be determined according to the copyright and neighboring rights in the first step. Based on this ratio, the royalties for the music usage will be distributed to individual rights holders. The reporting system could provide a sum of royalties, royalties from different OSPs, and other information as a chart or a diagram.

Once the music right holders register the copyright music and each OSP transmits the log information for settlement, we could realize transparent settlement and distribution structure because the music right holders have the right to request the sales details. Figure 3 shows an example of building a settlement and distribution system applying these three modules. We assumed that the metadata which was used in the system was generated by random generation from sample log data,

and song price settlement method was applied because the music usage log data was not disclosed in public.



**Figure 3.** An example of building a settlement and distribution system applying all three modules.

#### 4. Applications utilizing the music usage data based on AI

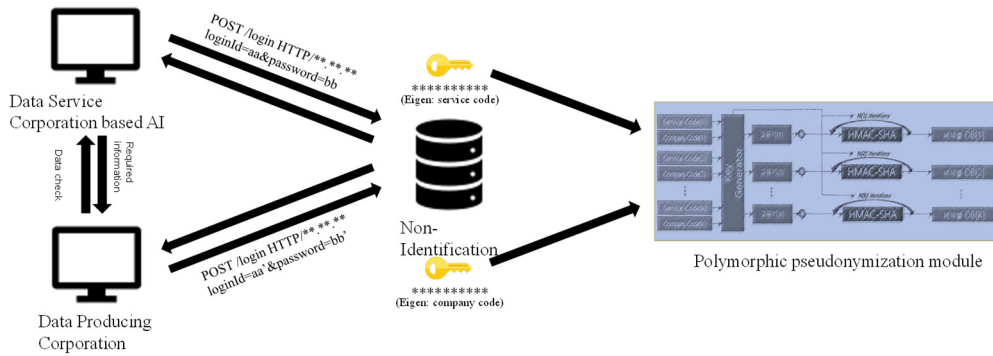
##### 4.1. Pseudonymization of music usage information

The music usage information could be used for various purposes other than settlement or distribution. The usage information data in various fields has a considerable asset value by itself such as MyData services in finance field. However, in the field of music industry, the utilization of music usage information is limited due to the Personal Information Protection Act. The music right holders gave the authority of the music rights to the Associations, and they permit the use of music to the music distributors. Therefore the personal information of the music was fragmented, and no one could try the applications by using the music usage information. However, in the proposed system, the agreement such as usage consent was already constructed by registering the music in the first step, so it would be immune from the legal issues. For the effective protection of personal information, it is more convenient that the pseudonymization of the music usage data was done. This chapter discusses the ways for public services or OSPs to utilize music usage data through pseudonymization before AI-based applications.

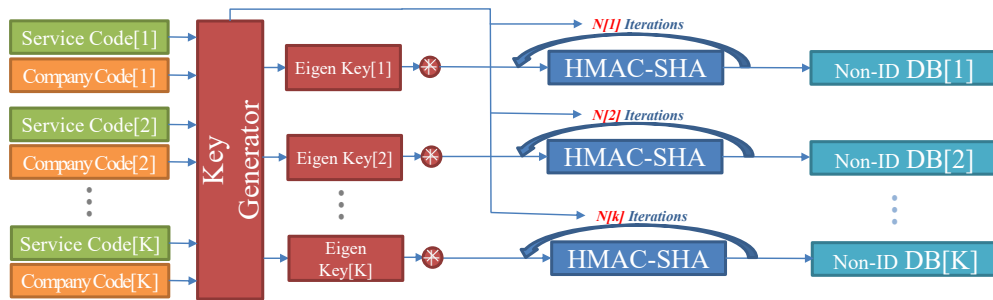
The music usage data contains personal information, it is necessary to de-identify the personal information. De-identification of personal information refers to deleting or modifying part or all of personal information so that a specific individual could not be identified. The related guidelines in Korea provide five methods: pseudonymization, aggregation, data reduction, data suppression, and data masking. These five methods could be divided into anonymization and pseudonymization depending on the deletion of personal identification. Anonymized information has a freedom to be utilized because of impossibility of re-identification, however, it lacks utility. On the other hand, pseudonymization data could not recognize the individuals without additional information. Therefore, it could be used for purposes such as statistics, research, and preservation of records in public interest [7,8].

The subject of pseudonymization of music usage information could be an OSP that produces the data; however, the popular way is to choose the independent organization who could deal with pseudonymization through de-identifying data. Figure 4 shows a process diagram for managing the unique key of the independent organization. The transmitted music usage data is pseudonymized using a JAVA-based high-speed one-way pseudonymization module. Here, the pseudonymization algorithm needs two key functions. First, since the amount of day-based music usage data is massive,

so it is easy to process pseudonymization only when the computing time of the algorithm is extremely short. Second, it is necessary that the same identifier (continuity) between previous and current data should exist. This is because the utility of music usage data increases only when there is continuity of the same music. Figure 5 shows an example of a polymorphic pseudonymization algorithm module which makes de-identified data stable. The advantages of polymorphic pseudonymization process are competitiveness, wide-spread of data usage, and independency of the data. Therefore, it could be helpful for effectiveness of human resource and service improvement.



**Figure 4.** A process diagram for managing the unique key of the independent organization.



**Figure 5.** An example of a polymorphic pseudonymization algorithm module.

#### 4.2. Applications using music usage data– detecting an act of hoarding

##### 4.2.1. Detecting an act of hoarding

There are various applications with pseudonymized music usage data. One can expect that this system could be applied to detect an act of hoarding because outliers in music usage records could be detected. The so-called act of hoarding refers to an illegal act of manipulating the music charts and real-time streaming rankings so that songs of specific singer are ranked. An AutoEncoder model could be developed for detecting abnormal music usage records by using de-identified music usage data [9]. This model discriminates the noise and differences from results that input data makes the output data equal.

We conducted the experiment with around 4 million songs usage data from genie music. This data includes number of listeners, streaming count for each hour, and accumulated number of listeners and streaming count. We added an index to detect fake streaming, *fake-stream-count*, which discriminate fake streaming count from real streaming. The added index (*fake-stream-count*) could be calculated from following equation:

$$(\text{fake-stream-count}) = (\text{play-in-term}/\text{person-max-stream-in-term}) - (\text{listen-in-term}), \quad (2)$$

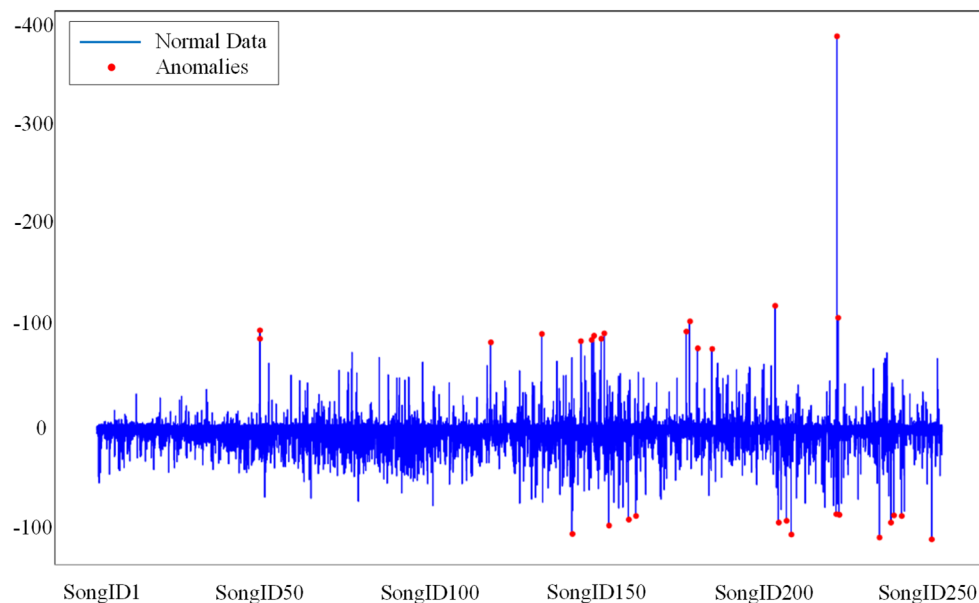
where (play-in-term) is the number of total streaming, (person-max-stream-in-term) is the maximum number of streaming for one person, and (listen-in-term) is the number of listeners per hour. From this index, we conducted the proposed AutoEncoder model through normalization of the index.



Figure 6 shows an example of sample songs from 4 million songs usage data of genie music with several indexes. Among the sample data, there were several anomaly points as dramatically increased as shown in Figure 7. In case of unsupervised deep learning AutoEncoder model, there was no final goal or objective, so it could analyze several negative value of (fake-stream-count) as anomaly points even though it was normal data. According to the results, we could decide if amount of people who streamed the song is smaller than (listen-in-term) amount it means that one person listened specific song repeatedly. Therefore, in this case we could pick them up one of the candidate songs of acting of hoarding.

	index	date	rank	change	sing_id	subject	singer	rank_out_term	play_in_term	listen_in_term	play_total	listen_total	person_max_stream	count_person_stream	fake_stream_count	
	763493	496	2020/05/03/1	247	NEW	89186138	BLUE	하성운	2322	1828341.0	13778.0	11697817	297853	20	91317.05	77339.05
	763496	499	2020/05/03/1	250	NEW	88656476	BIRD	하성운	2513	1986610.0	14350.0	22867617	491911	20	99330.50	84980.50
	1838965	3218	2020/10/29/12	219	NEW	87730738	논사발	장승환	5790	3411068.0	96025.0	47818879	2133539	20	170553.40	74528.40
	2077741	2744	2020/12/08/10	245	NEW	88421579	180도	변	6627	3483172.0	91301.0	54261232	2180574	20	173158.60	81857.60
	2219256	259	2021/01/01/1	10	NEW	81043108	내가 제일 잘 나가	2NE1	8779	1265986.0	155784.0	5540948	777700	20	63299.30	-92484.70
	2260739	5742	2021/01/07/22	243	NEW	87579322	그날처럼	장막철	4254	2657639.0	57417.0	89705216	2751023	20	132881.95	75464.95
	2310989	1992	2021/01/16/7	43	NEW	90612173	죽어 (Who U Are)	강다니엘	1992	1804754.0	13736.0	6978566	20245	20	90237.70	76501.70
	2320044	5047	2021/01/17/20	48	NEW	87538960	My Way	이수 (임비타믹스)	8772	5119215.0	176123.0	54742005	2282870	20	255960.75	79837.75
	2354965	3968	2021/01/23/15	219	NEW	89586709	차가워진 이 바람엔 우리가 버티어 (Bad Love)	HYUN (박혁준)	5395	3091663.0	77352.0	16304241	1018711	20	154583.15	77231.15
	2369383	386	2021/01/26/1	137	NEW	88773174	소우주 (Mikrokosmos)	별한스앤인	6859	3483395.0	92095.0	38532799	1587648	20	174169.75	82074.75
	2390687	5690	2021/01/29/14	191	NEW	90183913	Call Me When It's Over	James Smith	4226	1718742.0	171189.0	2291979	252754	20	85937.10	-85251.90
	2482982	5985	2021/02/13/23	236	NEW	79438991	결혼하 줄까 (Feat. BIZNIZ)	이승기	7623	1024725.0	131097.0	5059889	651043	20	51236.25	-79860.75
	2516708	3711	2021/02/19/14	212	NEW	16174326	내 여자라니까	이승기	7008	696260.0	111663.0	2057326	421460	20	34913.00	-76750.00
	2752232	3235	2021/03/20/20	236	NEW	89196118	UN Village	백현 (BAEKHYUN)	9468	3251717.0	79098.0	18482671	752226	20	162585.85	83487.85
	2767981	2984	2021/04/02/11	235	NEW	69406689	이별탈출	이우	8020	3987982.0	106914.0	14996563	894170	20	199399.10	92485.10
	2804730	2733	2021/04/08/14	234	NEW	88048450	너 없인 안 된다	비투비	7682	3341339.0	98179.0	46367513	1851819	20	167066.95	88887.95
	2871245	4248	2021/04/19/16	249	NEW	88909978	다섯 번째 계절 (SSFWL)	오마이걸 (OH MY GIRL)	6576	3186884.0	90551.0	24385115	1325529	20	159344.20	68793.20
	3167485	488	2021/06/08/1	239	NEW	91300664	그 섬 (Forbidden Island)	하성운	3066	2365685.0	12309.0	6778483	73717	20	118264.25	105975.25
	3182220	3223	2021/06/10/12	234	NEW	80657998	Shock	비스트 (Beast)	9916	1381805.0	151768.0	5088141	655200	20	69090.25	-82677.75
	3219491	4494	2021/06/16/17	245	NEW	72853368	Gee 소녀시대 (GIRLS' GENERATION)	7992	1304843.0	146278.0	7609824	1038059	20	65242.15	-81035.85	
	3241736	2739	2021/06/20/10	240	NEW	89783141	불타 (Spark)	태연 (TAEYEON)	12474	604317.0	123292.0	635877	141575	20	30215.85	-93076.15
	3454972	5994	2021/07/25/23	245	NEW	85561702	자갈가	피아노자갈가	3864	7192670.0	18891.0	35358159	217638	20	359633.50	340742.50
	3460975	5997	2021/07/26/23	248	NEW	89148273	괜찮아도 괜찮아 (That's okay)	디오 (D.O.)	6514	3512209.0	79729.0	31616990	1507860	20	175610.45	95881.45
	3465213	4235	2021/07/27/16	236	NEW	80918696	Supa Dupa Dupa	Dal★Shabet	13421	857060.0	118715.0	2484761	393009	20	42853.00	-75982.00
	3653977	999	2021/08/28/8	250	NEW	33730086	Kissing You	소녀시대 (GIRLS' GENERATION)	11615	1955861.0	193451.0	6444747	815155	20	97793.05	-95657.95
	3707709	1981	2021/09/06/7	232	NEW	82308415	너의 모든 순간 (Piano Ver.)	성시경	10056	965461.0	130741.0	3496174	521941	20	48273.05	-82467.95
	3720971	2343	2021/09/08/12	244	NEW	82617743	해야겠다 만났다	다비치	12363	882846.0	120540.0	7316388	776149	20	44142.30	-76397.70
	3756974	3246	2021/09/14/12	247	NEW	81066203	UGLY	2NE1	10310	2064889.0	180216.0	7966889	870954	20	103249.45	-76966.55
	3897474	3746	2021/10/07/22	247	NEW	80773659	꺼져 줄게 잘 살아 (Feat. 윤문형)	지나	13069	1236504.0	159013.0	3725963	503348	20	61825.20	-97187.80

**Figure 6.** An example of sample songs including usage data with several indexes (several songs and singers were written in Korean).

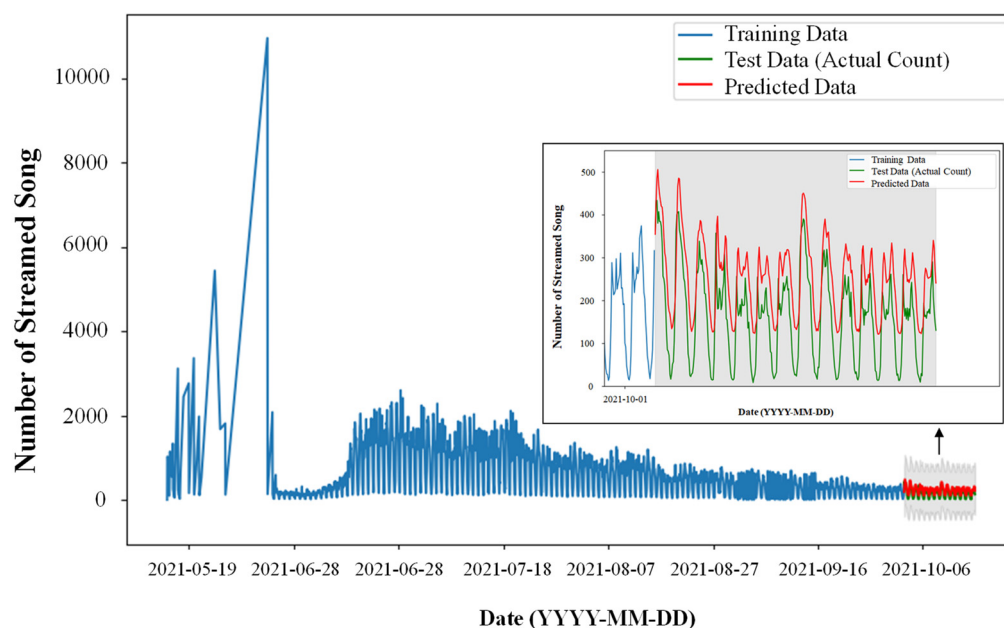


**Figure 7.** An example of anomaly detection of music streaming.  $x$  axis denotes each song and  $y$  axis denotes normalization results of the AutoEncoder model. Red dot indicates anomaly data of streamed song.

#### 4.2.2. Prediction of the marketability of music

Another utilization of music usage data is to predict the marketability of music. Previously, the Korean music industry was limited only to focus music sales. However, the music industry is a fairly large market, so it could be expanded into other fields. For example, by using the music usage data from the past, it could be possible to predict the marketability of specific music or artists, and this could help expand the industry to other fields through prediction based on data such as the Moneyball case in baseball or Opta in soccer.

The music usage data was time-series numerical data and it shows non-stationary so we use ARIMA model for prediction of the marketability of music. The ARIMA model means AutoRegressive Integrated Moving Average, which is effective for analyzing time-series data [10]. However, the music usage data shows non-stationary data, so we applied a difference on data to make it seems like to be stationary. The prediction data could be generated by ARIMA model from python library, and the number of listeners per hours in Figure 6 was needed for prediction. For proper ARIAM model, we adopted AIC (Akaike Information Criterion) score which is criteria that evaluate the quality of statistic model. By using this, we set boundaries of error and improved the accuracy of prediction in terms of distance between prediction data and actual data. Figure 8 shows the validity of our proposed model during five months after releasing specific song. The blue line indicates the number of streaming of the song as a training data. The green line and red line show the difference between actual streaming data of song and predicted data. It shows similar tendency after October of 2021 as shown in Figure 8(inset). As shown in Figure 8, the number of the streamed song shows sharp increment at the releasing time and it was stabilized. We concluded that our model is only effective when this model is independent for external valuables. For example, when the song was exposed by mass media or news, the number of streaming could be increased. Our model could not reflect this; however, it could be supplemented if we added an index of SNS data or news crawling data as an input data.



**Figure 8.** An example of prediction graph based on the music usage during five months (2021-05-19 to 2021-10-06) after releasing song. The blue line indicates training data, the green line indicates test data (actual data), and the red line indicates the predicted data after five months.

## 5. Legislation plan

As mentioned in Sec. 2, the current log collection system collects log data through MOU and there is no compulsion issue. The structural limitations of the current music data usage are due to the structure of the music streaming platform and fragmentation of music copyrights of individuals. In the most music streaming platform, the royalties were paid according to ratio by summing up all the

payments made by all users and dividing them by the number of streaming per song. This would be able to manipulate of sales volume or the music charts using a few accounts, and it may even distort royalty distribution. The fragmentation of music copyrights of individuals could be solved by individuals itself, because they could only request how many times the song was streamed.

Since several embezzlement issues, the Ministry of Culture, Sports, and Tourism publicized the integrated music network and announced that they will revise and improve the system. However, the amended Copyright Act did not have direct clauses on music usage information collection; it is not suitable that the operating plans for log collection system is prepared in the Copyright Act for the current legal system. However, there is a room for adequate law revision by the followings because there were many provisions in the Music Industry Promotion Act.

- Provision location (provision)
- Integrated log collection system operator and duty (Minister of Culture, Sports, and Tourism)
- Obligation to join the log collection system (compulsion)
- Obligation to transmit usage information (technical issues and detailed adjustment necessary)
- Establishment of support grounds
- Others (scope of obligated subscribers, scope and cycle of transmission data, and method)

However, there are several obstacles to this legislation. First, there could be a procedural burden about legislation [11,12]. Many steps will be required in terms of procedure, and the legislation could not be done unless there is anyone who plays a leading role. Second, there could be resistance from the obligated one, which is the music distributor. For example, the amendment of the Music Industry Promotion Act that intended to impose an obligation on music record producers to record performance information was proposed to the National Assembly; however, it ended up failing due to the strong resistance from music record/music-video production business operators. Third, the coordination among the stakeholders is required. Various stakeholders exist due to the nature of the log collection system, and the initiating the government-led system without gathering their opinions will cause many problems or even end up in rejection of participation in the worst-case scenario. Finally, since this system needs to be institutionalized and organized, it makes taxations and resulting in public opinion backlash. Therefore, the legislation should be carried on in consideration of those obstacles. One of the candidates is that this is for comparison like a black box only when there is a problem instead of for settlement.

## 6. Conclusions

This study discussed the limitations and development direction of the music log collection system for transparent settlement and distribution of music royalties. It first analyzed the log collection status and structural limitations of the current music usage information in Korea and suggested a transparent settlement and distribution structure for music usage based on the analysis. The results of constructing a pilot module were derived through the proposed settlement and distribution structure, and the music usage data was pseudonymized to examine several applications by using AI models. Finally this study described ways to legislate this structure as well as the procedural difficulties in legislation. Based on this study, the current log collection system in Korea will hopefully be used for public purposes in a more progressive way.

**Author Contributions:** Conceptualization, Y.K., D.H.K. and S.P.; Funding acquisition, Y.K., Y.K., J.H., S.H., J.J., B.L. and H.O., Investigation and methodology, Y.K., D.H.K., S.P. and Y.K.; Project administration, Y.K., Resources, D.H.K. and S.P.; Supervision, Y.K.; Writing of the original draft, Y.K., D.H.K. and S.P.; Writing of the review and editing, Y.K. and Y.K.; Validation, D.H.K., S.P. and Y.K.; Formal analysis, Y.K., D.H.K., S.P. and Y.K.; Data curation, S.P. and Y.K.; Visualization, Y.K., D.H.K. and S.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports, and Tourism in 2022 (Project: Development of copyright information verification platform utilizing massive sound source log information data, Project Number: CR202104005, Contribution Rate: 100%).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Anderson, C.; The Long Tail: *Why the Future of Business Is Selling Less of More*. New York:Hyperion, United States, 2006.
2. Easley, D.; Kleinberg, J. Power Laws and Rich-Get-Richer Phenomena. In *Networks, Crowds, and Markets: Reasoning about Highly Connected World*, 1<sup>st</sup> ed.; Cambridge University Press: England, 2010.
3. Salganik, M. J.; Dodds, P. S.; Watts, D. J.; Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **2006**, 311, pp.854-856.
4. Copyright Law in Korea. Available online: <http://copyright.or.kr/eng/laws-and-treaties/copyright-law/act.do> (accessed on 31<sup>st</sup> July 2023).
5. Distribution Rules in Korea (amended Sep. 20, 2019). Available online: <http://komca.or.kr/foreign2/eng2/K0403.jsp> (accessed on 31<sup>st</sup> July 2023).
6. Passman, D. S. *All You Need to Know About the Music Business*, 10<sup>th</sup> ed.; Simon and Schuster: New York, United States, 2019.
7. Aamot, H.; Kohl, C. D.; Richter, D.; Knaup-Gregori, P. Pseudonymization of Patient Identifiers for Translational Research. *BMC Med Inform Decis Mak* **2013**, 13, pp. 75-1-75-15.
8. Ribeiro, S. L.; Nakamura, E. T. Privacy Protection with Pseudonymization and Anonymization In a Health IoT System: Results from OCARIoT. In Proceedings of the IEEE 19<sup>th</sup> International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, 28<sup>th</sup>-30<sup>th</sup> October 2019.
9. Zhou, C.; Paffenroth, R. C. Anomaly Detection with Robust Deep Autoencoders. In Proceedings of the 23<sup>rd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada, 13<sup>th</sup>-17<sup>th</sup> August 2017.
10. Ho, S. L.; Xie, M. The Use of ARIMA Models for Reliability Forecasting and Analysis. *Comput Ind Eng* **1999**, 35, pp. 213-216.
11. The 2 Copyrights In a Song (and How To Make More Money from Them). Available online: <https://lawyerdrummer.com/2020/10/the-2-copyrights-in-a-song/> (accessed on 31<sup>st</sup> July 2023).
12. Dicola, P.; Touve, D. Licensing in the Shadow of Copyright. *Stanford Technology Law Review* **2014**, 17, pp. 397.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.