

# 머신러닝을 활용한 음원 차트와 뉴미디어 데이터를 활용한 K-POP 아이돌 인기 요인 분석

<sup>1</sup>최지원, <sup>2</sup>정다연, <sup>3</sup>최강규, <sup>4</sup>임태인, <sup>5</sup>김대훈, <sup>6</sup>정종균, <sup>7\*</sup>노승민

## Analyzing K-POP idol popularity factors using music charts and new media data using machine learning

<sup>1</sup>Jiwon Choi, <sup>2</sup>Dayeon Jung, <sup>3</sup>Kangkyu Choi, <sup>4</sup>Taein Lim and  
<sup>5</sup>Daehoon Kim, <sup>6</sup>Jongkyn Jung, <sup>7\*</sup>Seunmin Rho

### 요약

K-POP 시장은 문화를 넘어 외교, 환경 운동 등 사회 전반에 미치는 영향력이 지대해지고 있다. 이에 따라 아이돌의 성공 요인을 알아내고자 음원, 음반 등 전통적 데이터를 활용하여 머신러닝 기반으로 다양한 논문들이 수행되고 있다. 하지만, 기존의 선행 연구는 최근 아이돌의 인지도에 미치는 인스타그램 릴스, 유튜브 쇼츠, 틱톡, 트위터 등과 같은 뉴미디어 플랫폼의 영향을 반영하지 못했다는 한계점이 있다. 따라서 기존의 연구로는 매일 변화하는 미디어 트렌드를 고려하지 못하여 최근 아이돌 성공 요인의 인과관계를 뚜렷하게 밝히는 데 어려움이 있었다.

이러한 문제점을 해결하기 위해, 본 논문은 아이돌 관련 데이터의 수집 시스템과 분석 방법론을 제안한다. 아이돌 데이터의 특이성을 반영한 컨테이너 기반 실시간 데이터 수집 자동화 시스템을 개발해, 아이돌 데이터 수집의 안정성과 확장성을 확보하고 K-Means 클러스터링 기반 이상치 탐지 모델을 통해 성공 아이돌 군집을 비교, 분석한다.

그 결과, 성별, 앨범 발매 시기 후 성공 시점, 뉴미디어와의 연관성 등 성공 아이돌들의 공통점을 파악할 수 있었다. 이를 통해, 최종적으로 각 아이돌별, 앨범 형태별, 컴백 시기에 따른 최적 컴백 프로모션을 기획해 아이돌의 성공 가능성을 증진할 수 있을 것으로 기대한다.

### Abstract

The K-POP market has become influential not only in culture but also in society as a whole, including diplomacy and environmental movements. As a result, various papers have been conducted based on machine learning to identify the success factors of idols by utilizing traditional data such as music and recordings. However, there is a limitation that previous studies have not reflected the influence of new media platforms such as Instagram releases, YouTube shorts, TikTok, Twitter, etc. on the popularity of idols. Therefore, it is difficult to clarify the causal relationship of recent idol success factors because the existing studies do not consider the daily changing media trends.

To solve these problems, this paper proposes a data collection system and analysis methodology for idol-related data. By developing a container-based real-time data collection automation system that reflects the specificity of idol data, we secure the stability and scalability of idol data collection and compare and analyze the clusters of successful idols through a K-Means clustering-based outlier detection model.

As a result, we were able to identify commonalities among successful idols such as gender, time of success after album release, and association with new media. Through this, it is expected that we can finally plan optimal comeback promotions for each idol, album type, and comeback period to improve the chances of idol success.

**Keywords:** Idols, K-POP, data collection systems, new media platforms, success factors of idols

<sup>1</sup> 중앙대학교 산업보안학과, 학부연구원 (victoria3050@cau.ac.kr)

<sup>2</sup> 중앙대학교 산업보안학과, 학부연구원 (dayeon620@kakao.com)

<sup>3</sup> 중앙대학교 산업보안학과, 학부연구원 (c020602@naver.com)

<sup>4</sup> 중앙대학교 산업보안학과, 학부연구원 (taeng02@icloud.com)

<sup>5</sup> LLOYDK, 이사 (reonian@lloydk.co.kr)

<sup>6</sup> LLOYDK, CEO (david@lloydk.co.kr)

<sup>7\*</sup> 교신저자 중앙대학교 산업보안학과, 교수 (smrho@cau.ac.kr)

## I. 서론

2023 년 현재, K-POP 시장은 매년 빠르게 성장하고 있다. 손열(2020)에 따르면 유명 아티스트들이 문화를 넘어 외교, 환경 운동 등 사회 전반에 미치는 영향력이 크게 증가하고 있으며[1][2], 다수의 소속사들이 이러한 아티스트들을 육성하고자 큰 노력을 기울이고 있다. 하지만, 이러한 노력에도 불구하고 실제로 [3] 성공하는 아이들은 극소수에 불과하다. 이에 따라 아이들의 성공 요인을 알아내고자 하는 시도들이 다양하게 수행되고 있다. 그러나, 현재까지 수행된 연구에서는 음원 및 음반 데이터와 같은 전통적 데이터를 종속 변수로 사용할 때, 휘발성 데이터라는 이유로 주간 차트를 사용하여 예측하는 그쳤다. 또한, 아이돌 생존 가능성에 초점을 두어, 역설적으로 아이돌 생존 가능성에 원인이 되는 아이돌 성공 요인을 파악하지 못하는 한계가 존재했다[3]. 따라서, 기존의 연구로는 뉴미디어 시대 스포츠 콘텐츠를 중심으로 빠르게 트렌드가 확산되고, 각종 SNS 알고리즘으로 인해 아이돌 인지도의 급상승이 일어나고 있는 현재 트렌드를 반영하지 못하여 구체적인 아이들의 성공 요인을 파악하기에 한계가 존재한다 [4][5].

기존 선행 연구에는 음원 데이터(멜론 차트), 인터넷 기사, SNS 데이터 등을 수집하여 소셜 데이터 분석을 통한 음원 흥행 예측 지표 연구를 실행하고자 하였다[6]. 하지만, 예측하고자 하는 독립변수가 주간 차트 순위인 음원 데이터인 점에서, 매일 변화하는 현재의 뉴미디어 트렌드를 반영하는 것에 한계가 존재하였다.

또한, 머신러닝 기법을 활용한 아이돌 생존 가능성을 예측한 선행 연구도 존재하지만, 소속사 규모, 유닛 활동 여부 등 아이돌 소속사 및 아이돌 인원 구성과 관련된 데이터를 주로 독립변수로 활용하여, 역설적으로 최근 아이돌 생존 가능성의 원인이 되는 성공 요인을 파악하지 못하는 한계가 있다[7].

이에 본 논문에서는 데이터 수집 자동화 시스템을 통해 음원 및 음반 사이트 측에서 제공하지 않는 휘발성 성격의 일간 차트를 수집하고자 한다. 또한, 휘발성 데이터인 트위터(SNS) 일간 팔로워 수를 수집하여 동일한 시간에 휘발성 데이터를 주기적으로 수집하지 못하여 일간 데이터를 사용하지 못한 선행 연구의 한계를 보완하고자 한다. 나아가, 데이터 수집 자동화 시스템을 이용하여 수집한 데이터를 활용하여 아이돌 군집별 이상치 탐색 모델을 제안한다.

본 모델을 통해 현재 트렌드에 근간한 성공 아이돌의 공통점을 도출하고 새로운 인사이트를 제시하고자 한다. 이를 위해 아이돌의 미디어 데이터를 수집한 뒤 군집별 이상치 탐지를 위해 K-Means 기반의 클러스터링을 활용한다. 이후 각 군집별로 y 값에 대한 Linear Regression 의 가중치를 기준치로 활용해 각 노드간 변화율을 비교하여 이상치를 탐지한다. 마지막으로 해당 이상치에 대해 뉴미디어 데이터 등을 근간으로 새로운 마케팅 전략을 제안하고자 한다.

본 연구는 휘발성이 강하고 수집이 어려운 아이돌 데이터를 집계할 시스템을 구축하고, 아이돌의 이상치에 대한 원인을 해석함으로써 아이돌 성공에 대한 다양한 요소들의 상대적인 영향력을 더 정확하게 이해하고 분석할 수 있을 것으로 기대한다.

## II. 연구방법

### 2.1 데이터 선정(Selecting Data)

#### 1) 데이터 구성(Data composition)

본 연구에서는 아이돌의 성장 이상치를 파악하기 위해 2022 년 한터차트 기준 초동 판매량이 많은 남자 16 개 그룹, 여자 15 개 그룹 총 31 개의 그룹을 선정하여 연구를 진행하였다. 초동 판매량은 음반 발매 이후 7 일 이내 팔린 음반 판매량을 의미하며, 아이돌의 ‘팬덤 인기’를 나타내는 핵심 지표이다[8]. 해당 순위에서 아이돌 그룹이 아닌 솔로 가수는 제외하여 진행하였다. 선정된 가수들과 관련된 데이터는 멜론 차트[9], 한터 차트, 트위터, 인터파크 공연 예매 차트, yes 24 공연 예매 차트,[10] 네이버 프로필에서 크롤링하여 수집하였다.

Table 1. Collected dataset  
표 1. 수집된 데이터셋

	Variable name	description	collection
Independent variable	mln_chart_out_count	Number of chart outs for songs on Melon albums	melon
	release_after	Release date	melon
	album_type_encoded	album type (regular, mini, single)	melon
	genre_encoded	genre type	melon
	artist_encoded	artist type	melon
	mln_heart	Melon album heart count	melon
	group_type	group type(female, male idol)	melon
	twl_flwrs	Twitter follower count	twitter
	twl_flwrs_diff	daily Twitter follower change	twitter
	hnt_daily_sales	daily album sales(Hateo chart)	hanteo
	concert_bool	concert availability	Interpark, yes24
dependentvariable	mln_chart_num	Melon daily chart rank one day prior	melon

2) Data preprocessing

수집한 데이터 셋의 범주형 데이터는 모델 학습을 위해, 문자형(Character) 변수들을 원-핫 인코딩(One-Hot Encoding)을 수행하여 해당 속성에 해당 될 경우 1, 그렇지 않을 경우 0 으로 변수를 생성하였다. 부울(Boolean) 변수들은 Yes 인 경우 1, No 는 0 으로 변수값을 변경하였다. 이처럼 수집한 데이터를 통해 각 아이돌별 군집을 구성한 뒤 이상치를 탐지한 모델을 구축하였다.

2.2 데이터 수집 자동화 시스템

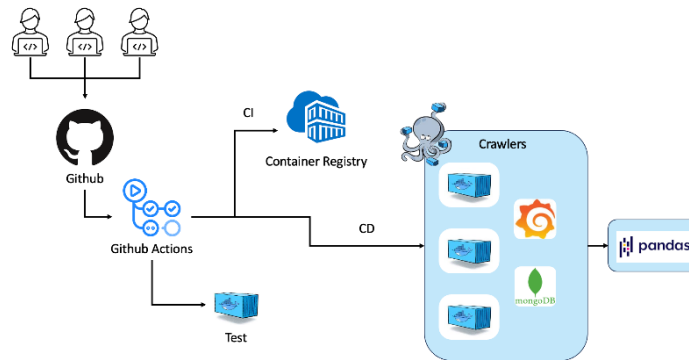


Figure 1. Purpose of idol status analysis - Large-scale data collection automation system  
그림 1. 아이돌 현황 분석의 목적 - 대규모 데이터 수집 자동화 시스템

본 논문에서 수집하고자 하는 데이터에는 주기적으로 수집되어야 하는 멜론 앨범과 관련된 데이터, 트위터와 관련된 데이터가 존재한다. 이에 따라 매일, 혹은 매시간마다 데이터를 빠짐없이 수집할 수 있도록 자동화 시스템을 설계하였으며 아이돌 상태분석을 위한 대규모 데이터 수집 자동화 시스템을 위한 구성도는 Figure 1 과 같다.

시스템의 기초를 이루는 크롤러는 Python 을 사용하여 개발되었다. 개발된 크롤러는 컨테이너 환경에서의 운영을 위해 처리해야 할 다양한 옵션들을 공통 모듈로 추상화함으로써, 개발자가 크롤링 함수 개발에 더욱 집중할 수 있도록 설계되었다. 데이터 수집 주기는 아이돌 데이터의 동향을 반영하여 일일 주기와 특정 시간대를 목표로 설정하였고, 이는 데이터베이스에 안정적으로 저장될 수 있도록 공통 모듈에 통합되었다.

공통 모듈과 개별 크롤러는 GitHub 을 통해 버전 관리되며, Pull Request 및 Merge 작업시 GitHub Actions 를 이용해 빌드 및 크롤링 테스트를 자동화함으로써 크롤러의 코드 통합 및 테스트 과정을 효율적으로 구성하였다. 코드가 Merge 되면, Docker Image 가 자동으로 빌드되어

Container Registry 에 업로드되며, 이후 새로운 크롤러 컨테이너가 타깃 서버에 자동으로 배포된다. Docker Compose 는 크롤러, 데이터베이스, 모니터링 도구 등을 통합 관리하여 연속적인 운영을 지원한다.

초기 인프라에서는 아이돌 데이터들의 특성에 따라 비정상적인 종료 없이 연속적으로 수집이 되어야 하지만 컨테이너 안의 크롤러 어플리케이션 작업 상태 파악 문제와 웹사이트 변경 시 재시작 조건 설정 문제, 컨테이너 및 어플리케이션 상태에 따른 재배포 문제 등의 운영상 문제가 발견되었다. 특히, 본 논문의 핵심 데이터인 음원 및 뉴미디어 데이터 수집 과정에선 요청 제한 및 캡차 시스템과 같은 외부 장벽에 직면하였고, 작업 진행 상태 모니터링에 어려움을 겪었다.



Figure 2. Improved large-scale data collection automation system  
그림 2. 대규모 데이터 수집 자동화 시스템 개선

이러한 문제를 해결하기 위해, Figure 2 와 같이 Docker Compose 에서 Kubernetes 로 이전해 관리하는 것을 제안한다. 이는 보다 세밀한 컨테이너 관리를 가능하게 하며, 컨테이너의 재시작 및 크롤러 상태 모니터링을 단순화시켜 줄 수 있다. 또한 Slack 및 KakaoTalk 을 통한 실시간 알림 기능을 통해 작업 진행 상황을 쉽게 파악할 수 있도록 개선할 수 있으며, ELK(Elastic Search, Logstash, Kibana) 스택을 활용하여 수집된 데이터의 로그 관리와 분석을 자동화함으로써, 데이터의 처리 및 대시보드 작성 과정을 더욱 효율적으로 만들 수 있다.

결과적으로, 이러한 시스템 설계 및 인프라의 개선을 통해 아이돌 데이터 수집 및 처리 작업의 자동화에 대한 신뢰성과 효율성을 크게 향상시킬 수 있다. 이와 함께, 수집 범위가 다양해질 수 있는 아이돌 데이터의 특성상, 새로운 컨테이너 배포와 기존 컨테이너의 수정 등이 원활히 진행될 수 있게 함으로써 본 논문에서 제안하는 모델의 신뢰성 및 안정성을 크게 향상시키는 데 기여할 수 있다.

### III. 모델 구성

#### 3.1 인기 척도 군집화 모델

##### 1) 목적

본 논문은 아이돌 그룹의 성공 여부에 따른 음원 발매 기간 차이를 이용하여 군집별 이상치 탐지 모델을 구성한다. 실제로 2013년 데뷔 직후 컴백 기간이 엑소(14개월), 소녀시대(14개월), 샤이니(11개월)에 비하여 2023년 아이돌 그룹의 경우 라이즈(1개월), 엔팀(3개월), 보이넥스트도어(3개월), 제로베이스원(4개월)으로 과거에 비해 그 빈도가 줄었다. 즉, 성공한 1군 아이돌의 경우 음원 발매의 폭이 크고, 데뷔한지 2년 이내의 아이돌의 경우 음원 발매를 빈번하게 한다는 특성을 볼 수 있다.

따라서, 데뷔 시기 및 성별 등에 따른 아이돌 별 음원 발매 특성이 다른 점을 고려하여 같은 특성끼리 묶도록 군집화 모델을 도입하고자 한다. 해당 군집화 과정에서는 비슷한 음원 데이터 군집으로 빠르게 분류할 수 있는 거리기반 알고리즘 중 하나인 K-Means 를 사용하였다.

## 2) 활용 데이터

2023년 4월 17일부터 2023년 6월 24일까지 약 3개월간 수집한 멜론 데이터를 기준으로 발매일로부터 60일 이내의 행만 추출하여 데이터를 재구성하였다. 발매일 기준 60일 이내의 723개의 행을 통해, 음원/반 발매일로부터 60일 이내 아이돌의 발매 앨범별 성공 지표인 음원 데이터를 사용하였다.

## 3.2 시계열 기반 아이돌 성장 이상치 탐지 모델

### 1) 목적

앞에서 얻어진 군집 결과를 토대로 음원 차트를 기준으로 한 군집별 성장에 대한 이상치를 탐지하는 모델을 구성한다. 본 모델은 음원차트 성장 추이에 대한 이상치를 파악해 군집별로 세분화된 인기 및 성공 변화 이상치를 최신 트렌드를 반영하여 파악한다.

### 2) 이상치 탐지 방식

본 논문에서는 이상치를 파악하는 방식으로 멜론 차트 내 하트 수의 증가 추이와 전체 평균 변화율을 통해 추출하여 본 수치 값이 특별히 높은 경우, 그 이유에 대해 트렌드 분석과 결부하여 해석한다.

### 3) 활용 데이터

K-Means 데이터와 마찬가지로 약 3개월간 수집한 멜론 데이터를 기준으로 재구성한 데이터를 사용하였다. 이때 이상치의 추이를 분석하는데 있어서 인기를 명시적으로 보여주는 음원 차트 순위와 관련된 `mln_chart_num`, `mln_heart`, `mln_heart_diff` 데이터를 기준으로 두어 해당 수치가 비정상적인 추이를 띄는 구간에 대해 분석하고자 하였다.

## IV. 연구 결과

### 4.1 인기 척도 군집화 모델

Table 2. Clustering results using K-Means

표 2. K-means 을 사용한 클러스터링 결과

Artist	Cluster
SEVENTEEN	0
THE BOYZ	0
SEVENTEEN	1
THE BOYZ	2
ONEUS	2
ONEUS	0
VERIVERY	0
TEMPEST	0
CRAVITY	0
BTOB	0
STRAY KIDS	1
P1Harmony	0
BTS	1
ATEEZ	0
NMIXX	0
TWICE	0
STAYC	0
IVE	0
Kep1er	0

Table 2는 K-Means 클러스터를 이용해 임의로 군집을 3으로 설정한 뒤 각 아이돌별 군집을 확인한 내용이다. 본 결과는 한 아이돌이 두 개의 군집으로 분류되는 결과가 소수 존재한다.

이는 수집된 기간 중 음원 데이터가 여러 종류일 때 두 가지 군집으로 분류되는 현상이 반영되었다.

위의 군집 데이터의 경우 비정상적인 추이를 보이는 음원만 따로 분류하여 음원과 비슷한 시기의 트렌드를 결합하여 해석하고자 한다.

## 4.2 아이돌 별 성장 추이 분석

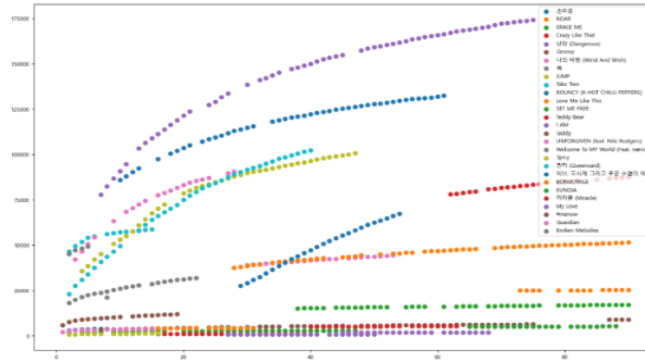


Figure 3. Visualization of mln\_heart based on release\_after  
그림 3. release\_after 에 기반한 mln\_heart 시각화

Figure 3 는 이상치의 존재 여부를 확인하고자 수집 데이터 기간 동안의 모든 앨범에 대해 mln\_heart 을 기준으로 시각화 한 결과이다. 이는 앞서 설명한 것과 같이 수집한 데이터의 release\_after 기준이 모두 다르기에 시작점이 다른 개형을 보인다는 문제가 발생한 것을 확인할 수 있다.

또한 파란색 점 그래프인 SEVENTEEN-손오공의 경우와 같이 타 변화율에 비해 비교적 가파른 증가 추이를 보이는 이상치를 탐지하고 이상치에 대해 뉴미디어 데이터를 기반으로 정성적 분석을 진행하는 것이 목적이다.

따라서 군집별 mln\_heart, mln\_heart\_diff, mln\_chart\_num 값의 비교를 통해 아이돌의 성장 이상치를 아래와 같은 단계로 파악하였다.

- ① mln\_heart, mln\_heart\_diff, mln\_chart\_num 의 누적합 그래프를 이용해 군집별 평균 변화율을 계산한다.
- ② 군집별 평균 변화율을 임계치로 설정한 뒤, 노드간 순간 변화율이 큰 경우에 대해 이상치로 탐지한다.
- ③ 모든 그래프에서 공통적으로 이상치로 탐지한 값을 아이돌 성장 변화 이상치로 탐색한다.
- ④ 탐색된 이상치를 기준으로 데이터 기반 정성평가를 진행한다.

Table 3. Average change rate by cluster  
표 3. 클러스터별 평균 변화율

Cluster	mln_heart	mln_heart_diff	mln_chart_num
0	459.809	98.468	354.080
1	123.982	2884.852	3298.743
2	542.416	1064.349	1243.8326
3	16 MB1475.451	1486.280	1512.446

클러스터별 평균변화율은 LinearRegression 을 이용해 클러스터별로 계산하여 변화율의 threshold 로 사용하였다.

위에서 우리는 시계열 변수로 사용 중인 release\_after 가 수집 앨범별 데이터에서 동일하지 않음을 확인하였다.

따라서 아이돌 앨범별로 데이터를 확인하여 이상치를 정리해 공통점을 분석하고, 정성적으로 해당 이상치를 분석을 추가적으로 진행하였다.

### 4.3 시계열 기반 아이돌 성장 이상치 탐지 그래프

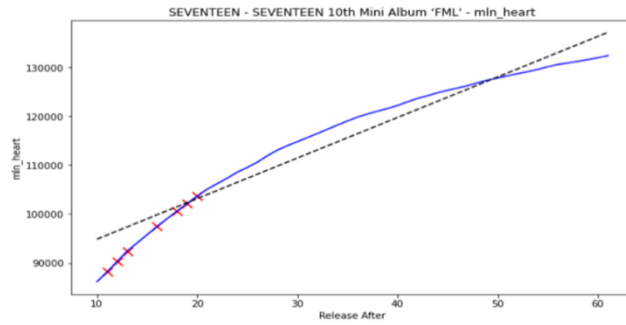


Figure 4. Outliers for mln\_heart in SEVENTEEN's 'FML' album  
 그림 4. 세븐틴의 'FML' 앨범에서 mln\_heart에 대한 이상값

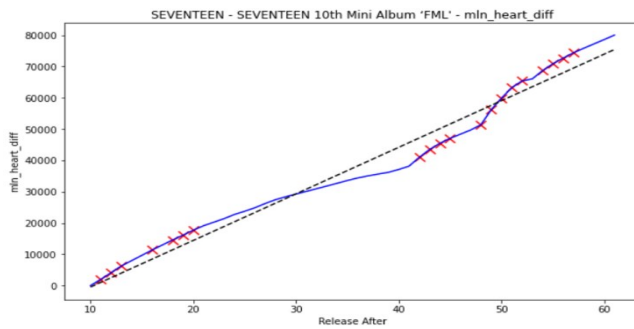


Figure 5. Outliers for mln\_heart\_diff in SEVENTEEN's 'FML' album  
 그림 5. 세븐틴의 'FML' 앨범에서 mln\_heart\_diff에 대한 이상값

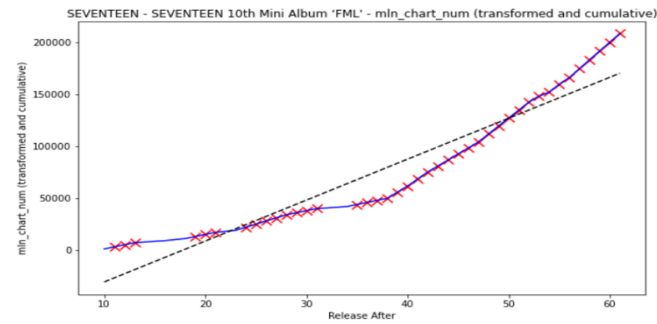


Figure 6. Outliers for mln\_chart\_num in SEVENTEEN's 'FML' album  
 그림 6. 세븐틴의 'FML' 앨범에서 mln\_chart\_num에 대한 이상값

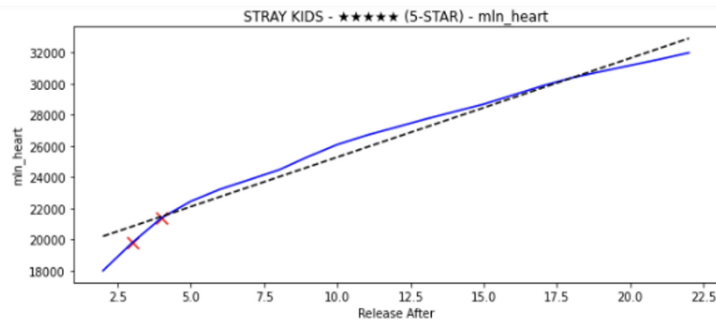


Figure 7. Outliers for mln\_heart in STRAY KIDS's '5-STAR' album  
 그림 7. 스트레이 키즈 '5-STAR' 앨범의 mln\_heart에 대한 이상값

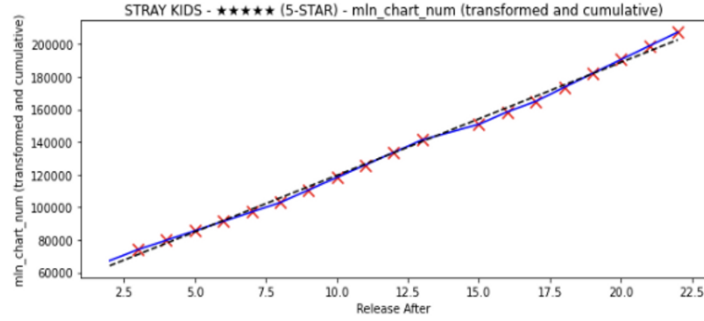


Figure 8. Outliers for mln\_chart\_num in STRAY KIDS's '5-STAR' album  
 그림 8. 스트레이 키즈의 '5-STAR' 앨범에서 mln\_chart\_num 에 대한 이상값

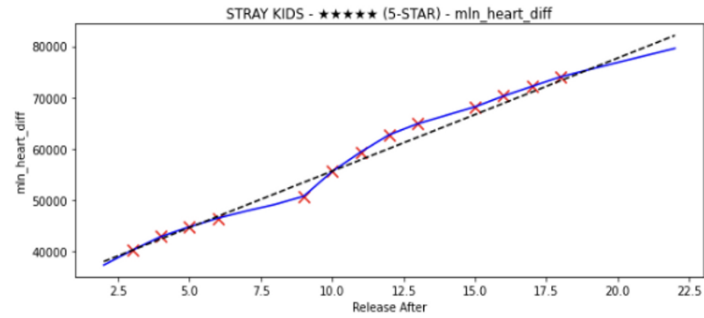


Figure 9. Outliers for mln\_heart\_diff in STRAY KIDS's '5-STAR' album  
 그림 9. 스트레이 키즈의 '5-STAR' 앨범의 mln\_heart\_diff 에 대한 이상값

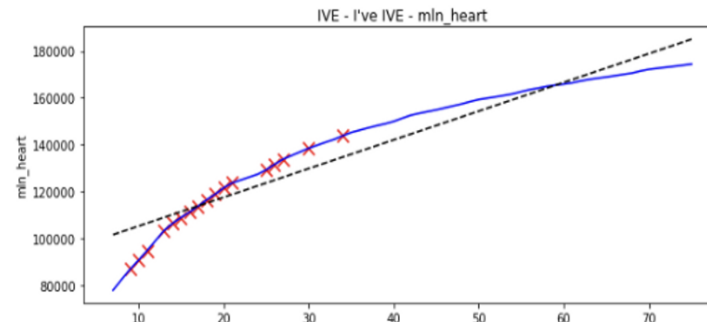


Figure 10. Outliers for mln\_heart in IVE's 'I've IVE' album  
 그림 10. IVE 의 'I've IVE' 앨범의 mln\_heart 에 대한 이상값

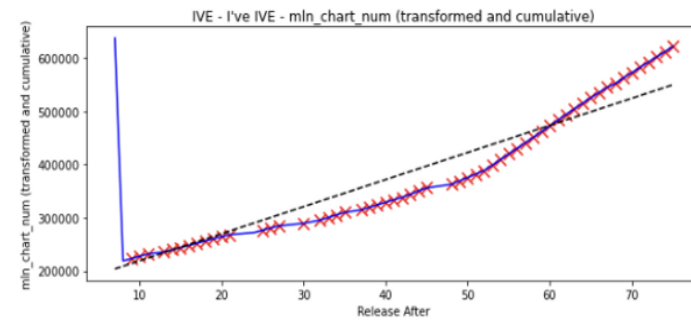


Figure 11. Outliers for mln\_chart\_num in IVE's 'I've IVE' album  
 그림 11. IVE 의 'I've IVE' 앨범의 mln\_chart\_num 에 대한 이상값



Figure4 ~ Figure12 에서의 x 는 release\_after, y 는 각각의 mln\_heart, mln\_heart\_diff, mln\_chart\_num 를 의미하며 이를 누적하고 합한 그래프이다. 해당 그래프에 대한 결과를 분석하였다.

Table 4. Set of idols with common anomalies  
표 4. 일반적인 이상치를 가지는 아이돌 집합

Type	result
SEVENTEEN - FML	11,12,13,19,20 days
STRAY KIDS - ★★★★★	3,4 days
BTS - Take Two	3,4,5 days
IVE - I've IVE	9, 10, 11, 13-21, 25-27, 30, 34 days
LE SSERAFIM - BASTIONS	11-14, 16, 20, 29, 30-44, 48, 49 days
aespa-MY WORLD	9-17, 20-21 days
G IDLE - I feel	9-10, 13-24, 27-28 days

Table 4 를 확인해보면 일반적으로 여자 아이돌의 경우 발매 이후 9~11 일 사이 최초로 이상치가 탐지되고, 남자 아이돌의 경우 3~4 일 내에 이상치가 탐지되는 것을 확인할 수 있다.

#### 1) SEVENTEEN

2023 년 4 월 24 일 발매된 남자 아이돌 그룹 SEVENTEEN 의 미니 10 집 'FML' 앨범의 경우 음원 발매일로부터 11~13 일, 19~20 일 사이에 공통 이상치가 탐지되었다. 남자아이돌의 경우, 음원 발매일로부터 7 일 이내 음반 판매량이 남자아이돌의 경우 팬덤에 더 편향되어 초동 기간 음반 판매량이 팬덤의 규모 즉, 팬덤 충성심을 증명하는 것이라고 여겨진다. 일반적으로 발매일로부터 1 주일 이내에 이상치가 파악되었다[11].

해당 앨범의 타이틀 곡 '손오공' 숏폼 챌린지가 열풍이 영향을 주었다는 것을 파악했다[12]. 또한, 발매일로부터 11 일 이후 유튜브 채널 '십오야' 에서 '세븐틴 편' 이 인기를 끌었음을 파악하였다. 이를 통해, '뉴미디어를 통한 직·간접적 홍보 효과가 상당한 영향을 준다는 것을 확인하였다.

#### 2) STRAY KIDS

2023 년 6 월 2 일 발매한 남자 아이돌 그룹 'STRAY KIDS'의 정규 3 집 '★★★★★(5-STAR)'의 경우 음원 발매일로부터 3~4 일 이내에 공통 이상치가 탐지되었다. 이는 음원 발매일로부터 7 일 이내 음반 판매량이 폭증하는 남자 아이돌의 특성에서 기인했음을 확인하였다.

#### 3) IVE

2023 년 4 월 10 일 발매한 여자 아이돌 그룹 'IVE' 의 정규 1 집 'I've IVE' 의 경우 음원 발매일로부터 9~11 일, 13~21 일, 25~27 일, 30, 34 일에 공통 이상치가 탐지되었다. 일반적으로 대중적인 여자 아이돌의 경우 음원 발매일 이후 10~14 일 이내 공통 이상치가 탐지된다는 특징을 파악했다. 이러한 패턴은 'I've IVE' 의 타이틀 곡 'I AM' 에도 동일하게 적용되었으며, 최근 젊은 세대 사이에서 유행하는 '틱톡'과 'YouTube Shorts' 같은 숏폼 챌린지가 이러한 이상치에 영향을 미친 것으로 확인할 수 있었다. 일반적으로 숏폼을 활용한 챌린지는 음원 발매 이후 7 일 이상 지난 경우 일반적으로 유행이 시작되는데, 이와 같은 유행은 'I've IVE'의 경우에도 관찰되었다.

또한, 앨범 발매일로부터 2 주 이후에 나타난 공통 이상치를 분석한 결과, '4 세대' 여자 아이돌 그룹의 경우에는 앨범 활동 이전에 수록곡 중 일부를 먼저 공개하는 '선공개 곡'이 활동에 미치는 영향을 확인할 수 있었다. 이를 통해, 선공개 곡을 발표한 후에 본 앨범의 타이틀 곡 활동을 진행하면, 대중의 관심을 오랜 기간 동안 유지할 수 있음을 밝혀냈다[13][14].

종합적으로, 현재는 데이터 부족으로 인한 release\_after 의 통일이 되지 않았고, 이로 인해 군집화와 지도/비지도 학습 모델, 본 논문의 최종 모델인 이상치 탐지 모델까지 모두 세부적인 시계열 분석의 한계를 갖지만, 수집한 데이터 범위 내에서 최선의 과정과 결론을 도출하였다.

#### 4) 분석 결과

공통 성장 이상치가 나온 그룹을 추출해서 분석해본 결과 각각 남자, 여자 그룹에서 현재 가장 인기가 많은 것으로 여겨지는 그룹들이 추출되었음을 확인할 수 있었다. 실제로 남자 그룹에서는 ‘세븐틴’, ‘스트레이키즈’, ‘방탄소년단’ 이 여자 그룹에서는 ‘르세라핌’, ‘아이브’, ‘에스파’, ‘여자 아이들’이 현재 트렌드를 통해 성장 이상치를 보이고 있음을 파악할 수 있다.

그룹별 비교·분석 결과, 이상치 탐색 모델에서 남자 아이돌과 여자 아이돌 간의 음원 발매 이후 이상치 발생 시기가 달랐다. 남자 아이돌의 경우 음원 발매일 이후 3~4 일 이내, 여자 아이돌의 경우 9~11 일 사이 최초로 이상치가 탐지됨을 파악하였다.

이러한 차이는 K-POP 산업에서 ‘아이돌 그룹의 성별’에 따른 특성 차이로 해석되는데, 남자 아이돌 그룹은 전통적으로 ‘팬덤’이 강조되어 음원 발매와 동시에 음반 판매가 집중되는 초동 기간과 동일하게 발매 후 7 일 이내에 이상치가 발생하는 경향을 보인다. 반면 여자 아이돌 그룹은 ‘대중’의 선택을 받아 음원 시장에서 강세를 보이며, 최근 10-20 대 사이에서 유행하는 ‘틱톡’과 ‘YouTube Shorts’ 등의 숏폼 챌린지의 영향을 받아 음원 발매 후 10 일 이후부터 이상치가 탐지되었다.

## V. 결론 및 한계점

본 논문은 극소수의 아이돌들이 보여주고 있는 성공 이상치를 탐지하기 위해 `mln_chart_num`의 추이와 전체 평균 변화율 모델을 토대로 트렌드를 반영한 성공 인사이트를 제공하였다. 인사이트를 제공하는 과정에서 우선적으로, 아이돌 음원 성적에 대한 군집화 모델을 통해 군집을 확인한 뒤, 군집 내에서의 이상치 탐지 모델을 도입하였다. 본 결과에는 최신 트렌드와 뉴미디어를 반영하여 어떤 상황에서 이상치가 주로 발생하는지에 대해 해석하였다.

실제로 `release_after`에 따른 `mln_heart`을 시각화 하였을 때 이상치를 보이는 그룹들은 타 변화율에 비해 비교적 가파른 증가 추이를 보이는 것을 확인하였다. 이를 시계열 기반 아이돌 성장 이상치 그래프로 자세히 확인하였을 때에는 남자아이돌의 경우 음원 발매일로부터 7 일 이내에 음반 판매량이 폭증한다는 사실을 확인할 수 있었다. 또한 발매일 이후 트렌드 분석과 같이 보았을 때 유튜브 콘텐츠나 틱톡 등 숏폼 챌린지의 사용이 직간접적으로 이상치에 영향을 미친다는 사실 또한 확인하였다.

이렇듯, 본 논문을 활용한다면 향후, 더 다양한 분석을 통해서 어느 시기에 챌린지 마케팅을 진행해야 하는지, 어떤 챌린지가 매체 이용자들 사이에서 호응을 얻을 수 있을지 등에 대한 추가적인 인사이트를 제시할 수 있을 것으로 기대할 수 있을 것이다.

본 논문을 확장한다면, 기존 프로모션으로 진행했던 방송 데이터(라디오 데이터, TV 예능 출연 빈도)와 최근 급 부상하게 된 오프라인 프로모션[15](오프라인 팝업스토어, 네컷 등과의 콜라보, 카페 콜라보[16]), 뉴미디어[17](틱톡 챌린지, 유튜브 예능 출연) 프로모션 등 데이터를 수집하여 기존과 변화한 프로모션 형태가 아이돌 인기에 미치는 영향을 분석할 수 있을 것이라 본다.

## VI. 감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1F1A1063134).

## VII. 참고문헌

- [1] Son, Y. (2020). [BTS Attraction] BTS's Global Attraction Story: Learning Korean Attraction Diplomacy from BTS. East Asia Institute, EAI Working Paper, 2020,12, 1 - 17 (17pages)
- [2] Lee, S., & Lee, S. (2019). "How Do Korean Idols Survive? Factors for Survival and Success in the Korean Idol Industry." The Journal of the Korean Women's Communication Association, Media,

- Gender & Culture, 34(2), 51-98. <http://doi.org/10.38196/mgc.2019.06.34.2.51>
- [3] Korea Creative Content Agency, 2021 Survey on the State of the Popular Culture and Arts Industry, 2022
- [4] Choi, M. (2021). "The Impact of Interaction on Simultaneity and Recommendation Intent of Communication Platforms." *Journal of the Korean Entertainment Industry*, 15(6). 23-33. 2021.
- [5] Seo, J. "New Girl Group 'Fifty Fifty' Directly to Billboard... The Secret Is TikTok Challenge." *Hankyoreh*. 2023.04.12., [https://www.hani.co.kr/arti/culture/culture\\_general/1087569.html](https://www.hani.co.kr/arti/culture/culture_general/1087569.html)
- [6] Kim, G., & Kim, M. (2018). "A Study on Predictive Indicators of Music Popularity Through Social Data Analysis." *Digital Content Society*, 19(6), 1105-1114. doi: 10.9728/dcs.2018.19.6.1105
- [7] Kim, S., Ahn, J., & Choi, B. (2020). "A Study on the Prediction of Idol Survival Possibility Using Machine Learning Techniques: Focusing on the Enhancement of Industrial Competitiveness." *Journal of the Korean Contents Society*, 20(5), 291-302
- [8] Han, M., & Na, E. (2022). "Changes in Fandom Culture of Idol by the Use Attitude of Social Media in Generations." *Journal of the Korean Contents Society*, 22(2), 605-616.
- [9] Korea Creative Content Agency. (2023). 2023 Music Industry White Paper.
- [10] Jung H. "Selling Over 100,000 Albums in Just 7 Days... A Record-breaking March for Album Initial Sales." *Hankyore*, 2022.05.06., <https://www.hani.co.kr/arti/culture/music/1041791.html>
- [11] "Loyalty over Popularity... The Heart of K-Pop Idols Is Their Fandom.", *Donga Ilbo*, 2021.7.21., <https://www.donga.com/news/Culture/article/all/20210721/108073511/1>,
- [12] Baek. J., "SEVENTEEN completes "FML" promotion...K-pop record "Unrivaled status", *Chosun ilbo*, 2023.05.08., [https://www.chosun.com/entertainments/enter\\_general/2023/05/08/Y5E3KDU67QLGEJCEIGFYHN6UTU/](https://www.chosun.com/entertainments/enter_general/2023/05/08/Y5E3KDU67QLGEJCEIGFYHN6UTU/)
- [13] Kim. T, "K-Pop Singers' Pre-release Songs Outperform Title Tracks.", *Donga Ilbo*, 2023.04.10., <https://www.donga.com/news/Culture/article/all/20230409/118754442/1>
- [14] Jung. B, "Better Performing Pre-release Songs than Title Tracks? A New Goldmine for K-Pop.", *Chosun ilbo*, 2023.04.13., <https://www.chosun.com/entertainments/music/2023/04/13/EUVKSM45F4CMD4YNEHHJBGJJDQ/>
- [15] Heo, J, "'Buy Limited Goods, Take Photos'... The Reason Why Fandom Welcomes 'Pop-up Stores'." *Seoul Economy*., 2023.08.27., <https://www.sedaily.com/NewsView/29TKJTZR5>
- [16] An, H. "Making Money from Starbucks 'Blackpink Drink' Becomes Popular." *Hankyung*. 2023. 8. 17., <https://www.hankyung.com/article/202308177674g>
- [17] Kim, T., & Han, K. (2021). "A Study on the Media Influence on Individual Music Consumption." *Journal of Popular Music Contents*, 5(1), 59-77

## 저자소개

---



**최지원(Jiwon Choi)**

2021년 3월 중앙대학교 산업보안학과 학사과정 재학

관심분야 : 데이터분석, 뉴미디어, 슛폼, 추천시스템



**정다연(Dayeon Jung)**

2021년 3월 중앙대학교 산업보안학과 학사과정 재학

관심분야 : 빅데이터, MLOps, DevOps



**최강규(Kangkyu Choi)**

2021년 3월 중앙대학교 산업보안학과 학사과정 재학

관심분야 : 데이터분석, 빅데이터, 이상데이터 탐지



**임태인(Taein Lim)**

2021년 3월 중앙대학교 산업보안학과 학사과정 재학

2023년 6월 Init Cloud CEO

관심분야 : 분산처리, DevOps



**김대훈(Daehoon Kim)**

2013년 ~ 2016년 고려대학교 정보통신기술연구소 연구교수

2013년 ~ 2021년 미국 실리콘밸리 스타트업 CEO 및 CTO

2021년 9월 ~ 현재 LLOYDK CEO

관심분야 : 제너레이티브 AI, 빅데이터, 머신러닝, IT 솔루션, 비즈니스 전략 기획



**정종균(JongKyun Jung)**

2005년 2월 경동대학교 컴퓨터공학과 학사

2005년 ~ 현재, 경력 19년 솔루션 개발 및 빅데이터 총괄

관심분야 : IT 보안, 제너레이티브 AI, 모니터링 서비스 자동화, 빅데이터 솔루션



**노승민(Seunmin Rho)**

2008년 8월 아주대학교 정보통신전문대학원 정보통신공학과 박사

2008년 10월 Carnegie Mellon University Postdoctoral Researcher

2021년 3월~현재 중앙대학교 경영경제대학 산업보안학과 교수

관심분야 : 인공지능, 생체인식, 개인정보보호

---