http://dx.doi.org/10.15813/kmr.2018.19.2.005

댓글이 음원 판매량에 미치는 차별적 영향에 관한 텍스트마이닝 분석[1]

The Impact of Comments on Music Download and Streaming: A Text Mining Analysis

박 명 석 (Myeong-Seok Park) 한양대학교 경영대학 석사[2]
권 영 진 (Young-Jin Kwon) 한양대학교 경영대학 석사과정[3]
이 상 용 (Sang-Yong Tom Lee) 한양대학교 경영대학 교수[4]

ABSTRACT

This study mainly focused on measuring the impact of comments for a particular song on the number of streamings and downloads. We modeled multiple regression equations to perform this analysis. We chose digital music market for the object of analysis because of its inherent characteristics, such as experience goods, high bandwagon effect, and so on. We carefully utilized text mining technique in accordance with the algorithm of Naïve Bayes classifier to distinguish whether a comment for a piece of music be regarded as positive or negative. In addition, we used 'size of agency' and 'existence of hit song' as moderating variables. The reason for usage of those variables is that those are assumed to affect users' decision for selecting particular song especially when downloading or streaming via music sites. We found empirical evidences that positive comments for a particular song increase the number of both downloads and streamings. However, positive comments may decrease the number of downloads when the size of agency of the artist is big. As a result, we were able to say that a positive comment for a particular song functioned as 'word-of-mouth' effect, inducing other users' behavioral response. We also found that other features of an artist such as size of the agency that the artist belongs to functioned as an external factor along with feature of the song itself.

Keywords: : Big Data Usage, Text Mining, Music sales, Regression Analysis.

[1] 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A3A2066740)
논문접수일: 2018년 3월 16일; 1차 수정: 2018년 5월 23일; 2차 수정: 2018년 5월 30일; 게재 확정일: 2018년 5월 31일
[2] 제 1 저자 (Tel: +82-2-2220-1814, E-mail: audtjr1019@naver.com)
[3] 제 2 저자 (Tel: +82-2-2220-1814, E-mail: bohemian4682@gmail.com)
[4] 교신저자 (Tel: +82-2-2220-1814, E-mail: tomlee@hanyang.ac.kr)

1. 서론

한국 오프라인 음반 시장은 지속적인 매출 감소로 인한 쇠퇴를 거듭하고 있는 반면, 온라인과 모바일 상의 디지털 형태로 된 음악 소비는 꾸준히 늘고 있는 추세이다. 즉, 음악 소비의 형태가 초기의 물리적 매체인 음반 구매를 통한 방식에서 온라인 접속을 통한 음원 다운로드, 혹은 온라인상에서의 음원 스트리밍으로 변화하고 있는 것이다. MP3 파일 형태의 음원이 다운로드 하기에 용이하다는 점과 웹과 모바일을 아우르는 음원 스트리밍 서비스의 보편화가 이러한 추세를 이끄는 동인으로 작용하고 있다.

Aguiar & Martens (2016)은 다양한 음악 플랫폼의 안정화로 인한 음원 스트리밍 이용자 수의 증가가 음원 판매량에 전반적으로 긍정적인 영향을 미친다는 사실을 발견하였다. 그리고 Strobl & Tucker (2000)는 영화 산업의 블록버스터 효과와 유사하게 이미 대중적 스타가 된 아티스트들이 음원 판매 순위 차트에 더 오래 머물러 있게 된다는 소위 “superstar effect”를 발견하였다. 하지만 이 연구의 superstar effect는 앞서 말한 음원시장의 변화 이전에 발견된 것이다. 이 시점 이후 MP3와 P2P라는 음악 감상 기기의 등장으로 음원시장은 이전과 전혀 달라졌고, 이에 Bhattacharjee et al. (2007)는 P2P 열풍 전과 후의 음원 판매 차트를 비교했다. 그리고 P2P 열풍 후 음원의 차트 수록 기간이 예전에 비해 평균 42% 단축된다는 사실을 알아냈다. 하지만 차트 수록 이전 100주 동안 상위 차트에 등장한 즉, 대중적 스타인 아티스트들은 다른 아티스트에 비해 약 35% 수록 기간이 길다는 사실을 발견하여, 새로운 기기의 등장으로 인한 음원시장의 변화에도 불구하고 Super Star effect는 변함없이 존재함을 증명하였다.

이와 더불어 Kretschmer & Peukert (2017) 는 유럽의 음원 판매 시장 데이터를 사용하여 음원 판매 요인을 연구했고 온라인상의 비디오(Youtube) 시청수가 디지털 음원 판매를 촉진시키고 물리적 음원 판매까지 영향을 미친다는 것을 발견했다. 또한, Hiller (2016)은 YouTube 음원관련 동영상의 공유는 잠재적 이용자들에게 더 많은 음원 노출의 기회를 제공하는 반면, 차트 상단에 위치한 곡들의 음원 판매에는 부정적인 영향을 줄 수 있고 이 영향은 광고효과에 의해 완화 될 수 있음을 알아냈다. 이와 비슷하게 Adler (1985)는 높은 음원 판매 실적을 경험한 아티스트들이 단순히 아티스트들의 재능 이외에도 음원사의 판촉 지원과 같은 외적인 부분이 성공 여부를 결정하는데 중요한 역할을 하는 것을 밝혀냈다.

한국인터넷진흥원의 보고서에 따르면, 온라인 쇼핑으로 특정 상품을 구매할 때 소비자중 79.3%는 다른 소비자의 이용 후기 댓글을 참고하는 것으로 나타났고 이 중 94.3%는 다른 소비자의 이용 후기 댓글이 본인의 구매의사결정에 영향을 미쳤다고 응답했다. Basuroy (2003)의 연구에서는 온라인 상의 소비자 이용 후기가 제품 매출액에 직접적인 영향을 미친다는 것을 밝혔다. 매출액은 매우 중요한 재무 성과 지표라는 것을 고려할 때 소비자의 이용 후기는 기업의 입장에서 무시할 수 없는 중요한 피드백이라는 것을 알 수 있다. 또한, 다수의 선행 연구에 따르면 온라인 상의 이용자가 특정 상품에 남긴 댓글은 방향성, 즉 긍정적 댓글과 부정적 댓글로 분류될 수 있고, 이 여부에 따라서 다른 이용자들은 구매의사결정에 영향을 받는 것으로 보고되고 있다(Duan et al. 2008; Elliott 2002; 박은아 등 2007; 손지연과 어숙희 2008; 손진아와 이은영 2007; 장리 등 2017). 특히 이용자들은 해당 제품에 대한 긍정적인 댓글보다 부정적인 댓글에 더 민감하게 반응하는 것으로 드러났다(Basuroy 2003; 박찬과 유창조 2006; 손진아와 이은영 2007).

따라서 본 연구에서는 이용자의 댓글이 기업의 성과에 영향을 미친다는 사실에 착안하여 미디어 콘텐츠 중 음원의 다운로드와 스트리밍 수에 댓글이 어떻게,

어느 정도 영향을 미치는지 분석했다. 온라인상의 댓글과 미디어 콘텐츠 판매의 관계를 분석한 선행 연구에서는 댓글이 미디어 콘텐츠 중 책의 매출에 영향을 미친다는 연구(Chevalier & Mayzlin, 2006)와 영화의 매출에 영향을 미친다는 연구(Chintagunta et al, 2010)만이 이루어졌었다. 하지만, 음원은 위의 다른 미디어 콘텐츠들과는 재화에 대한 접근성이나 가격, 소비형태 관점에서 성질이 매우 다르다. 지금까지의 음원에 대한 연구는 음원의 직접적인 소비량을 나타내는 다운로드 수와 스트리밍 수가 아닌, 음원 플랫폼에 대한 분석만이 다루어졌다. 더욱이, 유행에 따르는 인기 변화가 매우 극심한 음악 소비는 Bandwagon 효과가 강하게 나타나기 때문에, 사람들의 반응과 댓글이 다른 사람들의 소비에도 민감하게 작용하리라고 판단된다.

이에 본 연구에서는 음원 플랫폼 상의 댓글이 음원의 다운로드, 스트리밍 수에 미치는 영향을 분석함으로써 기존에 연구되지 않은 영역을 연구했다는 점에서 학문적으로 기여하는 바가 있다. 이에 더불어 음원 시장의 대/소형 기획사와 유/무명 아티스트의 위치에 따라 음원의 댓글이 다운로드 수나 스트리밍 수에 미치는 영향을 구체적으로 제시하여 실무적 공헌을 하고자 한다. 또한, 본 연구에서는 스트리밍 사이트상의 곡에 대한 좋아요'와 '앨범에 대한 좋아요' 수가 소비자들의 선호를 반영한다고 판단하였고, 종속변수에도 영향을 미친다고 가정한 뒤 가설 설립 시 반영하여 함께 검토하고자 한다. 다음 2장에서는 본 논문과 관련이 있는 음원 판매에 관한 선행연구를 살펴보면서 동시에 텍스트마이닝과 베이즈 분류법에 관한 배경이론을 간략히 살펴볼 것이다. 3장에서는 연구 모형과 가설을 설명하고, 4장에서는 연구 방법론에 대하여 설명하고자 한다. 이후 5장에서 회귀분석 결과를 살펴보고 논의할 것이며, 마지막 6장에서는 본 연구의 결론을 요약하고 한계점 및 향후 연구를 설명할 것이다.

2. 선행연구 및 배경이론

2.1 음원 판매에 관한 선행연구

서론에서 언급한 바와 같이 음원 판매에 대한 다양한 선행연구들이 존재하였고, <표 1>은 관련 연구들을 요약한 표이다. 다른 일반적 상품과는 달리 음원 판매에 관해서, 댓글 어느 정도 판매에 영향을 미치는지에 대한 연구가 아직 이루어지지 않았음을 알 수 있다. 따라서 본 연구는 <표 1>의 기존 연구들과 맥을 같이 하면서도 댓글의 영향력을 연구의 중심점에 놓는다는 측면에서 차별성이 있다고 할 수 있다.

Hiller (2016)는 빌보드 차트에 등재된 곡들을 상위 10곡, 상위 25곡, 상위 50곡으로 분류한 후 각 그룹의 음반 판매 총 매출을 종속 변수로 설정한 회귀 분석을 시행했다. 분석 결과 곡의 빌보드에 등재된 기간이 길어짐에 따라 총 매출에는 부정적인 영향을 미쳤고 그 정도는 상위 그룹에 갈수록 커졌다. 한편, 곡의 빌보드에 등재된 첫 주의 음반 판매량은 총 음반 매출에 긍정적인 영향을 미쳤다. 이것을 종합해보면 음원이 처음 공개되었을 때 음원 사이트의 이용자들은 해당 곡에 가장 관심을 가지고, 그 후 자연스럽게 관심이 식어간다는 것을 알 수 있다.

Aguiar & Martens (2016)는 음원 다운로드 웹 사이트와 음원 스트리밍 웹 사이트의 이용 수가 음원 판매 웹 사이트에 미치는 영향을 연구했다. 이 연구가 진행된 동기는 음원 다운로드와 음원 스트리밍이 대중화, 보급화 되면서 음원을 직접 구매하는 이용자의 수가 줄어들었을 것이라는 통념을 검증해보기 위함이다. 종속변수에 대한 회귀분석 결과, 음원 다운로드 웹 사이트의 이용 수와 음원 스트리밍 웹 사이트의 이용 수 모두 음원 구매 웹 사이트 이용 수에 통계적으로 유의한 양(+)의 영향을 미쳤고, 이 결과는 통념과는 상반된다. 즉, 음원을 직접 구매하는 것과 다운로드 및 스트리밍하는 것은 상충 관계가 아니라는 것이다. 이 연구는 본 연구

<표 1> 음원 판매에 영향을 미치는 요인에 대한 선행 연구

연구자	종속변수	주요 독립변수	연구방법	주요 연구결과
Aguiar & Martens (2016)	-디지털 음원 매출 웹 사이트의 클릭 수	(a)허가 받은 음원 스트리밍 웹 사이트 클릭 수 (b)무허가 음원 다운로드 웹 사이트 클릭 수	회귀 분석	(a)허가 받은 음원 스트리밍 웹 사이트의 총 클릭 수는 디지털 음원 매출 웹 사이트의 클릭 수에 통계적으로 유의한 양(+)의 영향을 미쳤음 (b)무허가 음원 다운로드 웹 사이트의 클릭 수 또한 디지털 음원 매출 웹 사이트의 클릭 수에 통계적으로 유의한 양(+)의 영향을 미쳤음.
Strobl & Tucker (2000)	-음원 차트에 등재되는 기간(생존기간)	(a)음원이 영화 사운드에 활용됐는지 여부 (b)아티스트의 인기 여부	카플란-마이어 생존 함수를 통한 Weibull 회귀분석	(a)음원이 영화의 사운드트랙(soundtrack)에 활용됐는지 여부는, 해당 음원의 생존 기간에 통계적으로 매우 유의한 양(+)의 영향을 미쳤음. (b)아티스트의 과거 인기 여부는, 해당 음원의 생존 기간에 통계적으로 매우 유의한 양(+)의 영향을 미쳤음.
Hiller (2016)	-음반 매출	(a)빌보드 차트 등재 누적 기간 (b)첫 앨범 여부 (c)첫 주 음반 판매량 (d)직전 앨범의 빌보드 차트 등재 여부 (e)직전 앨범의 총 매출	회귀 분석	(a)빌보드 차트에 등재된 기간이 길어질수록 음반 매출에는 음(-)의 영향을 주었음. (b)아티스트의 첫 앨범 여부는 음반 매출에 음(-)의 영향을 주었음. (c)첫 주 음반 판매량은 전체 음반 매출에 양(+)의 영향을 미쳤음. (d)차트 상위 25위의 곡 중, 직전 앨범의 빌보드 차트 등재 여부는 음반 매출에 음(-)의 영향을 미쳤음.

에서 사용된 종속변수(다운로드 수, 스트리밍 수) 설정에 영향을 주었다.

Strobl & Tucker (2000)는 카플란-마이어 생존 함수를 이용해서 특정 음원이 영국 음원 차트(NME, New Musical Express)에 등재된 생존 기간을 분석했다. 회귀분석 결과, 특정 음원이 영화의 사운드트랙으로 활용되었는지 여부는 해당 음원의 차트 등재 기간에 매우 긍정적인 영향을 미쳤다. 그리고 아티스트의 과거 인기 여부 또한 해당 음원의 차트 등재 기간에 긍정적인 영향을 미쳤는데, 이러한 사실은 음원 시장에서도 superstar effect가 존재한다는 것을 보여준다.

2.2 텍스트 마이닝

텍스트 마이닝이란 비 정형화된 텍스트 데이터를 처리하여 의미 있는 패턴을 추출하거나 그 텍스트에서 반영하고 있는 의미를 수치화한 뒤 그 값을 추출해 내는 과정이다. 이전부터 텍스트 마이닝을 이용한 다수의 선행 연구들이 있었지만, 2012년 미국 대선에서의 활용을 기점으로 이 기술의 효용성에 대한 논의가 확대되고 있다. 이러한 추세에 따라 한국의 연구에서도 텍스트 마이닝 기법을 활용하여 2012년 대선과 관련된 트위터 코멘트를 실시간 수집, 저장, 그리고 분석하였고 이 기법을 통해 매우 효과적으로 사회 현상을 분석할 수 있었다(배정환 등, 2013; 이소현 등, 2017).

이 기법은 일련의 과정을 거친다. 먼저 텍스트 데이터

<표 2> 텍스트 마이닝 실행 단계

단계	과정	내용
1단계	데이터 수집	인터넷에서 올라오는 모든 텍스트 데이터들이 그 대상이 됨
2단계	용어 추출	통계적 방법을 사용하여 첫 번째 단계에서 수집한 데이터를 정제하고 그중 관심있는 단어들을 문장에서 추출하는 과정
3단계	정보 추출	문서 내에 있는 특정한, 혹은 어떠한 의미를 갖고 있는 단어를 추출하는 과정
4단계	정보 분석	단계의 누적 데이터 들을 토대로 본격적으로 분석을 실시하는 단계

를 스크래핑(scrapping)한 뒤 전처리 작업을 통해 구조화된 형식의 원시데이터를 형성한다. 그리고 산출된 수치를 통해 모델링을 하거나 다양한 텍스트 마이닝 기법들을 활용하여 원시 데이터를 취합한 뒤 해석하는 과정을 거친다. 데이터 수집부터 분석까지는 일련의 4단계를 포함하고 <표 2>를 통해 알 수 있다.

1단계는 데이터 수집 과정이다. 인터넷상의 모든 텍스트 데이터들은 수집의 대상이 될 수 있다.

2단계는 용어 추출과정이다. 통계적 방법을 사용하여 첫 번째 단계에서 추출한 데이터를 정제하고, 그중 관심있는 단어들을 문장에서 추출하는 과정이다. 여러 가지 방법 중 단일의 문서에서 등장하는 특정 단어를 추출하는 TF(term frequency)와 전체 문서들 중 특정 단어를 포함하고 있는 문서들의 단어를 추출하는 DF(document frequency)가 대표적이다. 3단계는 정보 추출 과정이다. 문서 내에 있는 특정한, 혹은 어떠한 의미를 갖고 있는 단어를 추출하는 과정이다. 4단계는 정보 분석과정이다. 이 단계는 3단계까지 누적 축적된 데이터들을 토대로 본격적으로 분석을 실시하는 단계로서 대표적으로 클러스터링, 토픽모델링, 빈도수 측정 등의 방법을 통해 텍스트 데이터를 마지막으로 분석, 해석하는 단계이다.

이상훈 등(2015)은 위와 같은 단계를 거친 텍스트 마이닝을 활용해서 영화에 대한 인터넷 이용자들의 댓글을 수집 및 분석했고 이전의 연구에서 흥행 예측 변수를 사용했을 때보다 더 정확한 결과를 도출했다.

2.3 Naïve Bayes Classifier

기계 학습(machine learning) 과정에서 사용되는 분류 기준인 나이브 베이즈 분류(Naive Bayes Classifier)는 특히 스팸을 뚜렷하게 구분하는 텍스트 분류기로 잘 알려져 있다. 본 연구는 음원 사이트상에서의 특정 곡에 대한 댓글을 긍정/부정으로 분류했고 이때 나이브 베이즈 분류(Naive Bayes Classifier) 기준을 사용하였다. 댓글을 긍정/부정으로 분류할 때 이 분류 기준이 작용하는 알고리즘의 원리는, 사전에 미리 분류해 놓은 댓글들을 기준으로 긍정 댓글의 출현 빈도가 높거나, 혹은 부정 댓글의 출현 빈도가 높은 단어들을 탐색하여 그 차이를 보는 과정을 거친다. 예를 들어서 긍정 댓글들 중에서 어떠한 특정 단어가 다른 단어들보다 현저하게 출현빈도가 높다면 그 단어는 해당 댓글이 긍정적인 댓글인지 부정적인 댓글인지 구분하는 기준으로 작용한다. 이러한 알고리즘의 원리로 베이즈 분류기는 긍정으로 판단되는 댓글에서 어떤 특정한 단어를 발견할 확률과 부정 댓글이라 판단되는 댓글에서 같은 단어를 발견할 확률을 계산하고, 이를 토대로 긍정/부정 중 가능성이 높은 확률을 선택한 뒤 분류를 진행한다.

본 논문에서는 특정 댓글의 긍정/부정 판별 가능성을 계산할 때 기준으로 사용되는 모든 단어의 출현 빈도수를 다른 단어의 빈도수와 독립적으로 추정 가능하다고 가정한 뒤 텍스트 분류 알고리즘을 적용시킨다. 이러한 가정의 방법을 통계적 독립이라고 하는데,

가정을 할 때 많은 변수를 고려하지 않기 때문에 단순(Naive) 모형 방식이라고 부른다. 혹은 18세기에 처음으로 조건부 확률을 정의한 수학자 베이즈의 이름을 따서 베이즈(Bayes)모형이라고 부르기도 한다(Rish, 2001).

3. 연구모형과 가설

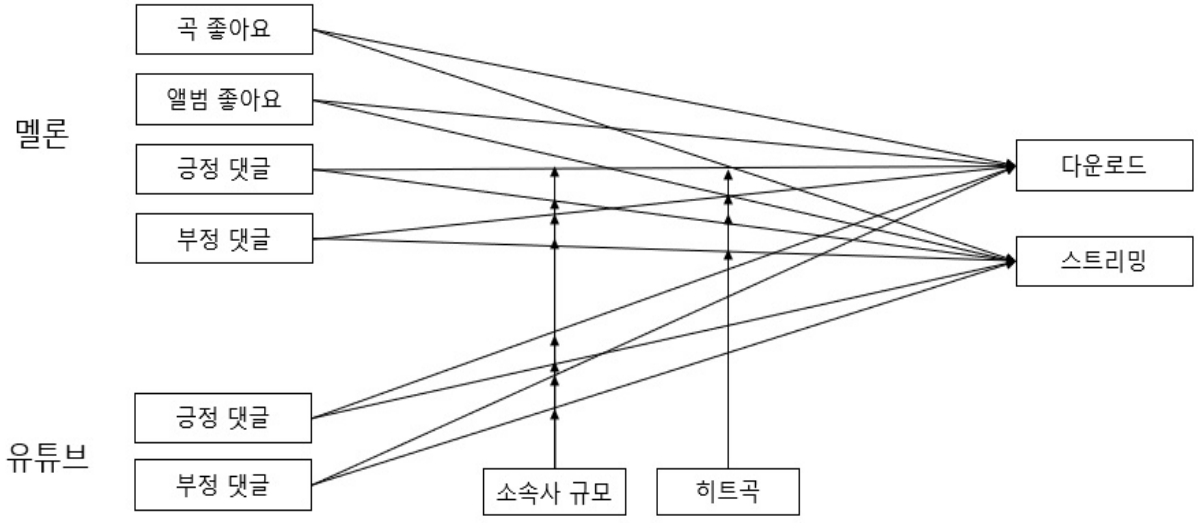
3.1 연구 모형

본 연구는 음원 스트리밍 사이트(멜론에서 곡이나 앨범에 대해 이용자들이 부여한 '좋아요'와 댓글(긍정 vs. 부정)을 독립변수로 설정한 후, 종속변수(다운로드 수, 스트리밍 수)에 미치는 영향을 분석했다. <그림 1>은 본 연구의 연구 모형(회귀 모형)을 도식화한 것이고, 그림에서 알 수 있듯 조절변수는 소속사의 규모와 히트곡의 수로 설정하였다. 이 회귀 모형을 기준으로 회귀분석을 시행했고 결과는 결과 섹션에서 확인할 수 있다.

3.2 연구 가설

3.2.1. 소비자 리뷰 감성 - 긍정 vs. 부정

Forman et al. (2009)는 중립적인 내용을 담은 온라인 리뷰에 비해 극단적인 내용을 담은 온라인 리뷰가 더욱 유용한 리뷰임을 밝혔고, Cao et al. (2011) 또한 도서판매 플랫폼과 소프트웨어 다운로드 플랫폼을 대상으로 한 연구에서 중립적인 리뷰보다 극단적 내용의 리뷰가 더욱 유용함을 밝혔다. 이 두 선행 연구를 포함한 많은 연구들의 결과를 바탕으로 본 연구에서는 소비자 리뷰가 상품 매출에 미치는 영향을 보다 잘 파악하기 위해서 리뷰 감성을 "긍정"과 "부정"으로 양분하였다. 한편, 긍정적 리뷰와 부정적 리뷰가 소비자의 구매 태도에 미치는 영향에 대한 연구 결과들은 다소 상충된다. 김창호 (2006)는 긍정적 리뷰보다 부정적 리뷰가 소비자가 상품에 대한 태도와 구매 의도에 더욱 유의한 영향을 미친다는 결론을 냈고, Chevalier & Mayzlin (2006)은 리뷰 별점이 상품 매출에 음(-)의 영향을 미친다는 것을 밝혔다. 하지만, 다수의 선행 연구들은 소비자의 긍정적인 리뷰가 상품 매출에 양(+)의



<그림 1> 연구모형

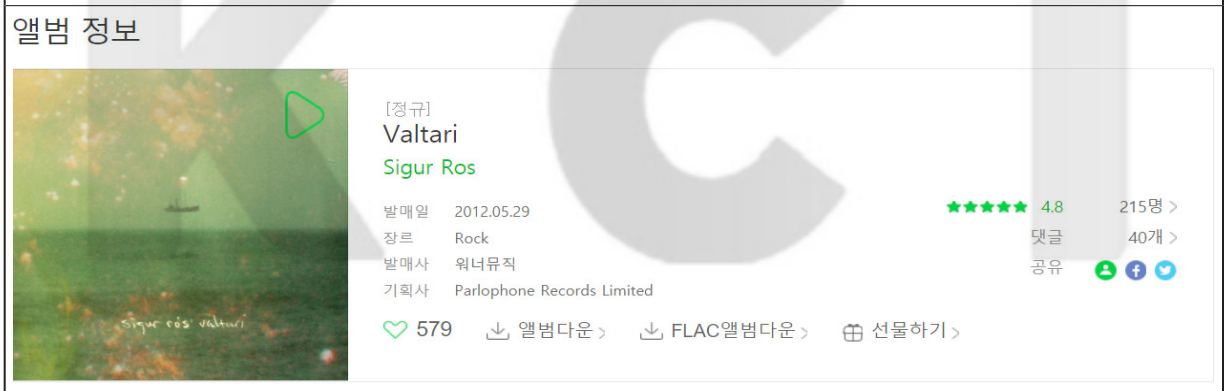
영향을 미친다는 결론을 냈다 (Chen et al. 2004; 박지영과 유요섭 2011). 따라서, 본 연구는 멜론 사이트 상의 소비자 리뷰가 상품 매출에 미치는 영향에 대한 다음과 같은 가설을 설정했다.

H1-a: 음원 스트리밍 채널 (멜론) 상의 긍정 댓글은 음원 다운로드 수에 양(+)의 영향을 미칠 것이다.
H1-b: 음원 스트리밍 채널 (멜론) 상의 부정 댓글은 음원 다운로드 수에 음(-)의 영향을 미칠 것이다.

한편, 소비자들은 댓글과 더불어 소위 "좋아요"를 클릭함으로써 특정 음원에 대한 리뷰를 나타낼 수 있다. <그림 2>는 멜론 사이트 상의 "좋아요"을 나타내는 예시이다. "좋아요"를 나타낼 수 있는 대상은 앨범과, 앨범에 수록된 개별 곡이다. 즉, 앨범에 대한 "좋아요"는 소비자들이 해당 앨범 전체를 평가하는 수단이고, 곡에 대한 "좋아요"는 개별 곡에 대한 평가를 반영한다. 우리는 "좋아요"가 긍정 댓글과 유사한 기능을 한다고 가정했고, 따라서 Chen et al. (2004)와 박지영과 유요섭 (2011)의 연구 결과에 따라 다음과 같은 가설을 설정했다.

H1-c: 음원 스트리밍 채널 (멜론) 상의 앨범에 대한 "좋아요" 수는 음원 다운로드 수에 양(+)의 영향을 미칠 것이다.
H1-d: 음원 스트리밍 채널 (멜론) 상의 곡에 대한 "좋아요" 수는 음원 다운로드 수에 양(+)의 영향을 미칠 것이다.

앨범 "좋아요" 예시



곡 "좋아요" 예시

<그림 2> 멜론 사이트 상의 "좋아요" 예시

3.2.2. 조절 변수 설정

기획사 규모

음원 소비자의 리뷰가 음원 매출에 미치는 영향을 보다 더 정확하게 알아보기 위해서는 음원 자체의 특성을 조절 변수에 반영한 뒤 그 영향을 파악할 필요가 있다. 본 연구는 특정 음원의 아티스트가 속한 기획사의 규모와 해당 음원의 아티스트가 과거의 히트곡 수를 조절 변수로 사용하였다. Graham et al. (2004)은 인터넷의 보급화에 따른 음반 시장의 변화를 연구했다. 그들의 연구 내용에 따르면, 인터넷의 보급화 이전에는 소위 "빅 5"라고 불리는 EMI, Sony, Universal-Vivendi, Time Warner, 그리고 Bertelsmann (BMG)가 음반 유통 산업을 지배한 반면, 이후에는 그런 견고한 과점 체제가 와해되었다고 한다. 따라서 본 연구에서는 다음과 같은 가설을 설정했다.

H1-e: 기획사 규모에 따라 음원 스트리밍 채널 (멜론)의 긍정 댓글은 음원 다운로드 수에 상이한 영향을 미칠 것이다.

H1-f: 기획사 규모에 따라 음원 스트리밍 채널 (멜론) 의 부정 댓글은 음원 다운로드 수에 상이한 영향을 미칠 것이다.

히트곡

특정 아티스트의 히트곡 수는 대중적 인지도를 대표할 수 있다. 앞서 언급했듯이, Strobl & Tucker (2000)는 영화 산업의 블록버스터 효과와 유사하게 이미 대중적 스타가 된 아티스트들이 음원 판매 순위 차트에 더 오래 머물러 있게 된다는 소위 "superstar effect"를 발견하였다. 또한 Kohli & Jaworski (1990)의 연구에 의하면 이러한 소위 대중적 스타들에게 마케팅이 보다 집중적으로 이뤄진다. 이러한 선행 연구 결과를 기준으로 본 연구는 히트곡이 많은 아티스트의 곡일수록 대중들의 음원 소비가 높을 것으로 가정했다.

H1-g: 아티스트의 히트곡 수가 많을수록 음원 스트리밍 채널 (멜론) 의 긍정 댓글은 음원 다운로드 수에 상이한 영향을 미칠 것이다.

H1-h: 아티스트의 히트곡 수가 많을수록 음원 스트리밍 채널 (멜론) 의 부정 댓글은 음원의 다운로드 수에 상이한 영향을 미칠 것이다.

3.2.3. 음원 판매량 측정 기준 - 다운로드 vs. 스트리밍

디지털 음원의 특성을 고찰한 유재원 등 (2018)의 연구는 음원 소비를 다운로드와 스트리밍으로 나누어 분석했고 그 배경은 다음과 같다. Bardhi & Eckardt (2017)는 소비의 개념을 고정적 소비와 유동적 소비로 양분하였다. 음원 스트리밍은 "소유하지 않고 일시적인 접근 권한을 얻는 소비" 라는 점에서 유동적 소비로 분류가 되며, 음원 다운로드는 일종의 물질적 소유로 개념화할 수 있다는 점에서 고정적 소비로 분류 가능하다. 디지털 음원이 태동하기 전에 사람들은 음악을 물질적 소비를 통해서만 소유할 수 있었지만, 이후에는 디지털 음원을 이용할 뿐만 아니라 소유할 수 있게 되었다. 한편, 디지털 음원은 탐색재와는 달리 직접 듣고 체험해봐야 그 효용을 가늠할 수 있는 경험재이기 때문에 (양성수 등 2016), 본 연구에서는 다운로드와 스트리밍을 분리해야 소비자의 행동을 파악할 수 있다고 가정했다. 따라서 종속 변수를 다운로드로 설정한 이전의 가설들에 더해서, 스트리밍에 미치는 영향을 알아보고자 했다. <표 3>은 해당 가설들을 제시한다.

4. 데이터 및 연구 방법

본 논문에서는 분석 시 멜론 사이트 상의 448186개 곡 댓글을 사용하였고, 분류 모형 사용 시 3만개의 댓글을 직접 분류하였다. 직접 분류 시 긍정/부정을 판단하기 모호한 댓글은 노래를 들었다고 판단하고 긍정으로 분류하였다. 모델링은 Python을 통해 진행하였다.

<표 3> 연구 가설 (스트리밍)

가설	가설 내용
H2-a	음원 스트리밍 채널 (멜론) 상의 긍정 댓글은 음원 스트리밍 수에 양(+)의 영향을 미칠 것이다.
H2-b	음원 스트리밍 채널 (멜론) 상의 부정 댓글은 음원 스트리밍 수에 음(-)의 영향을 미칠 것이다.
H2-c	음원 스트리밍 채널 (멜론) 상의 앨범에 대한 "좋아요" 수는 음원 스트리밍 수에 양(+)의 영향을 미칠 것이다.
H2-d	음원 스트리밍 채널 (멜론) 상의 곡에 대한 "좋아요" 수는 음원 스트리밍 수에 양(+)의 영향을 미칠 것이다.
H2-e	기획사 규모에 따라 음원 스트리밍 채널 (멜론) 의 긍정 댓글은 음원 스트리밍 수에 상이한 영향을 미칠 것이다.
H2-f	기획사 규모에 따라 음원 스트리밍 채널 (멜론) 의 부정 댓글은 음원 스트리밍 수에 상이한 영향을 미칠 것이다.
H2-g	아티스트의 히트곡 수가 많을수록 음원 스트리밍 채널 (멜론) 의 긍정 댓글은 음원 스트리밍 수에 상이한 영향을 미칠 것이다.
H2-h	아티스트의 히트곡 수가 많을수록 음원 스트리밍 채널 (멜론) 의 부정 댓글은 음원 스트리밍 수에 상이한 영향을 미칠 것이다.

직접 분류된 댓글은 3만개이지만, 긍정/부정 댓글의 비율을 고르게 맞추기 위해 실제 모델링에는 1만7천개의 댓글만 사용하였다. 그 중 training set을 1만개, test set을 7천개로 나누어 모델링하였고 사용된 feature는 1000개이다. Test set 7000개의 댓글을 통해 머신 러닝 기법으로 베이즈 분류기 모형을 설정했고, 3000개의 댓글로 테스트 했을 때 대략 80%의 정확도를 확인할 수 있었기에 해당 모형으로 나머지 분류를 진행했다. 아래의 <표 4>는 그 중 의미 있는 feature의 예시이다. exists(꼴)은 댓글에 꼴이라는 명사가 나왔을 때, True 0:1 즉, 부정과 긍정의 비율이 21.6:1.0임을 뜻한다. 반대로 exists(굿)은 댓글에 굿이라는 명사가 나왔을 때, 긍정과 부정의 비율이 8.8:1.0임을 뜻한다.

4.1 종속변수

본 논문에서는 2017년 1월부터 9월 28일까지 발표된 공식 음원을 상대로 연구를 진행했으며, 그 중 대한민국에서 가장 공신력 있는 음원 차트인 가온 차트에서 주차별로 집계된 다운로드 수와 스트리밍 수를 종속변수로 사용하였다. 각 곡당 수집 기간은 첫 가온 차트의 순위에 진입한 시점부터 순위에서 제외될 시점까지이다. 가온 차트는 주간 별로 Top100위까지의 다운로드 순위와 스트리밍 순위가 업데이트 된다. 그 이외의 데이터는 수집할 수 없고, 순위에서 제외된 곡들은 댓글의 활동이 현저히 줄어들어 순위에 유지되는 동안의 댓글들 만을 분석에 사용하였다.

<표 4> 분류 모델의 유의미한 Feature

Informative Feature		
Exists (꼴)	True 0 : 1	21.6 : 1.0
Exists (닥)	True 0 : 1	20.9 : 1.0
Exists (놈)	True 0 : 1	15.8 : 1.0
Exists (굿)	True 1 : 0	8.8 : 1.0
Exists (총)	True 1 : 0	5.9 : 1.0

4.2 독립변수

독립변수로 사용된 데이터들 중 댓글에 대한 데이터는 멜론을 크롤링(crawling)해서 수집하였다. 크롤링 도구로는 Python을 활용했고, 전처리 시에는 R을 사용하였다. 댓글 중 특수문자 혹은 'ㅜ' , 'ㅋ' 와 같은 의미를 갖지 않는 댓글들은 분석 대상에서 제외하였다. 형태소 분석과 품사 태깅(tagging)에는 Python에서 가장 많이 사용되는 형태소 태깅 라이브러리인 KoNLPy 라이브러리를 이용했고, 그 중 twitter라는 패키지를 사용하여 댓글 중 명사만을 태깅하여 분석에 사용했다.

4.3 회귀분석

본 연구는 음원에 대한 이용자들의 긍정/부정 댓글과 스트리밍 채널 (멜론) 의 구분이 종속변수로 설정한 음원의 다운로드 및 스트리밍 수에 미치는 영향을 연구하기 위해 회귀 모형을 수립한 뒤 회귀 분석을 실행하였다. 본 연구에서 사용된 회귀 분석 모형은 아래의 수식으로 표현될 수 있다.

$$= \beta_0 + \sum_{i=1}^{4} \beta_i X_i + \varepsilon$$

Y : 종속변수(다운로드 수, 스트리밍 수)
β_0 : y절편(intercept)
β_i : X_i각 변수의 회귀계수
X_i : 독립변수
ε : 잔차(residual)

아래 (1)번 식의 은 곡에 대한 좋아요 수, 는 앨범에 대한 좋아요 수, 은 스트리밍 채널의 긍정 댓글 수, 은 스트리밍 채널의 부정 댓글 수를 나타낸다. 또한, 기획사의 규모와 아티스트 히트곡 수를 분류하여 조절효과를 알아보기 위한 수식은 다음과 같다.

$$Y = \beta_0 + \sum_{i=1}^{2} \beta_i X_I + \sum_{i=1}^{2} \beta_{(2+i)} X_{(2+i)} M_{1k} + \sum_{i=1}^{2} \beta_{(4+i)} X_{(4+i)} M_{2k} + M_{1k} + M_{2k} + \varepsilon \quad (1)$$

$$Y = \beta_0 + \sum_{i=1}^{2} \beta_i X_I + \sum_{i=1}^{4} \beta_{(2+i)} X_{(2+i)} M_{1k} + \sum_{i=1}^{2} \beta_{(6+i)} X_{(6+i)} M_{2k} + M_{1k} + M_{2k} + \varepsilon \quad (2)$$

식 (1), (2)의 는 각각의 회귀 계수를 나타내고, 식 (1)에서의 은 각각 곡에 대한 좋아요 수와 앨범에 대한 좋아요 수를 의미한다. 조절변수들이 음원의 다운로드와 스트리밍의 수에 영향을 미치는지 알아보기 위한 식 (2)의 나머지 은 각 채널의 긍정, 부정 댓글 수의 회귀계수를 의미한다. 는 각각 조절 효과를 알아보기 위한 조절변수로서, 기획사의 규모를 나타내는 조절변수 는 대형일 경우 1, 소형일 경우 0으로 설정하였고, 는 해당 아티스트의 히트곡 수를 의미한다. 기획사의 규모에 대한 분류 기준은 2016년 기준 매출의 전체 70% 이상을 차지하는 기획사를 대형, 나머지를 소형으로 정의하였다. 히트곡 보유 수에 대한 기준은 아티스트별 지난 5년간 연 순위 100위 내로 진입한 히트곡의 수로 설정하였다.

5. 분석 및 결과

5.1 회귀분석 결과

<표 5>는 '앨범에 대한 좋아요 수', '곡에 대한 좋아요 수', '긍정/부정 댓글의 수'가 다운로드 수에 미치는 영향을 알아보기 위해 수립한 회귀 모형에 대한 분석 결과를 나타낸다. 분석 결과, 앨범에 대한 좋아요 수와 곡에 대한 좋아요 수가 다운로드 수에 유의 수준 5%미만으로 유의하게 영향을 미친다는 사실을 알 수 있었다. 아쉽게도 댓글(긍정/부정) 변수에 대한 베타 계수는 유의한 값을 나타내지 않았다. 흥미로운 점은 앨범에 대한 좋아요 수와 곡에 대한 좋아요 수가 반대 방향으로 다운로드 수에 영향을 미친다는 사실이다. 앨범에 대한 좋아요 수는 다운로드 수에 부정적인 영향(베타 계수: -0.264, p-value<0.05)을 미친 반면, 곡에 대한

<표 5> Base Model 회귀분석 (다운로드)

	독립변수	Coefficient	Robust SE	p
다운로드	상수	.005	.041	0.892
	앨범 좋아요	-.264**	.131	0.046
	곡 좋아요	.836***	.203	0.000
	멜론 긍정 댓글	.097	.142	0.494
	멜론 부정 댓글	.067	.132	0.612
	통계량	Adjusted = 0.6450, F = 91.41, p = 0.000		
*** p < 0.01, ** p < 0.05, * p < 0.1				

<표 6> Base Model 회귀분석 (스트리밍)

	독립변수	Coefficient	Robust SE	p
스트리밍	상수	-.000	.056	0.997
	앨범 좋아요	-.203	.154	0.192
	곡 좋아요	.882***	.203	0.000
	멜론 긍정 댓글	-.054	.154	0.723
	멜론 부정 댓글	.042	.111	0.707
	통계량	Adjusted = 0.6262, F = 44.13, p = 0.000		
*** p < 0.01, ** p < 0.05, * p < 0.1				

좋아요 수는 다운로드 수에 긍정적인 영향(베타 계수: 0.836, p-value<0.01)을 미친다는 사실을 실증적으로 밝혔다. 하지만 아쉽게도 댓글이 다운로드 수에 미치는 영향을 밝히지 못했다.

이 결과에 대한 해석은 다음과 같다. 멜론 사이트 상의 곡에 대한 "좋아요"는 소비자의 소극적이지만 긍정적인 리뷰를 나타낸다. 소극적이라고 표현한 이유는 직접 댓글을 남기는 것에 비해서는 시간과 노력이 덜 소모되기 때문이다. "곡 좋아요" 수가 다운로드 수에 긍정적인 영향을 미친다는 결과는 다수의 선행 연구 결과와 마찬가지로 (Chen et al. 2004; 박지영과 유요섭 2011), 소비자의 긍정적인 리뷰가 상품 매출에 영향을 미쳤음을 보여준다. 하지만 "앨범 좋아요" 수가 다운로드 수에 음(-)의 영향을 미친다는 연구 결과는 해석하기 모호하다. 특정 곡이나 앨범에 대한 긍정적인 리뷰가 다운로드 수에 상이한 영향을 미치는 이유에 대한 추가적인 연구가 필요하다고 판단된다.

<표 6>은 <표 5>에서 사용된 회귀 모형에 종속 변수만 스트리밍 수로 바꾼 회귀 모형에 대한 분석 결과를 나타낸 표이다. 모든 독립 변수에 대한 결과는 <표 5>와 대체적으로 유사하지만 "앨범 좋아요"는 통계적으로 유의한 회귀 계수를 갖지 않게 되었다. 눈에 띄는 점은 "곡 좋아요" 수가 스트리밍 수에도 매우 유의한 영향을 미쳤다는 것이다. Bardhi & Eckardt (2017) 가 주장한 유동적 소비(스트리밍)와 고정적 소비(다운로드) 모두에 소비자의 긍정적인 리뷰("곡 좋아요")가 양(+)의 영향을 미쳤다는 것을 실증적으로 밝혔다. 하지만 여전히 긍정/부정 댓글이 종속 변수에 미치는 영향을 파악할 수 없었다.

<표 7>은 음원 소비자의 리뷰가 음원 다운로드 수와 스트리밍 수에 미치는 영향을 파악하기 위한 회귀 모형 분석 결과를 나타낸다. <표 5>, <표 6>과는 "기획사 크기"와 "히트곡 수"라는 조절 변수를 추가한 모형이라는 차이가 있다. <표 7>의 두 모형 (다운로드, 스트리밍)

모두 표 5, 6의 기본 모형보다 수정된 R 제곱 값이 증가함을 확인할 수 있고, 이는 조절 변수가 조절 효과를 가지는 것임을 나타낸다.

먼저 다운로드 수를 종속 변수로 설정한 회귀 모형의 분석 결과, "앨범 좋아요"와 "곡 좋아요" 는 기본 모형의 결과와 마찬가지로 다운로드 수에 각각 음(-), 양(+)의 영향을 미쳤다. 한편, 기본 모형과 상반되는, 혹은 추가적인 결과는 다음과 같다. 첫째, "멜론 긍정 댓글"이 다운로드 수에 통계적으로 유의한 양(+)의 영향을 미치는 것으로 드러났다. 즉, 소비자의 긍정적인 리뷰가 제품 매출에 영향을 미친다는 선행 연구 결과를 뒷받침한다 (Chen et al. 2004; 박지영과 유요섭 2011). 둘째, Graham et al. (2004)의 연구 결과와 마찬가지로 기획사 (혹은 음반사)의 규모가 소비자에게 양(+)의 영향을 미치지 않는 것으로 드러났다. 이러한 사실은 본 연구에서 조절 변수로 사용한 "멜론 긍정 댓글*기획사 크기"가 다운로드 수에 음(-)의 영향 (회귀계수: -.464, p-value<0.05) 을 미친다는 것을 통해 알 수 있다. 셋째, Strobl & Tucker (2000)가 제시했던 아티스트의 소위 "Superstar effect"가 음원 스트리밍 사이트 상의 다운로드 수에 긍정적인 영향을 미치지 않는 것으로 드러났다. 오히려 "멜론 긍정 댓글*히트곡 수"와 "멜론 부정 댓글*히트곡 수"가 다운로드 수에 미치는 영향은 각각 음(-, 회귀계수: -.317, p-value<0.01)의 방향과 양(+, 회귀계수: .300, p-value<0.05)의 방향이었다.

<표 7> 조절효과 회귀분석

종속변수	독립변수	회귀 계수 (표준오차)	종속변수	독립변수	회귀 계수 (표준오차)
다운로드	히트곡	.0557 (.038)	스트리밍	히트곡	.075* (.042)
	기획사	-.109 (.071)		기획사	-.134 (.122)
	앨범 좋아요	-.373*** (.129)		앨범 좋아요	-.231 (.179)
	곡 좋아요	1.002*** (.136)		곡 좋아요	1.031*** (.157)
	멜론 긍정댓글	.523*** (.176)		멜론 긍정댓글	.311 (.309)
	멜론 부정댓글	-.142 (.138)		멜론 부정댓글	-.162 (.239)
	멜론 긍정댓글 * 기획사	-.464** (.223)		멜론 긍정댓글 * 기획사	-.489 (.360)
	멜론 부정댓글 * 기획사	.167 (.184)		멜론 부정댓글 * 기획사	.265 (.248)
	멜론 긍정댓글 * 히트곡	-.317*** (.116)		멜론 긍정댓글 * 히트곡	-.208 (.167)
	멜론 부정댓글 * 히트곡	.300** (.119)		멜론 부정댓글 * 히트곡	.156 (.184)
	상수	.089 (.062)		상수	.116 (.108)
	Adjusted =0.731			Adjusted = 0.675	

*** p < 0.01, ** p < 0.05, * p < 0.1

본 연구 결과를 최근의 연구인 “다운로드/스트리밍이 선호되는 디지털 음원 특성에 대한 고찰” (유재원 외 2018) 의 연구 결과와 연계시키면 다음과 같은 결론을 낼 수 있다. 인지도가 높은 대상에 대해 사람들은 평균적으로 애착심을 덜 갖고 (Beatty et al. 1988; Eisentraut 2012), 아티스트에 강한 애착심을 갖지 못한 음원 소비자는 “소유”의 개념이 강한 음원 다운로드 이용에 부정적일 수 있다 (유재원 외 2018). 본 연구는 아티스트의 대중적 인지도를 대변하는 “히트곡 수”가 음원 소비자의 댓글이 다운로드 수에 미치는 영향에 대해 양(+)의 조절 효과를 갖지 못하다는 사실을 밝혔고, 선행 연구 결과를 재확인 할 수 있었다는 점에서 시사하는 바가 있다. 다음으로 스트리밍 수를 종속 변수로 설정한 회귀 모형의 분석 결과, “곡 좋아요”는 다운로드를 종속 변수로 둔 모형과 마찬가지로 스트리밍 수에 통계적으로 유의한 양(+)의 영향을 미친다는 증거를 발견할 수 있었다. 하지만, 아쉽게도 음원 소비자의 댓글이나 조절 변수가 스트리밍 수에 미치는 영향을 파악할 수 없었다.

5.2 다중공선성 검증

조절 변수를 구성 변수의 곱으로 계산하여 회귀분석을 시행할 때 다중공선성(Multicollinearity) 문제가 발생할 수 있다 (이도화와 이종법, 2010). 따라서 본 연구는 분산 팽창 인수 (VIF, Variance Inflation Factor)를 통한 다중공선성 검증을 실시했고, 해당 결과는 <표 8>에 제시 돼있다. 일반적으로 공차(Tolerance)가 0.1(VIF=10) 이상의 값을 갖는 독립변수는 다중공선성 문제가 있는 것으로 간주된다.

본 연구에서 사용된 회귀 모형 분석에 대한 다중공선성 검증 결과, 종속변수가 다운로드인 회귀 모형 분석 시 “멜론 긍정 댓글” (VIF=10.11), “멜론 부정 댓글” (VIF=12.27), “멜론 부정 댓글*기획사” (VIF=11.94)는 다중공선성 문제가 있는 것으로 보인다. 따라서 회귀분석 결과 및 연구자들의 해석은 제한적으로 수용되어야 함을 밝힌다. 하지만, 종속 변수가 스트리밍인 회귀 모형은 분산 팽창 인수 (VIF) 값이 10을 초과한 독립변수가 없고, 회귀 분석 결과 및 연구자들의 해석에 제약이 없다고 판단된다.

<표 8> 다중공선성 검증

다운로드			스트리밍		
독립변수	공차	VIF	독립변수	공차	VIF
히트곡	0.793	1.26	히트곡	0.889	1.12
기획사	0.954	1.05	기획사	0.948	1.05
앨범 좋아요	0.239	4.17	앨범 좋아요	0.281	3.55
곡 좋아요	0.346	2.89	곡 좋아요	0.375	2.66
멜론 긍정댓글	0.098	10.11	멜론 긍정댓글	0.110	9.06
멜론 부정댓글	0.081	12.27	멜론 부정댓글	0.125	7.96
멜론 긍정댓글 * 기획사	0.124	8.04	멜론 긍정댓글 * 기획사	0.120	8.28
멜론 부정댓글 * 기획사	0.083	11.94	멜론 부정댓글 * 기획사	0.148	6.75
멜론 긍정댓글 * 히트곡	0.171	5.85	멜론 긍정댓글 * 히트곡	0.203	4.92
멜론 부정댓글 * 히트곡	0.191	5.21	멜론 부정댓글 * 히트곡	0.234	4.26

<표 9> 연구가설 검증 결과

다운로드				스트리밍			
가설	검증 결과	가설	검증 결과	가설	검증 결과	가설	검증 결과
H1-a	채택	H1-e	채택	H2-a	기각	H2-e	기각
H1-b	기각	H1-f	기각	H2-b	기각	H2-f	기각
H1-c	기각	H1-g	채택	H2-c	기각	H2-g	기각
H1-d	채택	H1-h	채택	H2-d	채택	H2-h	기각

6. 결론

6.1 연구 결과 요약 및 시사점

본 연구는 텍스트 마이닝을 이용해서 음원 사이트상의 곡에 대한 댓글을 긍정/부정으로 분류한 뒤, 긍정/부정 댓글들의 수가 다운로드 수와 스트리밍 수에 어떻게 영향을 미치는지 실증 분석 해보았다. 먼저 다운로드 수와 스트리밍 수로 각각 종속 변수로 설정했고, 이 두 변수에 영향을 미치는 요인들을 식별하기 위해서 다수의 설명 변수들을 설정한 뒤 회귀 분석을 시행했다. 아래의 <표 9>는 연구 가설에 대한 검증 결과를 나타내는 표이다.

연구 모형 분석 결과에 대한 요약은 다음과 같다.

첫째, 멜론 사이트 상의 "곡 좋아요" 리뷰는 다운로드를 종속 변수로 사용한 모형과, 스트리밍을 종속 변수로 사용한 모두에서 긍정적인 영향을 미치는 것으로 드러났다. 즉, 소비자의 긍정적인 리뷰가 상품 매출 (음원 매출)에 기여한다는 사실을 밝혔다는 점에서 실무적인 시사점을 갖고, 이는 기존의 연구 결과(Chen et al. 2004; 박지영과 유요섭 2011)와 부합한다는 것을 확인했다는 점에서 학문적인 의의를 지닌다.

둘째, 멜론 사이트 상의 긍정 댓글이 음원 매출에 미치는 영향은 종속 변수를 다운로드로 설정한 회귀 모형에서만 유의한 결과가 도출됐다. "곡 좋아요"와 마찬가지로 "멜론 긍정 댓글"은 다운로드 수에 양(+)의 영향을 미쳤다. 이를 통해 음원 소비자의 리뷰가 상품 매출에 기여함을 밝힐 수 있었다. 따라서, 선행 연구 (유재원 외 2018) 결과를 재확인할 수 있었다는 점에서 학문적인 의미를 지니고, 리뷰 시스템의 기능적 장점을 실증적으로 밝혔다는 점에서 멜론과 같은 음원 산업 기업에게 시사하는 바가 있다.

셋째, "기획사의 크기"와 "히트곡 수"가 조절 효과는 갖는다는 것을 발견했다. 실무적인 관점에서 이 결과를 상품 판매 전략에 활용할 수 있을 것이다. 즉, 기획사가 크거나 히트곡 수가 많은 대중적인 아티스트의 음원에 대한 음원 소비자의 긍정적인 리뷰가 상품 판매에 부정적인 영향을 미친다는 사실은 오히려 효과적인 음원 판매 전략을 용이하게 할 수 있다. 예컨대, 대중적이지 않은 아티스트의 음원에 대한 댓글 남기기 시스템을 보다 더 촉진시킨다면 다운로드로 발생되는 매출을 향상시킬 수 있을 것이다.

6.2 연구의 한계점 및 향후 연구 방향

본 연구는 실증적 분석을 통해 음원 스트리밍 사이트상의 긍정/부정 댓글의 수가 다운로드 수와 스트리밍 수에 미치는 차등적 영향을 알아낼 수 있었다. 또한, 아티스트가 소속된 기획사의 규모, 아티스트의 히트곡 수에 따른 조절 효과 분석은 실무적인 참고 자료가 될 수 있을 것이다. 추후의 연구에서는 음원의 다양한 특성을 고려하고, 더 많은 샘플을 수집해서 다채로운 결과를

이끌어 낼 수 있을 것이라 생각한다.

하지만 본 연구에서는 시계열데이터 또는 패널 데이터를 분석에 사용하지 못했기 때문에, 정확한 인과 관계를 규명하지 못한 한계점이 명확하다. 기본적으로 상관관계를 통한 분석이기 때문에 인과관계론적인 해석에는 제약이 존재한다. 또한 댓글의 특성상 신조어나 비속어 혹은 특정 팬들만 사용하는 어휘 등이 많아서, 많은 텍스트 데이터들이 제외된 채로 분석이 진행됐기 때문에 이러한 어휘들을 더 세밀하게 정제할 수 있는 방법이 개발된다면 좀 더 명확한 결과를 제시할 수 있을 것이다.

참 고 문 헌

[국내 문헌]

1. 김창호, “온라인 구전의 정보방향과 평가내용이 구전효과에 미치는 영향: 대학생 휴대폰 구매관련”, *한국통상정보학회지*, 제8권, 제4호, 2006, pp. 23-41.
2. 박은아, 김영우, 서현숙, “온라인 사용 후기 정보의 영향력: 제품 유형에 따른 차이 분석”, *한국심리학회 연차 학술발표논문집*, 2007, pp. 584-585.
3. 박지영, 유요섭, “온라인 구전정보의 메시지 방향성이 구전정보의 효과성에 미치는 영향: 브랜드 인지도의 조절효과를 중심으로”, *한국서비스산업학회지*, 제8권, 제3호, 2011, pp. 15-28.
4. 박찬, 유창조, “온라인 구전커뮤니케이션이 상품평가에 미치는 영향에 관한 연구: 상표사용후기와 답글을 중심으로”, *소비자학연구*, 제17권, 제1호, 2006, pp. 73-93.
5. 배정환, 손지은, 송민, “텍스트 마이닝을 이용한 2012 년 한국대선 관련 트위터 분석”, *지능정보연구*, 제19권, 제3호, 2013, pp. 141-156.
6. 손지연, 어숙희, “화장품 온라인 쇼핑 시 온라인 구전 정보의 방향성이 소비자 구매의도에 미치는 영향”, *한국미용학회지*, 제14권, 제3호, 2008, pp. 929-945.
7. 손진아, 이은영, “인터넷 의류쇼핑에서 온라인 구전정보 특성 중 방향성과 동의성이 소비자 구매활동 변화에 미치는 영향”, *한국의류학회지*, 제31권, 제8호, 2007, pp. 1157-1167.
8. 양성수, 김인호, 정철, “디지털음악 콘텐츠서비스의 수용 요인이 지각된 가치와 고객만족에 미치는 영향”, *한국콘텐츠학회논문지*, 제16권, 제11호, 2016, pp. 456-463.
9. 유재원, 형지현, 이재영, “다운로드/스트리밍이 선

호되는 디지털 음원 특성에 대한 고찰", *마케팅연구*, 제33권, 제1호, 2018, pp. 1-21.
10. 이도화, 이종법, "조직공정성과 조직구성원 성과 사이의 관계에 대한 성별과 통제위치의 조절효과", *인적자원관리연구*, 제17권, 제2호, 2010, pp. 121-137.
11. 이상훈, 조장식, 강창완, 최승배, "텍스트 마이닝을 활용한 영화흥행 예측 연구", *한국데이터정보과학회지*, 제26권, 제6호, 2015, pp. 1259-1269.
12. 이소현, 이연경, 송의령, 김희웅, "토픽모델링을 이용한 국가간 행복과 불행 토픽 비교 분석: 한국, 미국, 영국, 브라질", *지식경영연구*, 제18권, 제3호, 2017, pp. 101-124.
13. 장리, 최강준, 이재영, "온라인 구전량 및 평점과 시기별 영화 흥행과의 관계", *지식경영연구*, 제18권, 제2호, 2017, pp. 65-83.
14. 한국인터넷 진흥원, 인터넷주제심층조사 요약보고서, "*웹 2.0 시대의 네티즌 인터넷 이용 현황: 참여와 공유의 인터넷*", 2006.

[국외 문헌]

1. Adler, M., "Stardom and talent", *The American economic review*, Vol. 75, No. 1, 1985, pp. 208-212.
2. Aguiar, L. and Martens, B., "Digital music consumption on the internet: evidence from clickstream data", *Information Economics and Policy*, Vol. 34, 2016, pp. 27-43.
3. Bardhi, Fleura, and Giana M. Eckhardt, "Liquid Consumption", *Journal of Consumer Research*, Vol. 44, No. 3, 2017, pp. 582-597.
4. Basuroy, S., Chatterjee, S., and S. A. Ravid, "How critical are critical reviews? The box office effects of film critics, star-power and budgets", *Journal of Marketing*, Vol. 67, 2003, pp. 105-117.
5. Beatty, S. E., Pamela H., and Lynn R. K., "The involvement—commitment model: Theory and implications", *Journal of Business research*, Vol. 16, No. 2, 1988, pp. 149-167.
6. Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., and Telang, R.,"The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts", *Management Science*, Vol. 53, No. 9, 2007, pp. 1359-1374.
7. Cao, Q., W. Duan, and Q. Gan, "Exploring determinants of voting for the 'helpfulness' of online user reviews: A text mining approach", *Decision Support Systems*, Vol. 50, No. 2, 2011, pp. 511- 521.
8. Chen, P. Y., S. Y. Wu, and J. Yoon, "The impact of online recommendations and consumer feedback on sales", *Proceedings of the International Conference on Information Systems*, 2004, pp. 711-724.
9. Chevalier, J. A., and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews", *Journal of marketing research*, Vol. 43, No. 3, 2006, pp. 345-354.
10. Chintagunta, P. K., Gopinath, S., and Venkataraman, S., "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets", *Marketing Science*, Vol. 29, No. 5, 2010, pp. 944-957.
11. Duan, W., Gu, B., and A. B. Whinston, "The dynamics of online word-of-mouth and

product sales—An empirical investigation of the movie industry", *Journal of Retailing*, Vol.84, 2008, pp. 233-242.

12. Eisentraut, J., "The accessibility of music: participation, reception, and contact", 2012, *Cambridge University Press*.
13. Elliott, K. M., Understanding consumer-to consumer influence on the web (Doctoral dissertation), *Duke University,* 2002.
14. Forman C, Ghose A, and Goldfarb A., "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live", *Management Science*, Vol. 55, No. 1, 2009, pp. 47-57.
15. Graham, G., Burnes, B., Lewis, G. J., and Langer, "The transformation of the music industry supply chain: A major label perspective", *International Journal of Operations & Production Management*, Vol. 24, No. 11, 2004, pp. 1087-1103.
16. Hiller, R. S., "Sales displacement and streaming music: Evidence from YouTube", *Information Economics and Policy*, Vol. 34, 2016, pp. 16-26.
17. Kohli, A. K., and B.J. Jaworski "Market orientation: the construct, research propositions, and managerial implications," *The Journal of Marketing,* 1990, pp. 1-18.
18. Kretschmer, T. and C. Peukert, "Video Killed the Radio Star? Online Music Videos and Recorded Music Sales", *London School of Economics and Political Science*, 2017, LSE Library.
19. Rish, I., "An empirical study of the naive Bayes classifier", IJCAI 2001 workshop on empirical methods in artificial intelligence, 2011, IBM.
20. Strobl, E. A., and Tucker, C., "The dynamics of chart success in the UK pre-recorded popular music industry", *Journal of Cultural Economics*, Vol. 24, No. 2, 2000, pp. 113-134.

저 자 소 개

박 명 석 (Myeong-Seok Park)

한양대학교 비즈니스인포메틱스학과 대학원을 졸업했으며, 현재는 (주)디구루 데이터분석실에 재직중이다. 주요 관심 분야는 머신러닝, 시계열 분석, 데이터 전처리 및 시각화 등이다.

권 영 진 (Young-Jin Kwon)

한양대학교 경영학과를 졸업한 후, 현재 동 대학원에서 경영정보시스템 석사 과정 중이다. 주요 관심 분야는 정보 보안 투자, 정보시스템 투자, 소셜미디어 등이다.

이 상 용 (Sang-Yong Tom Lee)

현재 한양대학교 경영대학 교수로 재직 중이다. 서울대학교 경제학과를 졸업하고, Texas A&M University에서 박사학위를 취득하였다. 주요 관심 분야는 정보경제, 개인정보보호 (privacy) 및 보안, 소셜미디어, 정보통신정책, 기술경영 등이다. 관련 연구들을 MIS Quarterly, Management Science, Journal of Management Information Systems를 비롯한 다수의 저널에 관련 논문을 게재하고 있다.