

## 텍스트마이닝을 활용한 오프라인과 온라인 고객 경험 분석 및 O2O 전략 제안: 프로야구산업을 바탕으로\*

천 세 연(제1저자)

아주대학교 경영대학 비즈니스애널리틱스 (석사)

이 지 은(공동저자)

한양사이버대학교 경영정보·AI비즈니스학과(교수)

강 주 영(교신저자)

아주대학교 e-비즈니스학과 (교수)

## Analyzing Offline and Online Customer Experience Using Text Mining and Suggesting O2O Strategies: based on the Professional Baseball Industry

Cheon, Se Yeon(First Author)

Department of Business Analytics, Ajou University (Master)

Lee Ji-Eun(Co Author)

Department of Management Information·AI Business, Hanyang Cyber University

Kang, Ju Young(Corresponding Author)

College of Business Administration, Ajou University (Professor)

### Abstract

\* "본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음" (IITP-2023-2018-0-01424)

The COVID-19 has led to a shift in many industries from offline to online, with offline activity decreasing and online channel activity increasing, and this shift can be seen in the sports industry as well. However, professional baseball has been centered on offline revenue generation, and as regulations are relaxed due to COVID-19, each team should explore strategies to convert online viewers to offline viewers. This study aims to collect experience data from offline stadium visitors and online game viewers and analyze them to derive O2O strategies for professional baseball. This study analyzed the data through using text mining methods and based on this, we identified and compared the attributes of experience differences between offline and online game viewers. As a result of the study, there were significant differences in the experience attributes of offline and online game viewers, and O2O strategies were proposed based on the results.

*Keywords* : Professional baseball, O2O strategies, Text mining, BERTopic  
 접수일(2023년 02월 20일), 수정일(2023년 03월 16일), 게재확정일(2023년 03월 16일)

## I. 서론

코로나 19는 과거 경제 위기와는 달리 여러 산업 구조에 큰 변화를 일으키고 있으며, 대표적으로 오프라인 위주의 산업들의 온라인화이다. 코로나 19 여파로 인해 쇼핑, 교육 및 의료 등 여러 산업에서 오프라인의 활동은 감소하였으나 비대면 트렌드가 활성화됨에 따라 이러한 변화에 맞추어 산업 내 온라인 채널의 활동은 증가하였다(지방공기업평가원, 2020). 소비자들의 지출 변화를 통해 이를 확인할 수 있었으며, 온라인에서 이뤄지는 활동에 대한 소비자 지출은 2019년 대비 2020년 5월에 11% 증가했다. 그러나 대인 접촉이 필수적인 서비스업에서는 약 75%의 지출이 감소하였으며, 이는 대면 시장을 수요로 가지는 업종인 숙박, 음식업, 문화나 기타 산업이 코로나 19의 충격을 피하지 못한 것으로 볼 수 있다(이민우·유지은, 2020). 그럼에도 외식업에서는 배달 앱과 배달 대행 서비스를 도입하여 오프라인 매장에서의 방문 고객 수 감소로 인한 매출액 피해를 온라인을 통해 줄이고자 하였으며, 문화 산업 중 하나인 영화시장은 OTT 등 인터넷 VOD로의 전환을 통해 매출 피해를 완화하고자

하였다(한국농촌경제연구원, 2021; 국회입법조사처, 2022). 이렇듯 코로나 19는 여러 산업을 오프라인에서 온라인으로 빠르게 전환시켰다.

최근 코로나로 인해 생긴 규제들이 완화되면서 온라인화가 되었던 여러 산업에서 온라인 고객을 오프라인으로 이동시키고자 노력하고 있다(경제정보센터, 2021). 온라인에서 오프라인, 즉 O2O(Online-to-offline)은 온라인에서 잠재 고객에게 인지도를 높이고 이를 통해 고객을 오프라인으로 이동시켜 오프라인에서의 판매 및 사업 수익을 증가시키는 것을 목표로 한다(Leung et al., 2016).

이러한 변화는 스포츠산업에서도 확인할 수 있다. 스포츠산업은 오프라인에서의 수익 창출이 주를 이루고 있으며, 프로스포츠에서는 수익 창출의 주체인 구단이 입장 수익, 모기업의 지원, 광고, 응원 용품과 굿즈 판매 등의 활동을 통해 수익을 창출하고 있다(매일경제, 2021).

프로스포츠 산업에서 수입 규모가 가장 큰 프로야구는 코로나 19로 인한 오프라인 직관 규제로 무관중 경기가 시행되면서 오프라인으로부터의 수익이 절대적으로 감소하였다. 특히, 프로야구는 전 세계에서 최초로 코로나 이후 무관중 경기로 리그를 시작하였으며, 그로 인한 수익 피해가 더 클 것으로 예상된다(서효민·권영범, 2020). 다른 산업들과 마찬가지로 코로나 19로 인한 규제들이 완화에 따라 프로야구는 2022년 4월 새로 개막하는 시즌부터 무관중 및 제한적 관중 규제가 해제되었고, 온라인으로만 경기를 볼 수밖에 없던 시청자들은 오프라인을 방문하여 경기를 직관할 수 있게 되었다. 그러나, 입장 규제가 완전히 해제되었음에도 2022년 기준 야구 경기장을 방문한 총관람객 수는 약 600만 명으로 2019년 코로나 19 발병 이전 총관람객 수인 약 728만 명에 못 미치는 수치로, 이는 코로나 19 이전만큼 경기장을 찾는 관람객이 회복되지 못했음을 알 수 있다(연합뉴스, 2022; 월간조선, 2022).

따라서 각 프로야구 구단들은 코로나 19 이전만큼 관람객을 유치하기 위해 온라인에서 오프라인으로 방문할 잠재적인 관람객을 파악하고 이러한 시청자들을 오프라인으로 이동시키는 전략이 필요하다. 이와 같은 전략을 수립하기 위해 구단은 경기를 온라인으로 보는 시청자와 오프라인으로 보는 관람객의 경험 차이를 이해하고 온라인과 오프라인에서 차별화되는 속성을 파악하는 것이 필요하다. 이를 위해 온라인과 오프라인에서 시청자와 관람객이 어떠한 경험적 차이를 보이는지에 대한 분석이 이루어져야 한다.

지금까지 이루어진 연구들은 오프라인 경기장 관람객을 대상으로 경기 관람 만족 등의 경험이나 재관람 의사 등을 다룬 연구와 온라인 실시간 스포츠 중계 시청자들의 미디어 경험으로서의 실재감을 조사한 연구로, 이러한 오프라인 관람과 온라인 시청에 대한 경험 비교 연구는 부족한 실정이다(권웅·이현우, 2015; 이동연·이수영, 2014; 김대원·조용운, 2016). 또한, 설문 조사로 진행되는 연구는 조사 대상이 한정되어 있

어 그 수가 상대적으로 적으며, 미리 설계된 설문지로 인해 조사 대상의 다양한 경험을 반영하기 어렵다는 한계가 존재한다(Coughlan et al., 2009; 채승훈 등, 2015). 이에 본 연구에서는 코로나 19 이후 온라인 시청자와 오프라인 관람객의 경기 경험 비교 분석을 위해 텍스트 마이닝 기법을 활용하고자 한다.

온라인 실시간 채팅 데이터와 오프라인 직관 후기를 수집하고 이를 통한 경험 비교 분석을 바탕으로 온라인 시청자를 오프라인으로 이동시키는 전략을 도출하고자 한다. 수집한 빅데이터를 바탕으로 텍스트 마이닝을 통해 핵심 키워드를 도출하고자 하며, 도출된 핵심 키워드 간 관련성을 파악하고 온라인과 오프라인에서의 경기 관람 속성에 어떠한 차이가 존재하는지 살펴보고자 한다. 또한, 이를 활용하여 프로야구 구단에서의 온라인에서 오프라인으로의 마케팅 전략을 제시하고자 한다.

본 연구는 다음과 같은 절차로 진행된다. 2절에서는 프로스포츠의 온라인과 오프라인 채널에 관한 선행 연구와 이러한 시장에서의 경험 차이를 진행하기 위한 텍스트마이닝 사용 연구에 대해 살펴보고, 이러한 텍스트마이닝을 활용하여 온라인과 오프라인 경험을 비교한 연구로 개념을 정립하고자 한다. 3절에서는 온라인 시청자와 오프라인 관람객의 경기 시청 경험 데이터 수집 방법과 이를 활용하여 진행한 연구방법에 대해 다루며 4절에서는 온라인과 오프라인 고객 경험의 차이점을 텍스트마이닝 결과로 확인한다. 마지막으로 5절에서는 연구 결과와 기여 및 한계점을 언급하여 향후 연구 방향을 제시한다.

## II. 선행 연구

본 연구에서는 프로스포츠 산업의 온라인 및 오프라인 채널에 관한 선행 연구와 이러한 채널에서의 경험의 차이를 이해하기 위한 텍스트 마이닝의 활용에 대해 검토한다. 이를 통해 온라인과 오프라인 경험을 비교하고 텍스트 마이닝을 사용한 연구의 개념을 확립한다. 또한, 온라인 및 오프라인 고객 경험 데이터 비교를 위한 텍스트 마이닝 기법을 사용하여 수행된 연구방법에 대해 논의하고자 한다.

### 2.1 프로스포츠산업의 오프라인 및 온라인 채널

프로스포츠가 지속적이고 더 많은 이익을 얻기 위해 구단은 기존 관람객을 유지하는 전략을 도출하고 잠재 고객을 찾아 오프라인으로 유도하는 방안을 모색해야 한다. 스포츠산업은 스포츠 활동에서 요구되는 용품과 장비, 스포츠 시설과 경기 및 서비스, 이벤트 등의 서비스를 제공함으로써 부가가치를 창출하는 산업으로 개념화할 수 있다

(한국문화체육관광부, 2020). 그 중, 프로스포츠 구단은 입장 수익, 경기장 광고와 식음료 및 상품 판매 등으로 수익을 창출하고 있으며 이는 경기장을 방문하는 관람객을 통해 발생된다. 이러한 관람객들을 통한 직접적 수익은 스폰서 유치 및 광고료, 중계료 등의 간접수익으로 이어지므로 구장을 방문하는 관람객과의 접촉은 구단에게 있어 매우 중요하다(고경진 등, 2018).

경기장 방문 관람객을 지속적으로 유지하는 방안으로는 관람객이 경기장을 방문하여 느낀 경험을 더 나은 경험으로 제공함으로써 충성 고객으로 전환시키는 것이다. 김주영 등(2017)은 관람객의 관람 제약요인을 도출하여 관람객이 지속적으로 경기장을 방문하고 관람하게 함으로써 이탈을 방지하고, 잠재 고객을 창출할 수 있으며 타구단과의 경쟁력을 강화한다고 밝혔다. Da Silva and Las Casas(2017)는 프로스포츠 구단 관람객 충성도에 영향을 미치는 요인들을 밝혔으며, 8가지 스포츠 경기 관람 동기를 정의하여 이를 바탕으로 관람객의 충성도와 동기에 따른 지속적인 경기장 방문을 위한 마케팅 전략 믹스를 사용해야 한다고 언급했다.

인터넷과 뉴미디어의 발전으로 온라인 채널이 등장하였으며, 대표적으로 실시간 온라인 스포츠 스트리밍 사이트를 통해 기존 관람객은 시간 제약 없이 어디서든 경기 시청이 가능해졌다. 이러한 인터넷 플랫폼 경기 시청률이 매년 큰 폭으로 증가했으며, 경기장 관람이 불가능한 경우 하이라이트 시청 또한 온라인 콘텐츠를 이용하여 시청하는 것으로 나타났다(문화체육관광부, 2020). 구단은 초창기에 뉴미디어를 단순한 정보를 제공하는 홍보와 인지도 상승에 초점을 맞추어 이용하였으나 현재는 팀과 팬의 양방향 관계를 발전시킬 수 있는 효과적인 소통 수단으로 인식하여 다양한 방법으로 이용하고 있다. Leszczyński et al.(2021)은 설문 조사를 통해 폴란드 스포츠 시장에서 다양한 세대의 팬이 디지털화된 채널에서 제공하는 스포츠 제품 및 서비스를 소비하는 방식을 파악하였으며, 그 결과로는 후세대로 갈수록 TV 방식이 아닌 소셜미디어 등의 온라인 플랫폼을 이용하여 경기 시청 및 정보를 얻는다고 밝혔다.

소셜미디어의 발전은 구단이 스포츠 팬에게 다양한 콘텐츠를 홍보하고 제공하며, 스포츠 팬은 이를 소비함으로써 구단과 팬의 새로운 접점을 만들었다. Obaid and CK(2022)는 구단이 팬들에게 스포츠 이벤트와 다양한 제품과 서비스 광고 및 팀 홍보 등 다양한 스포츠 마케팅을 진행하고, 소셜미디어와 소셜네트워크를 통해 팬들과의 관계를 이해하며 이에 맞는 마케팅 전략을 수립하는 것이 중요하다고 언급했다. Norris et al.(2015)는 스포츠구단의 소셜미디어 영향력은 빠르게 증가하며 점점 더 많은 프로스포츠 리그가 인터넷, 소셜 네트워킹 사이트 및 모바일 기술을 통해 팀을 홍보하고, 이를 통해 효율적으로 더 많은 스포츠 팬 또는 소비자에게 다가갈 수 있다고 언급했다. 소셜미디어를 통해 프로스포츠 구단 팬들 간 소통이 이루어지면서 새로운 피드백 및 의견 공유 등이 진행되고, 이는 구단과 팬 사이의 경험을 제공함으로써

팬들을 위한 스포츠 경험을 변화시키고 있다(Da Silva and Las Casas, 2017).

온라인과 뉴미디어, 소셜미디어의 발전은 팬들이 새로운 방식으로 라이브 스포츠를 경험하고 상호작용하는 방식으로 변화시켰다. 그러나 프로스포츠에서 경기장을 방문하는 관람객이 많아질수록 입장 수입이 늘어나는 것과 별개로 온라인 플랫폼에서의 중계권료가 한정되어 있으며, 구단은 실시간 경기 시청자의 증가로 인한 부가적인 수익이 제한되어 있다. 소셜미디어는 주로 구단이 스포츠 팬에게 다양한 홍보를 진행하고, 양방향 소통을 할 수 있는 창으로써의 역할을 수행했으나, 구단과 관람객과의 관계 유지 및 잠재 고객 발굴의 수단으로도 활용되어야 하며, 궁극적으로는 부가적인 수익 창출을 위한 오프라인 관람까지 이루어질 수 있게 활용되어야 한다.

따라서 구단은 온라인에서 오프라인 잠재 관람객 유치 및 기존 시청자 유지를 위해 소셜미디어 등 다양한 온라인 플랫폼에서 이들의 경험을 이해하고 분석하여 이를 바탕으로 오프라인 관람 활성화 방안에 대한 전략을 모색해야 한다.

## 2.2 O2O(Online-to-Offline) 전략과 고객 경험

온라인에서 오프라인의 전략은 즉, O2O(Online-to-Offline)는 온라인에 존재하는 잠재 고객에게 기업 인지도를 높이고 이를 통해 잠재 고객을 오프라인으로 이동시켜 오프라인에서의 판매 및 사업 수익을 증가시키는 것을 목표로 한다. 인터넷의 발전으로 고객들이 온라인에서 활동하는 시간이 길어졌으며, 이에 따라 기업은 온라인과 오프라인 채널에 모두 관여하는 고객 행동 패턴을 도출함으로써 변화하는 고객의 니즈를 파악해야 한다. 기업이나 소매업자들은 고객을 끌어들이기 위한 온라인과 오프라인의 상호작용을 통해 온라인 고객을 오프라인으로 이끌거나 혹은 온라인 채널을 바탕으로 오프라인에서의 판매를 유도해야 한다(Shen et al., 2019; Yao et al., 2022).

즉, 기업은 O2O 플랫폼을 통해 고객을 온라인으로 끌어들이고, 실제 서비스 대비 소비는 고객이 오프라인에서 경험하도록 하는 것이 중요하다(Du and Tang, 2014). 그러나, 대부분의 O2O 관련 스타트업은 비즈니스를 유지하기 위해 고객과의 지속 가능한 관계를 발전시키기보다는 매출 증대에 집중하였으며, O2O에서 중요한 고객의 행동을 면밀하게 파악하지 않았기에 실패했다고 밝혔다(Yao et al., 2022). O2O를 통해 오프라인에서의 기대한 서비스를 경험한 고객은 온라인 고객으로 유치할 수 있으나, 온라인에서의 고객이 오프라인 매장에서 기대하는 서비스가 없다면 그 고객은 더 이상 O2O 플랫폼을 신뢰하지 못하게 된다(Shen et al., 2019). 따라서, 기업은 온라인에서 고객의 니즈를 파악한 후에 이를 오프라인으로 적용하여 고객이 서비스에 대한 기대를 충족할 수 있도록 해야 한다. Leung et al.(2016)은 뷰티 산업에서 O2O를 통해 고객 충성도를 높이기 위한 핵심 요인 및 속성을 규명하였으며, 이를 위해서는

고객이 불만을 가지는 요소를 줄이고 만족도를 높이는 요소를 극대화하는 O2O 마케팅 전략을 활용해야 한다고 언급했다. Wan and Chen(2019)은 O2O를 성공적으로 진행하기 위해 고객의 행동을 이해하는 것이 중요하며, 오프라인에서의 경험을 명확히 한 후 온라인에서의 다양한 서비스를 제공하여 고객이 만족할 수 있는 전략을 제시해야 한다고 밝혔다.

따라서, 성공적인 O2O를 위해 기업은 고객이 오프라인에서 만족하는 요인을 발견하고, 이를 온라인을 통해 다양한 방법으로 홍보 및 서비스를 제공함으로써 잠재 고객을 발굴하고 기존 고객이 지속적으로 기업에 관심을 가질 수 있도록 해야한다.

### 2.3 텍스트마이닝을 활용한 O2O 연구

기존 O2O 관련 연구에서는 설문 조사와 인터뷰 등의 방법론이 주로 사용되어 왔다. Liang et al.(2014)는 설문 조사를 통해 O2O에서 중요한 소비자 신뢰도 요인을 연구하여 이에 대해 평가하였으며, Gao et al.(2014)는 설문 조사를 통해 O2O 플랫폼으로 전환하는 비용의 결정 요인을 탐구하고 이러한 요인이 전환 비용에 영향을 미치는지를 연구하였다. 그러나 이러한 설문 조사와 같은 방법론은 고객의 응답을 즉각적으로 얻기 어려울 뿐만 아니라 많은 시간과 비용을 투자해야 한다는 한계점을 가지고 있다(Chang et al., 2020). 이를 해결하기 위해 텍스트마이닝 기법을 바탕으로 O2O 관련 연구들이 진행되고 있다.

텍스트 마이닝은 데이터 마이닝을 텍스트 데이터로 확장한 것이며, 구조화되지 않은 텍스트 데이터에서 의미 있는 정보를 추출하는 기술이다(He et al., 2013). 즉, 텍스트 마이닝은 주어진 텍스트 데이터에서 숨겨진 패턴이나 경향을 자동으로 식별하며 추세를 설명하는 해석 또는 모델을 만든다(Guo et al., 2012). 텍스트 마이닝은 핵심 단어를 활용한 빈도분석이나 토픽모델링, 감성 분석 등을 포함하는 분석이며, 뉴스, 웹사이트나 SNS 등 텍스트가 존재하는 다양한 영역에서 활용되고 있다(Hu and Liu, 2012; 김기탁, 2020).

Chang et al.(2020)은 텍스트마이닝 방법론을 활용하여 온라인 레스토랑 리뷰 감성을 바탕으로 고객의 O2O 신뢰성에 영향을 미치는 주요 요인을 파악했다. 이 연구에서는 기존연구에서의 O2O 신뢰도에 영향을 미치는 요소를 찾고, 고객의 텍스트 리뷰 감성을 분류하는 예측 모델을 구축하여 추후 O2O 시장 내 기업이 고객의 신뢰를 해하지 않으며 서비스 품질을 향상시키기 위해 고객 의견에 응대해야 한다고 언급했다.

Shen et al.(2019)은 소셜미디어에서 O2O 트렌드 분석을 연구하기 위해 텍스트 마이닝을 활용하였으며, 이러한 미디어 기반 추세 분석을 통해 기업에서의 비즈니스적 이점을 얻을 수 있다고 언급했다. 이 연구에서는 트렌드를 확인하기 위해 소셜미디어

에서 수집한 단어의 빈도와 가중치를 직접 계산하여 사용하였다. 이와 유사한 방법으로는 텍스트 마이닝 기법 중 TF-IDF(Term Frequency-Inverse Document Frequency)가 있다. TF-IDF는 단어의 빈도와 역 문서 빈도를 사용하여 단어마다 중요한 정도를 가중치로 주는 방법으로, 주로 문서의 유사도를 구하는 작업이나 문서 내에서 특정 단어의 중요도를 구하는 작업 등에 사용되고 있다. TF-IDF는 특정 문서에서 나타나는 단어의 등장 빈도수와 그 특정 단어가 등장한 문서의 역수를 곱하여 이를 곱한 값으로, 이를 통해 모든 문서에서 자주 등장하는 단어의 중요도를 낮게 평가하고 특정 문서에서만 등장하는 단어의 중요도를 높게 판단한다. 따라서 TF-IDF 값이 낮으면 그 중요도가 낮다고 판단되며 반대로 TF-IDF 값이 크면 중요도가 큰 것으로 평가된다(유원준·안상준, 2021).

텍스트 마이닝 분야 중 하나인 토픽모델링은 텍스트 내에 속한 단어를 분석하여 문서의 주제를 도출하는 방법론으로, 소셜미디어, 뉴스 데이터 및 저널 등의 텍스트를 통해 정보를 얻고자 하는 연구에서 다양하게 활용되고 있다(김성근 등, 2016). 토픽 모델링은 단어를 근간으로 하여 토픽을 추출하고, 그에 해당하는 텍스트들을 대표하는 특정 이슈나 주제들을 자동으로 추출하는 분석이므로, 빅데이터를 가지고 잠재적인 주제나 쟁점을 분석하는 데 유용한 기법이다(유소연·임규진, 2021). 토픽모델링은 대표적으로 잠재 디리클레 할당(Latent Dirichlet Allocation) 방법이 많이 사용되었으나, 이는 최적의 결과를 얻기 위해 토픽의 수, 불용어 목록, 형태소 분석이 필요하며 단어간 의미 관계가 무시된다는 약점을 가지고 있다(고영수 등, 2022). 이에 최근 클러스터링 기술과 TF-IDF를 활용한 BERTopic 기법이 제안되었다(Grootendorst, 2020; 고영수 등, 2022). BERTopic은 BERT 기반 트랜스포머와 c-TF-IDF를 활용하여 단어에 가중치를 할당한 다음, 쉽게 해석할 수 있도록 클러스터링을 진행하며 대량의 텍스트 데이터에서 잠재된 의미 있는 주제를 찾아낸다. BERTopic은 다음과 같은 단계로 진행된다. 먼저, 사전 훈련된 SBERT를 사용하여 텍스트 임베딩을 만든 후, UMAP를 통해 임베딩 차원을 축소한다. HDBSCAN으로 축소된 임베딩을 클러스터링하며, 마지막으로 c-TF-IDF를 사용하여 핵심 단어를 도출한다. 이러한 BERTopic은 다른 토픽모델링 기법과 비교했을 때 높은 주제 일관성과 주제 다양성을 보이는 것으로 확인되었다(Alhaj et al., 2022; Grootendorst, 2022).

Yun et al.(2021)은 디지털 환경에서의 그림자노동 관련 연구를 진행하기 위해 O2O 주문 배달 앱에 대한 사용자 리뷰를 수집한 후, 토픽 모델링을 활용하여 키워드에 따른 주제어를 도출하였다. 키워드와 사용자 리뷰 분석 결과와 설문조사를 통한 양적 연구를 동시에 수행하여 O2O 주문 배달 앱의 사용자가 인지하는 서비스 품질 요인들을 파악하고, 이 중 그림자노동의 특성으로 무엇이 있는지를 파악하며 개선이 필요한 속성들을 알아보았다.

### Ⅲ. 연구방법

#### 3.1 연구 개요

성공적인 O2O를 위해서는 고객의 행동을 이해하는 것이 중요하며, 이를 위해 오프라인에서의 경험을 명확히 밝히고 이를 온라인 서비스로 제공함으로써 고객이 오프라인과 온라인 모두에서 만족할 수 있게 해야 한다(Wan and Chen, 2019). 특히, 프로야구는 코로나 19로 인한 입장 제한 규제가 해제됨에 따라 온라인으로만 시청이 가능했던 시청자들을 오프라인으로 이동시키기 위한 전략을 모색해야 한다.

이에 각 구단은 오프라인 경기 관람객과 온라인 시청자의 경험을 파악하고 이를 차별화하는 전략이 필요하다. 기존 프로스포츠에서 직관 관람 만족이나 경기 관람 의도와 온라인 서비스에 관한 연구는 설문 조사 등의 방법론을 통해 이루어져 왔다(채한승·이종호, 2000; 서효민·권영범, 2020; 허단비 등, 2022). 그러나 이러한 방법론은 비용과 시간이 많이 들고, 설문 대상을 정확하게 설정하지 않으면 결과가 오염될 수 있다는 단점이 존재한다. 이에 따라, 소셜네트워크 데이터를 통해 프로야구 팬들의 관람 문화조사와 관련된 연구와 실시간 스트리밍 스포츠 중계 채팅 데이터를 분석한 연구로는 각각 진행된 바 있으나 이러한 오프라인 직관 경험과 온라인 시청 경험을 비교한 연구는 거의 찾아볼 수 없었다(김기탁, 2020; 김대원·조용운, 2016). 따라서, 본 연구의 목적은 오프라인 직관 후기와 온라인 실시간 채팅 데이터를 비교하고 분석하며 경기 관람에 대한 경험 속성 차이를 파악해 이를 활용하여 O2O 전략을 수립하는 것이다.

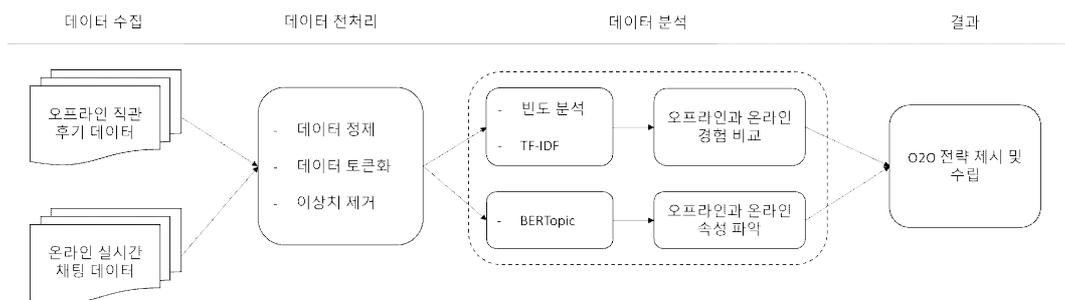
본 연구에서는 온라인에서 작성된 프로야구 오프라인 직관 후기와 온라인 스트리밍 실시간 채팅 데이터를 수집하여 텍스트 마이닝 기법을 통해 작성된 게시글과 채팅 데이터의 내용을 분석함으로써 오프라인에서의 경기 관람과 온라인 경기 시청의 경험 속성 차이를 파악하고자 한다. 이러한 텍스트 마이닝을 통해 도출된 경험 차이를 바탕으로 프로야구에서의 O2O 전략을 제안하고자 한다. 이를 위해 텍스트 마이닝 기법 중 빈도분석과 TF-IDF, BERTopic 토픽모델링을 활용하였다. 수집한 오프라인 직관 후기와 온라인 스트리밍 실시간 채팅 데이터에서 가장 많이 등장한 단어를 각각 추출하기 위해 빈도분석을 진행하였으며, 이 중 특정 문서에서만 등장하는 주요한 단어를 파악하기 위해 TF-IDF를 진행하였다. 이를 통해 오프라인 직관 후기와 온라인 실시간 채팅 데이터 내 전반적으로 언급된 경험 속성을 파악하고자 하였다. 또한, 오프라인 직관 후기와 실시간 채팅 데이터 내에서 대표하는 특정 이슈나 공통된 주제들을

자동으로 추출하는 분석인 토픽모델링을 진행하고자 하였다. 기존연구에서는 잠재 디리클레 할당 방법이 많이 사용되었으나 이는 단어 간 의미 관계가 무시된다는 약점을 가지고 있으며, 이에 따라 클러스터링 기술과 TF-IDF를 활용한 BERTopic 모델을 토픽모델링 기법으로 활용하고자 하였다. 이러한 BERTopic은 다른 토픽모델링 기법과 비교하였을 때 높은 주제 일관성과 주제 다양성을 보이는 것으로 알려져 있으며, 이에 따라 전술한 세 가지의 분석기법들을 바탕으로 오프라인과 온라인의 주요 경험 속성의 차이를 파악하고자 하였다.

### 3.2 연구 절차

본 연구는 다음 <그림 1>과 같이 진행된다. 첫째, 웹 크롤링을 통해 오프라인 직관 후기와 온라인 실시간 채팅을 수집한다. 둘째, 수집한 데이터에서 주제어와 관련이 없거나 이상치 단어를 제거하는 데이터 전처리 과정을 거친다. 셋째, 단순 빈도분석과 TF-IDF 분석을 통해 오프라인 후기 데이터와 온라인 채팅 데이터에서 많이 등장하는 단어 및 중요한 키워드를 추출하며, BERTopic 토픽모델링 분석을 진행한다. 마지막으로, 도출된 분석 결과로부터 얻은 인사이트를 바탕으로 O2O 전략을 제시한다.

<그림 1> 연구절차도



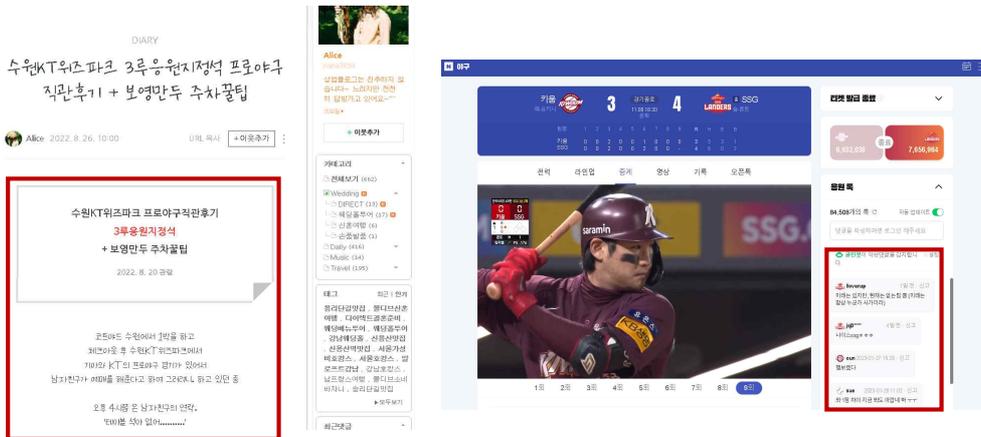
#### 3.2.1 데이터 수집

본 연구에서는 오프라인 관람객과 온라인 시청자의 경기 시청 방식에 따른 경험 차이를 확인하기 위해 파이썬 셀레니움 라이브러리를 사용하여 온라인에서 작성된 오프라인 경기 직관 후기와 실시간 온라인 실시간 채팅을 수집하였다. 온라인상에서 작성된 리뷰는 작성자의 다양하고 생생한 경험을 담고 있으며(채승훈 등, 2015), 실시간 채팅 데이터는 현재 경기를 시청하는 사람의 느끼는 경험을 바로 내포하고 있다는

점에서 이와 같은 데이터를 선정하였다. 조사 기간은 코로나 19 이후 무관중 및 제한적 관중 입장 규제가 해제된 2022년 4월부터 2022년 11월까지로, 코로나 19 이전과 같은 관중 100% 수용 상황이 만들어진 프로야구 한 시즌으로 설정하였으며 이는 정규시즌과 포스트시즌, 올스타 경기를 모두 포함한다.

본 연구에서 사용된 오프라인 경기 직관 후기 데이터는 야구 관련 커뮤니티와 'N'사 포털사이트에서 '프로야구 직관 후기'를 키워드로 한 게시글 중심으로 수집하였으며, 프로야구 시즌과 관련 없는 내용이 포함된 게시글을 제외하여 각각 252개와 260개의 데이터를 수집하였다. 다음으로는, 프로야구 한 시즌의 온라인 스트리밍 사이트 내 실시간 채팅을 수집하였다. 프로야구 정규시즌 경기는 월요일을 제외한 모든 요일에 10개의 팀 5경기로 진행되어 한 팀당 144경기를 시행한다. 2022년 포스트시즌은 총 16경기가 치러졌으며 올스타 경기 1개를 포함하여 총 737경기의 19,277,048개 실시간 채팅을 수집하였다.

<그림 2> 오프라인 직관 후기 게시글과 온라인 실시간 채팅 수집 화면 예시 ('N' 포털사이트 내 프로야구 직관 후기 게시글과 'N' 포털사이트 실시간 중계 화면)



<그림 2>와 같이, 온라인에서 작성한 오프라인 직관 후기 게시글과 온라인 스트리밍 사이트 내 실시간 채팅의 데이터 구조가 다른 양상을 가지고 있으며, 이에 따라 오프라인 직관 후기 게시글과 온라인 실시간 채팅 데이터 변수 수집에 차이를 보였다.

<표 1> 수집 데이터 예시

채널	변수명	변수설명	변수값 예제
오프라인	제목	게시글 제목	어제 한국시리즈 6차전 직관 후기
	작성일자	게시글 작성 일자	2022-11-09 16:18
	본문내용	게시글 본문 내용	아 일생에 있어 응원하는팀의 우승하는 장면을 직관으로 볼 날이 얼마나 있을까요...
온라인	경기ID	경기 고유 값	77771108WOSK02022
	작성일자	실시간 채팅 작성 시간	2022-11-08 21:28:13
	실시간 채팅 내용	실시간 채팅 댓글 내용	키움도 잘했어요 실책만 없었으면 재미 있었는데요

<표 1>은 전반적인 수집 데이터에 대한 예시이다. 본 연구에서는 온라인에서 작성된 오프라인 직관 후기와 온라인 스트리밍 사이트 내 실시간 채팅 데이터를 분석한 내용을 바탕으로 오프라인과 온라인 채널에서의 경험 차이 속성을 파악하고, 이를 활용하여 O2O 전략을 도출하고자 한다. 이에 따라 내용 분석이 가능한 직관 후기의 게시글 본문 데이터와 실시간 채팅 데이터를 수집하였다. 또한, 데이터 전처리 과정에서 데이터 수집 기간 내 작성되지 않은 데이터를 제거하기 위해 게시글과 온라인 채팅의 작성일자 변수와 온라인 실시간 채팅 데이터의 경기ID 변수를 수집하였다. 오프라인 직관 후기 데이터에서 수집된 제목 변수는 프로야구 직관 후기와 관련되지 않는 데이터를 제거하기 위해 사용되었다. 아래 <표 2>와 <표 3>은 수집한 오프라인과 온라인 데이터의 예시이다.

<표 2> 오프라인 직관 후기 수집 데이터 예시

작성일자	제목	본문 내용
2022-11-09 16:18	어제 한국시리즈 6차전 직관 후기	아 일생에 있어 응원하는팀의 우승하는 장면을 직관으로 볼 날이 얼마나 있을까요. 정...
2022-11-09 10:31	[스압][데이터] 인천SSG랜더스 2022시즌 우승 확정 및 시상식 한국시리즈 6차...	차전까지 전석 매진 위아래 마켓이 아쉽긴 하지만 어쨌든 기념 수건부터 챙기고...

## &lt;표 3&gt; 온라인 실시간 채팅 수집 데이터 예시

작성일자	경기ID	실시간 채팅 내용
2022-11-08 21:28:13	77771108WOSK02022	키움도 잘했어요 실책만 없었으면 재미 있었는데요
2022-11-08 21:28:58	77771108WOSK02022	멋찌다 트로피
2022-11-08 21:28:38	77771108WOSK02022	시리즈 mvp 폰트일까 최정일까??

**3.2.2 데이터 전처리**

본 연구에서 수집된 오프라인 직관 후기와 온라인 스트리밍 실시간 채팅 두 개의 데이터를 분석하기 위해 데이터 정제, 데이터 토큰화, 이상치 제거의 전처리 과정을 수행했다. 먼저, 오프라인 직관 후기 데이터는 포털 사이트와 커뮤니티 사이트에서 각각 수집하였으며 이를 바탕으로 내용 분석을 진행하기 위해 수집한 데이터를 하나의 데이터로 합치는 과정을 수행했다. 온라인 실시간 채팅 데이터 분석을 진행하고자 한 경기에 작성된 모든 실시간 채팅을 병합하여 하나의 문서로 만들었으며, 채팅 데이터라는 특수성으로 문장으로 이루어지지 않은 데이터는 모두 제거하였다. 또한, 두 데이터 모두 이모티콘과 특수 문자 등 분석과 관련 없는 단어를 제거하였다.

다음으로는 Konlpy의 Mecab 토큰라이저를 활용하여 명사를 추출하였다. 동사와 형용사 등 분석에 어려운 단어들을 제외하기 위해 명사만을 추출하였으며, 영어와 숫자 등의 분석과 관련 없는 단어들을 삭제하였다.

**3.2.3 빈도 분석 및 TF-IDF**

수집 및 전처리한 오프라인 직관 후기 데이터와 온라인 스트리밍 실시간 채팅 데이터를 바탕으로 빈도분석과 TF-IDF 분석을 진행하여 의미 있는 키워드를 파악하고자 하였다. 키워드 빈도 분석을 진행하기 위해 각 데이터의 토큰화된 명사를 사용하였으며, 가장 많이 등장한 상위 40개의 단어를 추출하였다. 상위 키워드를 빈도 순으로 추출한 후 TF-IDF 분석을 실시하여 해당 단어가 특정 문서에서 얼마나 중요한지를 파악하고자 하였다. TF-IDF 분석을 진행하기 위해 파이썬의 scikit-learn의 Feature extraction과 TfidfVectorizer 패키지를 활용하였다.

**3.2.4 BERTopic 토픽모델링**

오프라인 직관 후기 데이터와 온라인 스트리밍 실시간 채팅에서 나타나는 채널별 경기 경험 차이 관련 주제를 살펴보기 위해 기존에 많이 사용되었던 LDA 방식의 단점을 보완한 BERTopic을 활용하였다. 이를 위해 파이썬 3.8 버전에서 BERTopic

0.13.0 버전을 사용하였다. 기존 BERTopic은 내부적으로 scikit-learn의 CounterVetorizer를 사용하고 있으나 이는 단순한 띄어쓰기 토큰화를 수행하므로 이를 보완하고 한국어에 적용하기 위해 본 연구에서는 Mecab 형태소 분석기를 통해 토큰화를 진행하였다. 또한, 한국어에 대한 문서 임베딩 값을 얻기 위해 본 연구에서는 사전 훈련된 SBERT를 한국어 SBERT인 ‘KR-SBERT-V40K-klueNLI-augSTS’로 선택하였다.

## IV. 연구 결과

### 4.1 빈도분석 결과

본 연구에서 오프라인 직관 후기 데이터와 온라인 스트리밍 실시간 채팅 데이터를 텍스트 마이닝으로 도출한 결과 각각 7643개와 79238개의 단어가 수집되었으며, 수집된 데이터들을 통해 빈도분석 결과 상위 40개의 단어와 TF-IDF를 표와 워드클라우드로 시각화하여 정리하였다. 오프라인 직관 후기 빈도분석과 TF-IDF 결과는 <표 4>과 <그림 3>과 <그림 4>에 나타내었으며, 온라인 스트리밍 실시간 채팅 빈도분석 및 TF-IDF 결과는 <표 5>와 <그림 5>과 <그림 6>에 정리하였다.

먼저, 오프라인 직관 후기 단순 빈도분석 결과 “경기” 단어가 가장 많이 도출되었으며, 그 이후로 “야구”, “직관”, “선수”, “응원” 등의 단어가 많은 빈도로 도출되었다. 이러한 응원 활동이 관람객에게 있어 프로야구 경기 관람 몰입도를 높이며 관람 문화로서의 자리 잡고 있음을 확인할 수 있다(김용만, 2012). 또한, 오프라인 직관 후기 빈도분석 결과 중 “좌석”, “맥주”, “사진” 등 경기장 방문 전 사전 준비와 경기장을 방문하여 경험한 점을 나타내는 단어들이 등장하였다. 이와 같은 결과는 서비스스케이프 관점에서 프로야구장을 방문하는 관람객의 감정반응에 유의한 요인으로 도출된 좌석, 주차시설 및 먹거리 등의 편의시설 요인과 일치한다(이재호·전태준, 2017). 이는 관람객들이 경기장에서 이루어지는 경기 요소 외에도 다양한 경험을 한 것으로 판단된다.

온라인 실시간 채팅 단순 빈도분석 결과로는 “투수”, “선수”, “감독”, “경기” 등의 단어가 높은 빈도로 나타났으며, “선수”, “경기”, “야구” 등은 오프라인 직관 후기와 온라인 실시간 채팅 모두에서 높은 출현 빈도를 가지는 것으로 나타났다. 이는 김대원과 조용운(2016)의 실시간 스트리밍 스포츠 중계 채팅 내용분석 연구에서 경기의 주체인 선수, 감독, 팀에 대한 평가가 전체 채팅의 가장 큰 비중을 차지했다는 결과와 일치한다. 즉, 오프라인 직관 후기 빈도분석 결과로는 경기 상황과 더불어 응원, 먹거리

등의 다른 주요 경험들도 도출되는 반면 온라인 실시간 채팅 빈도분석 결과는 대체로 경기 내 상황과 관련된 단어들에 많이 도출된 것을 확인할 수 있었다.

다음으로는 TF-IDF 분석 결과로 도출된 점수를 통해 가중치를 부여하여 해당 단어의 중요도를 파악하고자 하였다. 오프라인 직관 후기 TF-IDF 분석 결과 “직관”이 가장 높게 도출되었으며, “경기”, “야구”, “선수”, “후기” 또한 모두 높은 순위를 보이는 것으로 보아 빈도분석 결과 높은 빈도의 단어들에 가중치를 고려함에도 중요한 요소임을 파악할 수 있었다. 온라인 실시간 채팅 TF-IDF 결과로는 “오늘”이 가장 높게 도출되었으며, “기아”, “롯데”, “경기”, “안타”, “삼성” 등의 단어들에 높은 순위를 보였다. 즉, “투수”, “선수”, “감독”, “야구” 등의 빈도가 높은 단어들은 많은 문서에서 도출되었음을 확인할 수 있다.

<표 4> 오프라인 직관 후기 빈도분석 결과

번호	단어	빈도	TF-IDF	TF-IDF 순위	번호	단어	빈도	TF-IDF	TF-IDF 순위
1	경기	1635	10.16	2	21	필드	318	0.47	40
2	야구	1407	10.00	3	22	시작	316	2.76	27
3	직관	1282	10.70	1	23	기아	312	4.02	17
4	선수	1252	6.66	5	24	화질	310	6.27	8
5	응원	870	4.92	15	25	읍선	307	6.24	9
6	야구장	621	5.88	12	26	자동	305	6.17	10
7	후기	572	6.29	7	27	재생	304	6.16	11
8	랜더스	485	4.56	16	28	좌석	297	2.60	29
9	롯데	462	5.78	13	29	파크	284	1.12	38
10	고척	438	2.42	32	30	시간	281	2.94	22
11	자리	428	2.56	30	31	맥주	280	2.88	23
12	사람	367	1.76	36	32	사진	271	2.80	26
13	잠실	360	5.27	14	33	원정	253	2.51	31
14	선택	352	6.32	6	34	오늘	253	8.00	4
15	티켓	348	3.41	20	35	친구	248	1.34	37
16	생각	341	0.81	39	36	타자	247	2.31	33
17	경기장	334	3.37	21	37	승리	247	1.91	34
18	삼성	325	3.77	19	38	다음	242	2.86	24
19	예매	322	1.80	35	39	투수	230	2.84	25
20	홈런	321	3.79	18	40	유니폼	221	2.62	28

<그림 3> 오프라인 직관 후기 빈도분석 결과 워드클라우드



<그림 4> 오프라인 직관 후기 TF-IDF 결과 워드클라우드



<표 5> 온라인 실시간 채팅 빈도분석 결과

번호	단어	빈도	TF-IDF	TF-IDF 순위	번호	단어	빈도	TF-IDF	TF-IDF 순위
1	투수	283008	14521.83	9	21	타격	105904	6155.155	25
2	선수	270895	10167.11	17	22	생각	91346	2293.851	38
3	감독	261983	14032.66	11	23	아웃	80046	5887.33	26
4	경기	260325	16561.44	4	24	응원	77863	3606.546	30
5	기아	258915	21277.25	2	25	포수	76495	6502.927	24
6	야구	242626	13108.59	13	26	이닝	72882	8430.189	20
7	롯데	219756	18431.42	3	27	시즌	72556	5310.755	29
8	엘지	183602	14893.46	8	28	불펜	71711	6754.034	23
9	삼성	163589	15779.29	6	29	문제	62988	2578.605	35
10	한화	163117	14479.49	10	30	올해	60686	7794.537	21
11	타자	159318	5328.308	28	31	연패	58646	6896.182	22
12	오늘	154865	39522.04	1	32	번트	58196	5845.95	27
13	홈런	135823	15346.13	7	33	상대	57640	2450.691	37
14	수비	132212	10470.8	16	34	중국	57610	578.7114	40
15	우승	129324	8558.775	19	35	정도	56870	1888.831	39
16	두산	117319	11949.22	14	36	구단	56788	3076.052	33
17	선발	116142	8883.981	18	37	프로	56547	2541.318	36
18	안타	115949	15962.05	5	38	교체	55646	3278.028	31
19	심판	113276	11485.77	15	39	타선	55235	3207.447	32
20	점수	110797	13286.68	12	40	코치	54804	2605.907	34



2번과 3번 토픽에서는 경기 상황과 관련된 단어들도 도출되었다. 2번 토픽에는 ‘선수’, ‘타자’, ‘투수’, ‘감독’ 등의 경기를 진행하는 플레이어와 ‘수비’, ‘타격’ 등 경기 상황에서 중요한 순간임을 나타내는 단어가 도출되었다. 이를 통해 이러한 경기를 진행하는 사람들과 경기 진행 중 특정 상황이 오프라인 경기에서 중요한 경험임을 알 수 있다. 3번 토픽에서는 주목할 만한 핵심어로 ‘타자’, ‘상황’, ‘타석’, ‘인터뷰’, ‘순간’ 등으로, 경기에서 특정 상황을 나타내는 단어들도 도출되었다. 2번과 3번 토픽모델링 결과를 통해 오프라인에서 경기를 직관한 관람객은 경기 플레이 상황을 보다 자세하게 기억하며 이를 후기로 작성한다는 점을 확인할 수 있었다. 이를 통해 오프라인에서의 경기 관람이 관람객들에게 깊은 인상을 남긴다는 특성을 파악할 수 있었다.

4번 토픽은 ‘경기’, ‘직관’, ‘선수’, ‘시즌’, ‘개막’ 등의 단어들도 도출됨으로써 이를 통해 새로운 시즌이 시작됨에 따라 경기장을 방문하는 사람들의 기대를 나타내고, 이를 경험하기 위해 경기장에 방문함을 알 수 있었다.

5번 토픽에서는 ‘와이프’, ‘아들’, ‘딸아이’, ‘오늘’ 등의 단어를 핵심어로 꼽을 수 있다. 경기장을 방문하는 관람객들은 가족이나 지인 등 주변인들과 함께이며, 오프라인 직관을 통해 추억을 경험한 내용으로 판단되었다.

<표 6> 오프라인 직관 후기 BERTopic 분석 결과

번호	토픽	문서 개수	비율	단어
0	-1	196	38.28%	야구, 경기, 선수, 직관, 응원, 고척, 후기, 기아, 야구장, 스카이
1	0	79	15.43%	야구, 직관, 경기, 잠실, 선수, 야구장, 후기, 응원, 잠실야구장, 사람
2	1	69	13.48%	랜더스, 경기, 야구, 선수, 필드, 직관, 야구장, 응원, 경기장, 수원
3	2	43	8.40%	야수, 선수, 타자, 투수, 타격, 감독, 수비, 생각, 선발, 정도
4	3	39	7.62%	타자, 상황, 타석, 안타, 최정, 랜더스, 김광현, 폰트, 인터뷰, 순간
5	4	39	7.62%	엘지, 경기, 직관, 선수, 시즌, 개막, 롯데, 박용택, 응원, 타자
6	5	32	6.25%	우승, 직관, 응원, 오늘, 와이프, 경기, 야구, 아들, 마지막, 딸아이
7	6	15	2.93%	롯데, 삼성, 한화, 야구, 대전, 기아, 한화이글스, 호텔, 광주, 원정

온라인 실시간 채팅 데이터 BERTopic 분석 결과는 아래 <표 7>과 같이 10개의 토픽으로 분류되었다. 오프라인 직관 후기 BERTopic 결과와 마찬가지로, -1 토픽은

이상치로 분류하여 결과 해석에 고려하지 않았다.

온라인 실시간 채팅 데이터 BERTopic 분석 결과는 오프라인 직관 후기 결과와는 다르게 특별한 단어들도 도출되지 않았으며, 주로 '롯데', '한화', '엘지', '삼성' 등과 같이 팀 이름과 선수 이름 및 '타격', '주루', '아웃', '심판', '홈런' 등과 같은 경기 상황과 관련된 토픽이 도출되었음을 확인할 수 있었다. 각 토픽별로 선수 이름과 관련된 단어들도 도출된 점은 경기를 진행하며 선수들로부터 특정 상황이 발생할 때 선수들의 이름을 언급하는 온라인 시청자의 행동으로 도출된 것으로 판단된다. 또한, 오프라인 직관 후기와 다르게 온라인 실시간 채팅 데이터에서는 경기 내적인 부분에 대해서만 단어들만 나타났으며, 이는 실시간 채팅을 남기는 시청자들이 경기 시청이라는 단일 경험만이 이루어지기 때문이라고 판단된다. 분석 결과를 통해 온라인으로 경기를 시청하는 시청자들에게 가장 중요한 점은 경기 내용임을 파악할 수 있었다.

<표 7> 오프라인 직관 후기 BERTopic 분석 결과

번호	토픽	문서 개수	비율	단어
0	-1	287	38.32%	삼성, 롯데, 경기, 투수, 선수, 야구, 수비, 선발, 엘지, 타자
1	0	98	13.08%	한화, 시환, 무적, 야구, 선수, 투수, 경기, 기아, 김도영, 감독
2	1	75	10.01%	엘지, 우승, 선발, 야구, 감독, 투수, 경기, 선수, 롯데, 지현
3	2	71	9.48%	야수, 선수, 타자, 투수, 타격, 감독, 수비, 생각, 선발, 정도
4	3	49	6.54%	롯데, 피터스, 동희, 심판, 반즈, 선수, 투수, 야구, 경기, 고승
5	4	47	6.28%	두산, 아웃, 심판, 엘지, 야구, 주루, 선수, 주자, 경기, 매수
6	5	33	4.41%	삼성, 감독, 델라, 선수, 투수, 경기, 김현곤, 야구, 타자, 삼영
7	6	27	3.60%	기아, 감독, 김도영, 양현종, 선수, 투수, 김종국, 경기, 대인, 박찬호
8	7	26	3.47%	선수, 투수, 기아, 경기, 김광현, 야구, 홈런, 감독, 롯데, 오늘
9	8	22	2.94%	기아, 감독, 대인, 중국, 선수, 김도영, 경기, 투수, 심판, 삼성
10	9	14	1.87%	엘지, 두산, 감독, 지현, 선수, 투수, 경기, 홈런, 타자, 야구

## V. 토의 및 결론

코로나 19 범유행은 스포츠산업을 포함한 다양한 산업 분야에서 오프라인에서 온라인 활동으로 전환되는 계기가 되었다. 특히 프로야구 구단은 코로나 19 규제로 관중 없이 경기가 진행되면서 오프라인 관중이 감소했다. 구단은 이러한 규제가 완화됨에 따라 온·오프라인 시청자의 경험 차이를 파악하고 이해함으로써 온라인 시청자를 다시 경기장으로 끌어들이는 전략을 도출해야 한다. 이를 위해 본 연구는 텍스트 마이닝 기법으로 오프라인 고객 직관 후기와 온라인 라이브 채팅 데이터를 수집하고 빈도 분석 및 TF-IDF, BERTopic 토픽모델링을 통해 분석하고 주요 경험 속성을 도출해 이를 활용하여 온라인 시청자를 오프라인으로 이동시키는 O2O 마케팅 전략을 수립하는 것을 목적으로 하였다. 텍스트 마이닝 분석 결과는 다음과 같다.

오프라인 직관 후기에서의 빈도분석 결과 응원과 직관 등의 경험을 나타내는 단어와 선수, 감독 등과 같은 경기 내에 중요한 인물과 경기 상황에 관련된 단어들인 높은 빈도를 나타냈다. 또한, 티켓 예매와 좌석, 맥주나 사진 등의 경기장을 방문 사전 준비와 경기장 서비스스케이프와 관련된 단어들도 다음으로 도출되었다. TF-IDF 분석 결과로는 빈도분석에서 높은 빈도의 단어들인 TF-IDF 점수 또한 높았으며, 이를 통해 이러한 단어들인 가중치를 고려함에도 중요하다는 점을 파악할 수 있었다.

온라인 실시간 채팅 분석의 빈도분석 결과로는 투수와 타자, 감독 등과 같은 경기를 진행하는 사람들과 관련된 단어들인 높은 빈도로 도출되었다. TF-IDF 분석 결과로는 팀 이름과 관련된 단어들인 높게 나타났으며, 빈도분석에서 높은 빈도로 나타난 단어들 또한 높은 TF-IDF 점수로 나타났다. 이러한 빈도분석과 TF-IDF 분석을 통해 오프라인 관람객과 온라인 시청자의 경험 속성의 차이를 확인할 수 있었다.

이후, BERTopic 결과로 오프라인 직관 후기에서는 경기장을 방문하여 경험할 수 있는 토픽들이 주로 도출되었으며, 이와 반대로 온라인 실시간 채팅 데이터에서는 경기 상황과 관련된 토픽들이 주로 나타났음을 확인하였다. 이를 통해 오프라인 관람객들은 경기장을 직접 방문하여 느끼는 다양한 경험이 존재하지만, 실시간 채팅을 남기는 시청자들은 경기 시청이라는 단일 경험만 가지고 있음을 확인하였다.

스포츠 온라인 시청자를 경기장으로 이동시키기 위해 오프라인 방문 관람객의 경험을 파악해야 하며 본 연구에서는 빈도분석과 토픽모델링 결과를 통해 다음과 같은 O2O 마케팅 전략을 제시한다.

첫째로, 빈도분석과 토픽모델링 결과 응원과 관련된 단어들인 많이 등장했다. 김용만(2012)의 연구에서는 응원 만족이 관람 태도와 재관람 의도에 영향이 있음을 언급하며, 이를 통해 기존 관람객의 충성도를 높일 뿐만 아니라 잠재 관람객에도 긍정적인 영향이 미친다고 밝혔다. 따라서, 각 구단은 경기장을 방문하여 생생한 경기 시청

과 온라인에서 경험할 수 없는 분위기를 강조하고, 이러한 경험을 하고자 하는 온라인 시청자에게 어필할 수 있는 전략을 세울 수 있다. 둘째로, 토픽모델링 결과를 통해 가족 및 지인과 관련된 단어들이 도출되었다. 프로야구 관람객 성향조사 보고서(2022)에 따르면, 경기 관람객은 경기장에 같이 방문하는 가족 및 지인들과 함께 즐길 수 있다는 점을 프로스포츠 경기 관람의 장점으로 생각하며, 이를 바탕으로 잠재 관람객이 유입된 구단이 많음을 확인할 수 있었다. 따라서 경기장에 방문하는 사람들에게 새로운 추억을 제공할 수 있는 마케팅 전략을 모색해야 한다.

마지막으로 빈도분석을 통해 좌석, 맥주와 사진 등의 단어가 등장했다. 경기 관람객들은 경기 관람 전 경기장 대표 먹거리와 현장에서의 구단 이벤트 등을 기대 요인으로 여기며 식음료 다양성, 좌석 편의성 등을 직관 결정 영향요인으로 여긴다(한국프로스포츠협회, 2022). 따라서, 구단은 오프라인 경기장에서의 먹거리를 개선하고, 사진을 찍을 수 있는 공간 및 이벤트를 다양하게 기획함으로써 오프라인 경기장으로 방문하게끔 하는 전략을 제시할 수 있다.

본 연구의 결과를 바탕으로 다음과 같은 시사점을 제안할 수 있다.

먼저, 본 연구에서는 설문 조사 기법을 바탕으로 시행된 기존 프로야구 관람객의 만족 조사를 텍스트 마이닝 기법을 활용하여 수행했다는 점에서 학문적 기여가 있다. 또한, 기존의 선행연구에서 소셜네트워크 데이터를 활용하여 프로야구 팬들의 관람 문화를 조사한 연구에서는 오프라인 내에서 고객 관람 문화에만 주목하여 추후 관람객의 태도 변화에만 중점을 두었으며, 실시간 스트리밍 스포츠 중계 데이터를 활용한 연구에서는 채팅 내용과 영상 간의 관계만 고찰하였음에 주목하여 오프라인 관람객과 온라인 시청자의 경험 차이에 중점을 두고 수행된 연구는 미흡하다고 할 수 있다. 따라서 본 연구는 코로나 19로 온라인에서 경기를 시청하던 시청자를 오프라인 경기장으로 이끌기 위한 O2O 전략을 빅데이터 분석 및 텍스트 마이닝 기법을 통해 수행함으로써 새로운 관점으로 프로야구의 오프라인 활성화 방안 연구에 기여할 수 있을 것으로 사료된다. 본 연구는 프로야구뿐만 아니라 이와 같은 상황으로 피해를 입은 다른 프로스포츠 산업에서도 활용할 수 있으며 더 나아가 타 스포츠산업으로도 확장될 수 있다.

다음으로, 본 연구를 통해 빅데이터 분석으로 온라인에서 오프라인으로의 고객 유치 마케팅 전략 수립을 하고자 하는 연구자들에게 좋은 자료가 될 수 있다. 본 연구에서는 프로야구 구단이 온라인 시청자를 오프라인 관람으로 이동시키기 위해 이들의 경험 차이를 이해하며 이를 바탕으로 마케팅이나 경영학적 전략을 수립할 수 있다는 점에서 실증적인 기여가 있다.

본 연구의 한계점으로는 첫째로, 데이터 수집 시 특정 포털 사이트와 커뮤니티만을 대상으로 조사하였다는 점으로, 추후 다양하고 광범위한 집단에서의 데이터 수집이

필요하다. 둘째로, 수집된 데이터는 코로나 19로 인한 입장 제한 규제가 해제된 직후 1년간의 데이터로 해당 시기는 코로나 19와 포스트 코로나가 공존했으며 이는 데이터 편향의 가능성이 있음을 의미한다. 마지막으로, 현재의 O2O는 오프라인 활성화를 위한 온라인 강화 뿐만 아니라 오프라인과 온라인 채널을 모두 상호보완하는 관계에 있으나, 본 연구에서는 프로야구 오프라인 경기 관람객과 온라인 경기 시청자의 경험 차이 속성을 파악하여 온라인 시청자를 오프라인 관람객으로 전환하고자 오프라인 활성화를 중심으로 전략을 제시하였다. 따라서 추후 연구에서는 완전한 코로나 종식 이후의 데이터를 수집하여 본 연구의 데이터와 비교분석을 통해 새로운 마케팅이나 경영학적 전략을 수립할 수 있을 것으로 기대하며, 연구방법과 분석결과를 보완하여 오프라인과 온라인 채널을 상호 보완하는 전략을 제시하고 옴니채널로써의 프로야구 발전을 기대할 수 있을 것으로 사료된다.

## 참고문헌

- 강승태, 반진욱 (2021), 프로 스포츠 구단이 돈이 된다고?, 매일경제, <https://www.mk.co.kr/economy/view.php?sc=50000001&year=2021&no=154942>
- 고경진, 박상현, 마윤성(2018), 프로야구 경기장 서비스스케이프가 재관람 의도에 미치는 영향: 소비자 감정의 조절 효과를 중심으로, *한국체육과학회지*, 27(6), 633-647.
- 고영수, 이수빈, 차민정, 김성덕, 이주희, 한지영, 송민 (2022), BERTopic을 활용한 불면증 소셜 데이터 토픽 모델링 및 불면증 경향 문헌 딥러닝 자동분류 모델 구축, *정보관리학회지*, 39(2), 111-129.
- 권웅, 이현우 (2015), 스포츠팬의 관람행동 반응, 몰입경험 및 재방문의도의 관계: 체화된 인지 이론 적용. *한국스포츠산업경영학회지*, 20(4), 67-80.
- 김기탁 (2020), 소셜네트워크 빅데이터를 활용한 코로나 19 에 따른 프로야구 관람 문화조사, *한국엔터테인먼트산업학회논문지*, 14(6), 139-150.
- 김대원, 조용운 (2016), 실시간 스트리밍 스포츠 중계 채팅창의 논의 내용과 양상에 대한 내용분석: 2015 년 한국 프로야구 포스트시즌의 경기를 중심으로, *사이버 커뮤니케이션학보*, 33(1), 5-46.
- 김성근, 조혁준, 강주영 (2016), 학술연구에서의 텍스트마이닝 활용 현황과 주요분석 기법. *정보기술아키텍처 연구*, 13(2), 317-329.
- 김용만 (2012), 프로야구 관중의 응원만족과 관람태도, 관람만족, 재관람의도 및 구 전의도와와의 관계, *한국스포츠산업경영학회지*, 17(2), 35-48.
- 김주영, 이지현, 김수잔 (2017), 프로야구 관람객의 관람제약요인과 관람만족, 관람몰입 및 관람태도의 관계, *스포츠사이언스 / Sport Science*, 34(2), 83-97.
- 김태완 (2022), [한국프로야구 출범 40주년 40장면] <21> 야구 관중의 역사, 월간조선 뉴스룸, [https://m.monthly.chosun.com/client/mdaily/daily\\_view.asp?idx=16602&Newsnumb=20221116602](https://m.monthly.chosun.com/client/mdaily/daily_view.asp?idx=16602&Newsnumb=20221116602)
- 문화체육관광부 (2020) , 2019 스포츠 산업 백서
- 서효민, 권영범 (2020), 프로스포츠 팬의 팀 동일시와 코로나 19 에 대한 스트레스 대처행동, 향후 경기관람의도의 영향 관계, *한국스포츠산업경영학회지*, 25, (4), 47-62.
- 유소연, 임규건 (2021), 텍스트 마이닝과 의미 네트워크 분석을 활용한 뉴스 의제 분

- 석: 코로나 19 관련 감정을 중심으로, *지능정보연구*, 27(1), 47-64.
- 유원준, 안상준 (2021), *딥 러닝을 이용한 자연어 처리 입문*, 위키독스.
- 이동연, 이수영 (2014), 실시간 스포츠 스트리밍 비디오 시청자의 미디어경험에 관한 연구: 의사사회적 상호작용과 사회적 실재감을 중심으로, *한국언론학보*, 58(1), 148-177.
- 이민우, 유지은 (2020), 코로나-19의 사회경제적 영향: 소비지출과 노동시장을 중심으로. *Asia-Pacific Journal of Business & Commerce*, 12(3), 121-141.
- 이재호, 전태준 (2017), 프로야구경기장의 서비스스케이프가 관중의 감정반응, 지각된 가치, 관람만족 및 재관람의도에 미치는 영향, *한국웰니스학회지*, 12(2), 165-178.
- 지방공기업평가원 (2020), 뉴 노멀 2.0 시대를 맞는 지방공기업
- 채승훈, 임재익, 강주영 (2015), 사용자 리뷰를 통한 소셜커머스와 오픈마켓의 이용 경험 비교분석, *지능정보연구*, 21(4), 53-77.
- 채한승, 이종호. (2000), 프로스포츠팬 성향 및 경기관람 결정요인에 관한 조사 연구, *한국스포츠산업경영학회지*, 5(2), 137-154.
- 천병혁 (2022), [천병혁의 야구세상] 관중·시청률 다 떨어졌는데...정말 WBC가 '특효약'일까, 연합뉴스, <https://www.yna.co.kr/view/AKR20221004074300007>
- 한국농촌경제연구원 (2020), 코로나 19 확산에 따른 농식품 소비분야 영향분석
- 한국프로스포츠협회 (2023), 2022 프로스포츠 관람객 성향조사 보고서\_프로야구
- 허단비, 이영린, 김태희 (2022), OTT 프로야구 생중계 서비스가 사용자의 커뮤니티스, 플로우, 인지적 경험 및 지속사용의도에 미치는 영향, *한국스포츠산업경영학회지*, 27(3), 124-145.
- Alhaj, F., Al-Haj, A., Sharieh, A., Jabri, R. (2022), Improving Arabic cognitive distortion classification in Twitter using BERTopic, *International Journal of Advanced Computer Science and Applications*, 13(1), 854-860.
- Chang, J.-R., Chen, M.-Y., Chen, L.-S., Chien, W.-T. (2020), Recognizing important factors of influencing trust in O2O models: an example of OpenTable, *Soft Computing*, 24(11), 7907-7923.
- Coughlan, M., Cronin, P., Ryan, F. (2009), Survey research: Process and limitations. *International Journal of Therapy and Rehabilitation*, 16(1), 9-15.
- Da Silva, E. C., Las Casas, A. L. (2017), Sport fans as consumers: An approach to sport marketing, *British Journal of Marketing Studies*, 5(4), 36-48.
- Du, Y., Tang, Y. (2014), Study on the Development of O2O E-commerce

- Platform of China from the Perspective of Offline Service Quality, *International Journal of Business and Social Science*, 5(4).
- Gao, Y., Kang, M., Wang, T., Wang, M. (2014), The role of switching costs in O2O platforms: Antecedents and consequences, *2014 International Conference on Management of e-Commerce and e-Government*
- Grootendorst, M. (2020). BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics, *Zenodo*, Version v0, 9\*.
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure, *arXiv preprint*
- Guo, Y., Ma, T., Porter, A. L., Huang, L. (2012), Text mining of information resources to inform forecasting innovation pathways, *Technology Analysis & Strategic Management*, 24(8), 843–861.
- He, W., Zha, S., Li, L. (2013), Social media competitive analysis and text mining: A case study in the pizza industry, *International journal of information management*, 33(3), 464–472.
- Hu, X., Liu, H. (2012), Text analytics in social media, *Mining text data*, 385–414.
- Leszczyński, M., Metelski, A., Rabczun, A. (2021), Digitalized Sports Products and Various Generations in the Era of Industry 4.0, *Sustainability*, 14(1), 95.
- Leung, P. P. L., Wu, C. H., Ip, W. H., Ho, G. T. S., Cho, V. W. S., Kwong, K. K. Y. (2016), Customer Loyalty Enhancement of Online-to-Offline Marketing in Beauty Industry, *2016 4th International Conference on Enterprise Systems (ES)*
- Liang, M., Yang, X., Ou, H. (2014), The Measurement of the Consumer Trust to O2O E-Commerce Based on Fuzzy Evaluation, *2014 Seventh International Joint Conference on Computational Sciences and Optimization*
- Norris, J. I., Wann, D. L., Zapalac, R. K. (2015), Sport fan maximizing: following the best team or being the best fan?, *Journal of Consumer Marketing*, 32(3), 157–166.
- Obaid, A. I., CK, K. K. (2022), The Impact of Social Media (SM) on Sport Marketing Strategy Development: A Case Study of Football Teams in Iraq, *Journal La Sociale*, 3(4), 155–159.

- Shen, C.-w., Min, C., Wang, C.-c. (2019), Analyzing the trend of O2O commerce by bilingual text mining on social media, *Computers in Human Behavior*, 101, 474-483.
- Wan, X., Chen, J. (2019), The relationship between platform choice and supplier's efficiency—evidence from China's online to offline (O2O)e-commerce platforms, *Electronic Markets*, 29(2), 153-166.
- Yao, P., Osman, S., Sabri, M. F., Zainudin, N. (2022), Consumer Behavior in Online-to-Offline (O2O) Commerce: A Thematic Review, *Sustainability*, 14(13), 7842.
- Yun, H., An, J., Park, S. C. (2021), User Perception about O2O Order·Delivery App Using Topic Modeling and Revised IPA, *Knowledge Management Research*, 22(3), 253-271.

**\* 저자소개 \***

· **천 세 연(cjstpdu9@ajou.ac.kr)**

아주대학교 경영대학 비즈니스애널리틱스학과에서 재학 중에 있다. 주요관심 분야는 빅데이터 분석, 머신러닝 등이다.

· **이 지 은(scall73@hanyang.ac.kr )**

한양대학교 정보기술경영학과에서 박사 학위를 취득하고 현재 한양사이버대학교 경영정보·AI비즈니스학과에 교수로 재직 중이다. 관심 분야는 인공지능, 에듀테크, 기술경영이다.

· **강 주 영(jygang@ajou.ac.kr)**

아주대학교 e-비즈니스학과에서 교수로 재직 중에 있다. 주요 관심분야는 텍스트마이닝, 빅데이터, 지능정보시스템이다.