
스포츠 팬들의 디지털 여론을 통한 경기력 예측 : KBO 리그 내 팀과 선수를 중심으로

2024. 06. 24

경희대학교 일반대학원 빅데이터응용학과 석사과정 강민석
경희대학교 일반대학원 빅데이터응용학과 석사과정 심건우
경희대학교 경영대학 경영학과 & 빅데이터응용학과 양성병 교수

I. 서론

- 연구배경
- 연구필요성 및 목적

II. 이론적 배경

- 스포츠의 가치와 프로스포츠
- 스포츠 팬덤과 디지털 참여
- 스포츠 심리학
- 세이버매트릭스
- 감성 분석
- 스포츠 경기력 예측

III. 연구방법

- 연구 프로세스
- 데이터 수집
- 데이터 전처리
- 탐색적 데이터 분석
- 모델 설계

IV. 연구결과

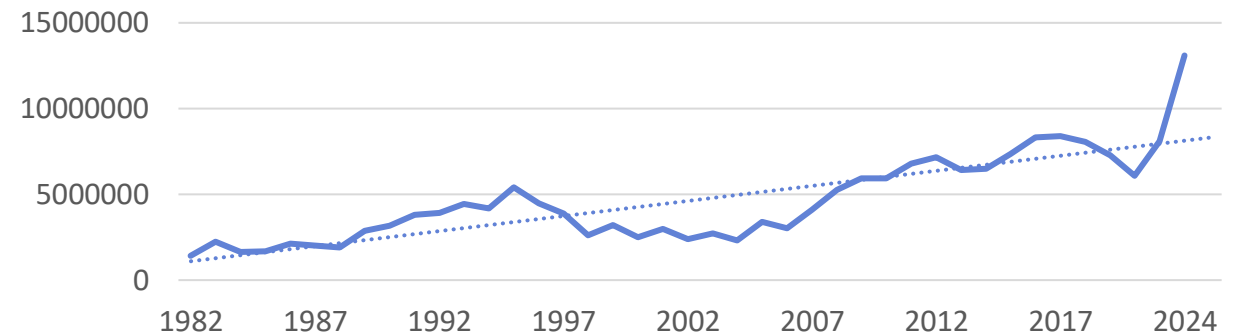
- 감성분석 결과
- 성능 비교분석

V. 결론

- 학술적 시사점
- 실무적 시사점
- 한계점

I. 서론 - 연구배경(1)

- ✓ 스포츠는 현대인들에게 삶의 활력을 제공하며, 신체적, 정신적, 사회적 가치를 지니고 있음(경동대, 2021)
 - ✓ 스포츠 활동은 기초체력 향상, 신체기관 기능 향상 등의 다양한 신체적 혜택을 제공하며, 스트레스 해소와 판단력 증진 등의 정신적 가치와 공동체 의식 강화 등의 사회적 가치를 제공함(경동대, 2021)
- ✓ 특히, 프로스포츠(Professional Sports)는 경제, 사회, 문화적 측면에서 더욱 큰 가치를 가짐(이준성, 2022)
 - ✓ 경기 입장권 판매, 방송 중계권 등으로 상당한 경제적 이익을 창출하고, CSR 활동을 통해 지역 사회에 기여하며, 같은 팀의 팬들을 중심으로 강력한 공동체 의식을 형성함(Bradbury, 2022; Oeckl & Morrow, 2022; Yuhei, 2019)
- ✓ 프로스포츠에서 가장 중요한 부분인 스포츠 팬덤은 특정 팀이나 선수에 대한 열정적이고 지속적인 지지를 보내는 집단을 의미하며(Ennis, 2020), 이들은 디지털 플랫폼을 통해 실시간으로 다양한 의견을 활발히 교류하고 있음(Sanderson & Jimmy, 2022)
- ✓ 최근에는 소셜 미디어와 같은 디지털 플랫폼의 발전으로 팬덤의 규모와 강도가 전 세계적으로 증가하고 있으며(Jessica, 2023), 이는 KBO 리그에서도 뚜렷하게 나타나고 있음(조형래, 2024)
 - ✓ 2024년 KBO 리그는 역대 최초 1,000만 관중을 달성할 추세를 보이고 있음
- ✓ KBO 리그의 팬들은 소셜 미디어와 다양한 온라인 커뮤니티를 통해 의견을 교류하고 있으며, 팬덤을 강화하고 있음(배동익 & 김대환, 2024)



<그림 1> KBO 리그의 연도별 총 관중 수

I. 서론 - 연구배경(2)

- ✓ 스포츠 심리학은 운동 선수의 정신적 건강과 경기력 향상을 위한 심리적 과정을 연구하는 학문으로, 선수들의 심리 상태가 경기력에 미치는 영향을 강조함(Lochbaum et al., 2022)
- ✓ 일반적으로 설문조사, 생리적 측정으로 선수의 심리를 측정하지만, 현실적으로 외부인이 시행하기 어렵고, 많은 시간과 자원이 소모됨(Yetton et al., 2019)
- ✓ 이러한 상황에서, 디지털 플랫폼에서의 팬들의 반응은 선수들의 심리 상태에 중요한 영향을 미침
 - ✓ 팬들의 긍정적인 반응은 자신감을 높이며, 부정적인 반응은 스트레스와 불안을 유발함(Weiner, 2021)
- ✓ 따라서, 팬들의 디지털 여론을 통해 간접적으로 선수들의 심리 상태를 추정할 수 있음(Kiler & Brixius, 2022)
- ✓ 특히, 야구는 선수들의 심리 상태가 좋지 않을 때 부정적인 영향이 많이 나오는 종목이며, 스포츠 심리학의 중요성이 더 부각됨(Sloan, 2023)
- ✓ 야구는 세이버매트릭스(Sabermetrics)와 같은 고급 통계 기법을 사용하여 선수의 성과를 평가하고 예측하며, OPS, WAR, GSC 등의 지표를 활용해 선수의 공헌도를 보다 정확하게 평가할 수 있고, 이는 팀의 전략적 의사결정에 중요한 역할을 수행함(Burroughs, 2020)
- ✓ 이러한 상황에서, 최근 OpenAI에 의해 개발된 GPT 모델은 팬들의 여론의 감성을 보다 정확하게 분석할 수 있는 주요 도구로 작용함(Kim et al., 2023)
- ✓ 한편, 스포츠 분야에서 경기력을 예측한 연구는 팀의 전략적 의사결정 등의 중요성으로 종목을 가리지 않고 활발히 수행되고 있으며, 경기력 뿐만 아니라 선수의 부상, 관중 수, 선수 몸값 등을 예측하는 연구도 많이 수행되었음(Huang & Li, 2021)

I. 서론 - 연구필요성 및 목적

연구필요성

- ✓ 기존 스포츠 팬덤에 관한 연구는 주로 팬덤의 문화와 소셜 미디어 등의 영향력 등에 초점을 맞추었으며, 팬들의 여론을 스포츠 심리학과 연계하여 여론이 선수 개인 및 팀 전체의 경기력에 미치는 영향을 분석한 연구는 미비하였음
- ✓ 기존의 스포츠 경기력 예측 연구는 주로 선수의 신체적 데이터나 경기 기록 등에 의존하고 있으며, 선수의 심리가 경기력에 큰 영향을 주는 기존 연구결과에도 불구하고, 팬들의 디지털 여론이 경기력에 미치는 영향을 체계적으로 분석한 연구는 미비하였음

연구목적

- ✓ 본 연구에서는 이러한 한계점을 극복하고자, 소셜 미디어와 온라인 커뮤니티에서 팬들의 여론을 수집하여 감성 분석을 수행하고, 이를 세이버매트릭스 지표와 결합하여 선수와 팀의 경기력을 예측하는 모델을 개발하고자 함
- ✓ 본 연구에서 설정한 연구 질문(RQ)은 다음과 같음
 1. 팬들의 여론이 선수의 경기력에 영향을 미치는지?
 2. 디지털 플랫폼의 종류에 따라 팬들의 여론이 상이한지?
 3. 팀, 선수마다 팬들의 여론이 경기력에 미치는 정도가 다른지?
 4. 선수의 신체적 데이터와 경기 기록을 통해 팀, 선수의 경기력을 예측한 모델과 팬들의 여론을 추가하여 경기력을 예측한 모델의 예측력이 차이를 보이는지?
- ✓ 이를 통해 팬 반응이 경기력에 미치는 구체적인 영향을 파악하고, 팀 관리, 멘탈 관리, 마케팅 전략 등에 실질적인 도움을 제공하고자 함

II. 이론적 배경 - 1) 스포츠의 가치와 프로스포츠

- ✓ 스포츠는 많은 현대인들에게 삶의 활력을 가져다 주며, 여러 **신체적·정신적 가치를 제공함**(경동대, 2021)
 - 스포츠 활동은 기초체력 향상, 신체기관 기능 향상, 신체 교정, 비만 예방, 조화로운 신체 발달 등 다양한 신체적 가치를 제공함
 - 정신적으로는 각종 스트레스 해소, 인간의 본능적 요구 충족, 판단력 증진, 지루함 해소 등 여러 정신적 가치를 제공함
 - ✓ 또한, 스포츠는 정서적 안정, 사회성 증진 등 여러 **심리적·사회적 가치를 제공함**(경동대, 2021)
 - 스포츠 활동은 의지력 및 정신력 배양 등의 심리적 가치와 조직 내에서 인간 행동에 대한 규칙, 역할의 학습을 통한 공동체 의식 강화 등의 사회적 가치를 제공함
 - ✓ 이러한 스포츠의 다양한 가치는 **프로스포츠(Professional Sports)**에서 특히 두드러지며, **경제·사회·문화적으로 다양한 가치를 제공함**(이준성, 2022)
 - ✓ 프로스포츠는 **상당한 경제적인 가치를 창출하며, 중요한 경제적 역할을 수행함**(Bradbury, 2022)
 - 경기 입장권, 방송 중계권, 스폰서십 계약, 관련 상품 판매 등을 통해 상당한 경제적 가치를 창출하며, 이를 통해 지역 경제 활성화, 일자리 창출에 기여함
 - ✓ 프로스포츠 팀들은 **여러 CSR(기업의 사회적 책임, Corporate Social Responsibility) 활동을 통해 중요한 사회적 역할을 수행함**(Oeckl & Morrow, 2022)
 - 스포츠 팀 내 선수들은 청소년 프로그램 운영, 지역 행사 참여 등 지역 사회에 참여하며, 자선 단체 기부, 자원봉사 활동을 통한 CSR 활동을 수행함
 - ✓ 프로스포츠는 **팬들의 팀에 대한 강력한 공동체 의식과 소속감을 통해 중요한 문화적 역할을 수행함**(Yuhei, 2019)
 - 프로스포츠 경기를 통해 팬들은 팀, 선수에게 강한 애정을 갖게 되며, 이를 통해 만들어진 커뮤니티는 지속적인 사회적 상호작용의 장을 형성함
- **본 연구에서는 일반적인 스포츠의 중요한 가치와 특히 프로스포츠에서 더욱 두드러지는 가치를 부각시킬 수 있는 연구를 수행하고자 함**

II. 이론적 배경 - 2) 스포츠 팬덤과 디지털 참여

- ✓ 스포츠 팬덤(Sports Fandom)은 특정 팀이나 선수에 대한 관심, 열정, 충성심 등의 행동을 보이는 집단을 의미함(Ennis, 2020)
- ✓ 스포츠 팬덤은 단순히 경기를 관람하는 것을 넘어, 다양한 디지털 플랫폼을 통해 의견을 활발히 교류하고 있음(Sanderson & Jimmy, 2022)
 - 과거의 스포츠 팬덤은 주로 팬클럽 등의 오프라인 모임을 통해 형성되었으며, 이는 팬덤의 형성과 유지에 중요한 역할을 수행하였음
 - 소셜 미디어, 커뮤니티, 포럼 등을 통해 팬들은 경기 중 실시간으로 반응을 보이며, 경기 후에도 팀과 선수에 대한 다양한 평가와 토론을 지속적으로 이어감
 - 딜로이트의 연구에 따르면, 스포츠 팬의 75%가 소셜 미디어 플랫폼을 통해 팀과 관계를 맺고 있음(Giorgio et al., 2023)
- ✓ 다양한 디지털 플랫폼이 등장하고 발전함에 따라, 전 세계적으로 스포츠 팬덤의 규모와 강도가 증가하고 있음(Jessica, 2023)
 - 41%의 스포츠 팬들은 디지털 플랫폼의 온라인 라이브 스트리밍을 통해 스포츠를 시청하며, 18-24세의 연령대는 55세 이상의 팬들보다 2배 이상 활용함
 - 미국 내 스포츠 팬들의 90%가 소셜 미디어를 통해 스포츠 콘텐츠를 소비하며, 이들 중 60%는 3년 전보다 충성심이 높아졌다고 응답함(Giorgio et al., 2023)
 - 2021년 스포츠 스폰서십 계약은 107% 증가하였음(Nielson, 2022)
- ✓ 이러한 스포츠 팬덤의 규모 증가는 2024년 한국의 KBO 리그에서도 뚜렷하게 나타나고 있음(조형래, 2024)
 - 2024년 KBO리그는 345경기 만에 500만 관중을 돌파했으며, 이는 10개 구단 체제 이후 가장 빠른 추세를 기록함
 - 전 구단의 전체 평균 관중 수는 14,558명을 기록하여 전년 대비 31% 증가하였으며, 역대 최초 1,000만 관중을 달성할 추세를 보이고 있음
- ✓ KBO의 팬들도 다양한 디지털 플랫폼을 통해 적극적으로 팀과 선수에 대해 실시간으로 활발한 의견을 표현하며, 팬덤을 강화하고 있음(배동익 & 김대환, 2024)
 - 소셜 미디어에서는 실시간으로 해시태그를 사용해 다양한 콘텐츠를 공유하며, 팀 계정의 게시물에서 댓글을 통해 여러 의견을 표현함
 - 네이버 카페, DC Inside, MLB Park, FM Korea 등의 온라인 커뮤니티를 통해 팀, 선수에 대한 토론을 나누며, 다양한 의견 및 정보를 교환하고 있음
- 본 연구에서는 KBO 리그 팬들의 디지털 여론을 중심으로 스포츠 팬덤의 특징을 분석하는 연구를 수행하고자 함

II. 이론적 배경 - 3) 스포츠 심리학

- ✓ 스포츠 심리학(Sport Psychology)은 운동 선수의 정신적 건강과 경기력 향상을 위한 심리적 과정을 연구하는 학문으로, 선수들의 동기부여, 스트레스 관리, 집중력 등을 다루며, 선수들이 최상의 경기력을 발휘할 수 있도록 기여하는 중요한 역할을 수행함 (Lochbaum et al., 2022)
- ✓ 스포츠 심리학에서 선수들의 심리를 측정할 수 있는 일반적 방법으로, 설문지와 생리적 측정 등이 있음(Trpkovici et al., 2023)
 - 스포츠 심리 상태 검사(ACSI-28), 경기 전 불안 질문지(CSAI-2), 스포츠 수행 심리 설문지(SAS-2) 등을 통해 심리 상태를 체계적으로 평가할 수 있음
 - 선수들의 심박수, 호르몬 수치, 뇌파 등을 통해 선수의 스트레스 수준과 심리적 반응을 평가해 객관적인 선수의 심리적 상태를 파악할 수 있음
- ✓ 그러나, 두 방법 모두 현실적으로 외부인이 측정하기 어렵고, 많은 시간 및 자원이 소모된다는 한계점이 존재함(Yetton et al., 2019)
- ✓ 이러한 상황에서, 디지털 플랫폼에서의 팬들의 여론은 선수들의 심리에 중요한 영향을 미침(Weiner, 2021)
 - 소셜 미디어 및 온라인 커뮤니티에서 격려, 응원 등 팬들이 보내는 긍정적인 메시지는 선수들의 자신감을 높이고, 경기력을 향상시키는 데 기여함
 - 반면, 부정적인 댓글이나 무분별한 비난은 선수들에게 심리적인 부담을 주고, 불안감과 스트레스를 유발할 수 있음
 - 이는 곧 선수들의 경기 집중력을 떨어뜨리고, 결과적으로 경기력 저하로 이어질 수 있음
- 따라서, 선수들의 심리를 직접적으로 측정하기 어려운 경우, 팬들의 디지털 여론을 통해 간접적으로 선수의 심리를 추정하는 것이 가능함 (Kiler & Brixius, 2022)
- ✓ 야구에서는 선수의 심리가 좋지 않을 때 여러 부정적인 현상이 나타날 수 있으며, 스포츠 심리학이 중요한 종목으로 볼 수 있음(Sloan, 2023)
 - 타자는 스트라이크존 판단이 흐려져 타격 성공률에 영향을 미치며, 투수는 제구력이 떨어져 볼넷을 많이 허용하게 되고, 이는 경기에 부정적 영향을 일으킴
- 본 연구에서는 KBO 팬들의 여론을 통해 선수의 심리를 간접적으로 추정하며, 이를 기반으로 선수의 경기력을 예측하고자 함

II. 이론적 배경 - 4) 세이버메트릭스

- ✓ 세이버메트릭스(Sabermetrics)는 야구 경기의 다양한 측면을 체계적으로 분석하고 평가하기 위한 통계적 접근 방법으로, "Society for American Baseball Research"의 약자인 SABR에서 유래됨(Burroughs, 2020)
- ✓ 세이버메트릭스는 기본적으로 전통적인 야구 통계보다 더 깊이 있는 분석을 사용하며, 중요한 역할을 수행함(Burroughs, 2020)
 - 전통적인 야구 통계는 타율, 홈런, 타점 등 기본적인 지표에 의존해 선수의 실제 가치를 완전히 반영하지 못함
 - 세이버메트릭스는 OPS, WAR 등의 지표를 통해 선수의 공헌도를 보다 정확하게 평가할 수 있음
 - 따라서, 팀은 데이터를 기반으로 보다 전략적인 의사결정을 내릴 수 있으며, 선수의 잠재력을 평가하고 성과를 모니터링하는 데 중요한 도구로 작용함
- ✓ 이러한 세이버메트릭스의 중요성으로 인해 메이저리그에서 전 세계적으로 확산되었으며, KBO에서도 최근 각광을 받기 시작하면서 각 팀들로부터 적극적으로 활용되어 리그의 경쟁력을 높이고 선수의 발전을 돕는 중요한 요소로 작용하고 있음(김태훈 외, 2020)
- ✓ 세이버메트릭스는 다양한 세부 지표를 포함하며, 본 연구에서 활용할 주요 지표와 그 특징은 다음과 같음(Huang & Li, 2021)
 - OPS (On-base Plus Slugging): 타자의 출루율과 장타율을 합산한 지표로, 타자의 공격력을 가장 종합적으로 나타냄

$$OPS = OBP + SLG, \quad OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF}, \quad SLG = \frac{1B + 2 \times 2B + 3 \times 3B + 4 \times HR}{AB}$$

- GSC (Game Score for Starting Pitchers): 선발투수의 경기 성과를 평가하는 지표로, 선발투수가 경기에서 얼마나 잘 던졌는지 나타냄
- $$GSC = 50 + Outs + 2 \times Innings Completed After the 4th + Strikeouts - 2 \times Walks - 2 \times Hits - 3 \times Earned Runs - 6 \times HomeRuns$$
- 본 연구에서는 팬들의 디지털 여론과 세이버메트릭스 지표(타자: OPS, 선발투수: GSC)를 결합하여 선수와 팀의 경기력을 보다 정확하게 예측하는 모델을 개발하고자 함

II. 이론적 배경 - 5) 감성 분석

- ✓ 감성 분석(Sentiment Analysis)은 텍스트 데이터에서 감정이나 의견을 자동으로 추출하는 자연어처리(NLP) 기술로, 주로 소셜 미디어, 설문조사, 리뷰 등의 텍스트 데이터를 분석하여 긍정, 부정, 중립 감정을 파악함(Tan et al., 2023)
- ✓ 감성 분석을 통해 **기업이나 조직의 의사결정을 지원**할 수 있고, 여론을 실시간으로 파악할 수 있으며, 고객 만족도를 평가하는 등의 중요한 역할을 수행함(Tan et al., 2023)
- ✓ 감성 분석은 **다양한 분야에서 응용되고 있으며**, 주요 응용 분야는 다음과 같음(Wankhade & Kulkarni, 2022)
 - **소셜 미디어 분석**: 트위터, 인스타그램 등의 소셜 미디어 데이터를 분석해 사용자들의 감정을 파악하고, 최신 트렌드를 예측함
 - **제품 및 서비스 리뷰 분석**: 제품 및 서비스에 대한 리뷰를 분석해 소비자의 만족도와 의견을 평가하며 서비스 개선에 활용됨
 - **정치 및 사회적 여론 분석**: 정치적 발언이나 사회적 이슈에 대한 대중의 반응을 분석하여 여론의 흐름을 파악함
- ✓ 감성 분석은 **다음 3가지의 기술적 접근 방법**을 활용해 텍스트 데이터를 분석함(Wankhade & Kulkarni, 2022)
 - **사전 기반 접근 (Lexicon-based Approach)**: 감성 사전을 이용해 텍스트에서 감정 단어를 매칭하고, 각 단어의 감정 점수를 합산함
 - **기계학습 기반 접근 (Machine Learning-based Approach)**: 텍스트 데이터를 학습 데이터로 사용해 감성 분석 모델을 훈련하여 새 텍스트의 감성을 예측함
 - **딥러닝 기반 접근 (Deep Learning-based Approach)**: LSTM, BERT, GPT 등의 모델을 통해 대량의 데이터를 학습해 높은 정확도의 감성 예측을 수행함
- ✓ **GPT는 OpenAI에서 개발한 최신 자연어처리 모델로, 대규모 데이터셋을 학습하여 텍스트의 맥락을 이해하고 감성을 보다 정확하게 분류할 수 있음**(Kim et al., 2023)
- **본 연구에서는 gpt-3.5-turbo 모델을 활용하여 소셜 미디어와 온라인 커뮤니티에서의 KBO 팬들의 여론의 감성을 보다 정확하게 분석하고자 함**

II. 이론적 배경 - 6) 스포츠 경기력 예측

- ✓ 스포츠 경기력 예측(Sports Performance Prediction)은 경기 결과를 예측하거나, 선수와 팀의 성과를 예측하는 과정으로, 통계 분석, 머신러닝 등의 기법을 통해 수행되며, 다양한 변수들을 고려하여 정확한 예측을 도출함(Huang & Li, 2021)
- ✓ 스포츠 경기력 예측은 다양한 종목에서 여러 이유로 스포츠 분야에서 연구가 활발하게 수행되고 있음(Huang & Li, 2021)
 - **전략적 의사결정 지원:** 팀과 코치는 경기력 예측을 통해 전략적인 결정을 내릴 수 있음
 - **스포츠 베팅 산업 활성화:** 경기 결과 예측은 스포츠 베팅 산업에서 중요한 역할을 수행하며, 산업의 신뢰성을 강화함
 - **팬 참여 증가:** 팬들은 경기력 예측을 통해 더 흥미롭게 경기를 관람할 수 있음
- ✓ 스포츠 경기력 예측에는 여러 접근 방법이 존재하며, 크게 3가지로 분류될 수 있음(Huang & Li, 2021)
 - **통계적 접근 방법 (Statistical Methods):** 과거 데이터를 기반으로 회귀 모델을 구축하거나, 경기 결과의 불확실성을 베이지안 모델을 통해 모델링함
 - **기계학습 접근 방법 (Machine Learning Methods):** 트리 기반 앙상블 모델 또는 SVM과 같은 모델을 통해 경기력 예측의 정확성을 높임
 - **딥러닝 접근 방법 (Deep Learning Methods):** 장기 의존성을 처리할 수 있는 LSTM, 또는 BERT, GPT 등의 모델을 활용한 텍스트 기반 경기력 예측을 수행함
- 그러나, 기존에 수행된 스포츠 경기력 예측 연구는 대부분 경기 내의 데이터만을 활용하였으며, 선수의 심리 상태가 경기력에 많은 영향을 미친다는 연구 결과에도 불구하고, 이를 고려하여 경기력을 예측한 연구는 미흡하였음
- 본 연구에서는 선수의 심리 상태를 확인할 수 있는 팬들의 여론 반응을 추가한 머신러닝 및 딥러닝 기반의 경기력 예측 모델을 개발하며, 이를 통해 팬 반응이 경기력에 미치는 영향을 정량적으로 분석하고자 함

II. 이론적 배경 - 6) 스포츠 경기력 예측

<표 1> 스포츠 경기력 예측 선행연구

종목(리그)	데이터	입력 변수	예측 변수	방법론	연구결과	저자(연도)
야구 (MLB, KBO)	2019년 MLB의 30팀 데이터	모든 투수를 포함한 2019년 시즌의 주요 경기 변수	승 or 패	1DCNN, ANN, SVM	특징 선택이 모델 성능 향상에 기여, 약 94%의 정확도 기록	Huang & Li (2021)
	1982 ~ 2020년 KBO 데이터	타율, 홈런 등의 6가지 변수	OPS	XGBoost, LightGBM, Random Forest	XGBoost에서 가장 높은 성능 기록	한정섭 외. (2022)
	2021년 MLB의 전체 투구 자료	타석 순서, 투구존, 투구 속도 등 25개 변수	투구 결과, 투구 후 점수 변화	CNN	1차 모델: 53%의 정확도 최적화 모델 성능: 84%의 정확도 기록	김주학 외. (2022)
	1982 ~ 2022년 KBO 데이터	타율, 홈런, 타점 등 10개 경기력 지표	미래의 타율, 홈런, 타점	ARIMA 시계열 분석	홈런과 타점은 과거 평균보다 높은, 타율은 과거 평균보다 낮은 수치 예측	오승욱 외. (2023)
	1961 ~ 2019년의 MLB 시즌 데이터	선수 연령, 키, 몸무게, 시즌 성적 등 21개 변수	시즌 홈런 수	LSTM, GRU, BiLSTM, AT-LSTM	LSTM에서 가장 높은 성능 기록	Sun et al. (2023)
기타 (EPL, NBA, NFL, e-sports)	1980 ~ 2017년의 NBA 결승 데이터	수비 리바운드, 3점슛 성공률, 자유투 성공 횟수 등 22개 변수	승 or 패	ANN, Naïve Bayes, Logistic Model Tree	수비 리바운드가 경기 결과에 가장 큰 영향을 준 변수, 83%의 정확도 기록	Thabtah et al. (2019)
	DotA 2의 5,744개 일반 경기, 186개 프로 경기	선택한 영웅, 플레이어 위치 등	승 or 패	LR, RF, LightGBM	RF 모델에서 77.5%의 정확도 기록, 경기 시작 5분 후에 85% 정확도 기록	Hodge et al. (2019)
	2005 ~ 2016년의 EPL 경기 데이터	골 차이, 유효 슈팅, 코너킥 등의 33개 변수	승 or 무 or 패	Naïve Bayes, SVM, Random Forest, GBM	모델별 52% ~ 58%의 정확도 기록, GBM에서 가장 높은 성능	Raboota & Kaur (2019)
	2015 ~ 2019년의 NFL 경기 데이터	득점, 야드, 공격/수비 플레이 수 등 42개 변수	승 or 패	SVM, Naïve Bayes, RF, AdaBoost, QDA, ANN	Naïve Bayes 모델에서 가장 높은 67%의 정확도 기록	Beal et al. (2020)

➤ 스포츠 경기력 예측 연구는 대부분 경기 내의 데이터만을 활용하였으며, 선수의 심리 상태를 고려하여 이를 활용해 경기력을 예측한 연구는 미흡함

Ⅲ. 연구방법 - 1) 연구 프로세스

1. 데이터 수집

1. 팬들의 여론 데이터: 소셜 미디어(인스타그램 - 팀 공식 계정의 댓글) + 온라인 커뮤니티(네이버 카페, DC Inside, FM Korea, MLB Park)
 2. 야구 경기 데이터: 스탯티즈, KBO 사이트
- 분석 팀: KBO의 10개 팀
 - 분석 선수: 시즌별 규정 이닝을 채운 선발투수 + 시즌별 규정 타석을 채운 타자 (많은 팬 반응 확보를 위함)
 - 데이터 기간: 2018시즌 ~ 2023시즌 데이터 (이전 시즌의 여론 데이터 수집 어려움)
 - 예측 단위: 경기별 예측

2. 데이터 전처리

1. 텍스트 전처리
2. 감성 분석

3. 탐색적 데이터 분석

4. 모델 설계

1. 팀 단위 경기력 예측 (팀의 승, 패 예측)
 1. 팬들의 여론 데이터 미포함: 팀의 이전 경기 데이터 -> 팀의 승패 예측
 2. 팬들의 여론 데이터 포함: 팀의 이전 경기 데이터 + **이전 경기 종료 후 경기 시작 전까지 플랫폼별 팀의 평균 감정** -> 팀의 승패 예측
2. 선수 단위 경기력 예측: 타자(OPS 예측) + 투수(GSC 예측)
 1. 팬들의 여론 데이터 미포함: 선수의 이전 경기 데이터, 신체 데이터 -> 선수의 경기력 예측
 2. 팬들의 여론 데이터 포함: 선수의 이전 경기 데이터, 신체 데이터 + **이전 경기 종료 후 경시 시작 전까지 플랫폼별 선수의 평균 감정** -> 선수의 경기력 예측

Ⅲ. 연구방법 - 2) 데이터 수집

Ⅲ. 연구방법 - 3) 데이터 전처리

Ⅲ. 연구방법 - 4) 탐색적 데이터 분석

Ⅲ. 연구방법 - 5) 모델 설계

1) 팀 단위 예측

X	x_sentiment1	x_sentiment2	x_sentiment3	x_sentiment4	x_sentiment5	y
팀 통계 데이터	Instagram_avg	FM Korea_avg	DC Inside_avg	NaverCafe_avg	MLB Park_avg	승 or 패

2) 선수 단위 예측 - 투수

X	x_sentiment1	x_sentiment2	x_sentiment3	x_sentiment4	x_sentiment5	y
투수 데이터 (통계, 팀, 이름, 신체 등)	Instagram_avg	FM Korea_avg	DC Inside_avg	NaverCafe_avg	MLB Park_avg	GSC

3) 선수 단위 예측 - 타자

X	x_sentiment1	x_sentiment2	x_sentiment3	x_sentiment4	x_sentiment5	y
타자 데이터 (통계, 팀, 이름, 신체 등)	Instagram_avg	FM Korea_avg	DC Inside_avg	NaverCafe_avg	MLB Park_avg	OPS

IV. 연구결과 - 1) 감성 분석 결과

Digital Platform	Aver.sentiment	Number of data
Instagram		
FM Korea		
Naver Café		
MLB Park		
DC Inside		

IV. 연구결과 - 2) 성능 비교분석

1) 팀 단위 예측

1-1) 팬의 여론 감성 점수 미포함

Model	Accuracy	F1 Score	ROC AUC Score
Random Forest			
XGBoost			
LightGBM			
LSTM			
1D CNN			

1-2) 팬의 여론 감성 점수 포함

Model	Accuracy	F1 Score	ROC AUC Score
Random Forest			
XGBoost			
LightGBM			
LSTM			
1D CNN			

IV. 연구결과 - 2) 성능 비교분석

2) 선수 단위 예측 - 투수

2-1) 팬의 여론 감성 점수 미포함

Model	MAE	MAPE	RMSE
Random Forest			
XGBoost			
LightGBM			
LSTM			
1D CNN			

2-2) 팬의 여론 감성 점수 포함

Model	MAE	MAPE	RMSE
Random Forest			
XGBoost			
LightGBM			
LSTM			
1D CNN			

IV. 연구결과 - 2) 성능 비교분석

3) 선수 단위 예측 - 타자

3-1) 팬의 여론 감성 점수 미포함

Model	MAE	MAPE	RMSE
Random Forest			
XGBoost			
LightGBM			
LSTM			
1D CNN			

3-2) 팬의 여론 감성 점수 포함

Model	MAE	MAPE	RMSE
Random Forest			
XGBoost			
LightGBM			
LSTM			
1D CNN			

1) 학술적 기여점

- ✓ **팬들의 여론의 감성을 기반으로 선수와 팀의 경기력 예측:** 기존 연구들이 주로 선수의 신체적 데이터와 경기 기록에 의존했던 것에 비해, 본 연구는 팬 반응과 같은 심리적 요인을 고려하여 경기력을 예측하는 시도였다는 점에서 의미가 있음
- ✓ **세이버매트릭스와 감성 분석의 결합:** 본 연구는 세이버매트릭스 지표와 감성 분석을 결합하여, 선수와 팀의 경기력을 예측하는 새로운 방법론을 제시하였음

2) 실무적 기여점

- ✓ **팀 관리와 멘탈 코칭, 마케팅 전략 등에 기여:** 소셜 미디어, 온라인 커뮤니티에서 팬들의 여론을 분석하여 선수와 팀의 심리 상태를 기반으로 경기력을 예측함으로써, 팀 관리와 멘탈 코칭, 마케팅 전략 등에 실질적인 도움을 줄 수 있음
- ✓ **전략적 의사결정 지원:** 팀과 코치는 경기력 예측, 심리 상태 등을 통해 전략적 결정을 내릴 수 있으며, 이는 경기 성과를 향상시키는 데 도움을 줄 수 있음

3) 한계점

- ✓ **팬 반응으로 심리 상태의 정확한 추정 어려움:** 팬 반응만으로 선수의 심리 상태를 완벽하게 추정하기 어려운 측면이 존재하며, 선수 심리는 매우 다양한 요인에 의해 결정됨
- ✓ **이전 경기 결과와 팬 여론의 높은 상관성:** 팀과 선수의 이전 경기의 경기력이 좋지 않을 경우, 자연스럽게 팬 반응이 부정적일 확률이 높고, 이는 모델이 지속적으로 하락된 경기력을 예측할 우려가 존재함

감사합니다