

# 머신러닝을 이용한 오디오북 플랫폼 기반의 웹케어 모형 구축에 관한 연구

## A Study on Developing a Web Care Model for Audiobook Platforms Using Machine Learning

정 다 훈 (Dahoon Jeong) 부산대학교 경제학부 학사과정  
이 민 혁 (Minhyuk Lee) 부산대학교 경영학과 조교수  
이 태 원 (Taewon Lee) 부산대학교 경영학과 BK21 교육연구단 연구교수, 교신저자

### 요 약

본 연구는 소비자 리뷰와 관리자 답변 간의 관계를 조사하여 소비자 리뷰를 효율적으로 관리하기 위한 웹케어의 필요성을 탐색하는데 목적이 있다. 효과적인 웹케어를 위한 방법론을 제안하고 오디오북 플랫폼 기반의 머신러닝을 이용한 웹케어 모형을 구축하고자 한다. 본 연구에서는 오디오북 플랫폼 4개를 선정하여 소비자 리뷰와 관리자 답변에 대한 데이터 수집 및 전처리 과정을 거쳐 토픽모델링, 주제불일치성, DBSCAN을 활용하고, 다양한 머신러닝 기법을 적용하여 분석을 시행하였다. 실험 결과 관리자 답변의 군집화 및 소비자 리뷰에 대한 답변 예측에서 유의미한 결과를 도출하였으며, 자원의 제한과 비용을 고려한 효율적인 방법론을 제안하였다. 본 연구는 머신러닝을 통해 웹케어 모형을 구축했다는 점에서 학술적인 시사점을 제공하며, 기업의 제한된 비용과 인력을 고려하여 웹케어 모형이라는 효율적인 방법론을 제시함으로써 실무적인 시사점을 지닌다. 본 연구에서 제안된 웹케어 모델은 개별화된 답변과 표준화된 관리자 답변을 제공하여 소비자 참여 및 유용한 정보 제공을 위한 전략적인 기초 자료로 활용될 수 있을 것이다.

**키워드** : 웹케어, 소비자 리뷰, 관리자 답변, 오디오북 플랫폼, 머신러닝, DBSCAN, 주제불일치성

## I. 서 론

최근 온라인에서 다양한 디지털 콘텐츠가 활발해지면서 소비자의 리뷰가 급속도로 증가하게 되었다. 소비자들은 온라인을 통해 쇼핑, 제품 구매, 정보 획득 및 공유 등 다양한 영역에서 활동하고 있으며, 소비자들이 직접 작성한 리뷰를 통해 보다 쉬운 서비스 정보 탐색과 의사결정으로 기대치

를 충족시키고 있다. 온라인 리뷰는 대부분 개인적인 의견으로 작성되어 있으며, 소비자의 만족과 불만사항, 요청사항, 감정 상태, 사용후기 등 다양한 정보가 내포되어 있다. 또한, 소비자와 소비자, 소비자와 관리자 간의 관계에서 소통의 일환으로 자리매김하게 되었다. 즉, 소비자의 경험, 생각 등 개인적인 의견을 작성하여 다양한 정보를 제공함으로써 추후 이용되는 소비자의 의사결정에 있어

서 매우 중요한 역할을 할 뿐만 아니라 관리측면에서도 매우 유용한 자료로 활용되고 있다(Lee *et al.*, 2015).

하지만, 온라인 리뷰는 높아진 소통의 수요에 따라 관리자 측에서는 신속하고 정확한 응대를 해야만 한다. 표준화된 답변은 소비자의 불만을 야기시키기도 한다. 소비자 리뷰에 대한 관리자의 빠른 대처와 올바른 답변은 해당 플랫폼을 이용하는 모든 소비자들이 확인 가능하므로 소비자의 요구에 따른 관리자의 답변은 신속하고 정확한 내용으로 전달되어야 한다. 하지만, 관리자의 답변이 소비자 리뷰에 대해 긍정적인 영향 보다 오히려 부정적인 영향을 미칠 수 있다. 즉, 표준화된 답변은 소비자에게 부정적인 영향을 미칠 가능성이 높고 개인화된 답변은 긍정적인 영향을 미칠 수 있다는 것이다(Wang and Chaudhry, 2018; Zhang *et al.*, 2022). Wang and Chaudhry(2018)의 연구에서는 소비자 리뷰에 대해 무성의한 답변이나 단순한 답변 내용이 부정적인 영향을 미칠 수 있는 결과를 언급하였고, Li *et al.*(2018)은 소비자 리뷰에 대해 관리자는 개인화된 답변을 제공해야 한다고 하였다. 관리자 측면에서는 소비자 리뷰를 관리함에 있어서 많은 시간과 비용, 인력이 동원되어야 하는 비효율적인 기업 운영이 되고 있는 상황이다. 본 연구에서는 웹케어 활동에서 발생하는 시간적·비용적 문제를 해결하기 위해 웹케어를 위한 관리자 답변을 위한 추천 모형을 제안하고자 한다. 웹케어는 온라인 리뷰의 소비자 의견을 바탕으로 문제를 발견하고 해결 방안을 모색하기 위해 소통하는 활동이라고 할 수 있으며, Van Noort and Willemsen(2012)에 따르면 웹케어는 소비자 리뷰에 기반한 적극적인 온라인 상호작용을 통해 문제를 식별하고 해결책을 찾는 과정이다. 본 연구에서는 제안하는 모형을 구축하기 위해 오디오북 플랫폼에서 생성된 소비자 리뷰와 관리자 답변을 활용하여 웹케어 활동에서 일어나는 문제를 효과적으로 해결하기 위한 모형을 제안한다.

최근 디지털 콘텐츠(웹툰, 게임, 오디오북 등)의

사용자 증대로 소비 활동에 대한 매출액이 증가하게 되었다. 특히, 오디오북 플랫폼은 소비자들이 시간에 구애받지 않고 언제 어디서나 사용 가능한 플랫폼으로 자리매김하여 많은 소비자들이 이용하고 있는 추세이다(성행남, 이태원, 2022). 오디오북은 이제 시각이 아닌 청각을 이용하여 보다 다양한 환경에서 이용이 가능하게 되었으며 모든 세대층의 소비자들이 선호하게 되었다. 오디오북은 디지털 콘텐츠로서의 구매 및 보관이 용이하고, 물리적 제약에서 벗어나 디지털 콘텐츠 기반으로 제공되어 소비자에게 편의성을 제공하고 있다. 또한, 직장인들의 효율적인 출·퇴근 시간 이용 등 많은 장점을 지니고 있어서 스마트폰으로 듣는 사용자가 많아지면서 선호하게 되었다. 하지만 오디오북 플랫폼 시장 활성화에 비해 이와 관련된 연구는 아직까지 미비한 실정이다(성행남, 이태원, 2022; 우세웅, 2019; 이윤아, 2020). 따라서 본 연구의 목적은 다음과 같다. 첫째, 오디오북 플랫폼에서 생성된 소비자 리뷰와 이에 대한 관리자 답변 간의 관계를 탐색하여 관리자측면에서 효율적인 웹케어의 필요성에 대해 살펴본다. 둘째, 효율적인 웹케어를 위한 방법론을 제안하고 기계학습기법을 활용한 웹관리 모형을 구축한다. 본 연구에서 제안하는 모형은 관리자측면에서 소비자 리뷰를 관리하는데 있어서 대응방법에 대한 의사결정에 도움이 되고, 신속하고 정확한 답변 추천을 통해 대응 시간 및 전체 비용을 절감 등 시너지 효과를 기대할 수 있다.

## II. 이론적 배경

### 2.1 웹케어(Webcare)

웹케어는 디지털 환경에서 소비자의 의견에 대한 기업 및 관리자의 참여를 의미한다. 과거의 소비자들은 기업 및 관리자가 제공하는 단편적인 정보를 통해 원하는 제품에 대한 정보만을 얻을 수 있었다. 하지만, 디지털 환경의 변화 및 활성화에

따라 소비자들은 광범위한 온라인 환경에서 자신이 직접 작성한 리뷰가 관리자 및 다른 소비자들에게 의견이 전달되고 유용한 정보를 제공할 수 있는 소통의 역할을 수행하게 되었다. 이와 같은 소비자 참여 활성화는 기업에 대한 긍정적인 영향과 부정적인 영향을 동시에 끼칠 수 있게 되었으며, 기업은 특히 부정적인 리뷰에 대해 모니터링, 직접적인 답변 등으로 관리를 하게 되었다. 온라인 상에서 관리자 답변은 소비자 리뷰에 직접적인 댓글을 작성하는 형식으로 이루어지며 소비자의 부정적인 의견에 대해 직·간접적으로 대처하고 소비자의 불만을 감소시킬 수 있게 되었다. 또한 웹케어 과정에서는 해당 환경에 접속한 다수의 대상에게 노출되기 때문에 관리자의 웹케어는 잠재적 소비자도 고려된다. 즉, 웹케어는 소비자의 피드백을 통해 문제를 발견하고 해당 문제에 대한 소통을 하는 활동이라고 정의할 수 있다.

Hong and Lee(2005)는 웹케어를 통해 소비자의 불만사항을 효율적으로 대처하고 소비자와의 소통 결과 브랜드 충성도를 높일 수 있다는 결과를 보고하였다. Van Noort and Willemsen(2012)은 웹케어를 '온라인에서 관리자의 적극적인 소비자 상호작용'이라 정의하였으며, 이를 통해 관리자가 웹케어를 통해 소비자의 해당 브랜드에 대한 인식을 높일 수 있다고 주장하였다. Zhang and Vasquez(2014)는 소비자 의견에 대한 관리자의 반응이 기업에 대한 온라인상의 평가에 긍정적인 영향을 미친다는 실험결과를 제시하였다. 황단(2018)은 온라인상에서 구매 경험자와 제품 판매자의 소통이 이루어진다는 주장을 통해 지속된 쌍방향적 소통의 중요성을 강조하였으며, 관리자는 소비자의 불만과 같은 문제 상황에 대처하고, 직접적인 케어를 통해 해당 불만이 다른 소비자에 대한 극단적인 영향을 줄일 수 있다는 연구 결과를 보였다.

이처럼 웹케어는 소비자에 대한 관리자의 소통이 기업 평판 및 이미지에 중요한 요소라는 연구 결과가 다수 존재한다. 그러나 소비자 리뷰와 관

리자 답변의 내용 흐름이 얼마나 일치하는지, 어떻게 하면 관리자가 효율적으로 소비자 리뷰에 대처할 수 있는지에 대한 연구는 충분하지 않다. 따라서 본 연구에서는 소비자 리뷰-관리자 답변에 대한 흐름의 일치성에 대한 탐색을 진행하고 기계 학습 기법을 통해 모형을 구축한 후 디지털 환경에서의 웹케어 실현에 관한 연구를 진행하고자 한다.

## 2.2 관리자 답변의 특성

Chen *et al.*(2019)는 관리자 답변을 온라인 플랫폼에서 관리자와 소비자가 소통하는 도구로 정의하였다. 관리자 답변은 소비자 리뷰와 상호작용하고 대응하기 위한 기업의 노력이라 볼 수 있으며 기존 연구에서는 관리자 답변이 소비자의 신뢰도, 만족도, 향후 소비자 리뷰의 양과 성향(긍·부정)에 영향을 미치고, 궁극적으로 기업의 매출과 같은 재무 성과에 영향을 미쳤다고 주장하였다(Herhausen *et al.*, 2019; Kumar *et al.*, 2018; Li *et al.*, 2018; Proserpio and Zervas, 2017; Sparks *et al.*, 2016; Wang *et al.*, 2020). 또한 관리자 답변을 통해 관리자들은 소비자가 지적한 서비스 결함을 복구할 수 있는 기회를 가지며 잠재적인 소비자가 구매 결정을 내리는데 도움을 줄 수 있다(Li *et al.*, 2018; Zhu *et al.*, 2020).

관리자 답변은 소비자와 관리자에게 긍정적인 영향을 끼칠 수 있으나, 디지털 환경이 확대되어 소비자 리뷰의 수가 폭발적으로 증가함에 따라 모든 소비자 리뷰에 관리자가 직접 응답하기에 어려운 환경이 되었다. 소비자 리뷰에 대한 답변을 제공하기 위해서는 상당한 시간과 인력을 필요로 한다. 이에 대해 기존 연구에서는 관리자가 일정한 템플릿을 사용하여 응답하는 사실을 발견하고 관리자 답변을 두 유형으로 구분할 수 있다고 주장하였다(Jin *et al.*, 2023; Wei *et al.*, 2013; Zhang *et al.*, 2019). 첫 번째 유형은 개별화된 관리자 답변이다. Jin *et al.*(2023)은 소비자 리뷰에 맞춤형 대응을

PMR(Personalized Managerial Responses)라 정의하며 관리자는 PMR을 통해 소비자 리뷰에 담겨있는 문제에 대해 적극적인 대응이 가능하다고 주장하고 소비자들에게 알맞은 정보를 제공하여 소비자의 불확실성과 인지적 비용을 줄일 수 있다는 연구결과를 보였다. Zhang *et al.*(2020)은 잠재적 소비자들이 구체적인 개별화된 관리자 답변을 통해 간접적인 정보를 얻고 관리자 리뷰가 도움이 된다는 인식을 통해 신뢰할 수 있다고 주장하였다. 또한 Li *et al.*(2018)은 기업들은 소비자 리뷰 성격에 맞추어 맞춤형 관리자 답변을 필요성을 강조하며 이를 통해 소비자의 판매 및 구매의도를 증진시킬 수 있는 효과를 기대할 수 있다는 결과를 보였다. 두 번째 유형은 표준화된 관리자 답변이다. 선행연구에서 표준화된 관리자 답변은 Standard Managerial Response, Stereotype Response, Untailored Response 등으로 정의되며 일정한 구조를 가지고 표준적인 메뉴얼에 따르는 관리자 리뷰를 의미한다(Li *et al.*, 2018; Wang *et al.*, 2020; Wang and Chaudhry, 2018; Zhang *et al.*, 2020). 선행연구에서는 표준화된 관리자 답변은 소비자 리뷰의 유용성과 후속 리뷰에 부정적인 영향을 미칠 수 있다(Wang and Chaudhry, 2018; Xie *et al.*, 2017). 또한 Chen and Xie(2008)와 Xie *et al.*(2017)은 표준화된 관리자 답변은 소비자에게 건설적인 정보를 제공하기 어렵고 잠재적 소비자가 적합한 제품을 식별하는 데 도움을 줄 수 없다고 하였으며, Jin *et al.*(2023)은 잠재적 소비자가 표준화된 관리자 답변을 형식적이고 무책임한 답변이라 여긴다고 보고하였다.

선행연구에 따라 개별화된 관리자 답변이 표준화된 관리자 답변보다 소비자에게 끼치는 효익이 크다는 연구 결과가 다수 존재하고, 관리자는 소비자 리뷰에 개별적으로 대응하는 방식이 이상적이라 할 수 있다. 하지만 실질적으로 소비자 리뷰를 처리하기 위한 비용과 시간을 고려하여야 하며 보다 효율적인 대응 방식을 통해 웹케어를 실현하여야 한다. Wang *et al.*(2020)은 소비자 리뷰 대응

방식에 있어서 시간과 비용을 고려하여 관리자 응답 자원 관리하여야 한다고 주장하였다. 이와 같은 상황에서 관리자 답변을 통해 소비자에게 최대한의 정보를 전달하고 잠재적 소비자의 효익을 고려하기 위한 방법의 필요성을 인지하여 본 연구에서는 웹케어 실현 과정에서 개별화된 관리자 답변과 표준화된 답변을 효율적으로 제공하는 전략을 제시한다.

### 2.3 개별화된 관리자 답변을 위한 웹케어 모형의 기능 및 역할

<그림 1>에서는 오디오북 플랫폼 ‘밀리의 서재’의 소비자 리뷰와 관리자 답변의 예시를 보여준다. 각 소비자 리뷰의 내용에 따라 관리자는 표준화된 관리자 답변과 개별화된 관리자 답변을 제시하고 있다. <그림 1>의 개별화된 관리자 답변을 통해 공급자는 소비자가 겪는 오류 사항에 대해 상세히 답변하고 있다. 공급자는 예시 1. 오프라인 접속문제와 같이 해결 가능한 상황에서는 즉각적으로 해결책을 제시하고, 예시 2. 구독 갱신 오류와 같은 추가적인 정보가 필요한 상황에서는 그에 알맞은 답변을 제시하고 있다. 이와 같이 공급자는 개별 소비자 리뷰에서 나타나는 내용 및 요구 사항에 대응하고 있음을 확인할 수 있다.

<그림 1>에서 살펴본 바와 같이, ‘밀리의 서재’에서의 관리자 답변은 소비자의 구체적인 요구와 문제에 대해 세심한 대응을 보여줌으로써 개별화된 관리자 답변의 효과를 명확하게 입증한다. 이러한 개별화된 접근 방식은 고객 만족도를 높이고, 브랜드에 대한 긍정적인 인식을 증진시키는 데 기여한다. 그러나 모든 소비자 리뷰에 개별적인 대응을 하는 것은 현실적으로 어려움을 가지며, 이는 관리자의 업무 부담과 운영 비용의 증가로 이어질 수 있다. Wang *et al.*(2020)은 소비자 리뷰 대응 방식에 있어서 관리자 응답 자원과 비용 측면에서 고려해야 한다고 강조하였다. 또한 소비자 리뷰와 관리자 답변의 내용 흐름이



표준화된 관리자 답변의 예시



개별화된 관리자 답변 예시 1. 오프라인 접속 문제

개별화된 관리자 답변 예시 2. 구독 갱신 오류

<그림 1> 소비자 리뷰와 관리자 답변의 예시

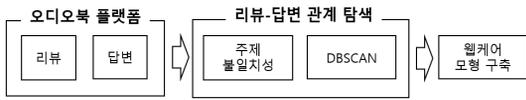
얼마나 일치하는지, 어떻게 하면 관리자가 효율적으로 소비자 리뷰에 대처할 수 있는지에 대한 연구는 충분하지 않다.

본 연구에서는 이러한 공백을 인지하고 효율적인 웹케어 모형의 필요성을 제기한다. 본 연구에서 제안하는 웹케어 모형은 개별화된 관리자 답변과 표준화된 답변을 적절히 조합하고, 디지털 환경에서의 최적화된 고객 서비스를 제공하는 방안을 모색하는 것에 목적을 가진다. 이러한 접근은 관리자의 시간과 자원을 효과적으로 관리하면서도, 고객의 만족도와 브랜드 충성도를 높이는 데 기여할 수 있다. Wang and Chaudhry(2018)는 소비자 리뷰에 대한 무성의한 답변이나 단순한 답변 내용이 부정적인 영향을 미칠 수 있다고 언급하였고, Li et al.(2018)은 관리자가 소비자 리뷰에 대해

개인화된 답변을 제공해야 한다고 하였다. 본 연구에서는 소비자 리뷰에 대해 관리자의 답변 유형을 분석하고 분류함으로써, 위와 같은 문제점을 보완하기 위해 웹케어를 위한 관리자 답변을 위한 추천 모형을 제안하고자 한다.

### Ⅲ. 연구 방법 및 프레임워크

본 연구에서는 오디오북 플랫폼을 이용하는 소비자를 대상으로 소비자의 리뷰와 관리자의 답변을 크롤링하여 수집한 후 전처리 과정을 거쳐 리뷰와 답변에 대한 관계를 탐색하고 분석한다. 이를 통해 웹케어 모형을 구축하고 최적의 모형을 선정하고 결과를 도출하는 과정으로 진행한다. 본 연구의 프레임워크는 <그림 2>와 같다.



〈그림 2〉 연구 프레임워크

### 3.1 데이터 수집 및 전처리

#### 3.1.1 데이터 수집

본 연구에서는 구글 플레이(Google Play)에 등록되어 있는 오디오북 플랫폼을 탐색하고 선정하여 연구를 진행한다. 오디오북은 보는 영상 콘텐츠가 아닌 귀로 듣는 콘텐츠로 최근 시대적 상황을 고려하면 소비자의 선호도가 높다고 할 수 있다. 오디오북 플랫폼에는 소설, 교양, 인기웹소설 등 다양한 콘텐츠를 제공하고 있으며, 소비자들은 시간과 장소에 구애받지 않고 이용이 가능하다는 장점으로 오디오북 시장에서 많은 이용과 높은 다운로드 수를 기록하고 있다. 성행남과 이태원(2022)은 4개의 오디오북 플랫폼(윌라, 밀리의 서재, 팟빵, 오디오클럽 등)을 선정하고 구글 트렌드 분석을 통해 오디오북 플랫폼의 관심분포와 현황에 대한 연구를 진행하였다. 본 연구에서는 기존 연구에서 사용된 4개의 오디오북 플랫폼을 선정하고 이에 해당하는 수집 가능한 소비자 리뷰와 관리자 답변 정보를 활용하고자 한다. 본 연구에서는 파이썬(Python)의 Google-Play-Scraper 라이브러리를 이용하여 자체적인 리뷰 크롤링 모델을 개발하여 4개의 오디오북 플랫폼의 리뷰와 답변 데이터를 수집한다. 리뷰와 답변 간의 관계를 탐색하기 위해 본 연구에서는 소비자 리뷰에 대해 관리자 답변이 존재하지 않는 경우를 제외한 모든 데이터를 선별하여 수집하였다. 관리자 답변의 경우 일반적으로 플랫폼을 이용하는 모든 소비자 리뷰에는 관리자의 답변이 존재하지 않는 경우가 많다.

#### 3.1.2 데이터 전처리

앱별 소비자 리뷰와 관리자 답변을 수집한 후 비정형화된 데이터에 대한 텍스트 전처리 과정을

진행한다. 본 연구에서 선정한 오디오북 플랫폼은 주로 국내 소비자들이 많은 이용을 하고 있다는 것을 고려하여 한국어를 대상으로 분석을 시행하였으며, 조사, 특수문자, 문장 부호, 불필요한 공백 등과 같은 불필요한 의미를 가지는 요소나 불용어 등을 제거한다. 이후 파이썬의 konlpy 라이브러리의 형태소분석기 Okt(Open Korean Text)를 사용하여 텍스트 요소들의 정규화(Normalization)와 어간화(Stemming) 분석을 시행하고, 소비자 리뷰와 관리자 답변의 키워드에 대한 중복처리, 토큰화, 형태소 태깅 등 전처리 과정을 거쳐 텍스트 분석에 용이한 형태로 데이터를 변환한다.

### 3.2 소비자 리뷰-관리자 답변 관계 탐색

#### 3.2.1 주제 불일치성

본 연구에서는 주제 불일치성을 통해 소비자 리뷰와 관리자 답변 간의 관계를 탐색하며, 주제가 유사한 리뷰와 답변을 선정하여 연구를 진행한다. 기존 연구에 따르면 Hu and Chan(2016)은 소비자가 작성한 리뷰는 하나의 주제를 다루지 않고 양면적인 메시지, 다양한 토픽을 중첩시킨 의미를 포함할 수 있다고 주장하였다. 이종원, 박철(2022)은 다면성이 내포된 소비자 리뷰의 주제에 대한 관리자 답변의 일치성을 정량화하여 소비자 리뷰-관리자 답변 간의 주제 거리를 산출하기 위해 토픽모델링 기법을 활용하였다. 이 연구는 LDA(Latent Dirichlet Allocation) 알고리즘을 사용하여 소비자 리뷰와 관리자 답변 전체의 주제를 구성하고 소비자 리뷰와 관리자 답변 간의 주제 일치성을 평가하기 위해서 토픽을 대상으로 사후 확률 차이의 절대값을 합산한 후 주제 거리를 산출하였다. 주제거리는 산출된 주제 거리에 역방향 코딩을 적용한 후 주제 일치성을 측정하였다. 본 연구에서는 기존 연구의 방법을 토대로 역방향 코딩을 진행하지 않고 주제 거리 계산 방법을 진행하고자 하며, 이를 주제 불일치성이라 명명하고 정의하여 연구를 진행한다.

본 연구에서는 주제 불일치성에 대한 결과를 도출하기 위해 데이터 전처리 후 소비자 리뷰와 관리자 답변을 TFIDF(Term Frequency Inverse Document Frequency)를 계산한다. TFIDF는 해당 키워드가 문서에서의 출현 빈도에 따라 키워드의 상대적 중요도를 정량화할 수 있으며, 텍스트 분석에 있어 주로 사용되는 기법으로 알려져 있다(이성직, 김한준, 2009). 다음으로 LDA 알고리즘을 적용하여 응집도(Coherence) 기반의 주제 개수 탐색을 진행한다. 응집도는 각 토픽이 의미론적으로 얼마나 일관성이 있는지 평가하는 지표로 사용되며, 문서 집단의 주제 개수를 선정하는데 기준이 되는 지표이다(Newman *et al.*, 2010). LDA 알고리즘은 주제의 개수에 따라 응집도의 결과가 다르게 출력되기 때문에 본 연구에서는 응집도의 변곡점에서 주제 개수를 선정한다. 최종적으로 본 연구의 오디오북 플랫폼별로 적합한 주제 개수를 선정하고 각 소비자 리뷰와 관리자 답변의 주제 거리를 계산하여 주제 불일치성을 산출한 후 주제 유사성이 높은 리뷰-답변을 선정하고 주제 불일치성이 높은 리뷰-답변을 제거한다.

### 3.2.2 DBSCAN(Density Based Spatial Clustering of Application with Noise)

DBSCAN은 동일 군집 내의 각 관측치들은 서로 거리 상의 밀도가 높게 형성되는 곳에 위치한다는 가정하에 각 관측치 주변의 밀도를 이용하여 군집을 형성하는 알고리즘으로 알려져 있으며, Ester *et al.*(1996)이 밀도 기반의 군집 방식으로 이를 제안하였다. DBSCAN을 이용하는 기존 연구에서는 소비자 리뷰를 군집화하여 소비자 리뷰-소비자 리뷰의 관계를 탐색하고, 이에 대한 통찰을 발굴한 연구가 많이 진행되었지만(Fauzi *et al.*, 2021; Indah *et al.*, 2019; Mohammed *et al.*, 2020;), 관리자 답변 간의 관계를 분석하는 연구는 극히 드물었다. 본 연구에서는 이를 보완하기 위해 기존 연구를 바탕으로 DBSCAN을 이용하여 소비자 리뷰-관리자 답

변과 관리자 답변 분류 시 개별화된 답변과 표준화된 답변으로 분류하기 위해 관리자 답변-관리자 답변의 관계 또한 탐색하고자 한다.

관리자 답변의 유형에는 개별화된 관리자 답변과 표준화된 관리자 답변으로 구분할 수 있는데 개별화된 관리자 답변은 소비자 리뷰에 대해 관리자가 개별적으로 작성한 답변을 의미하고, 표준화된 관리자 답변은 소비자 리뷰에 대해 관리자가 유형화된 내용으로 작성한 답변을 말한다. 개별화된 관리자 답변은 관리자 답변 집단에서 내용적 차별성이 있으며 표준화된 관리자 답변끼리는 유사한 의미와 유형화된 답변 구조를 가진다.

본 연구에서는 개별화된 관리자 답변은 관리자 답변 집단에서 차별성을 가지고 표준화된 관리자 답변은 내용적·구조적 유사성을 가져 유형화된다고 가정한다. 상기 과정을 거쳐 관리자 답변 집단을 대상으로 DBSCAN을 적용해 군집화를 시도하면 표준화된 관리자 답변의 유형끼리 군집이 형성되고 개별화된 관리자 답변은 비교적 밀도가 낮은 데이터로 판정되므로 표준화된 관리자 답변의 유형 군집에 포함되지 않고 이상치로 판정되어 해당 관리자 답변을 개별화된 관리자 답변으로 식별할 수 있다. 본 연구에서는 DBSCAN 분석 중 군집 과정에서 이상치 판정 방법론을 적용하여 탐색을 시도한다. 이를 위해 Indah *et al.*(2017)이 제안한 방법론에 따라 본 연구에서 전처리가 완료된 관리자 답변을 대상으로 TFIDF를 적용하고 코사인 유사도(Cosine Similarity) 계산, 코사인 거리(Cosine Distance)를 산출하여 DBSCAN으로 군집화 후 최적의 DBSCAN 파라미터( $\epsilon$ ,  $\min\_samples$ )를 탐색하기 위해 격자 탐색법(Grid Search)을 진행한다.

### 3.3 웹케어 모형 구축

본 연구에서는 소비자 리뷰에 따른 관리자 답변에 대한 유형을 예측하기 위한 웹케어 모형을 구축하는데 목적을 둔다. 웹케어 모형 구축을 위해 우선 소비자 리뷰의 TFIDF를 계산한 후 결과값을

토대로 독립변수를 선정하고 관리자 답변에 대한 군집된 결과를 이용하여 유형에 따른 개별화된 관리자 답변과 표준화된 관리자 답변을 종속변수로 설정한다. 본 과정에서는 다양한 기계학습기법(Naive Bayse, Support Vector Machine, Decision Tree, Random Forest, XGBoost)을 적용하여 Google Colab 환경에서 실험을 실시한다. 기계학습기법들은 텍스트 모형 구축 뿐만 아니라 다양한 분야에서 많은 연구에 적용되고 있으며, 본 연구에서도 다음과 같은 설명을 바탕으로 본 연구에서 사용할 적합한 웹케어 모형을 구축한다.

Naive Bayes는 개별 변수들이 서로 독립적이라는 가정하에 관측치를 이용하여 다항 생성식을 추정하고 개별 관측치의 독립변수가 포함될 확률이 가장 높은 범주로 분류하는 알고리즘으로 알려져 있고(김기룡, 2003). Support Vector Machine(SVM)은 훈련 데이터의 마진이 최대가 되는 최적화 식을 바탕으로 분류 함수를 추정하여 범주를 분류하는 기법이며 커널 함수를 통해 비선형 문제를 해결할 수 있고 일반화에 뛰어나다는 장점을 가지고 있다(Cortest and Vanick, 1995). Decision Tree는 의사결정 규칙을 바탕으로 집단의 불순도를 계산하여 집단의 유형을 소그룹으로 반복해서 나누는 알고리즘이다. 또한, 모형 해석력이 우수하며 비선형 관계와 변수 간에서 생겨나는 상호관계를 효율적으로 처리한다는 점에서 장점을 가진다(Hagenauer and Helbich, 2017). Random Forest는 임의성(random)을 기반으로 BootStrap을 생성하여 다수의 Decision Tree을 결합시, 상관관계를 낮추어 모형의 일반화를 높인다. 이를 통해 Decision Tree의 숫자를 늘려도 과적합을 방지할 수 있다는 장점을 가진다(Breiman, 2001). XGBoost는 Boosting 알고리즘을 기반의 모형이며 이전의 학습 모형의 결과를 새로운 관측치에 반영하여 성능을 더 향상시키는 작동 방식을 가진다. 다양한 하이퍼파라미터가 내재되어 있으며 유용성이 높은 장점을 가지고 있다(Chen and Guestrin, 2016). 이를 토대로 웹케어 모형을 구축하고 검증에 위해 정확성(Accuracy), 정밀성

(Precision), 재현성(Recall), F1-score 등을 통해 성과 지표를 사용하여 최적의 모형을 선정한다.

## IV. 실험 결과

### 4.1 데이터 및 시각화 결과

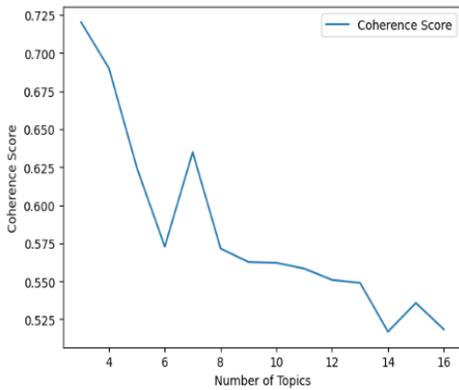
본 연구에서는 소비자들의 선호도가 높은 4개의 오디오북 플랫폼(밀리의 서재, 윌라, 팟빵, 오디오클럽 등)을 선정하고 해당 플랫폼별 소비자 리뷰 데이터를 수집하였다. 수집된 데이터는 소비자 리뷰, 리뷰 등록 일시, 리뷰 유용성, 소비자 만족도, 관리자 답변, 관리자 답변 등록 일시 등의 변수로 구성되어 있었으며, 본 연구에서는 이들 중 소비자 리뷰와 관리자 답변을 별도로 추출하여 실험하였다. 밀리의 서재는 15,613건, 윌라는 4,770건, 팟빵은 13,976건, 오디오클럽은 1,606건을 수집하여 총 35,965건에 대한 소비자 리뷰 데이터를 확보하였다. 또한, 관리자 답변으로는 밀리의 서재 14,955건, 윌라 2,088건, 팟빵 1,982건, 오디오클럽 1,534건으로 총 20,559건으로 수집되었다.



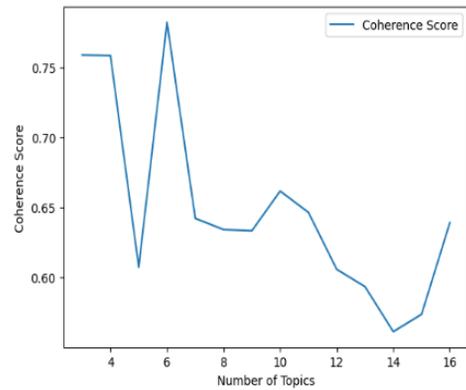
〈그림 3〉 오디오북 플랫폼별 관리자 답변에 따른 소비자 리뷰 시각화 결과

본 연구에서는 먼저 오디오북 플랫폼별로 수집된 소비자 리뷰와 관리자 답변 데이터를 토대로 텍스트 전처리 과정을 거친 후 빈번하게 등장하는 키워드를 선별하였다. 관리자 답변 유무에 따른 소비자 리뷰의 키워드 양상을 파악하기 위해 데이터 시각화 방법론 중 하나인 워드클라우드(WordCloud)를 이용하여 결과를 도출하였다. 오디오북 플랫폼에서 소비자 리뷰에 대한 관리자 답변이 있을 때 모든 플랫폼에서 공통적으로 ‘업데이트’, ‘사용’, ‘재생’, ‘다운로드’ 등의 키워드가 가장 많은 빈도값으로 도출되었고, 소비자 리뷰에 대한 관리자 답변이 없을 때 ‘결제’, ‘페이지’, ‘시간’, ‘사용’, ‘업데이트’, ‘방송’, ‘다운’, ‘재생’, ‘컨텐츠’ 등의 키워드가 빈번하게 등장하였다. 소비자의 답

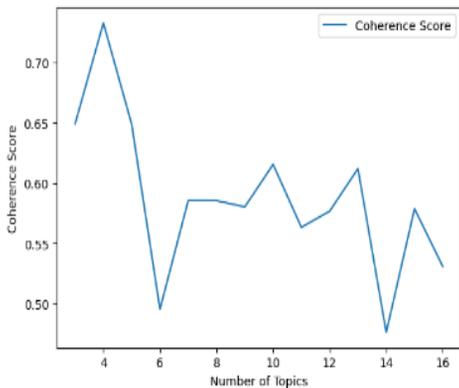
변이 있는 경우 소비자들이 주로 오디오북 플랫폼에서 제공되는 기능들에 대해 양호한 평가를 내렸음을 의미한다고 볼 수 있다. ‘업데이트’는 플랫폼이 지속적으로 서비스를 개선하고 새로운 기능을 도입하고 있음을 나타내며, ‘사용’, ‘재생’, ‘다운로드’는 플랫폼의 핵심 사용성과 기능에 대한 긍정적인 사용자 경험을 반영하고 있다고 파악된다. 이러한 피드백은 관리자가 미리 감지하고 고객들에게 응답하여 서비스 품질을 높일 수 있는 중요한 정보를 제공한다. 소비자의 답변이 없는 경우 ‘결제’와 ‘페이지’와 같은 키워드의 등장은 사용자들이 불만족을 표현하고 있다는 신호로 해석된다. ‘시간’이 강조되는 경우, 사용자들이 오디오북 플랫폼의 응답 시간 또는 서비스 속도에 대한 불만을 나타낼



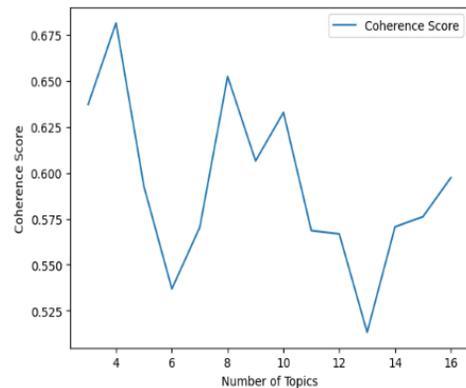
밀리의 서재(k=7)



윌라(k=6)



팟빵(k=4)



오디오클럽(k=4)

〈그림 4〉 오디오 플랫폼별 토픽 개수에 따른 응집도

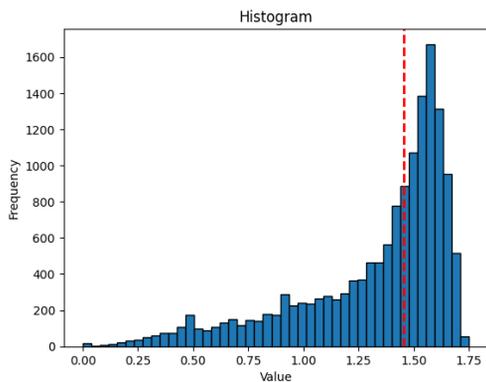
수 있으며, 이는 서비스의 신속성에 대한 향상이 필요하다는 시사점으로 이어진다. ‘방송’, ‘컨텐츠’ 등의 키워드는 사용자들이 플랫폼에서 제공하는 콘텐츠나 방송에 대한 관심을 나타내며, 이에 대한 불만이나 개선 사항이 존재할 수 있다.

#### 4.2 리뷰-답변 관계 탐색 결과

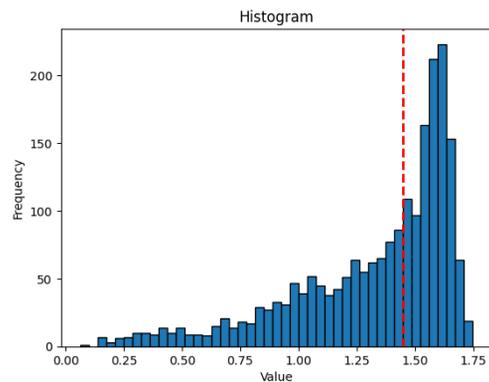
본 연구에서는 이중원, 박철(2022)의 연구를 토대로 텍스트 데이터의 주제 거리 계산 방식과 토픽모델링 기법인 LDA 알고리즘을 이용하여 소비자 리뷰와 관리자 답변 간의 주제 불일치성에 대한 결과를 도출하였다.

첫째, 토픽의 수(k)를 결정하기 위해 주제 개수

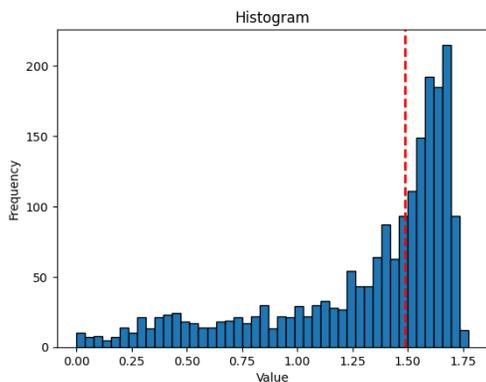
후보군을 3~17개로 설정하고 혼잡도(Perplexity)와 응집도(Coherence)를 통한 주제 개수와 이에 따른 응집도 변곡점을 이용하여 적합한 주제 개수를 오디오북 플랫폼별로 선정하였다. 본 과정에서는 소비자 리뷰와 관리자 답변의 사후 분포 확률을 산출하는 과정에서 동일 토픽의 사후분포 차를 구하고 토픽 분포 차이의 절대값에 대한 합을 산출하여 주제 불일치성을 도출하였다. 이때 소비자 리뷰와 관리자 답변의 주제 불일치성은 해당 소비자 리뷰-관리자 답변의 관계가 의미적으로 얼마나 불일치하는지 정량적으로 나타낸 값이다. 소비자 리뷰에는 다양한 의미를 가지며 양면적인 메시지를 지니고 있기 때문에 해당 리뷰에 대해 관리자는 여러 전체적인 흐름을 고려한 답변을 제시하여



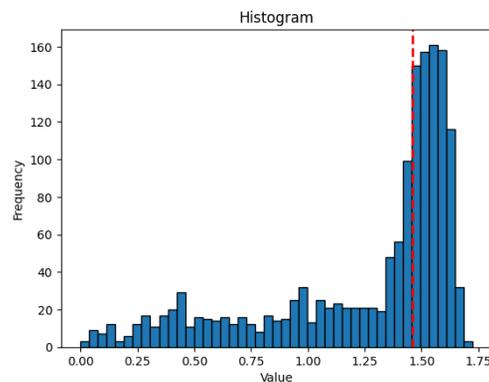
밀리의 서재



월라



팟빵



오디오클립

〈그림 5〉 오디오북 플랫폼별 주제 불일치성 막대 그래프(빨간 점선: 주제 불일치성 중위수)

야 한다(Hu and Chen, 2016). 따라서 본 연구에서는 높은 주제 불일치성(상위 50%)을 가지는 소비자 리뷰-관리자 답변 쌍을 제거하여 본 연구에 적합한 소비자 리뷰-관리자 답변 데이터를 바탕으로 연구를 진행하였다. 결과적으로 산출된 오디오북 플랫폼 주제 불일치성을 <표 5>에 나타내었다.

둘째, 본 실험에서는 개별화된 관리자 답변과 표준화된 관리자 답변에 대한 유형화를 위해 DBSCAN을 이용한 문서 군집을 실시하였다. 오디오북 플랫폼별 문서 군집에 필요한 DBSCAN 파라미터를 탐색하기 위해 격자 탐색법(Grid search)을 이용하여 실루엣 스코어(Silhouette Score)를 기준으로 선정하였으며, 최종결과는 <표 1>과 같다.

<표 1> 플랫폼별 최적의 DBSCAN 파라미터 선정

플랫폼	밀리의 서재	윌라	팟빵	오디오 클립
EPS	0.30	0.24	0.36	0.23
min samples	150	17	17	8
Silhouette Score	0.3736	0.2613	0.3920	0.4035
Num of Clusers	4	3	3	14

오디오북 플랫폼별로 개별화된 관리자 답변 비율과 표준화된 관리자 답변 비율에 대한 비교 분석시 차이를 나타내고 있었으며, 이는 각 오디오북 플랫폼별 소비자 리뷰 응대를 담당하는 관리자가 달라 생성된 관리자 답변의 특성이 상이함으로 인해 도출된 결과로 추측된다. 특히 오디오 클립

에서 73.53%의 개별화된 관리자 답변 비율에 비해 팟빵은 개별화된 답변 비율이 95.26%에 달하는 것을 확인할 수 있다.

셋째, 본 실험에서는 소비자 리뷰-관리자 답변과 관리자 답변-관리자 답변 관계 탐색 결과를 바탕으로 웹케어 모형 구축을 진행하였다. 웹케어 모형 구축을 위해 데이터셋을 6:2:2로 분할하여 진행하였으며, 기계학습기법 중 Naive Bayes, SVM, Decision Tree, Random Forest, XGBoost 등의 결과를 도출하고 정확성(accuracy), 정밀성(Precision), 재현성(Recall), F1-Score를 통해 성과 지표를 산출하였다. 각 오디오북 플랫폼별 모형 검증 과정에서 높은 성과 지표를 가지는 모형을 최적의 모형으로 선정하고, 이를 통해 평가 데이터를 통한 결과를 도출하였다. 실험 결과, 데이터 분할 과정에서 종속 변수(개별화된 관리자 답변/표준화된 관리자 답변)에 대해 층화 추출법을 이용해 종속변수의 비율을 고르게 분포시켰다.

평가 데이터셋 실험결과, 오디오북 플랫폼별로 구축된 모형의 성과지표는 <표 2>와 <표 3>과 같다. 일반적으로 SVM의 실행 시간이 가장 오래 걸렸으며, RandomForest, XGBoost 등 Tree 계열의 분류 모형이 높은 정확성(83.42%, 73.38%, 94.74%, 94.97%)을 보였다. 정밀성과 재현성을 동시에 고려할 수 있는 F1-score 기준에서는 오디오북 플랫폼별 성과가 상이하였다. 밀리의 서재는 SVM에서 62.92%의 성과를 보였으나 타 오디오북 플랫폼에서 SVM의 F1-Score는 비교적 높지 않은 것으로

<표 2> 밀리의 서재와 오디오클립의 개별 관리자 답변 탐지 모형 지표

Model	밀리의 서재					오디오클립				
	Accuracy	Precision	Recall	F1 Score	Time	Accuracy	Precision	Recall	F1 Score	Time
NB	47.93	58.91	64.42	46.21	0.0016	68.83	65.08	68.66	<b>65.16</b>	0.0007
SVM	83.22	68.45	60.96	<b>62.92</b>	0.1738	72.73	56.91	51.11	46.33	0.0136
DT	78.41	61.92	62.34	62.12	0.0007	62.99	47.96	48.36	47.82	0.0004
Random Forest	<b>83.42</b>	69.08	61.55	61.55	0.0476	<b>73.38</b>	62.86	55.44	54.62	0.0134
XGBoost	83.09	68.17	61.35	61.35	0.0048	70.78	56.94	53.67	52.83	0.0018

〈표 3〉 율라와 팻빵의 개별 관리자 답변 탐지 모형 지표

Model	율라					팻빵				
	Accuracy	Precision	Recall	F1 Score	Time	Accuracy	Precision	Recall	F1 Score	Time
NB	82.78	48.92	47.82	47.89	0.0012	83.42	59.69	86.02	<b>61.60</b>	0.0007
SVM	94.26	47.13	50.00	48.52	0.0336	92.46	47.67	48.42	48.04	0.0105
DT	91.39	65.10	71.95	<b>67.68</b>	0.0006	92.46	54.16	53.71	53.91	0.0003
Random Forest	<b>94.74</b>	97.36	54.17	56.33	0.0256	94.97	54.16	53.71	53.91	0.0201
XGBoost	94.74	77.79	61.99	66.28	0.0021	<b>94.97</b>	47.73	49.74	48.71	0.0018

관측되었다. 오디오클립과 팻빵에서는 Naive Bayse의 F1-score가 65.16%, 61.60%로 가장 높게 나타났으며 율라는 Decision Tree에서 F1-score가 67.68%로 나타났다. 이와 같이 오디오북 플랫폼마다 성과지표는 모형에 따라 상이하게 도출되었으며 전반적으로 트리 계열의 모형의 성능이 가장 우수하였음을 확인할 수 있다.

## V. 결 론

디지털 환경이 확대되고 소비자의 참여가 늘어감에 따라 온라인상에서의 기업 및 관리자의 소통 노력은 더욱 중요한 요소가 되었다. 소비자 리뷰는 제품에 대한 피드백 전달하는 역할 뿐만 아니라 잠재적 소비자의 인식에 영향을 끼치는 등 많은 역할을 하기 때문에 관리자의 답변이 기업의 평판과 신뢰도에 큰 의미를 가지게 되었다. 그러나 소비자 리뷰의 수는 급속도로 증가하고 있는 추세이며, 이를 관리하기 위한 비용과 인력을 고려할 때 개별적인 응대가 쉽지 않고 현실적으로 어려운 문제라 할 수 있다. 본 연구에서는 소비자 리뷰의 보다 효율적인 관리를 위한 웹케어의 필요성을 인지하여 소비자 리뷰와 관리자 답변의 관계, 관리자 답변 간의 관계에 대한 탐색을 시행하였다. 시각화 분석 결과 소비자의 답변이 있는 경우 오디오북 플랫폼은 계속해서 업데이트와 관련된 피드백을 수용하고 기능을 개선함으로써 사용

자들의 긍정적인 경험을 유지해야 한다. 또한, 소비자 리뷰에 빠르게 대응하여 사용자들에게 감사의 인사 또는 추가적인 설명을 통해 소통할 경우, 이는 고객 충성도를 높일 수 있다고 사료된다. 소비자의 답변이 없는 경우 플랫폼은 결제, 페이지, 시간과 관련된 불만족을 해소하기 위한 개선책을 모색해야 한다. 사용자들의 불편사항을 해결하고, 서비스의 신속성을 강조하는 데 초점을 맞추어야 한다. 콘텐츠와 방송에 대한 피드백은 새로운 콘텐츠 개발이나 방송 전략의 수정에 활용될 수 있으며, 사용자들이 원하는 방향으로 서비스를 발전 시키는데 중요한 역할을 할 것이다. 즉, 시각화 분석을 통해 플랫폼이 사용자 요구에 민감하게 대응하고, 서비스를 지속적으로 향상시키는 데 중요한 지침을 제공하며, 소비자와의 소통을 강화하여 긍정적인 사용자 경험을 유지하는 데 기여할 수 있다. 기존 연구에서 활용된 주제 불일치성과 DBSCAN을 적용하여 소비자 리뷰와 관리자 답변의 관계를 탐색하고 관리자 답변의 유형을 분리하였으며, 다양한 기계학습기법을 활용한 웹케어 모형을 구축한 후 오디오북 플랫폼에 적용 가능한 최적의 모형을 선정하였다. 밀리의 서재와 오디오클립, 율라는 Random Forest 기반의 웹케어 모형으로 각 83.42%, 73.38%, 94.47%의 정확성을 보이고 있었고, 팻빵은 XGBoost 기반의 웹케어 모형으로 94.97%의 정확성이 도출되었다. 각 오디오북 플랫폼에서 기계학습 모형의 성능이 다른 이유는

오디오북 플랫폼별 데이터의 구조적 차이와 소비자 리뷰의 내용 및 양상에서의 차이에 기인한 것으로 추측된다. 특히, Random Forest와 XGBoost는 비선형적인 데이터 특성을 잘 처리하고, 고차원의 데이터에서도 강력한 성능을 발휘하는 경향이 있어, 오디오북 플랫폼의 텍스트 리뷰 데이터를 분석하는데 효과적인 결과를 보였다.

본 연구의 학술적 시사점 및 실무적 시사점은 다음과 같다. 오디오북 플랫폼을 이용한 국내 연구가 활발히 진행되지 않은 소비자 리뷰와 관리자 답변 간의 관계를 탐색하여 웹케어를 위한 실질적 분석이 이루어졌다는 점에서 학술적 시사점을 가진다. 본 연구에서는 이종원과 박철(2022)이 제시한 주제 일치성을 소비자 리뷰와 관리자 답변에 적용하여 소비자의 기대와 서비스 간의 불일치성을 면밀하게 파악하였다. 이를 통해 본 연구에서 제시하는 웹케어 모형의 목적과 효율성을 향상시켰다는 점에서 학술적 기여가 있다. 그리고 기존에 많은 연구에서는 소비자 답변을 이용한 군집 방법을 활용한 사례가 많이 진행되었지만, 본 연구에서는 DBSCAN을 이용하여 관리자 답변 분류에 적용하였다는 점에서 차별점이 있으며, 관리자 답변 유형의 다양성을 효과적으로 분류하고 접근하였다는 점에서 의의를 가진다. 국내 연구에서는 주로 쇼핑물과 관광 및 여행 등 소비자 리뷰에 대한 연구가 활발히 진행되었지만, 관리자 답변을 포함한 소비자 리뷰 간의 관계를 살펴본 본 연구가 기초자료로 활용될 수 있으며, 보다 질적으로 향상된 연구로 진행될 것이라 기대된다. 이러한 관계에 대한 유형을 기계학습기법을 통해 웹케어 모형을 구축했다는 점에서 학술적인 시사점을 제공한다. 본 연구에서는 기업의 제한된 비용과 인력을 고려하여 웹케어 모형이라는 효율적인 방법론을 제시함으로써 실무적인 시사점을 지닌다. CRM(고객관계관리)의 측면에서는 소비자 리뷰에 대한 자동화된 대응과 처리를 통해 고객과의 소통을 강화할 수 있고, 개별적인 관리자 답변 및 표준화된 답변을 조합하여 효율적인 소비자 대응

전략을 구축할 수 있다는 장점을 가진다. 또한, RPA(로봇 프로세스 자동화) 기술을 적극 활용하여 CRM 업무를 자동화함으로써 비용과 시간을 효과적으로 절감할 수 있다. RPA는 소비자 리뷰에 대한 업무 프로세스를 자동화하고, 일관된 품질의 답변을 제공함으로써 고객 만족도를 높일 수 있게 된다. 특히, RPA를 통해 소비자 리뷰에 대한 트렌드나 패턴을 파악하여 소비자 대응에 대한 예측과 전략을 수립하는데도 많은 도움이 될 것이다. 이처럼 CRM과 RPA를 활용한다면 소비자 리뷰 관리를 개선할 뿐만 아니라 기업의 비용 효율성을 높이고 고객과의 효과적인 상호작용을 구축하는 데에 매우 유용할 것이다.

본 연구의 한계 및 향후 계획은 다음과 같다. 먼저, 본 연구에서는 실험 과정에서 실루엣 스코어를 기준으로 DBSCAN의 파라미터를 선정하였으나 실루엣 스코어만으로 표준화된 답변의 유형의 개수를 설정하는데 한계점이 있다. 향후 연구에서는 다양한 군집 방법과 기준지표를 탐색하여 답변의 유형화를 진행해야 할 필요성이 있다. 또한 소비자 리뷰에 따라 개별적인 관리자 답변 표준화된 관리자 답변을 예측하는 것 보다 표준화된 관리자 답변의 유형까지 예측하는 다중 클래스 문제로 접근하여 확장된 모형을 구축하여 웹케어 모형의 기능 확대를 통해 실무에 적합한 모형을 제시할 수 있을 것이다.

## 참 고 문 헌

- [1] 김기룡, *전자카탈로그 자동 분류기에 대한 연구* (석사학위논문), 서울대학교, 2003.
- [2] 성행남, 이태원, “오디오북 플랫폼 유료 사용자의 이용행태에 대한 연구: 지속이용의도 및 시장동향 분석을 중심으로”, *인터넷전자상거래연구*, 제22권, 제3호, 2022, pp. 115-135.
- [3] 우세웅, *오디오북 국내 사례와 출판시장의 인식 연구* (석사학위논문), 서울미디어대학원대학교, 2019.

- [4] 이성직, 김한준, “Keyword Extraction from News Corpus using Modified TF-IDF”, *한국전자거래학회지*, 제14권, 제4호, 2009, pp. 59-73.
- [5] 이윤아, *국내 오디오북 이용의 특성에 관한 연구: 이용행태, 동기, 만족도 및 재이용의사를 중심으로* (석사학위논문), 건국대학교, 2022.
- [6] 이중원, 박철, “웹케어(Webcare)는 온라인 리뷰 유용성에 영향을 미치는가? 주체 일치성과 정보원 신뢰의 효과”, *經營學研究*, 제51권, 제2호, 2022, pp. 437-460.
- [7] 황단, *소비자의 부정적인 댓글에 대한 웹케어가 브랜드 평가와 참여 의도에 미치는 영향* (석사학위논문), 중앙대학교, 2018.
- [8] Breiman, L., “Random forests”, *Machine learning*, Vol.45, 2001, pp. 5-32.
- [9] Chen, T. and C. Guestrin, “Xgboost: A scalable tree boosting system”, In *Proceedings of the 22nd acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, August 2016, pp. 785-794.
- [10] Chen, W., B. Gu, Q. Ye, and K. X. Zhu, “Measuring and managing the externality of managerial responses to online customer reviews”, *Information Systems Research*, Vol.30, No.1, 2019, pp. 81-96.
- [11] Chen, Y. and J. Xie, “Online consumer review: Word-of-mouth as a new element of marketing communication mix”, *Management Science*, Vol.54, No.3, 2008, pp. 477-491.
- [12] Cortes, C. and V. Vapnik, “Support-vector networks”, *Machine Learning*, Vol.20, 1995, pp. 273-297.
- [13] Ester, M., H. P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise”, In *kdd*, Vol.96, No.34, August 1996, pp. 226-231.
- [14] Fauzi, M. Z. and A. Abdullah, “Clustering of public opinion on natural disasters in Indonesia using DBSCAN and K-Medoids algorithms”, In *Journal of Physics: Conference Series*, Vol.1783, No.1, 2021, p. 012016.
- [15] Hagenauer, J. and M. Helbich, “A comparative study of machine learning classifiers for modeling travel mode choice”, *Expert Systems with Applications*, Vol.78, 2017, pp. 273-282.
- [16] Herhausen, D., S. Ludwig, D. Grewal, J. Wulf, and M. Schoegel, “Detecting, preventing, and mitigating online firestorms in brand communities”, *Journal of Marketing*, Vol.83, No.3, 2019, pp. 1-21.
- [17] Hong, J.-Y. and W.-N. Lee, “Consumer complaint behavior in the online environment”, *Web Systems Design and Online Consumer Behavior*, New Jersey, 2005, pp. 90-105.
- [18] Hu, Y.-H. and K. Chen, “Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings”, *International Journal of Information Management*, Vol.36, No.6, 2016, pp. 929-944.
- [19] Indah, R. N. G., R. Novita, O. B. Kharisma, R. Vebrianto, S. Sanjaya, T. Andriani, ... and R. Rahim, “DBSCAN algorithm: twitter text clustering of trend topic pilkada pekanbaru”, In *Journal of Physics: Conference Series*, Vol.1363, No.1 November 2019, p. 012001.
- [20] Jin, W., Y. Chen, S. Yang, S. Zhou, H. Jiang, and J. Wei, “Personalized managerial response and negative inconsistent review helpfulness: The mediating effect of perceived response helpfulness”, *Journal of Retailing and Consumer Services*, Vol.74, No.103398, 2023.
- [21] Kumar, N., L. Qiu, and S. Kumar, “Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies”, *Information Systems Research*,

- Vol.29, No.4, 2018, pp. 849-870.
- [22] Lee, J., S. B. Park, and S. Lee, "Are negative online consumer reviews always bad? A two-sided message perspective", *Asia Pacific Journal of Information Systems*, Vol.25, No.4, 2015, pp. 784-804.
- [23] Li, C., G. Cui, and L. Peng, "Tailoring management response to negative reviews: The effectiveness of accommodative versus defensive responses", *Computers in Human Behavior*, Vol 84, 2018, pp. 272-284.
- [24] Mohammed, S. M., K. Jacksi, and S. R. Zeebaree, "Glove word embedding and DBSCAN algorithms for semantic document clustering", In *2020 International Conference on Advanced Science and Engineering (ICOASE)*, December 2020, pp. 1-6.
- [25] Newman, D., J. H. Lau, K. Grieser, and T. Baldwin, "Automatic evaluation of topic coherence", In *Human language technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, June 2010, pp. 100-108.
- [26] Proserpio, D. and G. Zervas, "Online reputation management: Estimating the impact of management responses on consumer reviews", *Marketing Science*, Vol.36, No.5, 2017, pp. 645-665.
- [27] Sparks, B. A., K. K. F. So, and G. L. Bradley, "Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern", *Tourism Management*, Vol.53, 2016, pp. 74-85.
- [28] Van Noort, G. and L. M. Willemsen, "Online damage control: The effects of proactive versus reactive webcare interventions in consumer-generated and brand-generated platforms", *Journal of Interactive Marketing*, Vol.26, No.3, 2012, pp. 131-140.
- [29] Wang, L., X. Ren, H. Wan, and J. Yan, "Managerial responses to online reviews under budget constraints: Whom to target and how", *Information & Management*, Vol.57, No.8, 2020, p. 103382.
- [30] Wang, Y. and A. Chaudhry, "When and how managers' responses to online reviews affect subsequent reviews", *Journal of Marketing Research*, Vol.55, No.2, 2018, pp. 163-177.
- [31] Wei, W., L. Miao, and Z. J. Huang, "Customer engagement behaviors and hotel responses", *International Journal of Hospitality Management*, Vol 33, 2013, pp. 316-330.
- [32] Xie, K. L., K. K. F. So, and W. Wang, "Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach", *International Journal of Hospitality Management*, Vol.62, 2017, pp. 101-110.
- [33] Zhang, X., S. Qiao, Y. Yang, and Z. Zhang, "Exploring the impact of personalized management responses on tourists' satisfaction: A topic matching perspective", *Tourism Management*, Vol 76, 2020, p. 103953.
- [34] Zhang, Y. and C. Vasquez, "Hotels' responses to online reviews: Managing consumer dissatisfaction", *Discourse, Context & Media*, Vol.6, 2014, pp. 54-64.
- [35] Zhang, Z., H. Li, F. Meng, and Y. Li, "The effect of management response similarity on online hotel booking: Field evidence from Expedia", *International Journal of Contemporary Hospitality Management*, Vol.31, No.7, 2019, pp. 2739-2758.

## A Study on Developing a Web Care Model for Audiobook Platforms Using Machine Learning

Dahoon Jeong\* · Minhyuk Lee\*\* · Taewon Lee\*\*\*

### Abstract

The purpose of this study is to investigate the relationship between consumer reviews and managerial responses, aiming to explore the necessity of webcare for efficiently managing consumer reviews. We intend to propose a methodology for effective webcare and to construct a webcare model using machine learning techniques based on an audiobook platform. In this study, we selected four audiobook platforms and conducted data collection and preprocessing for consumer reviews and managerial responses. We utilized techniques such as topic modeling, topic inconsistency analysis, and DBSCAN, along with various machine learning methods for analysis. The experimental results yielded significant findings in clustering managerial responses and predicting responses to consumer reviews, proposing an efficient methodology considering resource constraints and costs. This research provides academic insights by constructing a webcare model through machine learning techniques and practical implications by suggesting an efficient methodology, considering the limited resources and personnel of companies. The proposed webcare model in this study can be utilized as strategic foundational data for consumer engagement and providing useful information, offering both personalized responses and standardized managerial responses.

**Keywords:** *Webcare, Consumer Reviews, Managerial Response, AudioBook, Machine Learning, DBSCAN, Topic Inconsistency*

---

\* Undergraduate Student, Department of Economics, Pusan National University

\*\* Professor, School of Business, Pusan National University

\*\*\* Corresponding Author, Research Professor, BK21 Digital finance Education and Research Center, Pusan National University

## ● 저 자 소 개 ●



**정 다 훈 (wjdekgns980@gmail.com)**

부산대학교 경제학부에서 학사과정 중이며, 현재 동 대학 핀테크융합전공 복수전공 중이다. 주요 관심분야는 데이터 마이닝, 자연어 처리, 생성형 AI, 리스크 관리 등이다.



**이 민 혁 (minhyuk.lee@pusan.ac.kr)**

KAIST에서 수리과학, 경영과학 학사를 서울대학교에서 산업공학 박사학위를 취득하였으며, 현재 부산대학교 경영학과 디지털금융 전공 교수로 재직하고 있다. 주요 관심 분야는 인공지능, 빅데이터, 시계열 분석, 블록체인 기술을 활용한 금융시장 분석, 투자 전략, 금융 혁신 전략 등이다.



**이 태 원 (twanny@pusan.ac.kr)**

부산대학교에서 경영정보 및 생산관리 전공으로 경영학 박사학위를 취득하였고, 현재 동 대학의 경영학과 BK21 디지털 금융 교육연구단 연구교수로 재직 중이다. 주요 관심분야는 비즈니스 애널리틱스, 빅데이터, 인공지능, CRM 등이다.

논문접수일 : 2023년 11월 27일

게재확정일 : 2024년 02월 12일

1차 수정일 : 2024년 01월 11일