

의사결정나무를 활용한 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드 도출 연구: 별점과 좋아요를 중심으로

민경수* · 유동희**

〈 목 차 〉

I. 서론	3.5 분류모델 구축
II. 문헌 연구	IV. 연구 결과
III. 연구 방법	4.1 핵심 키워드 및 규칙 도출
3.1 연구 진행 과정	4.2 규칙 해석
3.2 데이터 수집	V. 결론
3.3 전처리	참고문헌
3.4 데이터 균형화 및 변수 선택	<Abstract>

I. 서론

정보 기술의 발전이 가속화되고 전자상거래의 형태가 다양해지면서 온라인 쇼핑몰의 규모가 확장되고 있다. 특히 2019년 말에 시작된 코로나19의 확산으로 비대면 비즈니스가 급격하게 성장하면서 온라인 쇼핑몰의 시장 규모는 매년 20% 이상 성장하는 모습을 보이고 있다(김윤호 등, 2022). 일반적으로 온라인 쇼핑몰의 소비자들은 주로 상품을 구매하기 전에 온라인 소비자 리뷰를 확인하는데, 소비자들은 판매자들보다 상품의 품질과 관련된 정보가 부족하거나 제한적이기 때문이다(Mudambi and

Schuff, 2010). 따라서 소비자들은 상품을 구매한 소비자가 작성한 리뷰에 크게 의존하고 있으며, 생산자와 서비스 공급자는 이러한 소비자 리뷰를 분석하여 상품 및 서비스의 품질과 기준을 향상시키기 위해 노력하고 있다(Ravi and Ravi, 2015).

온라인 소비자 리뷰에 대한 의존성은 상품의 유형인 탐색재와 경험재에 따라 달라진다(Nelson, 1970). 탐색재는 상품 품질에 대한 정보를 비교적 쉽게 얻을 수 있고 주요 속성이 객관적이라 쉽게 비교가 가능하다. 반면에 경험재는 상품 품질에 대한 정보를 상대적으로 얻기 어려워 비용이 많이 들고, 주요 속성이 주관적이어서 비교하기 어려운 특징을 가지고 있다

* 국민대학교 경영학과 학사, lovable0131@gmail.com (주저자)

** 경상국립대학교 경영정보학과 및 경영경제연구소 교수, dhyoo@gnu.ac.kr (교신저자)

(Mudambi and Schuff, 2010). 탐색제는 소비자가 상품을 구매하기 전에 평가가 가능하며 경험제는 소비자가 상품을 구매한 이후 사용하는 과정에서 평가가 가능하다. 이러한 특징 때문에 일반적으로 경험제가 탐색제보다 온라인 소비자 리뷰에 더 많이 의존하는 경향이 있다(Bei et al., 2004).

온라인 소비자 리뷰는 다양한 소비자들에 의해 작성되고 그 양 또한 방대하기 때문에 기업들이 소비자의 의견을 분석하는데 많은 시간과 노력이 요구된다. 이를 해결하기 위해 최근 텍스트 마이닝 기법을 활용하여 온라인 소비자 리뷰로부터 소비자들의 의견을 추출하는 연구들이 진행되고 있다(김광국 등, 2018; 홍태호 등, 2018; 윤호민, 최규완, 2021). 기존 연구에서는 주로 온라인 소비자 리뷰의 유용성에 영향을 끼치는 요인을 분석하고 있으며 리뷰의 내용, 길이, 평점(Mudambi and Schuff, 2010), 리뷰어의 신뢰도(Li and Hitt, 2008), 그리고 리뷰의 감정(Chatterjee, 2020) 등을 분석대상으로 하고 있다. 기업은 텍스트 마이닝 기법을 활용하여 온라인 소비자 리뷰에 대한 특성을 빠르게 파악할 뿐만 아니라 많은 양의 소비자 리뷰로부터 발생하는 정보의 과부하 문제를 완화시킬 수 있다. 따라서 오늘날 기업들은 다양한 관점에서 온라인 소비자 리뷰를 분석하려고 노력하고 있다.

본 연구에서는 경험제인 로봇 청소기(robot vacuum cleaner)를 대상으로 데이터 마이닝 기법 중 하나인 의사결정나무 기법을 활용하여 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드를 도출하고자 한다. 이때 여러 독립변수들 중 키워드만을 대상으로 하여 자가평가에 해당되는 별점(star rating)과 타인평가에 해당

되는 좋아요(helpfulness) 결정 요인을 분석하고자 한다. 다시 말해 다른 독립변수들을 배제하고 오로지 리뷰의 키워드에만 초점을 둬서 어떤 키워드가 들어갔을 때 별점 평가에 영향을 주는지와 타인 리뷰가 유용하다고 평가하는지를 확인해보고자 한다. 본 연구는 데이터 마이닝에서 분류모델을 만들 때 주로 활용되었던 의사결정나무 기법을 텍스트 분석 분야에 활용한 점과 온라인 소비자 리뷰 평가 중 별점과 좋아요 결정에 영향을 주는 핵심 키워드를 도출하고 비교한 점에서 기존 연구들과의 차별성을 두고자 한다.

본 연구가 설정한 연구 문제(research question)는 다음과 같다.

- 연구 문제 1: 상품별로 자가평가와 타인평가에 영향을 주는 핵심 키워드는 서로 다르다.
- 연구 문제 2: 전체 상품의 경우 자가평가에 대한 핵심 키워드는 유사하며, 타인평가에 대한 핵심 키워드 또한 유사하다.

본 연구의 구성은 다음과 같다. 2장에서는 온라인 소비자 리뷰 정보를 대상으로 텍스트 마이닝을 적용한 기존 문헌들을 살펴본다. 3장에서는 본 연구에서 제안한 연구 진행 과정과 주요 분석 단계에 관해 기술한다. 4장에서는 연구 결과를 살펴본다. 마지막으로 5장에서는 본 연구의 요약과 함께 한계점 및 향후 발전 방향에 대해 다룬다.

II. 문헌 연구

온라인 소비자 리뷰를 대상으로 텍스트 마이

닝 기법을 적용한 연구는 크게 토픽 모델링, 감성 분석, 연관규칙 분석, 예측 모형 구축 연구로 구분할 수 있다.

먼저 토픽 모델링을 적용한 연구는 다음과 같다. Ma et al.(2013)에서는 중국 온라인 쇼핑몰의 디지털 카메라와 휴대폰 소비자 리뷰를 분석하였다. 상품의 특징을 추출하기 위해 LDA(Latent Dirichlet Allocation) 기법과 동의어 어휘 목록을 결합한 접근방법을 제안하였는데 이 방법이 기존의 연관규칙 분석보다 성능이 높은 것을 보여주었다. 김광국 등(2018)은 시장 점유율 상위 20위까지의 모바일 쇼핑 앱을 대상으로 LDA 기법을 적용하여 토픽을 분석하였고 분석된 토픽을 이용하여 설문문항을 구성하는 연구를 진행하였다. 설문 결과를 바탕으로 고객만족도가 고객 불평과 고객충성도에 어떠한 영향을 끼치는지 분석하였다. Wu et al.(2022)은 인스타그램과 페이스북의 성능과 품질에 대한 고객만족도를 확인하기 위해 온라인 소비자 리뷰를 분석하였다. LDA 기법을 활용하여 문서 내의 주요 토픽들을 파악하였고 사용된 단어와 단어 빈도 분석을 통해 두 가지 애플리케이션에 대한 고객만족도가 모두 높다는 것을 확인하였다.

온라인 소비자 리뷰의 감성 분석을 실시한 연구를 요약하면 다음과 같다. 연종흠 등(2011)은 유플러스 OZ스토어의 소비자 리뷰에 대한 감성 분석을 실시하였고, 상품 평가 외에 배송, 고객 불만과 같이 다양한 정보를 활용할 수 있는 온라인 감성 분석 처리 기법을 제안하였다. 야오즈옌 등(2021a)은 엘프닷컴 레스토랑 소비자 리뷰를 대상으로 감성 분석을 실시하였고 긍정적이거나 부정적인 감정이 많이 포함된 리

뷰일수록 유용성이 높다는 결과를 도출하였다. Luo et al.(2021)은 트립어드바이저에서 로봇 서비스가 있는 호텔들의 소비자 리뷰를 대상으로 로봇의 서비스 속성과 고객만족도 간의 관계를 조사하였다. 리뷰 및 개별 문장 수준에서의 감성 분석을 수행하여 소비자들이 로봇의 서비스 속성에 대해 긍정적인지 부정적인지를 파악하였다. 황호현 등(2022)은 무신사 소비자 리뷰에 대한 감성 분석을 실시하였는데 의류 카테고리나 소비자별로 리뷰가 다른 특성을 보이기 때문에 리뷰와 상품에 대한 추가적인 정보를 고려한 감성 분석 모델을 제시하였다.

온라인 소비자 리뷰에 토픽 분석과 감성 분석을 함께 실시한 연구는 다음과 같다. 홍태호 등(2018)은 트립어드바이저의 호텔 소비자 리뷰에서 여러 토픽을 추출한 뒤 감성 분석을 실시하여 호텔 관광객이 제일 중요하게 생각하는 요소를 파악하고자 하였다. 윤호민과 최규완(2021)은 트립어드바이저의 레스토랑 소비자 리뷰를 대상으로 토픽을 분류한 뒤 감성사전을 활용하여 토픽별 감성 분석을 실시하였다. 야오즈옌 등(2021b)은 아마존에서 판매되는 탐색재와 경험재 소비자 리뷰에서 LDA 기법을 적용하여 토픽을 도출하였고, 도출한 토픽별로 감성 분석을 실시하였다. 그 결과 긍정적이거나 부정적인 감정 크기가 클수록 리뷰 유용성에 영향을 주는 것을 확인하였다. Kwon 등(2021)에서는 스카이트랙스에서 아시아 항공사를 이용하는 소비자들의 온라인 리뷰를 분석할 때 토픽 모델링과 감성 분석을 활용하였다.

온라인 소비자 리뷰에 연관규칙 분석을 활용한 연구는 다음과 같다. Kim 등(2009)과 김근형(2011)의 연구에서는 오피니언 마이닝을 위

해 연관규칙을 활용하였다. RaziaSulthana and Ramasamy(2017)는 아마존 소비자 리뷰를 기반으로 단어들 간의 연관성을 파악하여 상품 추천 시스템 개발에 활용하였다. 김영길(2018)은 자동차 성능과 관련된 연관어를 파악하기 위해 연관규칙 분석을 실시하였으며, Jiang 등(2018)의 연구에서는 여러 온라인 플랫폼에 등록되어 있는 소형차 소비자 리뷰로부터 감성 디자인과 관련된 연관어를 분석하였다.

온라인 소비자 리뷰에서 의사결정나무를 활용한 연구는 주로 리뷰의 감성 예측 모형을 구축하거나 다른 기계학습 기법과 예측 성능을 비교하는 것으로 진행되었다(Pranckevičius and Marcinkevicius, 2017; Noori, 2021; Choudhary and Chhabra, 2021; Trivedi et al., 2022; Kausar et al., 2023). 또한 Lee and Choeh(2017)는 아마존 소비자 리뷰를 대상으로 의사결정나무를 사용하여 더 유용한 타인평가를 추천하는 시스템을 만들었다. 온라인 소비자 리뷰를 분석한 연구는 많지만 리뷰의 유용성에 관한 연구는 현저히 적으며, 리뷰의 유용성 결정요인에 대한 종합적인 분석이 없다는 것에 착안하여 상품에 대한 리뷰 데이터와 리뷰의 텍스트 특성에 대하여 종합적인 분석을 진행하였다.

앞서 살펴본 바에 의하면 지금까지 의사결정나무를 활용하여 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드와 규칙을 도출하는 연구는 찾아볼 수 없었다. 따라서 온라인 소비자 리뷰를 평가하는 주체를 자가와 타인으로 구분한 뒤 평가 결과에 영향을 주는 핵심 키워드를 비교한 연구는 처음이라는 점이 본 연구의 차별점이라 할 수 있다.

Ⅲ. 연구 방법

3.1 연구 진행 과정

본 연구에서는 기존 문헌들(김효곤, 유동희, 2022; 김윤승, 유동희, 2023)을 참고하여 연구 진행 과정을 <그림 1>과 같이 설계하였으며, 이는 온라인 소비자 리뷰 평가에 영향을 주는 키워드와 규칙을 도출하는 과정을 보여준다. 먼저 웹 크롤러를 활용하여 온라인 쇼핑몰인 아마존에 있는 상품 중 로봇 청소기에 대한 온라인 소비자 리뷰 정보를 수집한다. 다음으로 전처리 과정을 통해 비정형인 소비자 리뷰 정보를 정형 데이터 형태로 변환한다. 이때 독립변수는 문서 벡터화를 통해 도출된 키워드들이며, 목표변수는 소비자 리뷰 평가에 사용된 별점과 좋아요를 바탕으로 명목형으로 구성하고자 한다. 또한 분류모델의 편향된 학습을 방지하고자 목표변수의 클래스 비율을 맞추는 데이터 균형화 작업을 실시한다. 그 후, 의사결정나무 기법을 활용하여 분류모델을 구축하고, 변수 선택을 통하여 목표변수인 별점과 좋아요에 영향을 주는 키워드들을 최종 독립변수로 선택한다. 끝으로, 구축된 분류모델들 중 가장 성능이 좋은 분류모델을 통하여 별점과 좋아요 평가에 영향을 주는 핵심 키워드와 규칙들을 도출하고자 한다.

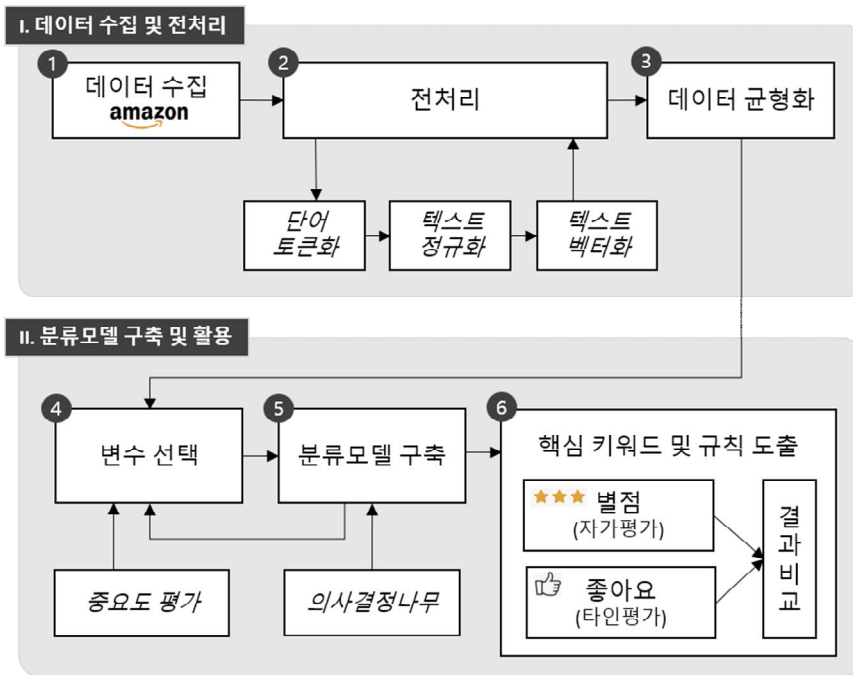
3.2 데이터 수집

본 연구에서는 웹 크롤러를 활용하여 아마존(<http://www.amazon.com>)에 등록된 온라인 소비자 리뷰를 수집하였다. 분석 상품으로는 하이테크 기술이 포함된 로봇 청소기를 선택하였다.

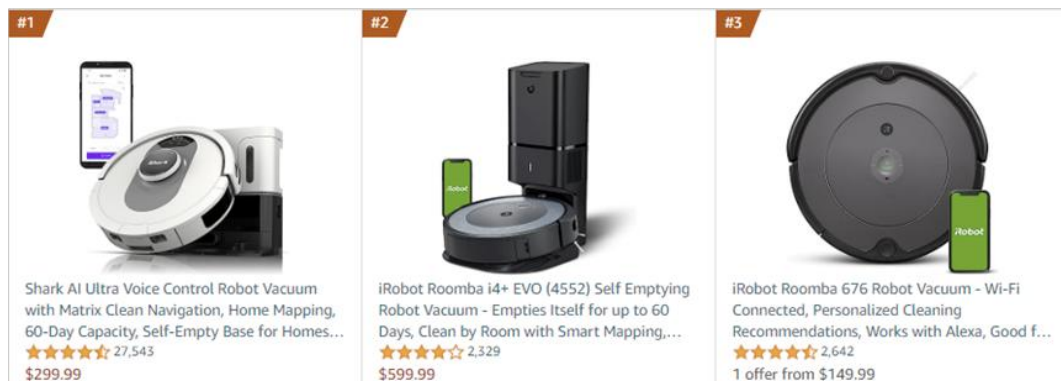
—의사결정나무를 활용한 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드 도출 연구: 별점과 좋아요를 중심으로

<그림 2>와 같이 데이터 수집일을 기준(2023년 7월 6일)으로 아마존에서 판매 순위가 높은 상위 3개의 로봇 청소기에 대한 온라인 소비자 리뷰를 수집하였다. 이때 영어로 작성된 리뷰만을 수집하였으며, 그 외 언어로 작성된 리뷰는 수

집하지 않았다. 수집된 리뷰 관련 정보에는 상품명, 사용자아이디, 날짜, 제목, 리뷰내용, 별점, 좋아요 수가 포함된다. 최근 아마존에서는 특정 검색 기준에 따라 수집할 수 있는 소비자 리뷰의 수를 100개로 제한하고 있어서 별점을



<그림 1> 연구 진행 과정



<그림 2> 아마존에서 제공되는 로봇 청소기 판매 순위 화면

<표 1> 수집된 데이터 정보

구분	상품명	전체 리뷰 수	수집된 별점별 리뷰 수					계
			1	2	3	4	5	
P1	Shark AI Ultra	27,543	100	100	100	100	100	500
P2	iRobot Roomba i4+EVO	2,329	100	59	52	50	100	361
P3	iRobot Roomba 676	2,642	100	70	70	100	100	440

기준으로 최대 100개의 리뷰를 수집하였다. 실제 아마존을 방문하는 소비자 또한 100개까지만 리뷰 검색이 가능하다. 따라서 실제 소비자가 경험할 수 있는 리뷰의 양과 크롤러로 수집된 리뷰의 양이 동일한 점을 토대로 <표 1>과 같이 수집된 데이터를 분석 데이터로 활용하고자 한다.

3.3 전처리

수집된 온라인 소비자 리뷰 데이터는 주로 텍스트 형태이기 때문에 전처리 과정을 통해 데이터 분석에 적합한 정형 형태로 변환하고자 한다. 먼저 단어 토큰화 단계에서는 제목과 리뷰내용의 단어를 모두 소문자로 변환한 뒤 WordTokenizer를 사용하여 단어를 구분하였다. 다음으로 텍스트 정규화 단계에서는 Rainbow에 정의된 불용어(stopword)를 기준으로 불용어를 제거하고 SnowballStemmer를 통해 어간 추출(stemming)을 실시하였다. 그 후, 문서 벡터화 단계를 통해 최소발생빈도를 기준으로 문서에 있는 키워드들을 벡터화하였다. 최소발생빈도는 모든 상품에 동일하게 분석 데이터의 20%에 해당하는 값으로 설정하였고, 최소발생빈도 기준을 만족하는 키워드가 특정 리뷰

에서 1회 이상 발생할 경우 그 값을 Y로 표시하고 그렇지 않은 경우 N으로 표시하였다. 여러 문서 벡터화 기법 중 최소발생빈도를 선택한 이유는 본 연구의 목적이 온라인 소비자 리뷰에 있는 특정 키워드의 존재 유무가 목표변수에 미치는 영향을 파악하는 것이기 때문이다. 여기에서 문서 벡터화로 표현된 키워드들 중 의미적으로 동일한 키워드를 병합하고 불필요한 키워드를 제거하는 작업은 수작업으로 진행하였다. 이렇게 전처리를 통해 벡터화로 표현된 키워드들은 분석에 사용할 초기 독립변수로 활용하였다.

본 연구에서는 온라인 소비자 리뷰의 별점은 자가평가를 그리고 좋아요는 타인평가를 나타내는 지표로 생각하고 이를 목표변수로 선택하였다. 목표변수의 값으로 별점은 Good(5점, 4점)과 Bad(3점, 2점, 1점)로 변경하였다. 이러한 기준으로 별점을 구분한 이유는 분석 상품들이 아마존에서 가장 인기가 높은 상품들이기 때문에 4점 또는 5점에 속한 별점이 많은 점과 다른 유사 연구들(임영서 등, 2020; 김준겸 등, 2021; 최혜선, 연규필, 2022)에서 본 연구와 같은 기준을 사용한 점을 들 수 있다. 또한 목표변수의 값으로 좋아요가 1개 이상인 경우는 Y로 그렇지 않으면 N으로 변경하여 분류모델 구축에 활

—의사결정나무를 활용한 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드 도출 연구: 별점과 좋아요를 중심으로

<표 2> 독립변수와 목표변수 정보

구분	상품명	초기 독립변수 수	목표변수(별점)의 클래스 수	
			Good	Bad
P1	Shark AI Ultra	64	200	300
P2	iRobot Roomba i4+EVO	15	150	211
P3	iRobot Roomba 676	17	200	240
구분	상품명	초기 독립변수 수	목표변수(좋아요)의 클래스 수	
			Y	N
P1	Shark AI Ultra	65	238	262
P2	iRobot Roomba i4+EVO	14	165	196
P3	iRobot Roomba 676	26	128	312

용하였다. 실험에 사용된 상품별 독립변수와 목표변수에 대한 내용을 요약하면 <표 2>와 같다.

3.4 데이터 균형화 및 변수 선택

일반적으로 목표변수 내의 특정 클래스에 속하는 비율이 서로 다를 경우, 관측 대상이 많은 클래스 방향으로 학습이 편향되어 이루어지기 때문에 특정 클래스만 잘 분류하는 분류모델이 만들어 질 수 있다(Chawla, 2009). 따라서 분류모델을 구축하기 전 목표변수에 있는 클래스의 비율을 맞추는 데이터 균형화(data balancing) 작업이 필요하다. 본 연구에서는 데이터 균형화 기법으로 랜덤 언더샘플링(random under-sampling) 기법을 사용하였다. 랜덤 언더샘플링 기법은 다수 클래스의 표본이 소수 클래스의 표본 크기가 될 때까지 랜덤으로 데이터를 제거하여 클래스 간의 표본 수를 동일하게 균형화시키는 기법이다. 그 결과 P1 상품 별점의 경우 400개(Good: 200, Bad: 200), 좋아요의 경우 476개(Y: 238, N: 238)로 데이터셋을 구성하였

다. 그리고 이렇게 구성된 데이터셋을 최종 데이터셋으로 선택하여 분석을 진행하였다. 나머지 P2 상품과 P3 상품에 대해서도 동일한 방법으로 데이터 균형화를 실시하였다.

다음으로 예측률을 높이기 위해 초기 독립변수를 구성하는 키워드들 중에서 분류모델의 구축에 필요 없는 키워드는 제거하는 변수 선택 작업이 필요하다(Dash and Liu, 1997). 본 연구에서는 이득비(gain ratio) 알고리즘을 사용하여 변수의 중요도를 평가하였고, 역방향 제거 방법을 활용하여 분류모델들을 구축하였다. 역방향 제거는 중요도가 낮은 순서대로 변수를 제거하는 방법이며, 구축된 분류모델 중에서 예측률이 가장 높은 모델에 사용된 변수를 최종 변수로 사용하고자 한다(Witten and Frank, 2005).

3.5 분류모델 구축

본 연구에서는 데이터 마이닝 분야에서 널리 사용되고 있는 의사결정나무 기법을 활용하여 별점과 좋아요 평가에 관한 분류모델을 구축하

고 그 결과를 바탕으로 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드와 규칙을 도출하고자 한다.

의사결정나무는 나무 형식으로 노드가 구성되는데 여기에서 노드는 독립변수로 구성되고, 가지는 독립변수의 값으로 표현된다. 이때 목표 변수의 클래스를 잘 구분하기 위한 조건에 따라 부모 노드는 여러 자식 노드로 분할되며 분할에는 자식 노드의 평균 순수도가 활용된다 (Witten and Frank, 2005). 일반적으로 의사결정나무를 기반으로 만들어진 분류 규칙은 사람이 이해하기 쉬운 형태로 만들어지고 해석력 또한 매우 우수하여 분류 결과에 대한 해석이 필요한 연구에서 널리 사용되고 있다.

본 연구에서는 웨카(Weka) 버전 3.8.6을 사용하여 의사결정나무 기반의 분류모델을 구축하였다. 실험에 사용된 데이터는 학습데이터 66%와 검증데이터 34% 비율로 분할하였고, C4.5기반의 의사결정나무를 활용하였다. 하이퍼파라미터를 조정하기 위해 잎 노드의 수는 전체 실험 데이터의 2.5% 수준으로 동일하게 설정하였고 나머지 파라미터 값들은 웨카에서 제공하는 기본 값을 사용하여 실험을 진행하였다.

전처리 후 선택된 초기 독립변수들을 중요도

가 낮은 변수 순으로 하나씩 제거해 가면서 분류모델의 성능을 평가하였다. 그 결과 각 상품의 온라인 소비자 리뷰에 대한 별점과 좋아요를 가장 잘 분류하는 모델을 발견할 수 있었다.

IV. 연구 결과

4.1 핵심 키워드 및 규칙 도출

본 연구에서와 같이 분류모델의 목표변수가 명목형인 경우, 분류모델의 성능은 예측률(hit ratio)로 평가된다. <표 3>은 예측률 계산에 사용되는 혼동 행렬(confusion matrix)을 보여준다. 이때의 예측률은 식(1)과 같으며 분류모델이 예측한 전체 클래스 수 중에서 정확히 예측한 클래스 수의 비율을 의미한다.

<표 4>에서 보는 바와 같이, P1 상품의 별점 분류모델에서는 의사결정나무를 26가지 중요 변수를 사용하여 구성하였을 때 가장 높은 예측률인 66.91%를 기록하였다. 이때의 별점 예측에 관한 규칙은 <그림 3>의 a)와 같다. P1 상품의 좋아요 분류모델은 의사결정나무를 4가지 중요 변수인 ‘house’, ‘make’, ‘run’, ‘update’를

<표 3> 혼동 행렬

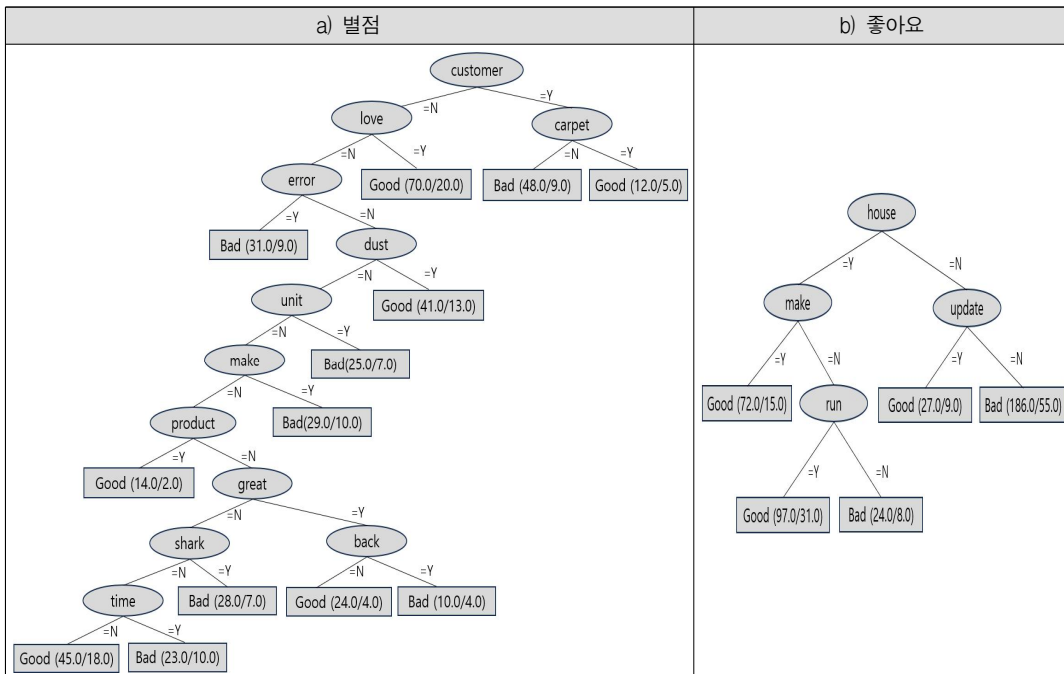
		예측	
		Good(Y)	Bad(N)
실제	Good(Y)	a	b
	Bad(N)	c	d

$$\text{적중률}(\%) = \frac{(a + d)}{(a + b + c + d)} \times 100 \quad \text{식(1)}$$

—의사결정나무를 활용한 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드 도출 연구: 별점과 좋아요를 중심으로

<표 4> 최고 예측률을 기록한 분류모델에 사용된 중요 변수

구분		중요 변수 (중요도순)	중요 변수 수	예측률
P1	별점	customer, support, service, job, love, error, return, dust, pet, unit, great, carpet, update, product, hair, months, back, pretty, bought, time, good, wifi, dirt, make, shark, doesn't	26	66.91%
	좋아요	house, make, run, update	4	69.75%
P2	별점	back, work, great, job, time	5	73.53%
	좋아요	vacuum, good	2	62.5%
P3	별점	love, floor, app, great, hair, job	6	71.32%
	좋아요	app, room	2	66.67%



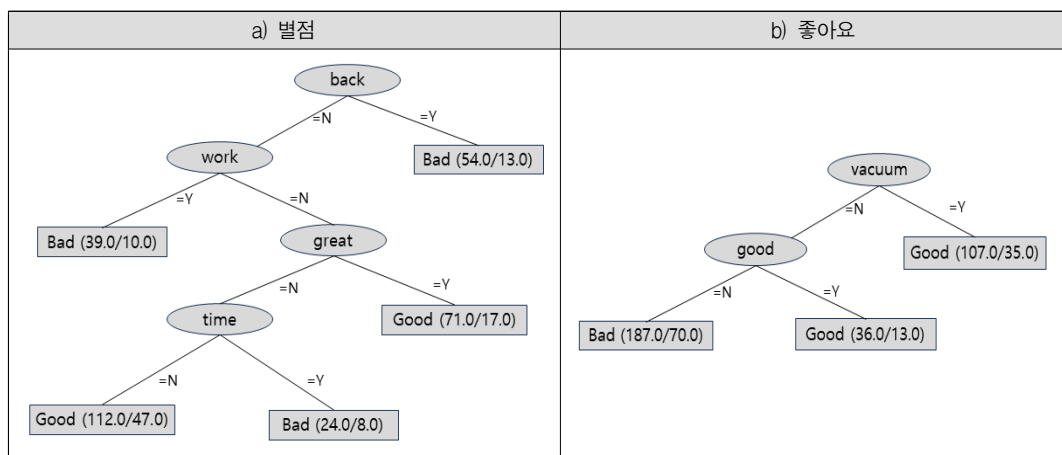
<그림 3> P1 상품에 대한 의사결정나무 규칙

사용하여 만들었을 때 가장 높은 예측률인 69.75%를 기록하였다. 이때의 좋아요 예측 규칙은 <그림 3>의 b)와 같다. 별점과 좋아요 예측에 관한 세부 규칙들을 구성하는 조건과 각 규칙의 예측률은 <표 5>에서 살펴볼 수 있다. 예를 들어, 규칙1에서 ‘customer가 Y이고 carpet이 N이다’인 조건을 만족하면 ‘Bad’로 예측하

는데, 이 규칙에 속한 클래스는 총 48개이며 그 중 39개가 실제 ‘Bad’에 속하여 예측률이 81%로 기록되었다. 규칙14에 속하는 클래스는 총 27개이며 그중 18개가 실제 ‘Good’에 속하여 예측률이 67%로 기록되었다. 나머지 규칙들도 동일한 방법으로 해석할 수 있다.

<표 5> P1 상품의 별점 및 좋아요 예측 규칙

구분	규칙	조건	예측결과		
			클래스	수	예측률
별점	1	customer=Y & carpet=N	Bad	48	81%
	2	customer=Y & carpet=Y	Good	12	58%
	3	customer=N & love=Y	Good	70	71%
	4	customer=N & love=N & error=Y	Bad	31	68%
	5	customer=N & love=N & error=N & dust=Y	Good	41	68%
	6	customer=N & love=N & error=N & dust=N & unit=Y	Bad	25	72%
	7	customer=N & love=N & error=N & dust=N & unit=N & make=Y	Bad	29	66%
	8	customer=N & love=N & error=N & dust=N & unit=N & make=N & product=Y	Good	14	86%
	9	customer=N & love=N & error=N & dust=N & unit=N & make=N & product=N & great=Y & back=N	Good	24	83%
	10	customer=N & love=N & error=N & dust=N & unit=N & make=N & product=N & great=Y & back=Y	Bad	10	60%
	11	customer=N & love=N & error=N & dust=N & unit=N & make=N & product=N & great=N & shark=Y	Bad	28	75%
	12	customer=N & love=N & error=N & dust=N & unit=N & make=N & product=N & great=N & shark=N & time=N	Good	45	60%
	13	customer=N & love=N & error=N & dust=N & unit=N & make=N & product=N & great=N & shark=N & time=Y	Bad	23	57%
좋아요	14	house=N & update=Y	Good	27	67%
	15	house=N & update=N	Bad	186	70%
	16	house=Y & make=Y	Good	72	79%
	17	house=Y & make=N & run=Y	Good	97	68%
	18	house=Y & make=N & run=N	Bad	24	67%



<그림 4> P2 상품에 대한 의사결정나무 규칙

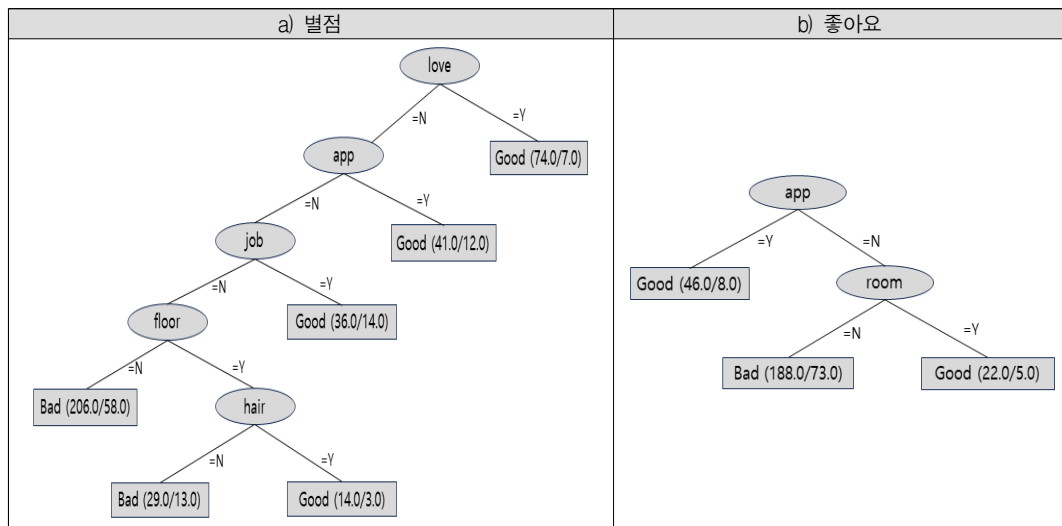
—의사결정나무를 활용한 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드 도출 연구: 별점과 좋아요를 중심으로

<그림 4>는 P2 상품의 별점과 좋아요 분류에 가장 높은 예측률을 보인 모델을 보여주고 있으며 해당 모델에 대한 세부 규칙은 <표 6>에서 확인할 수 있다. 예를 들어, <표 6>의 규칙 1에서 'back은 Y이다'인 조건을 만족하면 'Bad'로 예측하는데, 이 규칙에 속한 클래스는 총 54개이며 그중 41개가 실제 'Bad'에 속하여 예측률이 76%로 기록되었다. 규칙6에 속하는 클래스는 총 107개이며 그중 72개가 실제 'Good'에 속하여 예측률이 67%로 기록되었다.

또한 P3 상품에서 가장 높은 예측률을 보인 분류모델은 <그림 5>와 같고 그 세부 규칙을 요약하면 <표 7>과 같다. 예를 들어, <표 7>의 규칙1에서 'love는 Y이다'인 조건을 만족하면 'Good'으로 예측하는데, 이 규칙에 속한 클래스는 총 74개이며 그중 67개가 실제 'Good'에 속하여 예측률이 91%로 기록되었다. 규칙7에 속하는 클래스는 총 46개이며 그중 38개가 실제 'Good'에 속하여 예측률이 83%로 기록되었다.

<표 6> P2 상품의 별점 및 좋아요 예측 규칙

구분	규칙	조건	예측결과		
			클래스	수	예측률
별점	1	back=Y	Bad	54	76%
	2	back=N & work=Y	Bad	39	74%
	3	back=N & work=N & great=Y	Good	71	76%
	4	back=N & work=N & great=N & time=N	Good	112	58%
	5	back=N & work=N & great=N & time=Y	Bad	24	67%
좋아요	6	vacuum=Y	Good	107	67%
	7	vacuum=N & good=N	Bad	187	63%
	8	vacuum=N & good=Y	Good	36	64%



<그림 5> P3 상품에 대한 의사결정나무 규칙

<표 7> P3 상품의 별점 및 좋아요 예측 규칙

구분	규칙	조건	예측결과		
			클래스	수	예측률
별점	1	love=Y	Good	74	91%
	2	love=N & app=Y	Good	41	71%
	3	love=N & app=N & job=Y	Good	36	61%
	4	love=N & app=N & job=N & floor=N	Bad	206	72%
	5	love=N & app=N & job=N & floor=Y & hair=N	Bad	29	56%
	6	love=N & app=N & job=N & floor=Y & hair=Y	Good	14	79%
좋아요	7	app=Y	Good	46	83%
	8	app=N & room=N	Bad	188	61%
	9	app=N & room=Y	Good	22	77%

4.2 규칙 해석

먼저 상품별로 자가평가와 타인평가에 영향을 주는 핵심 키워드와 그들의 관계를 비교하고자 한다. P1 상품의 경우, 자가평가에서는 ‘customer’가 키워드 중 1순위가기 때문에 고객과 관련된 키워드가 가장 중요하다는 것을 알 수 있다. 총 400개의 리뷰 중 ‘customer’는 Bad에서는 44번, Good에서는 15번으로 별점이 높은 리뷰보다 별점이 낮은 리뷰에 약 3배 이상 포함되어 있다. 흥미로운 점은 별점을 Good이라고 평가하였음에도 고객서비스에 불만을 가진 리뷰가 많았다는 점이다. 이와 비교하여 타인평가에서는 ‘house’가 가장 중요한 키워드인 것을 파악할 수 있다. 476개의 리뷰 중 271개의 리뷰에서 ‘house’라는 단어가 포함되었으므로 절반이 넘는 소비자가 로봇 청소기를 집에서 사용하기 위하여 구입하였다는 것을 알 수 있다. 그리고 그중 좋아요라고 평가한 리뷰는 171개이다. 다음으로 자가평가에서 사용되는 핵심 키워드들을 살펴보면 ‘carpet’, ‘dust’와 같이 청

소 대상에 대한 특성과 ‘love’, ‘great’과 같은 감성 관련 특성으로 파악할 수 있다. 이와 비교하여 타인평가에서는 ‘update’, ‘run’과 같이 청소기 기능과 관련된 특성을 파악할 수 있다.

P2 상품의 경우, 자가평가에서 ‘back’이 가장 중요한 키워드로 나타나는데 소비자들은 충전기 연결 문제나 작동 문제와 같이 다양한 문제상황들을 경험하며 상품에 대한 불만과 실망을 표현하였다. 그렇기 때문에 환불을 요청하거나 이미 완료한 소비자들의 리뷰가 포함되어 있었다. 한편 타인평가에서는 ‘vacuum’이 키워드 중요도 1순위로 나타나는데 이를 통하여 청소기 본연의 기능을 비중 있게 평가한다는 것을 알 수 있다. 다음으로 자가평가의 핵심 키워드는 ‘work’, ‘time’과 같이 기능에 관련된 특성과 ‘great’과 같이 감성 관련 특성으로 파악할 수 있다. 이와 비교하여 타인평가는 ‘good’과 같이 감성 관련 특성으로 파악할 수 있다.

P3 상품의 경우 자가평가에서 ‘love’가 가장 중요한 키워드를 차지하였는데 별점이 높은 리뷰에서는 만족하였다는 의미로 사용되었지만,

별점이 낮은 리뷰에서는 처음에는 만족하였으나 시간이 지나면서 실망하였다는 의미로 사용되었다. 반면 타인평가에서는 ‘app’이 가장 중요한 키워드를 차지하였으므로 기능적 측면에 대하여 중요하게 여긴다는 것을 확인할 수 있었다. 다음으로 자가평가의 핵심 키워드들을 살펴보면 ‘app’, ‘job’과 같이 기능에 관련된 특성과 ‘floor’, ‘hair’와 같이 청소 대상에 대한 특성으로 파악할 수 있다. 반면 타인평가는 ‘room’과 같이 청소 대상에 대한 특성으로 파악할 수 있다.

상품별로 자가평가와 타인평가에 공통으로 등장하는 핵심 키워드의 경우, P1 상품의 ‘make’와 P3 상품의 ‘app’을 제외하고는 없는 것으로 나타났다. 따라서 ‘상품별로 자가평가와 타인평가에 영향을 주는 핵심 키워드는 서로 다르다’라는 연구 문제 1과 동일한 결과가 나온 것을 확인할 수 있다.

다음으로 전체 상품의 자가평가나 타인평가에 영향을 주는 핵심 키워드들의 관계를 비교하고자 한다. 자가평가에서는 공통으로 ‘great’이라는 감성 키워드만 포함되어 있는 것을 알 수 있다. 반면 타인평가에서는 공통된 키워드가 포함되어 있지 않다. 이를 통하여, ‘전체 상품의 경우 자가평가에 대한 핵심 키워드는 유사하며, 타인평가에 대한 핵심 키워드 또한 유사하다’는 연구 문제 2와는 다른 결과가 도출된 것을 확인할 수 있다.

추가로 분석한 내용을 요약하면 다음과 같다. 첫째, 의사결정나무를 구성하는 규칙의 경우 P1 상품의 별점에 관한 규칙이 13개인 것을 제외하고 나머지 상품들은 6개 이내의 규칙으로 구성되는 것을 알 수 있다. <표 1>을 살펴보면

P1 상품이 다른 상품들에 비하여 실제 소비자 리뷰의 수가 많다. 각 상품별 별점 또는 좋아요를 기준으로 100개 내외의 리뷰를 비슷하게 추출하였지만 P1 상품의 소비자는 좀 더 다양한 핵심 키워드를 사용하여 리뷰를 평가하는 것을 알 수 있다. 둘째, 기능적 측면으로 세 가지 상품을 비교하였을 때, P1 상품과 P2 상품은 ‘self empty’ 기능이 있고, P3 상품에는 해당 기능이 없다. 따라서 기능이 비슷한 P1 상품과 P2 상품의 키워드에 ‘self empty’와 관련된 키워드가 등장할 것으로 예상하였지만 핵심 키워드로 나타나지 않았다. 셋째, P2 상품과 P3 상품은 동일한 브랜드이기 때문에 자가평가와 타인평가에 유사한 핵심 키워드가 등장할 것으로 예상하였지만 이를 확인할 수 없었다. 따라서, 기능이 비슷하거나 동일한 브랜드여도 자가평가와 타인평가의 핵심 키워드는 동일하지 않다는 것을 알 수 있다.

지금까지의 연구결과를 통하여 자가평가와 타인평가에 영향을 주는 상품별 핵심 키워드는 서로 다르며, 전체 상품의 자가평가 핵심 키워드와 타인평가 핵심 키워드 역시 다르기 때문에 각각의 상품별로 맞춤형 세부 전략이 필요하다는 것을 알 수 있다. 또한 기능이 비슷하거나 동일한 브랜드여도 상품마다 소비자가 중요시하는 키워드가 다르기 때문에 생산자 및 마케팅 담당자들은 각 상품의 핵심 키워드들을 바탕으로 장점을 더 부각시키고 단점은 보완하여 소비자들로부터 긍정적인 평가를 얻는 방안을 마련해야 한다. 예를 들어, P1 상품은 자가평가 1순위 키워드인 ‘customer’를 고려하여 고객서비스를 개선하고, 타인평가 1순위 키워드인 ‘house’를 고려하여 집에서 사용할 때의 장

점을 부각시키는 활동을 한다면 소비자로부터의 긍정적인 평가를 얻을 수 있을 것으로 예상된다.

V. 결론

본 연구는 의사결정나무 기법을 활용하여 온라인 쇼핑몰인 아마존에서 판매 중인 로봇 청소기의 소비자 리뷰 중 자가평가와 타인평가에 영향을 미치는 핵심 키워드를 확인하였다. 그리고 이를 바탕으로 온라인 소비자 리뷰 평가에 영향을 주는 규칙을 도출하였다. 그 결과, 첫째, 자가평가와 타인평가에 등장하는 핵심 키워드들이 상품별로 상이하다는 것을 알 수 있었다. 이와 같이 동일한 상품에 대한 자가평가와 타인평가가 각각 다르게 이루어질 수 있기 때문에 생산자 및 마케팅 담당자들은 이러한 점을 심층적으로 파악하여 상품을 생산하고, 마케팅에 활용할 필요가 있을 것이다. 둘째, 기능이 비슷한 상품의 경우 관련 기능에 대한 키워드가 존재하지 않았으며 동일한 브랜드의 상품에서도 공통된 키워드가 존재하지 않는다는 것을 확인하였다. 따라서 기능이 비슷하거나 동일한 브랜드라도 각 상품에 대한 심도 있는 소비자 리뷰 분석이 매우 중요하다는 사실을 파악할 수 있었다.

본 연구는 온라인 소비자 리뷰 평가에 의사결정나무 기법을 활용한 첫 번째 연구로, 기존 연구와 다르게 리뷰의 키워드에 초점을 맞추어 자가평가와 타인평가를 분석하고, 핵심 키워드를 도출하였다는 점에서 의의가 있다. 이러한 연구를 바탕으로 향후 다양한 상품을 대상으로

하여 온라인 소비자 리뷰 분석, 즉 자가평가와 타인평가에 영향을 미치는 핵심 키워드를 도출하고 그 규칙을 마련하여 상품 생산과 마케팅에 도움이 될 수 있기를 바란다.

그러나 본 연구에서는 다음과 같은 한계점이 존재한다. 먼저 영어로 작성된 온라인 소비자 리뷰만을 분석 대상으로 고려하였기 때문에 보다 다양한 나라의 소비자 의견을 반영하지 못하였다. 또한 아마존에서 수집할 수 있는 온라인 소비자 리뷰가 별점을 기준으로 100개까지만 가능하였기 때문에 해당 상품에 대한 전체 온라인 소비자 리뷰를 분석하지 못하였다.

향후 연구에서는 보다 다양한 언어로 작성된 온라인 소비자 리뷰를 분석 데이터에 반영하고, 업데이트되는 상품에 관한 리뷰를 주기적으로 수집 및 분석하여 좀 더 일반화된 분석 결과를 도출하는 연구를 진행하고자 한다.

참고문헌

- 김광국, 김용환, 김자희, “사용자 리뷰 토픽분석을 활용한 모바일 쇼핑 앱 고객만족도에 관한 연구,” 한국전자거래학회지, 제23권, 제4호, 2018, pp. 41-62.
- 김근형, “연관성 모델에 기반한 오픈년마이닝 시스템의 설계 및 구현,” 한국정보통신학회논문지, 제15권, 제1호, 2011, pp. 133-140.
- 김영길, “A Study on Car Reviews Using Text Mining and Association Techniques,” 서울과학기술대학교 석사학위논문, 2018.

- 김윤승, 유동희, “머신러닝 기반 가치투자를 통한 주식 종목 선정 연구: 내재가치를 중심으로,” 정보시스템연구, 제32권, 제1호, 2023, pp. 179-199.
- 김윤호, 조재환, 장성원, 김혜란, 윤영호, “온라인 고객리뷰(OCR) 효과 결정요인에 관한 체계적 문헌연구,” 한국식품유통학회 하계학술대회, 2022, pp. 474-496.
- 김준겸, 최은솔, 윤수현, 이유빈, 김동환, “치킨 리뷰의 이면: 텍스트 마이닝을 통한 리뷰의 탐색적 분석을 중심으로,” 한국콘텐츠학회논문지, 제21권, 제11호, 2021, pp. 30-40.
- 김효곤, 유동희, “BERT를 활용한 미국 기업 공시에 대한 감성 분석 및 시각화,” 정보시스템연구, 제31권, 제3호, 2022, pp. 67-87.
- 야오즈옌, 박지영, 홍태호, “레스토랑의 온라인 리뷰를 통해 감성과 감정이 리뷰유용성에 미치는 영향에 관한 연구,” 지식경영연구, 제22권, 제1호, 2021a, pp. 243-267.
- 야오즈옌, Chernyaeva, Olga., 홍태호, “빅 데이터 분석을 이용한 온라인 리뷰유용성에 대한 감정적 특성과 정보적 특성의 영향,” 한국경영정보학회 학술대회논문집, 2021b, pp. 288-297.
- 연종흠, 이동주, 심준호, 이상구, “상품 리뷰 데이터와 감성 분석 처리 모델링,” 한국전자거래학회지, 제16권, 제4호, 2011, pp. 125-137.
- 윤호민, 최규완, “감성 분석을 활용한 레스토랑 선택속성과 만족도에 관한 연구: 외래 관광객 온라인 리뷰 중심으로,” 호텔경영학연구, 제30권, 제6호, 2021, pp. 99-117.
- 임영서, 이소영, 이지나, 류보경, 김현희, “온라인 고객 리뷰를 활용한 제품 효과 분석 기법,” 정보처리학회논문지, 소프트웨어 및 데이터 공학, 제9권, 제9호, 2020, pp. 259-266.
- 최혜선, 연구필, “밀키트 제품 리뷰 데이터를 이용한 텍스트 분석 사례 연구,” 한국콘텐츠학회논문지, 제22권, 제5호, 2022, pp. 1-14.
- 홍태호, 니우한잉, 임강, 박지영, “LDA를 이용한 온라인 리뷰의 다중 토픽별 감성 분석 - TripAdvisor 사례를 중심으로,” 정보시스템연구, 제27권, 제1호, 2018, pp. 89-110.
- 황호현, 이경찬, 유진이, 이영훈, “리뷰 데이터와 제품 정보를 이용한 멀티모달 감성 분석,” 한국전자거래학회지, 제27권, 제1호, 2022, pp. 15-28.
- Bei, L. T., Chen, E. Y., and Widdows, R., “Consumers’ Online Information Search Behavior and the Phenomenon of Search vs. Experience Products,” *Journal of Family and Economic Issues*, Vol. 25, 2004, pp. 449-467.
- Chatterjee, S., “Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach,” *International Journal of Hospitality Management*, Vol. 85, 2020, pp. 102356.
- Chawla, N. V., “Data Mining for Imbalanced

- Datasets: An Overview”, *Data Mining and Knowledge Discovery Handbook*, Springer: Boston, 2009.
- Choudhary, S., and Chhabra, C., “Sentiment Analysis of Amazon Food Review Data,” *2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, Sonapat, India, 2021, pp. 116-120.
- Dash, M., and Liu, H., “Feature selection for classification,” *Intelligent Data Analysis*, Vol. 1, No. 1, 1997, pp. 131-156.
- Jiang, H., Kwong, C. K., Park, W. Y., and Yu, K. M., “A multi-objective PSO approach of mining association rules for affective design based on online customer reviews,” *Journal of Engineering Design*, Vol. 29, 2018, pp. 381-403.
- Kausar, M. A., Fageeri, S. O., and Soosaimanickam, A., “Sentiment Classification based on Machine Learning Approaches in Amazon Product Reviews,” *Engineering, Technology & Applied Science Research*, Vol. 13, No. 3, 2023, pp. 10849-10855.
- Kim, W. Y., Ryu, J. S., Kim, K. I., and Kim, U. M., “A method for opinion mining of product reviews using association rules,” *In Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, 2009.
- Kwon, H. J., Ban, H. J., Jun, J. K., and Kim, H. S., “Topic Modeling and Sentiment Analysis of Online Review for Airlines,” *Information*, Vol. 12, 2021, pp. 78.
- Lee, S. and Choeh, J. Y., “Exploring the determinants of and predicting the helpfulness of online user reviews using decision trees,” *Management Decision*, Vol. 55, No. 4, 2017, pp. 681-700.
- Li, X., and Hitt, L. M., “Self Selection and Information Role of Online Product Reviews,” *Information Systems Research*, Vol. 19, No. 4, 2008, pp. 456-474.
- Luo, J. M., Vu, H. Q., Li, G., and Law, R., “Understanding service attributes of robot hotels: A sentiment analysis of customer online reviews,” *International Journal of Hospitality Management*, Vol. 98, 2021, pp. 103032.
- Ma, B., Zhang, D., Yan, Z., and Kim, T., “An LDA and Synonym Lexicon Based Approach to Product Feature Extraction from Online Consumer Product Reviews,” *Journal of Electronic Commerce Research*, Vol. 14, 2013, pp. 304-314.
- Mudambi, S. M., and Schuff, David., “What Makes a Helpful Online Review? A

—의사결정나무를 활용한 온라인 소비자 리뷰 평가에 영향을 주는 핵심 키워드 도출 연구: 별점과 좋아요를 중심으로

- Study of Customer Reviews on Amazon.Com,” *MIS Quarterly*, Vol. 34, No. 1, March 2010, pp. 185-200.
- Nelson, P., “Information and Consumer Behavior,” *Journal of Political Economy*, Vol. 78, No. 2, 1970, pp. 311-329.
- Noori, B., “Classification of Customer Reviews Using Machine Learning Algorithms,” *Applied Artificial Intelligence*, Vol. 35, 2021, pp. 567-588.
- Pranckevičius, T., and Marcinkevicius, V., “Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification,” *Baltic. J. Modern. Computing*, Vol. 5, No. 2, 2017, pp. 221-232.
- Ravi, K., and Ravi, V., “A survey on opinion mining and sentiment analysis: Tasks, approaches and applications,” *Knowledge-Based Systems*, Vol. 89, 2015, pp. 14-46.
- RaziaSulthana, A., and Ramasamy, S., “Context Based Classification of Reviews Using Association Rule Mining, Fuzzy Logics and Ontology,” *Bulletin of Electrical Engineering and Informatics*, Vol. 6, 2017, pp. 250-255.
- Trivedi, S. K., Singh, A., and Malhotra, S. K., “Prediction of polarities of online hotel reviews: an improved stacked decision tree (ISD) approach,” *Global Knowledge, Memory and Communication*, 2022.
- Witten, I. H., and Frank, E., “Data mining: Practical machine learning tools and techniques”, *Morgan Kaufmann Publishers*, San Francisco, 2005.
- Wu, Y. H., Li, W. Q., Tao, Z. J., and Cui, R., “Text Mining for Analyzing User Comments of APPs-Centered on Instagram and Facebook,” *In Proceedings of KIIT Conference*, December 2022, pp. 225-228.

민 경 수 (Min, Kyeong Su)



국민대학교에서 경영학사를 취득하였다. 현재 인공지능을 활용한 데이터 분석에 관심을 가지고 있으며 스타트업 창업을 준비하고 있다. 주요 관심분야는 지능형 정보시스템, 생성형 인공지능, 텍스트마이닝 등이다.

유 동 희 (Yoo, Dong Hee)



고려대학교 경영학사와 경영학 박사학위를 취득하였다. 현재 경상국립대학교 경영정보학과에서 교수로 재직하고 있으며, 주요 관심분야는 인공지능, 빅데이터 분석, 온톨로지, 지식 그래프, 지능형 정보시스템 등이다.

<Abstract>

Core Keywords Extraction for Evaluating Online Consumer Reviews Using a Decision Tree: Focusing on Star Ratings and Helpfulness Votes

Min, Kyeong Su · Yoo, Dong Hee

Purpose

This study aims to develop classification models using a decision tree algorithm to identify core keywords and rules influencing online consumer review evaluations for the robot vacuum cleaner on Amazon.com. The difference from previous studies is that we analyze core keywords that affect the evaluation results by dividing the subjects that evaluate online consumer reviews into self-evaluation (star ratings) and peer evaluation (helpfulness votes). We investigate whether the core keywords influencing star ratings and helpfulness votes vary across different products and whether there is a similarity in the core keywords related to star ratings or helpfulness votes across all products.

Design/methodology/approach

We used random under-sampling to balance the dataset. We progressively removed independent variables based on decreasing importance through backwards elimination to evaluate the classification model's performance. As a result, we identified classification models that best predict star ratings and helpfulness votes for each product's online consumer reviews.

Findings

We have identified that the core keywords influencing self-evaluation and peer evaluation vary across different products, and even for the same model or features, the core keywords are not consistent. Therefore, companies' producers and marketing managers need to analyze the core keywords of each product to highlight the advantages and prepare customized strategies that compensate for the shortcomings.

Keyword: Decision tree; Data mining; Online consumer reviews; Star rating, Helpfulness vote

* 이 논문은 2023년 8월 14일 접수, 2023년 9월 4일 1차 심사, 2023년 9월 12일 게재 확정되었습니다.