

# 머신러닝 기반 온라인 리뷰 감성 분석 모델링에 대한 연구

김민수 (한성대학교 IT공과대학 컴퓨터공학부 교수)\*

김주희 (동덕여자대학교 문화지식융합대학 교수)\*\*

## 국문 요약

온라인 리뷰는 시장 내에서의 기업의 가치를 평가하는 데 있어 중요한 역할을 하며, 기업의 수익에 큰 영향을 미치는 요인 중 하나이다. 따라서 온라인 리뷰의 감성 분석 지표는 사업의 성공을 예측할 수 있는 중요한 지표 중 하나이다. 본 연구에서는 대표적인 온라인 리뷰 플랫폼 중의 하나인 Yelp 플랫폼에 있는 레스토랑 리뷰 텍스트를 연구대상으로 선정하였고, Yelp Open Dataset에서 제공하는 리뷰 데이터 세트를 활용하였다. 본 연구에서는 레스토랑 리뷰의 Polarity Prediction을 위해 Logistic Regression, SVM, Random Forest, Gradient Boosting Machine(GBM), XGBoost, LightGBM 총 6가지 머신러닝 알고리즘을 사용하여 연구를 진행하였다. 각 모델의 성능평가 결과, Logistic Regression, SVM, LightGBM 알고리즘이 0.91로 가장 정확도가 높게 나타났다. 본 연구는 비정형화된 형태로 작성된 텍스트의 리뷰 데이터를 정량화하여 평점으로 예측할 수 있도록 하여 스타트업을 포함한 기업이 고객 피드백을 효과적으로 분석할 수 있도록 한다는 점에서 공헌점이 있다. 나아가 비즈니스 운영자들이 소비자 행동을 예측하고, 마케팅 전략 수립에 활용할 수 있는 유용한 인사이트를 제공할 수 있을 것으로 기대된다.

핵심주제어: 온라인리뷰, 머신러닝, 감성분석, 평점예측

## 1. 서론

정보통신기술의 발전과 더불어 디지털 경제 메커니즘이 확산되면서, 플랫폼 생태계는 이제 전 산업과 비즈니스를 구성하는 핵심 요소로 자리 잡았다. 예를 들어, 상거래나 서비스 분야에서 상품과 서비스의 공급자와 수요자가 거래하는 경제 활동이 플랫폼을 중심으로 이루어지는 플랫폼 경제가 빠르게 성장하였다. 제품을 판매하는 기업들은 이커머스 플랫폼을 주요 판매 채널로 활용하고 있으며, 고객들 역시 플랫폼을 주요 구매 경로로 선택하고 있기 때문이다(Mariani et al., 2023). 이처럼 기업과 고객을 포함한 주요 이해 관계자들이 플랫폼 생태계에 적극적으로 참여함으로써, 플랫폼은 현대 비즈니스 환경에서 필수적인 역할을 하고 있다. 특히, 코로나19로 인해 '비대면 언택트'가 사회문화적 트렌드로 자리 잡으면서, 플랫폼 경제가 더욱 빠르게 팽창할 수 있는 계기가 되었다(Choi & Leon, 2020).

디지털 경제의 확산과 함께 플랫폼의 중요성이 커지면서, 정보에 접근하고 활용하는 방식도 변화하고 있다. 특히, 고객들은 온라인 플랫폼을 통해 다양한 정보를 손쉽게 얻을 수 있게 되었으며, 이로 인해 정보의 유형과 특성에 대한 이해가 기업의 마케팅 전략과 소비자 의사결정에 중요한 영향을 미

치고 있다. 이러한 변화는 플랫폼을 중심으로 한 경제 활동의 성장뿐만 아니라, 고객들이 제품과 서비스에 대해 객관적 정보와 주관적 정보를 수집하여 보다 합리적인 판단을 내릴 수 있도록 돕는 역할을 하고 있다. 이에 따라, 정보의 특성과 이를 활용한 의사결정 과정에 대한 연구는 플랫폼 경제에서 더욱 중요해지고 있다.

한편, 정보의 유형에는 크게 두 가지가 있는데, 제품과 서비스의 객관적인 특성(예, 제품의 기능)을 제시하는 사실적 정보와 무형적인 특성(예, 제품에 대한 느낌)을 제시하는 주관적 정보가 존재한다(Deutsch & Gerard, 1955). 인터넷의 발전과 함께 온라인 플랫폼 기반의 여러 매체의 발달은 고객이 사실적 정보 이외에 이전 구매자의 사용 후기와 같은 주관적 정보에 보다 쉽게 접할 수 있는 환경을 조성하였다. Frederick et al.(2002)는 불확실한 상황에서 고객이 위험과 불확실성을 줄이기 위해 객관적인 정보(예: 양적 지표나 물리적 속성)와 개인적으로 느끼는 효용성 같은 주관적인 정보를 바탕으로 논리적인 판단 과정을 거친다고 하였다. 이들은 이러한 의사결정을 이중정보처리 모형(Dual-Process Model)으로 설명하면서, 비인적 정보(impersonal sources)와 달리 주관적 정보가 선호에 영향을 미쳐 구매 의사결정에 관여한다고 하였다.

Raju et al.(1995)는 객관적 지식, 주관적 지식, 그리고 사용 경험이 고객의 의사결정에 미치는 영향력을 살펴본 결과, 각

\* 주저자, 한성대학교 IT공과대학 컴퓨터공학부 교수, mskim@hansung.kr

\*\* 교신저자, 동덕여자대학교 문화지식융합대학 교수, kimjh@dongduk.ac.kr

· 투고일: 2024-09-20 · 수정일: 2024-10-14 · 게재확정일: 2024-10-21

각의 요소들은 서로 연관성이 있기는 하지만 독립적으로 결정 과정에 영향을 미친다는 것을 밝혀냈다. 온라인 플랫폼을 기반으로 하는 디지털 경제 패러다임에서 흥미로운 점은, 과거에는 유통이 어려웠던 ‘사용자 경험’과 ‘사용 후기’와 같은 주관적 정보의 접근성이 높아지면서 관련 정보의 유통이 활성화되었다는 것이다(나희경·이희우, 2016; 정화영·양영석, 2007).

제품과 서비스의 종류는 다양해지는 반면, 기능상의 차별점을 찾아보기가 어려워졌기 때문에 고객은 이전 고객의 경험과 사용 후기와 같은 정보를 기반으로 구매 결정을 하는 경향이 더욱 커졌다. 특히, 온라인 리뷰에는 사용자 경험과 관련된 여러 요소들의 평가가 담겨 있기 때문에 다양한 정보가 포함되어 있다는 점에서 그 의미가 더욱 크다고 할 수 있다(Thakur, 2018). 예를 들어, 레스토랑에 대한 평가에서 ‘맛’을 중요하게 생각하는 고객은 맛과 관련된 후기를 남기겠지만, ‘분위기’나 ‘서비스’를 중시하는 고객은 맛보다는 자신이 관심 있는 요소에 대해 평가할 것이다. 이러한 주관적 평가의 특성은 고객마다 중요하게 생각하는 측면이 다르다는 것을 의미하며, 이는 다음 고객들이 참고할 수 있는 다양한 정보를 생성하는 결과를 가져다준다. 온라인 리뷰는 기업입장에서도 중요한 정보의 원천이 된다. 온라인 리뷰에는 제품이나 서비스에 대한 평가가 담겨 있기 때문에, 고객의 행동을 탐색하고 시장의 반응을 조사하는 데에 중요한 자료가 되기 때문이다(Mariani et al., 2023).

하지만, 온라인 리뷰는 정보 비대칭을 해소하는 효과를 제공하면서도 동시에 정보 과부하로 인해 새로운 정보 탐색 비용이 발생하는 문제도 초래하고 있다. 고객이 플랫폼을 통해 다양한 측면의 정보를 얻을 수 있는 편리함이 있지만, 증가하는 온라인 리뷰의 증가로 인해 오히려 어떠한 리뷰가 원하는 정보(가령, 레스토랑에서 맛이나 가격, 분위기와 같은 요소)를 포함하고 있는지를 찾는 데에 시간이 소요되는 상황이 발생하게 된 것이다(Wang et al., 2023). 기업입장에서도 온라인 리뷰와 같은 비정형화된 자료의 증가는 방대한 데이터 분석 과정에서 많은 비용과 시간이 소요된다는 문제점을 야기하였다. Hong et al.(2017)은 리뷰의 길이와 내용의 모호성 등이 리뷰의 유용성에 영향을 미칠 수 있다고 하였으며, Mariani et al.(2023)는 길이가 긴 리뷰일수록 정보 가치로써의 유용성이 감소된다고 설명하였다.

최근에는 이러한 문제를 해결하기 위해 머신러닝을 활용하여 온라인 리뷰를 고객 평점으로 정량화함으로써 유용성을 높이려는 시도가 이루어지고 있다. Alslaity & Orji(2024)는 정량화된 평점을 통해 제품에 대한 평가 정보를 더 쉽게 파악할 수 있다고 하였다. 이러한 논의가 대두되면서, 정량화된 평점이 온라인 리뷰의 내용을 정확하게 반영하는지에 대한 연구가 중요한 이슈로 떠올랐다.

머신러닝 기법은 리뷰 텍스트를 평점으로 예측하는 모델을 개발함으로써 이 문제를 해결할 수 있는 방법을 제시하고 있다. 이에 본 연구는 머신러닝 알고리즘을 통해 온라인 리뷰를 평점으로 예측하는 모델을 제안하고자 한다. 이를 위해 Yelp

의 레스토랑 리뷰 데이터를 사용하여 감성 분석(Sentiment Analysis)을 진행하고, 고객 피드백을 효과적으로 분석해 예측 가능한 인사이트를 제공하는 데 기여하고자 한다.

구체적으로, ‘Support Vector Machine(SVM)’, ‘Logistic Regression’, ‘Random Forest’, ‘Gradient Boosting Machine(GBM)’, ‘XGBoost(eXtreme Gradient Boosting)’, ‘LightGBM(Light Gradient Boosting Machine)’의 6가지 머신러닝 알고리즘을 활용하여 예측 성능이 높은 모델을 탐색하고, 모델의 정확도를 개선하고자 한다. 이를 통해 고객 피드백 분석이 필요한 다양한 산업 분야에서 활용 가능한 정교한 감성 분석 모델 개발에 기여할 수 있을 것이다.

## II. 이론적 배경

### 2.1. 온라인 리뷰와 정보 비대칭 해소

선행연구에서는 온라인 리뷰가 정보 비대칭을 해소하고 유용성을 높이는 데에 중점을 두고, 다음과 같은 주제들을 다루고 있다.

첫째, 정보 경제학의 맥락에서 온라인 리뷰의 유용성을 다룬 연구들이 있다(Majumder et al., 2022). 정보 경제학적 관점에서 온라인 고객 리뷰는 정보 비대칭 문제를 해소하고, 소비자들이 더 나은 구매 결정을 내릴 수 있도록 돕는 중요한 역할을 한다. 정보 비대칭 상황에서는 고객이 제품의 품질이나 대안에 대한 완전한 정보를 얻기 어렵기 때문에, 정보 비대칭에서 오는 한계를 보완하기 위해 많은 시간과 노력이 필요하다(Stigler, 1961). 또한, 정보의 부족은 구매 과정에서 불확실성을 증가시켜(Nelson, 1970) 소비자들이 결정을 내리는 데 어려움을 겪게 만드는데, 온라인 리뷰는 이러한 불확실성을 상당 부분 해소할 수 있기 때문에 정보 비대칭 문제를 해결하는데 효과적이다. Shukla & Mishra(2023)는 온라인 리뷰는 고객 입소문이 형성되는 메커니즘의 핵심 요소로, 잠재 고객들이 구매 결정을 내릴 때 중요한 정보 원천으로 작용한다고 설명한다. 즉, 고객들은 구매 결정을 내리는 데 어려움을 겪을 때 다른 고객들의 리뷰를 통해 유용한 정보를 얻으려는 행동 패턴을 보이는데(Majumder et al., 2022; Choi & Leon, 2020), 소비자들이 공급업체가 제공하는 정보보다 기존 고객이 작성한 리뷰를 더 신뢰하는 경향이 나타나는 경우도 종종 발생한다(Lopez-Lopez & Parra, 2016).

이처럼 유용한 리뷰는 소비자들이 구매 결정을 내릴 때 중요한 도구가 되는데, 이러한 과정을 ‘단서-합산 이론’에 따르면 고객들은 이러한 단서를 바탕으로 자신의 지식을 확대하고 결정을 내리게 된다고 한다(Severin, 1967). 또한, Jahoda(1959)는 사회적 영향력이 개인의 행동에 중요한 영향을 미친다고 하였는데, 고객의 사용 경험이 담긴 리뷰는 신규 고객의 구매 결정에 사회적 영향력을 지니는 요소로 작용하게 된다(Cialdini & Goldstein, 2009). Deutsch & Gerard(1955)가 제안한 이중 과정 이론에 따르면, 사회적 영향력은 정보적 영향력과

규범적 영향력으로 나눌 수 있으며, 경험이 담긴 고객의 리뷰는 규범적 사회적 영향력으로 분류된다(Latane, 1981). 따라서 규범적 요소를 지닌 온라인 리뷰는 잠재적 구매자에게 강력한 사회적 영향을 미치며, 이러한 영향력은 소비자들의 온라인 리뷰를 통해 구매 결정을 내리는 과정에서 중요한 역할을 한다.

둘째, 빅데이터 분석 방법론적인 측면에서 고객의 구매 의도, 만족도, 제품 선호도 등을 예측하기 위하여 리뷰 텍스트 분석을 통해 제품의 장단점과 서비스 품질 수준을 정량적으로 평가하는 것이 중요하다는 것을 강조하는 연구들이 있다. 감성 분석 및 기계 학습을 위한 연구에서는 다양한 유형의 알고리즘과 방법론이 개발되었다. 예를 들어, Tripathy & Rath(2017)는 단일 주제 텍스트 리뷰를 분류하고 긍정 또는 부정 평가를 표현하기 위해 영화 리뷰 데이터셋에서 품사 태그를 사용하는 새로운 알고리즘인 Sentiment Fuzzy Classification을 제안하여 분류 정확도를 높이고자 하였다. Singla et al.(2017)는 소비자 제품 리뷰에 대한 통계 및 감성 분석을 수행하였으며, 온라인 리뷰 데이터를 긍정과 부정의 두 가지 범주로 필터링하여 사용자 리뷰를 분석하고자 하였다.

이와 같은 선행 연구들을 통해서 온라인 리뷰가 고객에게는 의사결정 과정에서 발생할 수 있는 불확실성을 감소시켜주며, 기업에게는 고객 행동을 예측하고 제품 및 서비스 평가의 신뢰성을 높이며, 마케팅 전략과 추천 시스템 개발에 유용한 데이터를 제공하는 데 기여하고 있다는 것을 알 수 있다.

## 2.2. 머신러닝 기반 감성 분석

최근 연구에서는 머신러닝 알고리즘을 활용한 감성 분석이 활발히 이루어지고 있다. Jemai et al.(2021)은 SVM과 Random Forest가 특히 높은 정확도를 보이며 감성 분석에 효과적임을

입증하였다. Naive Bayes는 계산 속도가 빠르다는 장점이 있지만, 데이터의 특성에 따라 알고리즘 성능이 달라질 수 있음을 밝혔다. Veena et al.(2021)은 VADER 알고리즘을 통해 감성 분석 정확도를 0.84까지 향상시켰다. VADER는 문장의 감성 강도와 극성을 세밀하게 측정하는 기능을 가지고 있어 독특한 강점을 지닌 것으로 평가된다. Yelp review data를 활용한 연구도 활발히 이루어지고 있다.

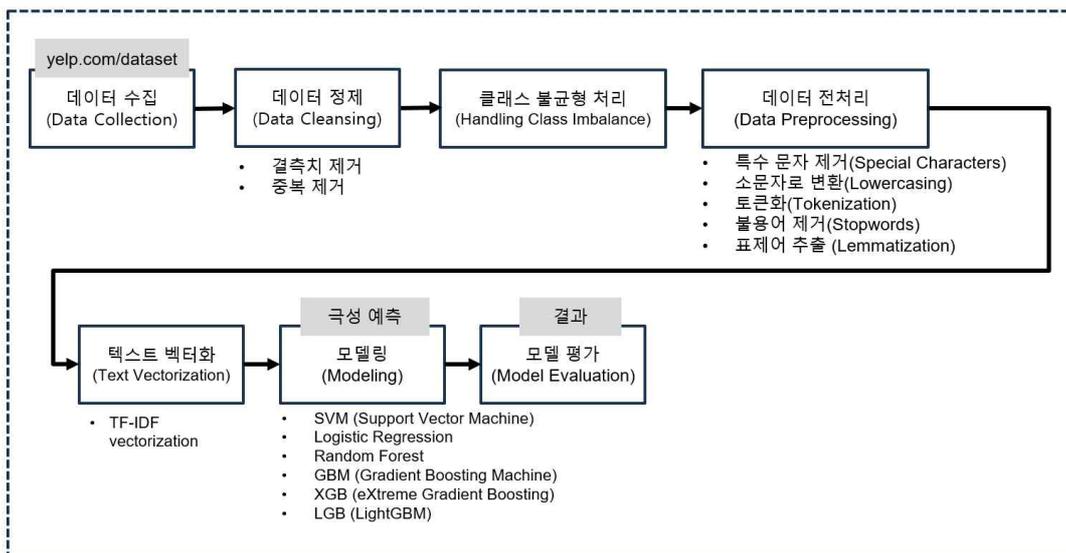
Xu et al.(2015)은 Yelp 리뷰 데이터를 활용해 Naive Bayes와 Multiclass SVM을 비교한 결과, Naive Bayes가 대규모 텍스트 분류에 더 적합하다는 결론을 내리며, 데이터 특성에 따른 알고리즘 선택의 중요성을 강조하였다. 또한, Hemalatha et al.(2019)은 Yelp 리뷰 데이터를 활용한 감성 분석 연구에서 Naive Bayes가 단순하면서도 강력한 예측력을 보여주었으나, Linear SVC가 특정 상황에서 더 나은 성능을 발휘하는 것을 밝혔다.

이와 같은 연구들은 머신러닝 알고리즘의 선택이 데이터의 특성과 연구 목적에 따라 달라질 수 있음을 시사한다. 본 연구에서는 이러한 알고리즘들을 활용해 온라인 리뷰 데이터를 분석하고, 클래스 균형 조정 및 텍스트 전처리와 파라미터 튜닝을 통해 가장 높은 정확도를 보이는 모델을 탐색하는 것을 목표로 한다.

## III. 연구 방법

### 3.1. 연구 프레임워크

본 연구에서는 디지털 플랫폼 보완자를 위한 리뷰 감성 분석 모델링을 위한 연구 절차를 <그림 1>과 같이 진행하였다.



<그림 1> 연구 프레임워크

### 3.2. 데이터 수집 및 데이터 전처리 과정

#### 3.2.1. 데이터 수집

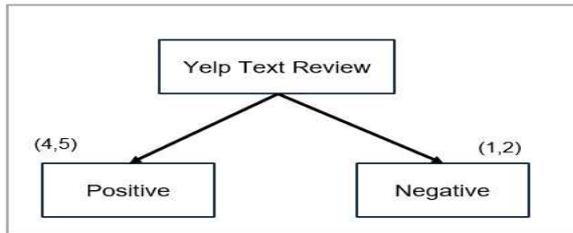
Yelp.com은 레스토랑 리뷰를 공유하는데 널리 사용되는 대표적인 온라인 플랫폼이다. 본 연구에서는 Yelp에서 제공하는 Open Dataset(<https://www.yelp.com/dataset>)을 활용하였으며, 이 데이터 셋은 전 세계 다양한 비즈니스에 대한 리뷰 데이터를 제공한다. 이 중 600만 건 이상의 리뷰 데이터에서 연구 목적에 맞게 1만 건을 무작위로 추출하여 분석에 사용하였다. Yelp.com의 리뷰 형태는 <그림 2>와 같다.



<그림 2> Yelp 온라인 리뷰 사례

#### 3.2.2. 클래스 균형 조정(Balancing Classes)

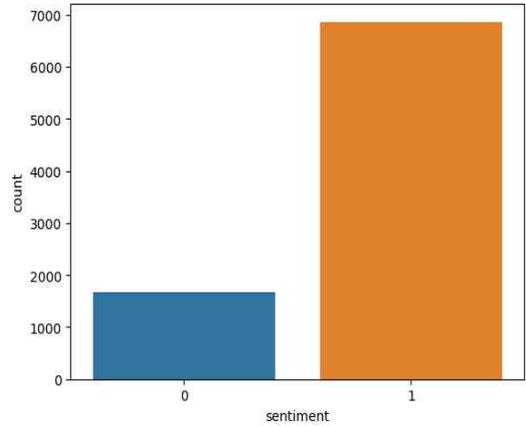
클래스 균형을 맞추기 위해 먼저 리뷰 텍스트의 감성 분류와 라벨링을 <그림 3>과 같이 진행하였다. 각 리뷰의 평점에 따라 긍정 감성 클래스와 부정 감성 클래스로 구분하였으며, 3점은 중립으로 간주해 분석에서 제외하였다.



<그림 3> 리뷰 텍스트의 감성 분류와 라벨링

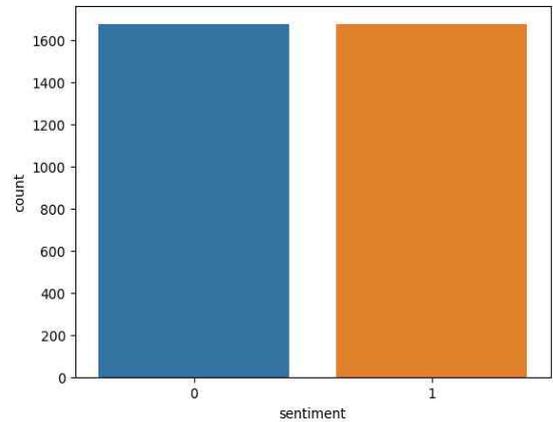
- 긍정 감성 (Label 1): 4점과 5점
- 부정 감성 (Label 0): 1점과 2점

불균형한 클래스 비율은 많은 기계 학습 작업, 특히 분류 문제에서 중요한 단계이다. 한 클래스가 다른 클래스에 비해 대표성이 부족할 경우, 모델은 편향될 수 있으며 다른 다수의 클래스를 예측하는 데 어려움을 겪어 소수 클래스에 대한 성능이 떨어질 수 있다. 이를 해결하기 위해 클래스 비율을 균형 있게 조정하여 <그림 5>와 같이 클래스 간 균형을 맞추었다.



(긍정: 1, 부정: 0)

<그림 4> 클래스 비율 조정 전 데이터셋의 감성 비율 분포



(긍정: 1, 부정: 0)

<그림 5> 클래스 비율 조정 후 데이터셋의 감성 비율 분포

#### 3.2.3. 텍스트 데이터 전처리

텍스트 데이터는 그 자체로는 머신러닝 알고리즘에 적용하기 어렵기 때문에 적절한 형태로 전처리하는 과정이 필수적이다. 본 연구에서는 Spacy 라이브러리를 활용하여 리뷰 데이터에 대한 전처리를 수행하였다. 구체적으로 다음의 단계를 거쳐 데이터를 정제하였다.

첫째, **Removing Special Characters**를 수행하였다. 텍스트 데이터에는 일반적으로 분석에 불필요한 특수 문자가 포함될 수 있다. 이러한 특수 문자는 텍스트의 의미 분석에 기여하지 않으며, 오히려 모델의 성능을 저하시킬 수 있다.

둘째, **Lowercasing**를 수행하였다. 대소문자 구분은 텍스트 분석에서 불필요한 혼란을 초래할 수 있다. 따라서, 모든 텍스트를 소문자로 변환하여 데이터의 일관성을 유지하는 것이 중요하다. 텍스트 데이터 내에서 중복되는 단어가 단일 형태로 통합되어 분석 효율성을 향상시킬 수 있다.

셋째, **Tokenization**을 수행하였다. 이는 텍스트 데이터를 개별 단어 또는 문장 단위로 분리하는 과정이다. 문장을 단어 단위로 쪼개는 것은 텍스트 분석에서 필수적인 과정으로, 머신러닝 알고리즘이 단어 단위로 데이터를 처리할 수 있게 해준다.

넷째, Stop Words Removal를 수행하였다. Stop Words는 "the"와 같이 분석에 큰 의미가 없는 단어들을 의미한다. 텍스트 내에서 자주 등장하지만, 문맥상 중요한 정보를 제공하지 않기 때문에 이를 제거함으로써 중요한 정보에 집중할 수 있도록 한다.

다섯째, Lemmatization을 수행하였다. Lemmatization은 단어를 그 기본 형태 어근으로 변환하는 과정이다. 단어의 변형된 형태를 어근으로 통합함으로써 데이터의 일관성을 높이고, 분석 대상의 핵심적인 의미를 더 잘 반영할 수 있게 한다.

### 3.2.4. 텍스트 벡터화

텍스트 데이터를 머신러닝 알고리즘에 활용하기 위해 TF-IDF(Term Frequency-Inverse Document Frequency)를 사용해 수치형 데이터로 변환하였다. TF-IDF는 단어의 빈도와 문서 내에서의 중요도를 동시에 반영해 각 단어의 가중치를 계산하는 방법이다. 이 방법을 통해 텍스트에서 중요한 키워드를 추출하고, 벡터 형태로 변환하여 머신러닝 모델의 입력으로 사용하였다.

## 3.3. 모델링 (Modeling)

본 연구에는 Yelp의 온라인 리뷰를 평점으로 예측하기 위하여 다음의 'Support Vector Machine(SVM)', 'Logistic Regression', 'Random Forest', 'Gradient Boosting Machine(GBM)', 'XGBoost(eXtreme Gradient Boosting)', 'LightGBM(Light Gradient Boosting Machine)'의 총 6개의 머신러닝 알고리즘을 활용하여 모델링을 진행하였다.

### 3.3.1. Support Vector Machine(SVM)

SVM은 지도 학습 알고리즘 중 하나로, 주어진 데이터를 특정 클래스에 분류하는 문제에 사용된다. SVM은 데이터를 고차원 공간으로 매핑하여 클래스 간의 가장 큰 마진을 찾는 최적의 결정 경계를 찾는 알고리즘이다. 기본 목표는 마진, 즉 초평면과 각 클래스의 가장 가까운 점 사이의 거리를 최대화하는 것이다. 새로운 데이터 포인트가 주어지면, 분류기는 이 데이터 포인트의 위치를 기준으로 최적의 초평면에 따라 특정 클래스에 할당한다.

### 3.3.2. Logistic Regression

로지스틱 회귀는 이름에 '회귀'라는 단어가 포함되어 있지만, 분류 문제를 해결하기 위한 알고리즘이다. 이 알고리즘은 선형 회귀를 기반으로 하되, 그 결과값을 로지스틱 함수(Sigmoid 함수)를 통과시켜 0과 1 사이의 값을 가지도록 한다. 결과적으로, 이진 분류 문제에 주로 사용되며, 확률적인 예측을 제공한다.

### 3.3.3. Random Forest

Random Forest는 여러 개의 결정 트리를 조합하여 만들어진 앙상블 학습 방법 중 하나이다. 각 트리는 부트스트랩 샘플을 사용하여 훈련되며, 트리의 분기에서는 임의의 속성 집합이 선택된다. 예측 시, 모든 트리의 예측을 집계하여 최종 결과를 도출한다. 이 방법은 과적합을 방지하고 일반적인 성능을 향상시킨다.

### 3.3.4. Gradient Boosting Machine (GBM)

GBM은 약한 학습기들을 순차적으로 학습시키며, 이전 학습기의 오차를 줄이는 방향으로 새로운 학습기를 추가하는 방식으로 동작한다. 주로 결정 트리를 기반으로 한다. 각 단계에서의 손실 함수의 그래디언트를 사용하여 오차를 보정하며, 성능을 향상시킨다.

### 3.3.5. XGBoost (eXtreme Gradient Boosting)

XGBoost는 GBM의 효율성과 속도를 개선한 알고리즘이다. 병렬 처리 기능, 교차 검증 내장, 누락된 값 처리 등의 다양한 기능을 제공한다. 또한, 규제를 포함하여 과적합을 방지하는 데 도움을 준다.

### 3.3.6. LightGBM (Light Gradient Boosting Machine)

LightGBM은 XGBoost와 유사하게 그래디언트 부스팅 프레임워크를 기반으로 하지만, 더 빠르고 메모리 사용량이 적다. 트리의 성장 방식이 깊이 기반(deep)이 아닌 리프 기반(leaf-wise)으로 진행되며, 더 큰 데이터셋에 효과적이다. 하지만 적은 데이터셋에서는 과적합될 가능성이 높아진다.

## 3.4. 모델 평가 (Evaluation)

### 3.4.1 평가 지표

기계 학습 모델의 성능 평가는 다양한 방법들이 존재한다. 본 연구에서는 Confusion matrix를 이용하여 모델을 평가하였다. Confusion matrix는 분류 모델의 성능을 평가하기 위한 대표적인 도구로, 실제 값과 모델이 예측한 값을 비교하여 분류 성능을 시각적으로 확인할 수 있다. 해당 내용은 <표 1>에 제시되어 있으며, 아래와 같이 4가지 요소로 구성된다.

True Positive(TP)은 모델이 실제로 양성인 데이터를 올바르게 양성으로 예측한 경우이고, True Negative(TN)는 모델이 실제로 음성인 데이터를 올바르게 음성으로 예측한 경우이다. False Positive(FP)는 실제로는 음성이지만 모델이 잘못해서 양성으로 예측한 경우 (Type I Error)이고, False Negative(FN) 실제로는 양성이지만 모델이 잘못해서 음성으로 예측한 경우 (Type II Error)를 뜻한다.

<표 1> Confusion matrix

		Predicted	
		Positive	Negative
Observed	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

이러한 confusion matrix의 4가지 요소에 기반하여 분류 모델의 성능 지표들을 계산할 수 있다. 본 연구에서는 Accuracy, Precision, Recall, F1 score를 활용하였는데, 이 지표들의 계산 방식은 <표 2>에 기술되어 있다. Accuracy는 전체 데이터 중 올바르게 분류된 데이터의 비율을 나타낸다. Precision은 예측된 양성 데이터 중 실제로 양성인 데이터의 비율을 나타내며, Recall은 실제 양성 데이터 중에서 얼마나 많은 데이터가 양성으로 올바르게 예측되었는지를 나타낸다. F1 Score는 Precision과 Recall의 조화평균이다.

<표 2> 모델 성능 평가 지표

Metric	Formula
Accuracy	$TP+TN / (TP+FP+TN+FN)$
Precision	$TP / (TP+FP)$
Recall	$TP / (TP+FN)$
F1 Score	$2 * (Precision * Recall) / (Precision + Recall)$

#### IV. 실증분석 결과

데이터 전처리 후 머신러닝 모델의 성능을 극대화하기 위해서는 각 모델에 적합한 하이퍼파라미터(hyperparameter)를 선정하고, 각각의 파라미터를 최적화하는 과정이 필요하다. 하이퍼파라미터는 모델 학습 과정을 제어하는 중요한 변수로, 모델 유형에 따라 설정이 달라진다. <표 3>은 본 연구에서 사용한 각 모델의 하이퍼파라미터와 그에 대한 최적 설정값이다.

<표 3> 하이퍼파라미터(Hyperparameter) 설정

Model	Hyperparameter	Range	Result of tuning
Logistic Regression	C	0.01, 0.1, 1, 10, 100	10
	solver	'newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'	'saga'
SVM	C	0.1, 1, 10	1
	gamma	'scale', 'auto'	'scale'
	kernel	'linear', 'rbf', 'sigmoid'	'linear'
Random Forest	n_estimators	100, 400, 1000	100
	max_depth	4, 6, 10	10
	min_samples_split	2, 5, 10	10
	min_samples_leaf	1, 4, 10	1

GBM	max_depth	4, 6, 10	4
	learning_rate	0.1, 0.15, 0.2	0.1
	n_estimators	100, 400, 1000	1000
XGB	max_depth	4, 6, 10	4
	learning_rate	0.1, 0.15, 0.2	0.1
	n_estimators	100, 400, 1000	1000
LGB	max_depth	4, 6, 10	4
	learning_rate	0.1, 0.15, 0.2	0.1
	n_estimators	100, 400, 1000	400

<표 4>는 본 연구에서 수행한 감성분석 성능평가 실험결과를 요약한 것이다. 이 실험은 Yelp 리뷰 데이터를 기반으로 다양한 머신러닝 모델의 예측 성능을 비교하고, 하이퍼파라미터 튜닝과 3-Fold 교차 검증(3-Fold Cross Validation)을 통해 최적의 모델을 선정하기 위해 진행하였다. 교차 검증은 모델 성능의 신뢰성을 높이기 위해 사용된 방법으로, 데이터셋을 3개의 서브셋(folds)으로 나누어 각 폴드에서 반복적으로 학습과 평가를 수행하는 방법이다. 이를 통해 모델의 일반화 성능을 검증할 수 있다. 또한, Accuracy, Precision, Recall, F1 Score 값은 가중 평균(weighted average)을 적용하여, 불균형한 데이터에서도 성능평가가 정확하게 반영되도록 실험을 진행하였다.

<표 4> 하이퍼파라미터 튜닝 후 3-Fold 교차 검증의 결과

Algorithm	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.91	0.91	0.91	0.91
SVM	0.91	0.91	0.91	0.91
Random Forest	0.82	0.85	0.82	0.74
GBM	0.89	0.89	0.89	0.88
XGB	0.90	0.90	0.90	0.90
LightGBM	0.91	0.91	0.91	0.90

(Accuracy, Precision, Recall, F1 score는 weighted avg의 값을 표기)

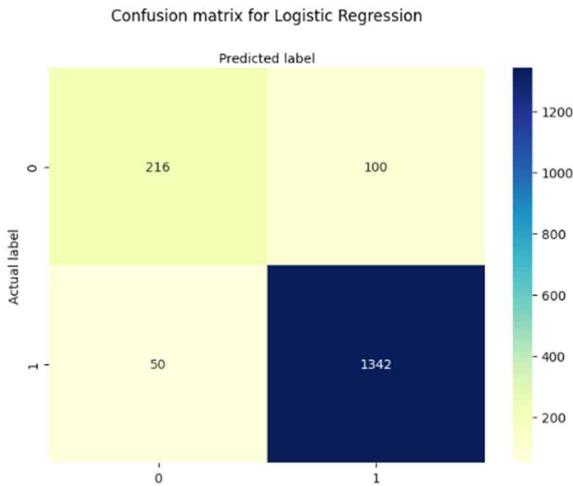
<그림 6>~<그림 11>은 본 연구에서 사용한 6개의 머신러닝 알고리즘의 Confusion matrix를 시각화한 것이다. 이를 통해, 각 알고리즘이 예측한 결과를 실제 레이블과 비교해 봄으로써, 해당 알고리즘이 얼마나 정확히 분류했는지, 오류가 발생한 패턴이 무엇인지 시각적으로 파악하고자 하였다.

#### 4.1. Logistic Regression, SVM, LightGBM

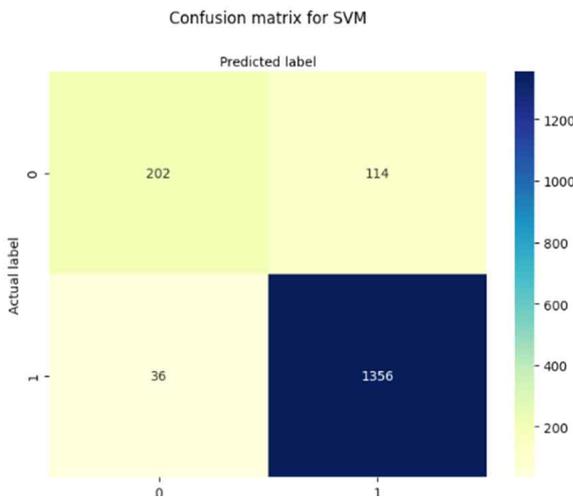
이 세 가지 알고리즘은 0.91의 Accuracy를 기록하며 가장 높은 성능을 보여주었다. Precision, Recall, F1 Score 모두 0.90~0.91 사이의 일관된 성능을 나타냈으며, 이는 양성 과 음성 클래스에 대해 균형 잡힌 예측 성능을 의미한다. <표 4>와 <그림 6>~<그림 8>에서 해당 실험결과를 확인할 수 있다.

실행 속도 측면에서는, Logistic Regression과 SVM이 LightGBM보다 빠르게 실행이 완료되었다. 이를 통해 컴퓨팅 자원

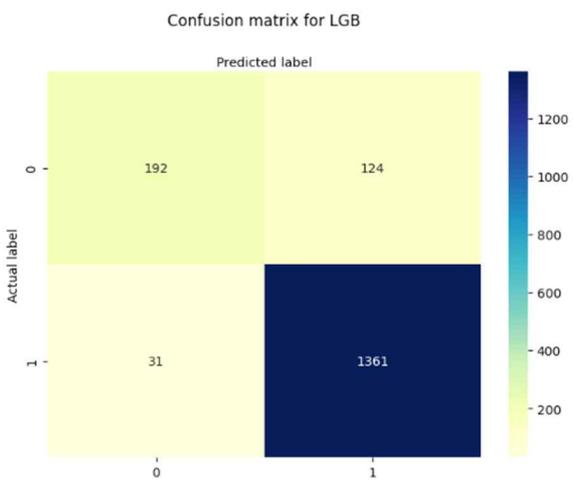
의 활용성 측면에서 Logistic Regression, SVM을 사용하는 것이 유리하다는 것을 확인하였다.



<그림 6> Confusion matrix for Logistic Regression



<그림 7> Confusion matrix for SVM

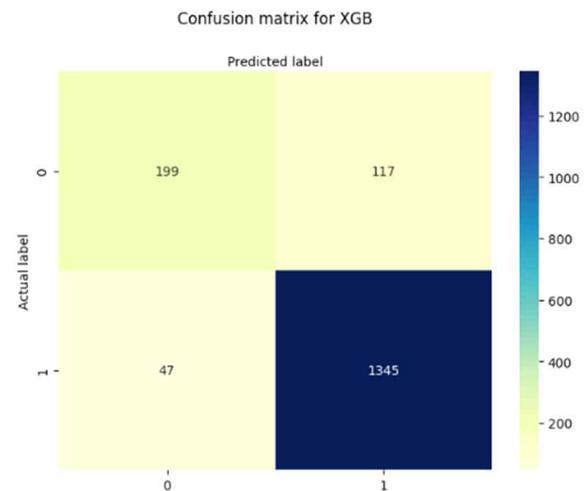


<그림 8> Confusion matrix for LGB

## 4.2 XGBoost (XGB)

XGBoost는 그 다음으로 좋은 성능을 보였으며, 0.90의 Accuracy와 함께 다른 주요 지표에서도 0.90의 일관된 성능을 보였다(<표 4>). 이는 XGBoost가 양성과 음성 클래스 모두에서 안정적인 예측을 수행했음을 의미한다.

<그림 9>를 보면, False Negative (FN) 값이 47로 나타나 일부 양성 클래스를 음성으로 잘못 예측한 경향이 있었다. 또한, False Positive (FP) 값은 117로 나타나, 음성 클래스(0)를 양성(1)으로 오분류한 사례도 다소 발생하였다. 이와 같은 결과는 Logistic Regression(100 FP)과 비교했을 때는 FP 값이 조금 더 높은 수준인 것을 확인할 수 있다.

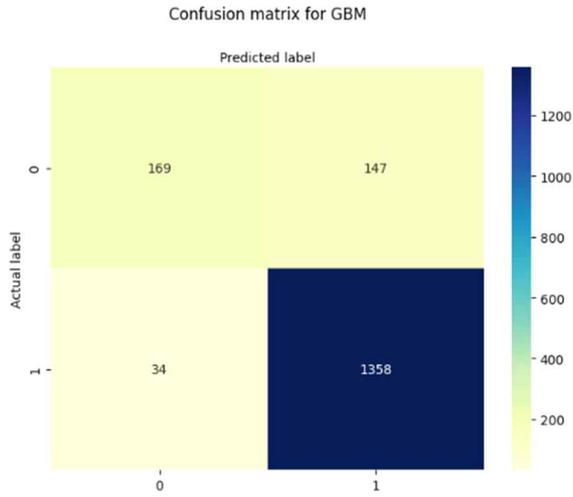


<그림 9> Confusion matrix for XGB

## 4.3. Gradient Boosting Machine (GBM)

GBM은 0.89의 Accuracy를 기록하였으며, Precision, Recall, F1-score 지표 역시 0.88~0.89 범위 내의 성능을 보였다(<표 4>). <그림 10>을 보면, False Positive (147)와 False Negative (34)의 비율이 4.1~4.2에서 분석했던 다른 알고리즘에 비해 상대적으로 높게 나타난 것을 알 수 있다.

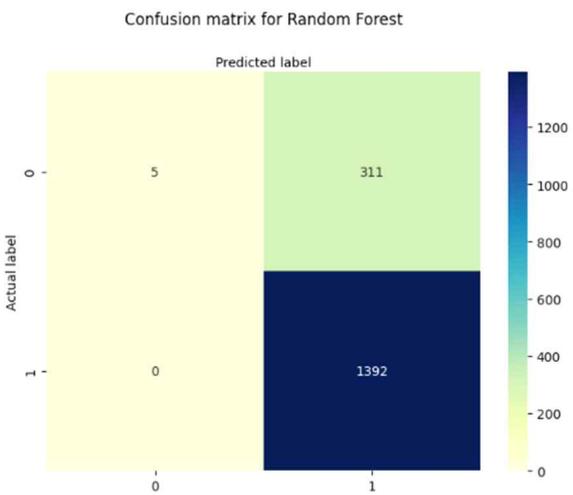
이와 같은 결과는 GBM이 양성과 음성 클래스 모두에서 예측의 일관성에 어려움이 있다는 것을 의미한다. 특히, False Positive 비율이 4.1~4.2의 다른 알고리즘(Logistic Regression, SVM, XGB, LGB)보다 높아 음성 클래스를 정확하게 분류하는 데에 어려움이 있었음을 확인할 수 있다.



<그림 10> Confusion matrix for GBM

#### 4.4. Random Forest

Random Forest의 Accuracy는 0.82로, 다른 알고리즘에 비해 비교적 낮은 성능을 기록했다. Precision은 0.85로 높았으나, F1 Score(0.74)가 낮아 예측의 일관성에서 한계가 있었다. False Positive (FP) 값이 311로 높게 나왔는데, 이것은 음성 클래스를 양성으로 잘못 예측한 비율이 높다는 것을 의미하며, 모델이 음성 클래스(0)를 제대로 구분하지 못하고 있다는 보여준다. <표 4>와 <그림 11>에 해당 실험결과가 나타나 있다.



<그림 11> Confusion matrix for Random Forest

### V. 결론 및 시사점

플랫폼 경제가 활성화되고 전 산업 부문에서 플랫폼과 앱 기반의 소비행태가 보편화되면서, 고객의 구매 경험이 텍스트 리뷰로 축적되고 있다. 이러한 리뷰는 기업에게는 제품과 서비스의 품질에 대한 고객들의 평가와 중요 속성에 대한 파악을 하기 위한 도구로써, 소비자에게는 구매 결정에 도움이 되

는 정보로써의 의미를 지니고 있다. 하지만, 고객의 텍스트 리뷰 자체는 비정형화된 데이터의 형태를 지니고 있으며 또한 특정 플랫폼마다 다른 형태로 남겨지기 때문에 고객과 시장 반응의 예측하는 지표로 삼기에는 많은 어려움이 있었다 (Taherdoost & Madanchian, 2023). 이에 텍스트 리뷰를 정량화된 지표인 평점으로 예측하고, 그 예측의 정확도를 높이기 위한 분석들이 활발하게 진행되어 왔다.

본 연구는 온라인 리뷰 데이터를 활용하여 감성 분석 모델링을 통해 고객 피드백을 정량적으로 평가하고 예측하는 방법을 제안하였다. Logistic Regression, SVM, LightGBM 등의 머신러닝 알고리즘을 사용하여 극성 예측 모델을 구축하였으며, 그 결과 Logistic Regression과 SVM 알고리즘이 높은 정확도를 보였다. 또한, hyperparameter tuning과 cross-validation을 통해 모델의 성능을 향상시켰으며, binary classification에서는 높은 정확도를 보인 반면, multiclass classification에서는 상대적으로 낮은 정확도를 보이는 한계점을 확인하였다. 이러한 결과는 다양한 감성 분석 모델의 특성과 성능을 비교하고, 각 알고리즘의 장단점을 파악하는 데 기여하였다.

온라인 리뷰 분석을 통해 기업은 고객의 요구와 시장의 변화를 빠르게 파악할 수 있으며, 긍정적인 리뷰를 유도하거나 부정적인 리뷰를 효과적으로 관리함으로써 고객 만족도를 높일 수 있다. 본 연구의 결과는 온라인 리뷰를 통해 기업이 고객 피드백을 보다 체계적으로 이해하고, 이를 바탕으로 마케팅 전략을 수립하는 데 실질적인 도움을 줄 수 있는 모델을 제시하는 데에 유용하게 활용될 수 있을 것이다. 특히, 본 연구에서 제안한 감성 분석 모델은 영화, 도서, 음악 등 다양한 분야에서 평판 분석과 개인화된 추천 시스템 개발에 활용될 수 있다는 점에서 의의가 있다.

본 논문의 연구 결과는 신생 스타트업에게도 시사점을 제공한다. 신생 스타트업은 창업 이후, 초기 비즈니스 모델 수립 시 끊임없이 고객 가설을 검증하고, 고객과 시장 반응을 측정하고 학습해 나가고 이를 토대로 고객을 만족시킬 수 있는 가치를 제안하는 비즈니스 모델을 구축하는 과정을 겪게 된다(Bortolini et al., 2021). Blank, & Eckhardt(2023)은 ‘고객개발(Customer Development)’을 통해 스타트업의 경우 고객을 발굴하고 입증하는 단계가 중요하다고 설명하였다. 특히, 고객개발의 첫 번째 단계인 고객발굴에서 고객에 대한 가설을 수립한 후, 고객의 반응을 확인하는 것이 필요하다고 하였다. 이때 고객의 피드백을 정량화된 수치로 예측할 수 있다면 가설 검증도의 정확도를 높일 수 있을 것이다. 평점 예측 모델은 ‘린 스타트업’에서도 유용하게 활용될 수 있다. 린 스타트업 모형은 빠른 사이클로 제품이나 서비스를 ‘구축(Build)’한 다음, 고객 반응을 ‘측정(Measure)’하고, 그 결과를 학습(Learn)’하는 과정을 반복함으로써 시장에서 통하는 비즈니스 모델을 찾아내는 피드백 프레임워크를 지니고 있다(Ries, 2011).

초기 스타트업은 최소 기능 제품을 선보이고 제품이나 서비스가 시장에 나왔을 때, 고객들이 어떻게 반응하는 지를 측정하고 이를 바탕으로 학습한다. 이 과정을 통해 획득한 정보는

제품이나 서비스를 개선하는 데에 활용되는데, 고객과 시장의 반응을 왜곡하지 않고 정확하게 파악할수록 고객에게 높은 가치를 제공하는 비즈니스 모델을 구축할 수 있다(Frederiksen, & Brem, 2017; Shepherd & Gruber, 2021). 제품과 서비스에 대해 다양한 관점에서 평가한 텍스트정보를 정량화된 점수인 평점으로 예측할 수 있는 머신러닝 모델은 고객이 중요시하는 가치를 스타트업이 어느 정도 제공하고 있는지 검증할 수 있는 지를 유효한 지표로 제공할 수 있다는 점에서 중요한 역할을 한다. 또한, 평점 예측 모델링은 제품과 서비스에 대하여 고객이 인지하고 평가하는 품질 속성을 파악할 수 있게 하여 고객 이탈을 방지하는 방안을 마련할 수 있다는 점에서도 장점이 있다.

이처럼 본 연구 결과는 6가지 머신러닝 알고리즘을 사용하여 예측 모델의 정확도를 향상시키고 이를 통해 레스토랑 분야 뿐만 아니라 다양한 산업 분야에서 고객 피드백을 분석할 수 있는 발판을 마련하였다는 점에서 의의가 있다. 또한 벤처 스타트업이 초기 아이디어의 타당성과 고객 가설을 효과적으로 검증할 수 있는 수단을 제공하였다는 점과 고객들이 평가하는 제품과 서비스 품질을 정량화하여 파악할 수 있기 때문에 시장 니즈를 파악하여 경쟁사와의 차별화 방안을 수립하고 경쟁 우위를 확보하기 위한 전략을 마련하는 데에도 도움이 된다는 점에서 실무적 시사점이 있다.

향후 연구에서는 더욱 정교한 감성 분석을 위해 다양한 데이터 전처리 기법과 추가적인 머신러닝 알고리즘을 적용하여 모델의 성능을 개선할 필요가 있다. 또한, 온라인 리뷰 데이터 외에도 소셜 미디어나 사용자 생성 콘텐츠와 같은 다양한 데이터 소스를 통합하여 고객의 행동과 선호도를 더 깊이 이해할 수 있는 연구가 필요하다. 이를 통해, 다양한 산업 분야에서 실질적인 인사이트를 제공하고, 기업의 경쟁력을 강화할 수 있는 모델을 개발하는 데 기여할 수 있을 것이다. 마지막으로, 머신러닝 분석시 연구 설계 과정에서 최근 이슈가 되고 있는 데이터의 개인정보보호와 가짜 리뷰를 고려하는 것이 필요하다. 이를 위해 데이터 수집 및 분석 과정에서 익명화와 데이터 최소화 원칙을 적용할 수 있는 방법과 가짜 리뷰를 필터링할 수 있는 방안 모색에도 노력을 기울이는 것이 필요하다.

## REFERENCE

나희경·이희우.(2016). 린 스타트업 방법론의 적용: 한국 '카다' 사례를 중심으로: 한국 '카다' 사례를 중심으로. *벤처창업연구*, 11(5), 29-43.

정화영·양영석.(2007). 창업기업의 비즈니스 모델 타당성 평가방안의 이론적 고찰. *벤처창업연구*, 2(2), 1-22.

Alslaity, A., & Orji, R.(2024). Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. *Behaviour & Information Technology*, 43(1), 139-164.

Blank, S., & Eckhardt, J. T.(2023). The lean startup as an

actionable theory of entrepreneurship. *Journal of Management*, 1-15.

Bortolini, R. F., Nogueira Cortimiglia, M., Danilevicz, A. D. M. F., & Ghezzi, A.(2021). Lean Startup: a comprehensive historical review. *Management Decision*, 59(8), 1765-1783.

Choi, H. S., & Leon, S.(2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems*, 139, 1-12.

Cialdini, R. B., & Goldstein, N.(2009). Normative influences on consumption and conservation behaviors. *Social Psychology of Consumer Behavior*, 273-296.

Deutsch, M., & Gerard, H. B.(1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51(3), 629-636.

Frederiksen, D. L., & Brem, A.(2017). How do entrepreneurs think they create value? A scientific reflection of Eric Ries' Lean Startup approach. *International Entrepreneurship and Management Journal*, 13, 169-189.

Frederick, S., Loewenstein, G., & O'donoghue, T.(2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2), 351-401.

Hemalatha, S., & Ramathmika, R.(2019). Sentiment analysis of yelp reviews by machine learning. In 2019 *International Conference on Intelligent Computing and Control Systems (ICCS)* (pp. 700-704). IEEE.

Hong, H., Xu, D., Wang, G. A., & Fan, W.(2017). Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems*, 102, 1-11.

Jahoda, M.(1959). Conformity and independence: A psychological analysis. *Human Relations*, 12(2), 99-120.

Jemai, F., Hayouni, M., & Baccar, S. (2021, June). Sentiment analysis using machine learning algorithms. In 2021 *International Wireless Communications and Mobile Computing (IWCMC)* (pp. 775-779). IEEE.

Latane, B.(1981). The psychology of social impact. *American psychologist*, 36(4), 343.

Lopez-Lopez, I., & Parra, J. F.(2016). Is a most helpful eWOM review really helpful? The impact of conflicting aggregate valence and consumer's goals on product attitude. *Internet Research*, 26(4), 827-844.

Majumder, M. G., Gupta, S. D., & Paul, J.(2022). Perceived usefulness of online customer reviews: A review mining approach using machine learning & exploratory data analysis. *Journal of Business Research*, 150, 147-164.

Mariani, M. M., Borghi, M., & Laker, B.(2023). Do submission devices influence online review ratings differently across different types of platforms? A big data analysis. *Technological Forecasting and Social Change*, 189, 1-12.

Nelson, P.(1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311-329.

Raju, P. S., Lonial, S. C., & Mangold, W. G.(1995). Differential effects of subjective knowledge, objective

- knowledge, and usage experience on decision making: An exploratory investigation. *Journal of Consumer Psychology*, 4(2), 153-180.
- Ries, E.(2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Currency.
- Severin, W.(1967). Another look at cue summation. *AV Communication Review*, 15(3), 233-245.
- Shepherd, D. A., & Gruber, M.(2021). The lean startup framework: Closing the academic-practitioner divide. *Entrepreneurship Theory and Practice*, 45(5), 967-998.
- Shukla, A., & Mishra, A.(2023). Role of review length, review valence and review credibility on consumer's online hotel booking intention. *FIIB Business Review*, 12(4), 403-414.
- Singla, Z., Randhawa, S., & Jain, S.(2017). Sentiment analysis of customer product reviews using machine learning. In *2017 international conference on intelligent computing and control (I2C2)* (pp. 1-5). IEEE.
- Stigler, G. J.(1961). The economics of information. *Journal of Political Economy*, 69(3), 213-225.
- Taherdoost, H., & Madanchian, M.(2023). Artificial intelligence and sentiment analysis: A review in competitive research. *Computers*, 12(2), 37-52.
- Thakur, R.(2018). Customer engagement and online reviews. *Journal of Retailing and Consumer Services*, 41, 48-59.
- Tripathy, A., & Rath, S. K.(2017). Classification of sentiment of reviews using supervised machine learning techniques. *International Journal of Rough Sets and Data Analysis*, 4(1), 56-74.
- Veena, G., Vinayak, A., & Nair, A. J.(2021). Sentiment Analysis using Improved Vader and Dependency Parsing. *2021 2nd Global Conference for Advancement in Technology (GCAT)*. IEEE.
- Wang, F., Du, Z., & Wang, S.(2023). Information multidimensionality in online customer reviews. *Journal of Business Research*, 159, 1-15.
- Xu, Y., Wu, X., & Wang, Q.(2015). Sentiment analysis of yelp's ratings based on text reviews. *2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 17(1).

# A Study on Machine Learning-Based Modelling of Online Review Sentiment Analysis

Minsu Kim\*  
Juhee Kim\*\*

## Abstract

Online reviews play a crucial role in assessing a company's market value and are a significant factor influencing profitability. As such, sentiment analysis of online reviews has emerged as a key indicator for predicting business success. This study focuses on restaurant reviews from Yelp, one of the leading online review platforms, utilizing the Yelp Open Dataset. Six machine learning algorithms were applied to predict the sentiment polarity of these reviews: Logistic Regression, Support Vector Machine (SVM), Random Forest, Gradient Boosting Machine (GBM), XGBoost, and LightGBM. Performance evaluations demonstrated that Logistic Regression, SVM, and LightGBM achieved the highest accuracy, with a score of 0.91. The primary contribution of this study is its ability to transform unstructured review text into quantifiable data, enabling businesses, especially startups, to effectively analyze customer feedback and predict ratings. These insights are expected to assist business owners in forecasting consumer behavior and developing strategic marketing approaches.

*KeyWords: Online review, Machine learning, Sentiment analysis, Restaurant valuation, Review score prediction*

---

\* First Author, Professor, Hansung University, mskim@hansung.kr

\*\* Corresponding Author, Professor, Dongduk Women's University, kimjh@dongduk.ac.kr