감성과학
제27권 1호, 2024
<연구논문>

pISSN 1226-8593
eISSN 2383-613X
Sci. Emot. Sensib.,
Vol.27, No.1, pp.69-80, 2024
https://doi.org/10.14695/KJSOS.2024.27.1.69

소셜 미디어 앱 리뷰에서의 감성 분석 연구: 인스타그램 중심으로

Research on Sentiment Analysis in Social Media App Reviews: Focusing on Instagram

이문기[1] · 우위항[2†]

Wen-Qi Li[1] · Yu-Hang Wu[2†]

Abstract

This study aimed to gain valuable insights into the performance and user satisfaction of applications (apps) through a thorough analysis of Instagram user reviews collected from Google Play. The study utilized text mining and sentiment analysis techniques and systematically identified emotions and opinions embedded in user reviews to deeply understand the areas of improvement and user experiences of the app. It analyzes how Instagram reviews reflect the diverse experiences of users and how they reveal the strengths and weaknesses of the app. For this purpose, sentiment analysis using the naive Bayes algorithm was conducted, and the results were expected to aid in the improvement of Instagram's services. In addition, the study aimed to assist developers in better understanding and utilizing user feedback, ultimately contributing to enhanced user satisfaction. This study explored the complex relationship between social media usage patterns and user opinions by seeking ways to provide a better user experience through these insights.

Key words: Sentiment Analysis, Social Media Analysis, Naive Bayes Algorithm, Instagram User Reviews, Text Mining

요 약

본 연구는 Google Play에서 수집된 Instagram 사용자 리뷰에 대한 심층 분석을 통해, 이 연구는 애플리케이션의 성능과 사용자 만족도에 대한 중요한 통찰력을 얻고자 한다. 텍스트 마이닝과 감성 분석 기술을 활용하여 사용자 리뷰에 담긴 감성과 의견을 체계적으로 파악하며, 이를 통해 앱의 개선점과 사용자 경험을 깊이 이해하려고 한다. 인스타그램 리뷰가 사용자들의 다양한 경험을 어떻게 반영하는지, 그리고 앱의 장단점을 어떻게 드러내는지를 분석한다. 이를 위해 나이브 베이즈 알고리즘을 사용한 감성 분석을 수행하며, 이 결과는 인스타그램 서비스 개선에 도움이 될 것으로 기대된다. 연구는 또한 개발자들이 사용자 피드백을 더 잘 이해하고 활용하는 데 도움을 주며, 결국 사용자 만족도를 향상시키는 데 기여할 것으로 예상된다. 이 연구는 소셜 미디어 사용 패턴과 사용자 의견의 복잡한 관계를 탐색하고, 이를 통해 더 나은 사용자 경험을 제공하는 방안을 모색한다.

주제어: 감성 분석, 소셜 미디어 분석, 나이브 베이즈 알고리즘, 인스타그램 사용자 리뷰, 텍스트 마이닝

[1] 이문기: 부경대학교 미디어케뮤니게이션학과 박사과정

[2] † (교신저자) 우위항: 부경대학교 미디어케뮤니게이션학과 박사과정 / E-mail: wuyuhang@naver.com / TEL: 0504-340-9059

1. 서론

소셜 미디어와 온라인 마켓의 부상에 따라 사용자 리뷰와 피드백은 브랜드와 개발자에게 중요한 통찰력을 제공하는 핵심 자원이 되었다. Google Play와 같은 대규모 앱 스토어에서는 풍부한 사용자 리뷰가 연구자와 개발자에게 애플리케이션의 성능 및 사용자 만족도에 대한 귀중한 연구 기회를 제공한다. 이러한 리뷰들의 감성 정보를 깊게 파악하기 위해서는, 최신의 텍스트 분석 기법을 도입하는 것이 필요하다.

Instagram 은 전 세계적으로 인기 있는 소셜 미디어 앱 중 하나로, 그 사용자들의 피드백은 단순한 앱의 성능과 기능 이상의 광범위한 소셜 미디어 사용 패턴과 그들의 다양한 요구를 반영한다. 이런 맥락 속에서, 본 연구는 Google Play 스토어의 Instagram 사용자 리뷰를 중심으로 깊이 있게 수행되었다. 다양한 플랫폼에서의 사용자 리뷰와 피드백은 감성과 관련된 많은 정보와 의견이 포함되어 있으며, 이를 통해 사용자의 기대와 만족도, 그리고 애플리케이션의 단점 및 개선점에 대한 인사이트를 얻을 수 있다. 특히 Instagram 과 같은 인기 있는 소셜 미디어 애플리케이션의 경우, 사용자들의 다양한 경험과 의견이 리뷰에 집약되어 있기 때문에, 이를 깊게 분석하는 것은 매우 중요하다.

그럼에도 불구하고, 이러한 대규모의 텍스트 데이터를 수동으로 분석하는 것은 현실적으로 불가능하다. 따라서, 기계 학습 및 자연어 처리 기술의 도입이 필요하며, 이를 통해 대규모 데이터에서 유의미한 통찰력을 추출하고 분석할 수 있다.

인스타그램 사용자 리뷰에 대한 본 연구의 목적은 나이브 베이즈 알고리즘을 활용하여 감성 분석을 수행하고 사용자 평가에서 나타나는 주요 키워드의 빈도와 알고리즘의 예측 빈도를 비교 분석하는 것이다. 이는 인스타그램의 서비스 개선과 사용자 만족도 향상에 대한 깊은 이해를 도모하기 위함이다. 나이브 베이즈 알고리즘과 사용자 리뷰 사이의 연관성을 분석하는 데 있어서의 이론적 배경과 실제 응용 사례를 탐구한다. 본 연구의 비교 분석은 감성 분석의 정확성을 높이는 동시에 인스타그램과 같은 소셜 미디어 플랫폼에 대한 이해를 심화시키는데 기여할 것이며, 소셜 미디어 분석의 향후 연구 방향과 실용적 응용에 중요한 정보를 제공할 것이다. 본 연구가 제공하는 분석은 소셜 미디어 사용 패턴과 사용자 의견의 복잡한 관계를 탐색하고 이를 통해, 사용자 경험을 극대화하고 앱의 기능을 사용자의 요구에 맞추어 개선하는 데 필요한 세부 사항을 식별할 수 있게 될 것이다. 이러한 방법으로, 연구 결과는 개발자들에게 사용자 기반의 피드백을 분석하고 응용하는 데 도움을 줄 수 있으며, 사용자 만족도를 높이는 데 직접적으로 기여할 수 있을 것이다.

본 연구에서는 나이브 베이즈 알고리즘을 도입하여 Instagram 사용자 리뷰의 감성 분석을 실시하였으며, 그 결과는 Instagram의 개발과 최적화를 위한 중요한 참고 자료로 활용될 수 있을 뿐만 아니라, 텍스트 마이닝 및 감성 분석 연구 분야에서 사용자 의견과 반응의 정확한 분류와 이해를 통해, 이러한 데이터가 가지는 의미와 가치를 깊이 있게 파악하고자 하다. 이를 통해, 소셜 미디어 플랫폼에서의 사용자 경험 개선, 제품 및 서비스의 품질 향상, 그리고 효과적인 커뮤니케이션 전략 수립에 기여할 수 있는 실질적인 방안을 제시하고자 하는 것이 본 연구의 목적이다. 이 연구의 구성은 다음과 같다.

제1장 서론에서는 연구의 배경, 목적 및 중요성을 설명하며, Instagram 사용자 리뷰의 의미와 그 분석의 중요성에 대해 강조하였다.

제2장은 연구의 이론적 배경을 다루며 텍스트 마이닝 및 감성 분석의 기본 개념, 나이브 베이즈 분석 방법을 설명하고, 인스타그램 리뷰를 사용한 감성 분석의 예를 제시한다.

제3장의 연구 설계 및 방법론에서는 인스타그램 사용자 리뷰 데이터의 수집 및 전처리 방법을 다룬다. 연구에서는 Instagram 리뷰를 수집, 정제하고, 사용자 평점에 기반하여 감성을 분류한 후, 나이브 베이즈 알고리즘을 사용하여 분석한다.

제4장 분석 결과는 사용자 평가를 기반으로 한 단어 빈도와 단어 클라우드를 통해 주요 키워드를 도출했고, 나이브 베이즈 분류기를 이용한 결과를 비교하며, 사용자 평가와 예측 사이의 일치도와 차이를 평가했다.

제5장 결론 및 논의에서, 인스타그램 이용자 평가 데이터를 활용한 감성 분석의 목적과 중요성을 정리하고 강조했다. Kaggle 플랫폼에서 수집한 데이터를 기반으로 전처리 과정을 진행하고, 텍스트 마이닝 기법을 적

용하여 사용자 리뷰 내의 주요 감성 단어를 시각화하는 방법론을 소개하고 이를 통해 소셜 미디어 데이터 분석을 통한 사용자 감성 이해와 이를 기반으로 한 의사결정 지원 시스템 개발에 기여할 수 있는 방법을 탐구하다. 이러한 연구 과정과 결과를 종합하여, 소셜 미디어 감성 분석의 이론적 및 실용적 응용 가능성을 탐색하고, 향후 연구 방향에 대한 제언을 제공하다.

2. 이론적 배경

2.1. 텍스트 마이닝과 감성 분석

다수의 연구에서 텍스트 마이닝 및 자연어 처리(NLP) 기술을 활용하여 소셜 미디어 텍스트 데이터로부터 감성을 추출하는 다양한 방법들이 탐구되었다. 이러한 연구들은 소셜 미디어 데이터 분석의 깊이를 더하고, 텍스트 기반 감성 인식의 정확성을 향상시키는데 중요한 기여를 하였다. 텍스트 마이닝을 활용한 미디어 의제 분석에 대한 연구로 정지원 외는 신문기사를 수집하여 장애인 노동의제를 분석하였으며, 이종혁·길우영은 토픽모델링 방법을 활용하여 대통령 신년 기자회견 관련 뉴스기사를 분석하고 의제 다양성과 미디어 다양성에 대한 연구를 수행하였다. 감성 분석에는 매우 다양한 응용 분야 가 있다. 예를 들어 소비자 산업에서 제품리뷰 감성분석을 통해 다양한 상품에 대한 사용자 선호도를 확 보하여 회사가 판매 전략을 조정하고 의사결정을 내릴 수 있다. 소셜 미디어에서 이벤트 댓글에 대한 감 성분석은 사용자가 필요로 하는 정보를 빠르게 찾아내고, 의미 있는 정보를 유추해내는데 중요한 역할을 한다(Salas-Zarate et al., 2016; Alashri et al., 2016; Mertiya & Singh, 2016). 2000년초 이래로 감성 분 석은 자연어처리(National Language Processing; NEP)에서 가장 활발한 연구 분야 중 하나가 되었다(Hussein, 2018). 기존의 감성분석을 위한 접근방법으로는 어휘기반(Lexicon-based approach)과 기계학습 기반(Ma-chine learning approach)의 방법으로 나눌 수 있다(Chung & Ahn, 2019). 어휘기반의 감성분석은 전 통적인 감성분석 방법으로 감성사전을 구축하고 사전에 기록된 감성단어의 점수를 이용하여 감성의 극 성을 판단한다(Turney & Littman, 2003; Hu & Liu, 2004; Bravo-Marquez et al., 2016; Yang et al, 2013).

이전 연구에서는 텍스트 마이닝 및 자연어 처리 기술을 이용하여 소셜 미디어의 텍스트 데이터에서 감성을 추출하는 방법에 대해 탐구하였다. ‘Sentiment analysis’는 사람들의 태도, 의견, 성향과 같은 데이터를 활용하여 특정 주제에 대해 긍정적인지 혹은 부정적인지를 분류하는 기술이다. ‘sentiment analysis’는 ‘감성 분석’ 으로 해석된다(Chung, 2022; Lee et al., 2020). 데이터 마이닝 기술 중 하나인 텍스트 마이닝은 자연어 처리와 정보 추출 등의 분야를 연구하는데 유용한 기술 중 하나이다. 소셜 미디어에서 흔히 찾을 수 있는 데이터들은 구조가 완전하지 않는 형태로 구성되어 있고 가공되지 않은 데이터로 그 안에서 불분명한 형태 안에 필요한 키워드 추출하는 작업은 중요하다(Kim et al., 2020; Lee & Kim, 2009).

텍스트 마이닝과 감성 분석은 현대 데이터 분석 분야에서 가장 빠르게 성장하고 있는 방법 중 하나로, 대량의 텍스트 데이터에서 의미 있는 정보를 추출하고 분석하는 데 큰 도움을 제공한다. 텍스트 마이닝은 정보 검색과 패턴 인식이라는 두 가지 주요 과정을 포함한다. 정보 검색은 대량의 데이터에서 관련성 높은 정보나 문서를 신속하게 찾아내는 과정이며, 패턴 인식은 이러한 데이터에서 일정한 패턴이나 경향을 파악하는 과정이다. 따라서 텍스트 마이닝을 통해, 우리는 문서 내용의 자동 분류, 특정 정보의 추출, 그리고 여러 문서 간의 상호 연관성을 명확하게 파악할 수 있다(Lee & Hong, 2020).

또한, 감성 분석은 텍스트 마이닝의 한 분야로서 빠르게 주목받고 있다. 감성 분석의 주요 목적은 텍스트 데이터 내에서 작성자의 감성, 의견, 태도 등을 정확하게 파악하는 것이다. 소셜 미디어 상의 다양한 피드백, 예를 들면 상품 리뷰, 영화 평가 또는 뉴스 기사에 대한 리뷰 등에서, 작성자가 긍정적인지 부정적인지, 또는 중립적인지의 반응을 감성 분석을 통해 분석할 수 있다. 이로써 기업, 연구자, 개발자 등은 소비자나 대중의 반응과 트렌드를 실시간으로 파악하여, 제품 개발이나 전략을 조정하는 데 큰 도움을 받을 수 있다(Jacobi et al., 2015).

2.2. 나이브 베이즈와 감성 분석

감성분석에서 중요히 여겨지는 언어의 구문론적, 형식적 부분은 대부분 촘스키의 1950-60년대 연구에 기반을 두고 있다. 그러나 실질적으로 감성분석 연구가 활발히 이루어지고 많은 발전을 보인 것은 2000년대 이후이며, 감성분석에 접근하는 다양한 연구 방법들과 이론이 정리됐고, 이를 통해 발전된 감성분석 방법론과 적용 방향에 대한 파악이 가능해졌다. 감성분석은 다양한 기계학습을 기반으로 이루어지고 있으며, 이때 기계학습은 크게 지도학습과 비지도학습(unsupervised learning) 기법에 바탕을 두는 것으로 나누어 볼 수 있다(Kang et al., 2022; Kim & Park, 2012). 감성분석은 뉴스, 블로그, SNS, 상품 리뷰 등의 텍스트로부터 긍정, 부정, 중립 등의 감성정보를 추출하는 방법론으로서 분류 결과는 조사 대상에 대한 전반적인 의견, 관점, 인식 등에 관심이 있는 분석가들에게 유용하게 활용될 수 있다. 감성분석은 주로 분석 자동화 및 감성사전 구축을 통한 방법론의 정확도 향상에 관한 연구를 시작으로, 최근에는 범위를 확장하여 주식가격 예측 등 광범위하게 활용되고 있다. 감성분석이 자연어 처리를 위한 접근법이라는 점에서 트위터와 같은 마이크로 블로그 자료들이 사용되었는데, 그러한 연구들은 나이브 베이즈, SVM (Support Vector Machine) 등의 모델 성능을 비교하거나, 어휘/문장/이모티콘 등을 구분하여 감성분석을 시도하였다. 한편, 마케팅 부문에서 서비스 향상을 위해 감성분석 결과를 만족도 혹은 선호도와 관련시키기도 하였다.

머신러닝(machine learning)은 인공지능의 한 분야로 데이터를 통해 얻어지는 경험으로부터 특정한 목표 작업에 대한 성능을 향상시키는 일련의 과정으로 정의된다. 수많은 머신러닝 기법 중 적절한 알고리즘을 선정하기 위해서는 목적에 대한 정의와 데이터에 대한 이해가 필요하다. 머신러닝 기법 중 하나인 나이브 베이즈 분류기는 주로 스팸 필터나 키워드 검색을 활용한 문서 분류에 사용된다(Bengio et al., 2003). 나이브 베이즈 분류기는 클래스 변수(class variable)의 값이 주어진 경우에 모든 자질(feature)들은 조건부독립임(conditionally independent)을 가정하는 모델이다(Kim et al., 2016). 나이브 베이즈 분류기의 기본 원리는 조건부 확률에 베이즈 정리를 적용하고, 문서나 데이터를 구성하는 각각의 요소들이 등장할 확률에 대한 독립성을 가정하여 입력 벡터를 확률적으로 분류하는 것이다(Kim et al., 2019; Byeon & Shim, 2020; Choi et al., 2016).

본 연구는 소셜 미디어의 사용자 피드백 및 감성 패턴을 나이브 베이즈 분류를 이용해 확인하는 것을 목표로 한다. 소셜 미디어 상에서 감성 표현은 개인의 주관성이 강하며, 각기 다른 맥락과 상황에 의해 영향을 받는다. 따라서 본 연구는 특정 기간 동안의 긍정적, 부정적, 중립적 피드백 데이터를 수집하여 분석함으로써, 보다 객관적인 감성분석 기준을 설정하고자 한다. 이 데이터를 통해, 사용자들의 일반적인 선호도와 소셜 미디어 콘텐츠에 대한 반응을 유추할 수 있으며, 이는 해당 콘텐츠의 인기를 가늠하는 지표로 사용될 수 있다.

2.3. 인스타그램 리뷰를 이용한 감성 분석

Instagram은 2010년에 시작된 사진 및 비디오 공유 소셜 미디어 플랫폼이다. 사용자들은 개인 프로필을 생성하고 사진이나 비디오를 게시하며, 다른 사용자들의 게시물을 '좋아요'하거나 리뷰를 달 수 있다. 또한, 스토리라는 기능을 통해 24시간 동안만 보여지는 사진이나 비디오를 공유할 수 있다. Instagram은 그 시각적 특성 때문에 사진작가, 아티스트, 브랜드, 연예인 등 다양한 사용자들에게 인기가 있다(Kim & Han, 2016; Byeon & Shim, 2020).

Instagram은 시각 중심의 플랫폼 특성으로 인해 특히 감성 표현이 활발한 플랫폼 중 하나로 간주되며 Instagram 사용자의 행동, 선호 및 반응에 관한 다양한 연구가 진행되었다. '세상의 순간들을 포착하고 공유한다'는 슬로건을 보유한 인스타그램은 온라인을 통한 사진 공유 및 전달이 가능한 소셜미디어이다. 한국의 경우 인스타그램 월별 사용자가 1,000만 명을 넘어섰으며, 더욱이 이미지에서 동영상 스트리밍 서비스로 영역을 확장하면서(Kim, 2018), 국내 SNS 앱 이용자 수 4위(2017년)에서 2위로(2018년) 상승함을 확인할 수 있다(Inside, 2019). 이러한 데이터와 연구 결과는 Instagram의 성장과 인기, 그리고 그 플랫폼에서의 사용자 활동의 활발성을 잘 나타내준다(Byeon & Shim, 2020).

Instagram은 그 특징적인 필터와 스토리 기능, 그리고 직관적인 사용자 인터페이스로 급속도로 인기를 얻어왔다. 이러한 요소들은 사용자들에게 쉽고 빠르게 그들의 순간을 아름답게 포착하고 공유할 수 있는 도구를 제공한다. 이 플랫폼은 단순히 사진과 비디오를 공유하는 공간을 넘어서, 개인 브랜딩, 비즈니스 마케팅, 그리고 사회적 활동에 필수적인 도구로 자리 잡았다. Instagram의 Influencer 마케팅은 특히 주목받고 있다. 전세계적으로 수많은 Influencer들이 자신의 팔로워들에게 제품, 여행, 라이프스타일 등 다양한 주제로 소통하며 큰 영향력을 행사하고 있다. Instagram은 최근에 쇼핑 기능을 도입하여, 사용자들이 게시물에서 직접 제품을 구매할 수 있게끔 하는 등의 혁신을 지속적으로 시도하고 있다. 이러한 변화는 Instagram이 단순한 소셜 미디어 플랫폼을 넘어, 전자상거래 플랫폼으로의 발전 가능성을 보여준다.

이런 다양한 기능과 변화 속에서도 Instagram의 핵심 가치는 여전히 “사람들과 연결”하는 것에 있다. 사용자들은 Instagram을 통해 가까운 친구, 가족, 그리고 세계 각지의 사람들과 그들의 순간들을 공유하며, 새로운 경험과 문화를 접하게 된다.

감성분석 연구는 컴퓨터 과학 분야의 한 주제에서 시작됐다고 할 수 있다. 그러나 현재 감성분석연구에 사용되는 데이터의 특성으로 인해 미디어 커뮤니케이션 분야에서도 많은 관심을 받는 연구 분야로 자리잡았다. 이에 따라 본 연구에서는 인스타그램 리뷰 데이터를 이용한 감성분석연구에 있어 컴퓨터 과학과 미디어 커뮤니케이션이라는 두 학문분야를 주로 다루고 있다. 컴퓨터 과학 분야에서 연구자들은 일반적으로 리뷰에서 감성 태도를 추출하고 채점 관련 정보를 자동으로 활용할 수 있는 감성 분석 모델 개발 및 채점 추정에 중점을 둔다. 이 접근 방식은 수동적으로 코멘트를 긍정적이거나 부정적으로 분류할 필요가 없다. 리뷰 데이터는 일반적으로 사용자가 직접 남긴 평가 및 채점 구조를 가지고 있어 자동화된 분석을 용이하게 한다. 반면 미디어커니션 분야에서는 리뷰에 있는 감성 정보를 활용한 앱으로 연구의 초점을 넓혔다. 감성 분석은 인스타그램과 같은 소셜 미디어 플랫폼에서 사용자의 커뮤니케이션 및 피드백을 해석하고 분석하여 대중의 감성과 관점의 추세를 이해하는 데 사용된다. 이러한 분석 결과는 콘텐츠 창작, 광고 포지셔닝, 브랜드 이미지 형성 등을 개선하고 대상 고객과의 커뮤니케이션 효과를 높이는 데 사용할 수 있다(Hong et al., 2016; Chung et al., 2017; Yun et al., 2017).

3. 연구방법

3.1. 데이터 컬렉션 및 전처리

구글 스토어의 인스타그램 사용자 리뷰에서 감성 패턴을 추출하는 것에 초점을 맞춘 연구 방법은 데이터 소스 및 전처리, 사용자 등급에 따른 감성 분류, 나이브 베이즈 알고리즘 적용의 세 가지 중요한 단계로 구성되었다.

이 연구는 2018년 9월 12일부터 2023년 7월 27일까지 Google Play 스토어에서 수집한 인스타그램 사용자 리뷰 데이터를 대상으로 감성 분석을 수행하였다. 원본 데이터 세트에는 총 210,543건의 리뷰가 포함되어 있었다. 데이터의 정확성과 분석의 신뢰도를 높이기 위해, 누락된 리뷰 또는 평점 정보가 있는 데이터 항목을 제거하는 전처리 과정을 거쳤다. 또한, 분석 가치가 낮다고 판단되는 너무 짧거나 긴 리뷰도 필터링하여 제외하였다. 이와 더불어, 평점 데이터를 감성 표현(예를 들어, 긍정적 또는 부정적)으로 활용하기 위해 특정 평점 기준에 따라 데이터를 필터링하였다. 텍스트 길이에 대한 필터링은 분석 가치가 낮은 과도하게 짧거나 긴 리뷰를 제거함으로써 데이터의 질을 높이는 데 기여하였다. 이러한 전처리 및 필터링 과정을 거친 후, 총 209,956건의 리뷰가 분석을 위해 남게 되었다. 이러한 전처리 과정은 데이터의 질을 향상시키고, 분석 결과의 정확도를 보장하기 위해 필수적인 단계였다.

데이터 전처리 단계에서는 데이터의 질을 보장하기 위해 주의 깊은 검토와 선별 과정을 거쳤으며, 이는 분석의 신뢰성을 높이는 데 기여하였다. 사용자 등급에 따른 감성 분류 과정에서는 리뷰의 긍정적 또는 부정적 성향을 평가하기 위해 사용자가 남긴 평점과 리뷰 내용을 상세히 분석하였다. 마지막으로, 나이브 베이즈 알고리즘을 적용하여 감성 분류의 정확도를 높이고, 리뷰 데이터에서 감성 패턴을 효과적으로 추출했다.

이 데이터 세트는 사용자의 텍스트 리뷰와 1~5점 점수를 포함하여 다양한 문화적 배경과 생활 경험을 가진 수천 명의 전 세계 사용자를 대상으로 한다. 이러한 데이터는 거시적 및 미시적 측면에서 인스타그램 애플리케이션 사용자의 실제 평가, 사용 습관 및 관심사에 대한 통찰력을 제공한다. 데이터를 처리하기 전에 후속 데이터 전처리 및 분석 방향에 대한 중요한 참고 자료인 데이터의 구조, 결측값 및 이상값을 이해하기 위해 탐색적 분석이 먼저 수행되었다. 데이터 전처리 단계에서는 기본적인 데이터 세척 작업 외에도 리뷰에 포함된 URL (Uniform Resource Locator), 숫자, 특수 문자 및 HTML (HyperText Markup Language) 태그에 특별한 주의를 기울이는데 이러한 요소는 감성 분석을 방해할 수 있으므로 제거해야 한다. 텍스트 표준화는 핵심 단계이며 모든 리뷰 텍스트를 소문자로 균일하게 변환하고 다른 표현으로 인한 오해를 피하기 위해 일반적인 약어를 확장한다. 또한 텍스트의 심층 의미 정보를 탐색하기 위해 NLP (Natural Language Processing) 도구를 사용하여 어간 추출 및 어형 환원을 수행하고 기본 형태로 다른 어형을 통합한다. 텍스트의 주요 내용을 처리한 후 일반적으로 사용되는 영어 중단어도 텍스트에 자주 나타나지만 감성 분석에 대한 실제 가치는 제한적이기 때문에 스크리닝 및 삭제되었다. 데이터 전처리 과정에서 감성 라벨도 결정해야 한다. 데이터 세트에는 사용자의 점수 데이터가 포함되어 있기 때문에 이러한 점수 데이터는 모델의 훈련 및 예측을 위해 감성 범주로 변환된다. 이 방법은 공개 저장소에서 얻은 사용자 리뷰 및 채점 데이터, Python에서 Pandas 라이브러리의 효율적인 데이터 가져오기 및 운영, 텍스트 세척 및 표준화에 대한 엄격한 처리를 결합하여 본 연구의 데이터 처리 및 분석 기반을 구성한다.

3.2. 사용자 등급에 따른 감성 분류

본 연구의 감성 분류 과정은 Instagram 사용자 리뷰 데이터에 대한 깊이 있는 이해를 바탕으로 진행되었다. 사용자들이 남긴 원시 평가를 통해 얻어진 데이터셋을 준비한 후, 감성 분석의 예비 단계로 나아갔다. 이 과정에서는 Instagram 사용자가 제공한 평점을 실질적인 감성 지표로 활용하여, 감성을 세 가지 명확한 범주로 분류하는 방법을 모색하였다. 구체적인 라벨링 방법은 다음과 같이 결정되었다: 사용자가 1 또는 2점을 부여한 리뷰는 '부정적' 감성을 나타내는 것으로 판단되었고, 3점을 부여한 리뷰는 '중립' 감성을, 4점 또는 5점을 부여한 리뷰는 '긍정적' 감성을 나타내는 것으로 분류되었다. 이러한 분류 기준은 실제 사용자 평가와 감성의 상관관계를 명확히 반영하며, 감성 분석 모델 훈련에 있어 정확한 라벨 참조를 보장하다. 이 접근 방식은 (Jang et al., 2015)의 연구에서도 언급되어 있으며, 사용자 리뷰 데이터를 활용한 감성 분석에 있어 효과적인 방법론으로 평가받고 있다. 이러한 분류 과정을 통해, 본 연구는 사용자 리뷰의 숫자 평점을 기반으로 한 감성 태그 변환의 정확성을 높이고, 데이터셋의 감성 분류를 더욱 명확하게 진행할 수 있었다.

이 접근 방식의 효율성은 기계 학습 알고리즘이 이해하고 처리할 수 있는 수치 데이터로 일반 텍스트 데이터를 변환하는 과정에서 중요한 역할을 하다. 텍스트 벡터화는 기계 학습 모델이 텍스트 데이터의 패턴을 인식하고 예측을 수행하는 데 필수적인 과정이다. 이 과정에서 사용된 TF-IDF(단어 빈도-역문서 빈도) 방법은 특정 리뷰 내에서의 단어 빈도뿐만 아니라 전체 데이터셋에서의 단어 빈도를 고려하여 각 단어에 가중치를 부여함으로써, 문서 내에서 중요한 단어를 강조하다. 이러한 방법은 텍스트 데이터를 수치 형식으로 표현하는 데 중요하며, 기계 학습 모델의 훈련과정에서 데이터의 특성을 보다 잘 반영할 수 있게 하다. 이 접근법의 유효성은 이전의 연구에서 입증되었으며, 예를 들어(Salton & McGill, 1983)의 연구에서 TF-IDF의 기본 원리와 응용에 대해 상세히 논의하고 있다. 이러한 텍스트 데이터의 수치화는 기계 학습 모델이 텍스트 내의 패턴을 인식하고, 이를 바탕으로 감성을 분류하는 데 필수적이다. 모델이 텍스트 데이터의 복잡성을 효과적으로 처리하고, 높은 정확도로 감성을 예측하기 위해서는 텍스트를 수치 형식으로 변환하는 과정이 중요하다. 따라서, TF-IDF 기법을 사용하는 이유는 모델의 성능을 최적화하고, 감성 분석의 정확도를 높이기 위한 핵심적인 단계로 간주된다.

이러한 라벨링과 텍스트 벡터화 기술을 결합함으로써, 우리는 Instagram 사용자 리뷰와 평가를 기반으로 더 정교한 감성 분석의 기초를 마련하였다. 이 접근 방

식은 데이터에 표현된 감성을 이해하고 분류하는 체계적인 방법을 제공한다.

3.3. 나이브 베이즈 알고리즘의 적용

감성 분석을 수행할 때, 핵심 방법으로서 나이브 베이즈 알고리즘과 TF-IDF 벡터화를 선택하였으며, 텍스트 분류 작업에서 나이브 베이즈 알고리즘의 효율성을 감안하여 단어 빈도 통계 방법인 TF-IDF 벡터화와 비교하였다.

원시 텍스트 리뷰를 처리하기 위해 TF-IDF 벡터화 방법을 사용하는 것은 중요한 단계였다. 이러한 단계는 텍스트 감성을 알고리즘이 효과적으로 처리하고 이해할 수 있는 숫자 형태로 변환했다. TF-IDF 방법은 특정 리뷰 내에서 단어 빈도뿐만 아니라 전체 데이터 세트에서의 빈도를 고려하여 각 단어에 합리적인 가중치를 할당한다.

데이터 집합 처리 측면에서 80%를 교육 세트로, 나머지 20%를 테스트 세트로 사용하여 분석의 견고성을 보장했다. 데이터 집합을 처리하는 방법에 대한 80/20 분할(학습 세트 80%, 테스트 세트 20%)이 널리 받아들여지는 규칙 중 하나라는 것은 기계 학습 및 데이터 과학 커뮤니티에서 잘 알려져 있다. 이러한 분할 비율은 데이터 과학 분야에서 일반적인 베스트 프랙티스로 간주되며, 다양한 학술 연구와 데이터 과학 프로젝트에서 성능 평가를 위한 표준 접근법으로 사용된다. 교육 데이터는 나이브 베이즈 분류기에 대한 필수 학습을 제공하고, 그 후에 테스트 세트에서 모델의 예측 능력을 테스트하였다. 평가 지표(정밀도, 재현율, F1 점수 포함)는 모델의 정확성을 포괄적으로 나타내었다. 더 나아가 이해를 높이기 위해 감성 범주별로 단어 빈도를 통합했다. 일반적인 불용어를 걸러내어 사용자 감성의 본질을 추출하였으며, 이를 워드 클라우드 형태로 생생하게 표현하여 제공되었다.

이 연구에서는 Python을 주요 프로그래밍 언어로 사용했으며 관련 도구 라이브러리에는 Scikit-learn(사이킷런) 및 NLTK (Natural Language Toolkit)가 포함된다. 모델 평가 단계에서 우리는 정확도, 회수율 및 F1 점수와 같은 주요 지표에 중점을 둔다. 동시에, 모델을 더욱 최적화하기 위해, 우리는 또한 모형이 최상의 성능을 달성할 수 있도록 매개변수를 조정하고 최적화했다.

4. 분석 결과

4.1. 사용자 평가에 따른 단어 빈도 분석 결과

사용자 평가별로 분류된 사용자 감성을 이해하는 독특한 시각을 단어 빈도 결과가 제공한다.

본 연구는 2018년 9월 12일부터 2023년 7월 27일까지의 기간 동안 Google Play 스토어에서 수집한 인스타그램 사용자 리뷰 데이터를 대상으로 감성 분석을 수행하였다. 이 데이터는 전 세계 다양한 지역의 Instagram 사용자들로부터 얻어진 피드백을 포함하고 있으며, 분석을 통해 긍정적인 리뷰에서 ‘good’이라는 단어가 14,338회로 가장 빈번하게 등장하는 것을 확인할 수 있었다. 이는 사용자들이 플랫폼에 대해 느끼는 만족감을 직관적으로 나타내는 결과이다. ‘love’(9,979회)와 ‘great’(7,430회) 같은 단어들이 뒤따라 긍정적인 감성을 더욱 강조하다. 흥미롭게도, 긍정적인 리뷰에도 불구하고 ‘problem’(6,438회), ‘please’(5,311회)와 같은 단어들이 자주 언급되었는데, 이는 사용자들이 작은 개선점이나 사소한 문제들을 제기할 수 있다는 것을 시사하다. ‘friends’(7,183회)와 ‘people’(8,050회)의 언급은 인스타그램의 사회적 특성과 개인들을 연결하는 역할을 강조하다. 이 분석은 2018년 9월부터 2023년 7월까지의 리뷰 데이터를 기반으로 하여, 인스타그램과 같은 소셜 미디어 플랫폼이 사용자 경험과 감성에 미치는 영향을 더 깊이 이해하는 데 기여하고자 한다.

중립적인 리뷰 카테고리에서는 ‘fix’라는 단어가 4,439회로 가장 많이 등장했다. 이는 상당수의 중립적 감성이 사용자들이 해결책이나 개선을 추구하는 데에 초점을 맞추고 있음을 시사한다. ‘problem’(4,368회), ‘account’(4,156회)가 뒤따라, 사용자들이 겪은 가능한 문제들을 나타내지만 전반적인 경험을 부정적으로 만들 정도는 아니었다는 것을 나타낸다. ‘update’(3,619회), ‘story’(3,204회) 같은 단어들은 애플리케이션의 새 기능이나 변경에 대한 사용자들의 중립적 반응을 나타낼 수 있다.

부정적인 리뷰에서는 완전히 다른 그림이 나타났다. ‘Account’가 이 카테고리에서 29,371회로 가장 많이 등장했다. 이 빈도는 사용자들 사이에서 계정과 관련된 중대한 문제가 있음을 시사하며, 계정 접근성 문제, 보안

문제, 또는 계정 기능에 대한 문제를 암시할 수 있다. ‘update’(21,495회), ‘fix’(19,505회), ‘problem’(17,778회) 같은 단어들의 높은 빈도는 애플리케이션의 특정 업데이트나 지속적인 문제에 대한 사용자 불만을 밝혀준다. ‘reels’(14,348회)의 출현은 이 기능에 대한 특정한 우려나 사용자 경험에서의 불쾌감을 나타낼 수 있다.

결과를 보면 사용자 평가별로 분류된 단어 빈도 분석은 사용자 감성에 대한 세밀한 정보를 제공하다. 긍정적인 리뷰는 전반적인 만족도를 강조하지만, 약간의 개선이 필요한 영역도 지적한다. 중립적인 리뷰는 해결책을 찾는 경향이 강하며, 부정적인 리뷰는 계정과 관련된 문제를 비롯한 즉각적인 주의가 필요한 영역을 강력히 시사한다.

4.2. 단어 클라우드 시각화 및 그들의 함의

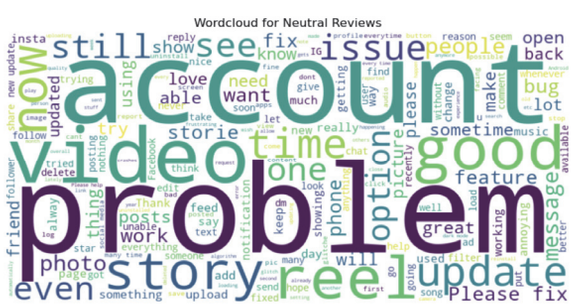
단어 클라우드는 글꼴 크기를 다르게 표현하여 단어의 빈도를 나타내는 강력한 시각 도구로서, 텍스트 데이터의 본질을 포착한다. 인스타그램 사용자 리뷰에서 파생된 단어 클라우드는 각 등급 범주 내에서 우세한 감성을 밝혀준다.

Fig. 1에서는 긍정적인 리뷰의 경우, ‘good’이 두드러져 나오며, 단어 빈도 분석 결과와 일치한다. ‘love’, ‘friend’, ‘story’와 같은 주목할 만한 용어들은 플랫폼이 연결을 촉진하고 이야기를 공유하는 역할을 시사한다. 훌륭한 칭찬들 속에서 ‘problem’이 두드러지는 위치를 찾아, 긍정적 리뷰어들도 가끔 걱정 사항을 가질 수 있음을 시사한다. ‘social’, ‘media’, ‘photo’와 같은 단어들은 인스타그램의 핵심 본질, 즉 사진 공유 소셜 미디어 플랫폼임을 강조한다.

Fig. 2에서는 중립적인 리뷰에서는 ‘problem’과 ‘good’가 여전히 확인할 수 있지만, ‘issue’, ‘update’, ‘fix’와 같은 용어들이 두드러진다. 이는 중립적 리뷰가 만족과 불만 사이에서 헤매는 변동적인 성격을 강조한다. ‘video’, ‘story’, ‘photo’의 빈번한 등장은 사용자가 이러한 기능을 평가하면서 가끔 문제를 겪을 수 있음을 나타낸다. ‘please’라는 용어는 사용자들이 개선이나 수정을 기대하며, 그들의 피드백에 희망적인 미소가 있다는 것을 보여준다.

Fig. 3에서는 부정적인 리뷰의 단어 클라우드는 더 긴급한 상황을 묘사한다. ‘issue’, ‘account’, ‘update’와 같은 단어들이 시각화를 지배하며, 단어 빈도 데이터와 일치하며 중요한 우려사항을 나타낸다. ‘video’가 두드러지면 비디오 기능이나 재생과 관련된 문제가 있을 수 있음을 나타낼 것이다. 특히 ‘fix’라는 용어가 두드러져 나와 사용자들이 수정을 간절히 원한다는 것을 반영한다. ‘login’, ‘password’, ‘support’와 같은 단어들은 사용자 접근성 또는 도움을 요청하는 데 관련된 잠재적인 문제를 시사할 수 있다.

이러한 시각화와 단어 빈도 데이터를 대조하면 사용자 감성을 다면적으로 이해할 수 있다. 단어 빈도는 용어의 일반적인 등장 빈도를 계량화하고, 단어 클라우드는 이러한 용어의 감성적 가치를 강조하여 인스타그램에서의 사용자 경험에 대한 통찰력을 풍부하게 한다.



4.3. 나이브 베이즈 분류 결과

사용자 리뷰의 숫자 평점을 감성 태그로 변환한 후, TF-IDF 방법을 사용하여 리뷰 텍스트를 벡터 형태로 변환한다. 나이브 베이즈 모델을 활용하여 감성 분류를 수행하고, 테스트 데이터에서 모델의 성능을 평가한다. 마지막으로, 코드는 각 감성 카테고리에 대한 ROC (Receiver Operating Characteristic)곡선을 작성하며, 다양한 임계값에서 모델의 성능을 평가하기 위해 AUC (Area Under the ROC Curve)값을 계산한다. 이러한 시각화 방법은 모델의 성능과 잠재적인 개선점을 파악할 수 있는 다양한 범주에서 모델의 예측 효과를 이해하는 데 도움을 주는 직관적인 방법을 제공한다.

나이브 베이즈 분류기를 통한 감성 분석 작업에서 나온 결과는 그 능력과 한계에 대한 포괄적인 시각을 제공한다. 전반적으로 이 분류기는 각 감성 범주에 대한 지표로 나타난 것처럼 혼합된 성능을 나타낸다. 전체적인 정확도가 0.74로, 부정적인 감성을 식별하는 데 우수한 능력을 보여주며 정밀도가 0.72, 인상적인 재현율은 0.96로 나타난다. 정밀도와 재현율의 균형인 F1 점수는 이 범주에 대해 0.82이다. 반면, 중립적인 감성은 분류기에게 상당한 도전임을 보여준다. 이 범주의 지표는 전반적으로 부족하며 개선이 필요한 부분을 강조한다. 반면, 긍정적인 감성에 대해서 모델은 좋은 성과를 보이며 F1 점수가 0.72이고 정밀도가 0.81, 재현율이 0.65로 뒷받침된다.

ROC 커브 통찰:

ROC 커브로 넘어가면 이 그래픽 표현은 참 긍정 비율을 거짓 긍정 비율에 대해 나타내며, 다양한 임계값을 통해 모델의 효능을 보여준다. 곡선 아래 영역(AUC)은 더 나은 분류를 나타내는 더 높은 값으로서 기준 역할을 한다. 부정적인 감성에 해당하는 곡선은 AUC가 0.85로, 이 범주에서 모델의 능숙성을 입증한다. 반면, 중립적인 감성 곡선은 0.68의 AUC로, 이전에 언급한 모델의 어려움을 강조한다. 긍정적인 감성 곡선은 인상적인 0.88의 AUC를 획득하여 강력한 성능을 시사하지만 향상의 여지가 있다는 것을 나타낸다.

혼동 행렬 분석:

혼동 행렬(Confusion Matrix)은 머신러닝 모델의 성능을 평가하는 데 중요한 도구로 간주된다. 이는 모델이 실제로 어떻게 예측하였는지와 예측 결과가 얼마나 정확한지를 시각적으로 표현하는 행렬이다.

혼동 행렬을 사용하면 각 카테고리에 대한 예측의 정확성뿐만 아니라, 잘못 분류된 예측도 쉽게 확인할 수 있다.

더 자세히 살펴보면 혼동 행렬은 각 감성에 대한 분류기의 성능에 대한 심층적인 통찰력을 제공한다. 부정적인 감성을 예측하는 분류기의 견고함은 분명하며, 24,021개 중 23,006개가 정확하게 분류된다. 그러나 중립적인 감성에 대한 분류기의 도전은 뚜렷하다. 5,402개의 중립적인 인스턴스 중 3개만 정확하게 분류되었

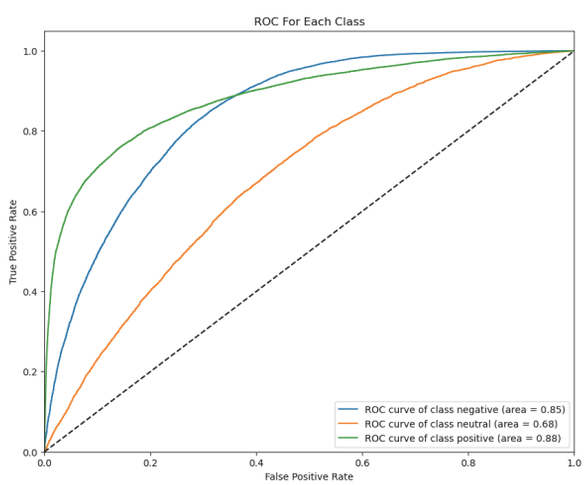


Fig. 4. Comparative ROC curves for negative, neutral, and positive sentiment classes

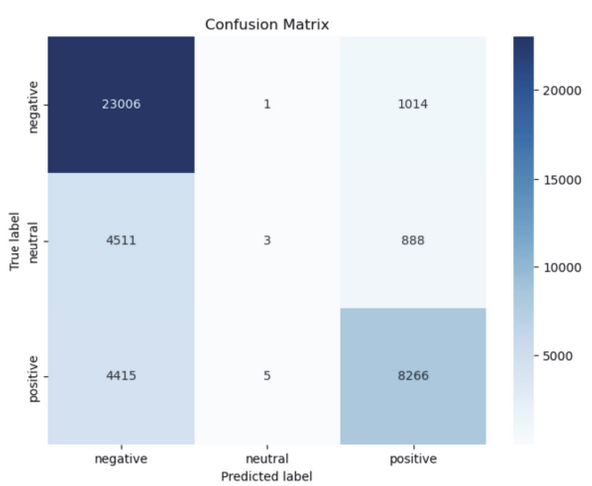


Fig. 5. Sentiment classification confusion matrix

으며, 대부분이 부정적이거나 긍정적으로 잘못 표시되었다. 이는 중립 및 다른 범주를 구분하는 데 모델이 사용하는 기능에서 잠재적인 중첩이나 유사성을 나타낸다. 긍정적인 감성의 경우 분류기의 성능은 훌륭하지만 완벽하지는 않다. 12,686개 중 8,266개를 올바르게 식별했지만 상당수인 4,415개가 부정적으로 잘못 분류되었다.

4.4. 비교 분석: 사용자 평가 VS 나이브 베이즈 예측

나이브 베이즈 분류기에서 생성된 키워드 빈도를 분석하고 실제 사용자 평가에서 얻은 빈도와 대조하면 분류기의 성능에 대한 중요한 통찰을 얻을 수 있다.

부정적인 리뷰의 경우, “account”라는 용어가 예측 및 실제 사용자 평가에서 모두 상위 목록에 있으며 나이브 베이즈 분류기는 약간 더 높은 빈도를 예측한다. 마찬가지로 “update”, “fix”, “problem”, “even”과 같은 용어가 두 데이터 세트 모두에서 두드러진다. 그러나 순위에서 차이가 있다. 예를 들어, “reels”는 분류기의 예측에서 10번째로 가장 빈번한 용어이지만 사용자 평가에서는 12위이다. 흥미로운 것은 “story”와 “still”과 같은 용어가 실제 사용자 평가보다 나이브 베이즈 예측에서 더 자주 나타난다.

중립적인 리뷰 카테고리에서 나이브 베이즈 예측과 사용자 평가 사이에 뚜렷한 차이가 있다. “fix”라는 단어가 사용자 평가에서 상위를 차지하지만 나이브 베이즈 예측에서는 훨씬 낮다. 이 분류에서 분류기는 “feature”, “story”, “able”과 같은 단어에 대한 예측 빈도가 사용자 평가에서 “fix”, “problem”, “account”와 같은 용어에 비해 상당히 낮다는 것을 보여준다.

긍정적인 리뷰의 경우, 나이브 베이즈 예측과 사용자 평가 사이에 더 가까운 일치가 있다. “good”, “love”, “people”, “great”와 같은 용어가 두 목록 모두에서 두드러진다. 그러나 “social”이라는 용어는 분류기의 예측에서 7위지만 사용자 평가에서는 15위이다. 반면에 부정적인 감성과 관련된 “problem”이 긍정적인 리뷰에 대한 사용자 평가에서 예상치 못하게 7위에 랭크되어 있어 데이터의 모호성이나 감성 해석의 복잡성을 나타낼 수 있다.

나이브 베이즈 분류기는 부정적인 리뷰와 긍정적인 리뷰의 키워드 빈도에 대한 합리적인 근사치를 제공하지만 중립적인 범주에서의 성능이 개선이 필요하다. 또한 예측된 빈도와 실제 빈도 사이의 불일치는 감성 분석의 본질적인 어려움과 더 나은 정확도를 위해 모델을 지속적으로 정제해야 한다는 중요성을 강조한다.

5. 결론

본 연구는 Kaggle에서 획득한 대량의 Instagram 리뷰 데이터를 상세하게 전처리한 후 텍스트 마이닝을 수행하여 긍정적, 중립적, 부정적인 감성을 표현하는 단어들로 구성된 워드 클라우드를 생성했다. 이러한 데이터를 바탕으로, Google Play 스토어에서의 Instagram 사용자 리뷰에 대한 감성 분석을 나이브 베이즈 알고리즘을 이용해 수행했으며, 그 결과를 사용하여 Instagram 서비스의 품질 및 사용자 만족도를 평가했다. 이 과정에서 각 감성 상태별로 사용자들이 자주 언급하는 키워드를 시각적으로 도출했다. 나이브 베이즈 알고리즘을 적용한 결과, 이 알고리즘은 Instagram 사용자 리뷰 데이터에 잘 맞는 것으로 나타났다. 모델의 성능은 뛰어났으며, 특히 부정적 및 긍정적 리뷰의 구분에 매우 효과적이었다. 나이브 베이즈의 가정, 즉 특성 간의 독립성이 항상 실제 상황에서 유효하지는 않지만, 이 연구에서는 그럼에도 불구하고 높은 성능을 보였다.

분석 결과, 나이브 베이즈 알고리즘의 적용을 통해 리뷰 데이터에 대한 긍정적 및 부정적 감성의 구분을 성공적으로 수행했으며, 이 결과는 나이브 베이즈가 감성 분석에 적합한 도구임을 나타내는 강력한 증거가 되었다. 그러나 중립 리뷰의 정확한 분류는 어려웠으며, 이는 중립 리뷰가 명확한 감성을 나타내지 않아 알고리즘이 정확히 분류하기 힘든 경우가 많기 때문이다. 이러한 한계에도 불구하고, 이 연구는 감성 분석의 중요한 단계를 성공적으로 수행하였으며, 이 결과를 통해 Instagram 서비스 향상 및 사용자 만족도를 높이기 위한 방향을 제시하였다.

이 연구의 실질적인 기여도는 다음과 같다. 첫째, 복잡한 소셜 미디어 환경 속에서 사용자 리뷰를 통한 감성 분석이 기업의 제품 개발과 마케팅 전략 수립에 중요한 데이터 기반을 제공할 수 있음을 보여준다. 둘째,

다양한 소셜 미디어 데이터를 분석하고 이해함으로써, 사람들의 관심사, 문화적 경향 및 사회적 상호작용에 대한 더 깊은 통찰을 얻을 수 있다. 마지막으로, 이 연구는 감성 분석 방법론의 효과성을 입증하며, 후속 연구에 중요한 참고 자료를 제공한다.

그럼에도 불구하고 이 연구에는 몇 가지 제한 사항이 있다. 첫째로, 데이터 소스가 Kaggle에 국한되어 있어서, 더 넓은 Instagram 사용자 리뷰 데이터를 포괄하지 못할 수 있다는 점이다. 또한, 나이브 베이즈 알고리즘은 본 연구에서 높은 정확도를 보였지만, 그 성능은 데이터 세트의 특성이나 다른 상황에 따라 달라질 수 있다는 점에서 일반화의 한계가 있다. 이러한 문제를 해결하기 위해, 미래 연구에서는 다양한 데이터 출처와 알고리즘을 사용하여 연구의 범위와 정확도를 더욱 향상시킬 수 있을 것으로 기대된다.

이 연구는 인스타그램 사용자 리뷰 데이터의 감성 분석을 통해 소비자의 선호도와 사회적 상호작용에 대한 통찰을 제공한다. 이러한 분석은 제품 개선에 도움을 주고, 기업의 커뮤니케이션 전략을 개발하는 데 유용한 정보를 제공할 수 있다.

REFERENCES

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research, 3*, 1137-1155. DOI: 10.1162/153244303322533223

Byeon, H. M., & Shim, S. W. (2020). The relationship between engagement with social media and social media advertising: The differentiating role of platform type (Instagram, Facebook and Youtube). *Journal of Advertising & PR Practical Research, 13*(2), 58-90. DOI: 10.21331/jprapr.2020.13.2.003

Chung, D. (2022). A case study of visualizing emotions with social media emotion analysis – focused on media art cases. *Archives of Design Research, 35*(1), 237-257. DOI: 10.15187/adr.2022.02.35.1.237

Chung, J., Yoo, K., & Koo, C. (2017). A study on social media sentiment analysis for exploring public opinions related to education policies. *National Information Society Agency*, *24*(4), 3-16. DOI: 10.22693/ NIAIP.2017.24.4.003

Choi, S.-J., Son, M.-Y., & Kim, Y.-H. (2016). Associated keyword recommendation system for keyword-based blog marketing. *KIISE Transactions on Computing Practices*, *22*(5), 246-251.

Hong, T., Kim, J., & Hyun, S. J. (2016). A user sentiment classification using instagram image and text analysis. *Journal of Smart Media, 5*(1), 61-68.

Jacobi, C., van Atteveldt, W., & Welbers, K. (2015). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism, 4*(1), 1-18. DOI: 10.1080/21670811.2015.1093271

Jang, K., Park, S., & Kim, W. J. (2015). Automatic construction of a negative/positive corpus and emotional classification using the internet emotional sign. *Journal of KIISE, 42*(4), 512-521. DOI: 10.5626/JOK.2015.42.4.512

Jeong, J, W., Lee, J. M., & Choi, S. Y. (2018). Analysis of news regarding the disabled labor using text mining techniques. *Reinterpretation of Disability*, *1,* 48-100.

Kang, C. M., Eo, K. S., & Lee, K. C. (2022). Investigating the performance of bayesian-based feature selection and classification approach to social media sentiment analysis. *Journal of Management Information Systems Research, 24*(1)*,* 1-19.

Kang, S., Kwon, B., Kwon, C., Park, S., & Yun, I. (2018). Development of incident detection algorithm using naive bayes classification. *The Journal of The Korea Institute of Intelligent Transport Systems, 17*(6), 25-39. DOI: 10.12815/kits.2018.17.6.25

Kim, I.-G., Kim, H.-M., Lim, B., & Lee, K.-K. (2016). Relationship between result of sentiment analysis and user satisfaction - the case of Korean meteorological administration. *The Journal of the Korea Contents Association, 16*(10), 393-402. DOI: 10.5392/JKCA.2016.16.10.393

Kim, J. I., Park, S. J., Kim, H. J., Choi, J. H., Kim, H. I., & Kim, P. K. (2020). Sensitivity identification method for new words of social media based on naive bayes classification. *Smart Media Journal*,

9(1), 51-59. DOI: 10.30693/SMJ.2020.9.1.51

Kim, N., & Park, J. (2012). Personal information detection by using naive bayes methodology. *Journal of Intelligence and Information Systems, 18*(1), 91-107. DOI: 10.13088/jiis.2012.18.1.091

Kim, R. A., & Han, E. K. (2016). The impacts of use motivation of instagram brand account on satisfaction and continuous use intention focusing on the mediating effect of flow. *Journal of Advertising Research, 111*, 5-39. DOI: 10.16914/ar.2016.111.5

Kim, S., Lee, Y. J., Shin, J., & Park, K. Y. (2019). Text Mining for Economic Analysis. *Bank of Korea WP, 2019*(18), 1-66.

Lee, J. H., & Gil, W. Y. (2019). News agenda classification and media diversity analysis using topic modeling: Through the analysis of news on the presidential new year press conference. *Korean Journal of Broadcasting Studies, 33*(1), 161-196. DOI: 10.22876/kab.2019.33.1.005

Lee, S., & Kim, H. J. (2009). Keyword extraction from news corpus using modified TF-IDF. *Journal of the Korean Electronic Commerce Association, 14*(4), 59-73.

Lee, S.-M., & Hong, S.-G. (2020). Policy agenda proposals from text mining analysis of patents and news articles. *Journal of Digital Convergence*, *18*(3), 1-12. DOI: 10.14400/JDC.2020.18.3.001

Lee, S. M., Ryu, S. E., & Ahn, S. (2020). Mass media and social media agenda analysis using text mining: focused on '5-day rotation mask distribution system'. *Journal of Korean Contents Association, 20*(6), 460-469. DOI: 10.5392/JKCA. 2020.20.06.460

Salton, G., & McGill, M. J. (1983). Introduction to modern information retrieval. *McGraw-Hill Book Co.*, New York.

Yun, Y., Jo, J., Hur, Y., & Lim, H. (2017). A comparative analysis of cognitive change about big data using social media data analysis. *Journal of Software & Data Engineering (Information Processing Society Journal),* 6(7), 371-378. DOI: 10.3745/KTSDE.2017.6.7.371

원고접수: 2023.12.20
수정접수: 2024.02.24
게재확정: 2024.02.27

ⓒ 2024 (by) the authors. This open access article is distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original work is properly cited.