

**언어·의미적 특성과 네트워크 특성을 활용한
불법 바이럴 리뷰 탐지 모델
-X의 텍스트 리뷰를 중심으로-**

*신한솔, **조선하, ***이서윤, ****안교영, *****성경선, *****최예림

**Illegal viral review detection model using linguistic
semantic characteristics and network characteristics
-Focusing on X's text reviews-**

*Han-Sol Shin, **Sun-Ha Cho, ***Seo-Yun Yi, ****Kyo-Young Ahn, *****Kyung-Sun Sung
and *****Ye-rim Choi

요 약

인터넷 기술의 발전과 함께 서비스 산업에 대한 온라인 리뷰가 증가하면서, 소비자들의 구매 의사결정 구조가 변화하였다. 이에 많은 기업들은 사용자들의 리뷰를 하나의 마케팅 방법으로 사용하고 있다. 그러나 상품에 대한 리뷰 작성 대가로 특정 혜택을 받은 사실을 명시한 리뷰는 마케팅 효과가 떨어지기 때문에, 이를 공개하지 않은 불법 바이럴 리뷰가 확산되고 있다. 기존의 온라인 리뷰 데이터를 이용한 연구는 사용자의 텍스트 데이터에서 유의미한 정보를 도출하고자 하는 연구가 주를 이루었으며, 사용자의 네트워크에 대해서는 고려하지 않았다. 따라서 본 연구는 텍스트의 언어적·의미적 특성을 반영한 KoELECTRA 모델과 네트워크 특성을 반영한 TabNet 모델을 활용하여 불법 바이럴 탐지 모델을 제안한다. 연구 결과, 두 모델을 결합한 양상블 모델이 기존 단일 모델들에 비해 가장 우수한 성능을 보였다. 이는 소비자 보호와 공정한 온라인 리뷰 생태계 조성에 기여할 것으로 기대된다.

Key words

*Viral Review, Linguistic Feature, Semantic Feature, Network Feature, Tri-training,
Ensemble Modeling*

* 서울여자대학교 데이터사이언스학과, 학부생 (foxgreen11@swu.ac.kr)
** 서울여자대학교 데이터사이언스학과, 학부생 (juliacho0604@swu.ac.kr)
*** 서울여자대학교 데이터사이언스학과, 학부생 (winterlike13@swu.ac.kr)
**** 서울여자대학교 데이터사이언스학과, 학부생 (ann120440}@swu.ac.kr)
***** 서울여자대학교 데이터사이언스학과, 학부생 (biblesun03@swu.ac.kr)
***** 서울여자대학교 데이터사이언스학과, 교수, 교신저자 (yerim.choi@swu.ac.kr)

I. 서 론

제품에 대한 구전 정보가 포함된 온라인 리뷰는 정보화 시대에 널리 활용되고 있다[1]. 전자상거래 플랫폼의 지속적인 발전으로 온라인 제품 리뷰의 양이 증가했으며 시간이 지남에 따라 더욱 확대될 것으로 예상된다[2]. 이러한 온라인 리뷰는 현대 소비 문화에서 중요한 역할을 하며, 소비자 의사결정 과정에 큰 영향을 미치고 있다[3].

'바이럴 마케팅'은 이러한 온라인 리뷰의 특징을 활용하여 상품에 대한 긍정적인 입소문 및 확산 효과를 내도록 하는 마케팅 기법이며, 온라인 및 SNS에서 자주 사용되고 있다[4]. 이와 같은 바이럴 마케팅성 리뷰 및 게시글을 적을 때는 '추천·보증 등에 관한 표시·광고 심사지침'에 따라 광고임을 알려야 하나, 이를 알리지 않고 바이럴 마케팅을 하는 소위 '불법 바이럴 리뷰'의 흐름이 증가하고 있다[5]. 불법 바이럴 리뷰는 작성자가 실제로 해당 제품을 사용한 것처럼 표방하여 소비자에게 직접 사용했다는 신뢰감을 주어 문제가 되고 있다[6]. 이로 인해 의견을 자유롭게 공유하고 사용자에게 양질의 정보를 제공하고자 하는 소셜 미디어 플랫폼의 신뢰성이 훼손되고 있다[7].

대표적인 소셜 미디어 플랫폼 중 한 곳인 X는 방대한 정보량, 쉬운 접근 및 빠른 확산이 가능하다는 특징으로 인해 잘못된 정보가 빠르게 퍼지는 현상이 발생하고 있다[8]. X의 바이럴 리뷰에 관한 연구는 꾸준히 발전되어 왔다. 초기에는 언어적 특성을 활용한 바이럴의 여부 판별이 주를 이루었다면 이후에는 언어적 및 의미적 특성을 활용하는 등 다양한 연구가 이어졌다[9]. 그러나 이러한 기존 연구는 텍스트 분석에만 그치는 한계점이 있다.

소셜 미디어 내 바이럴은 네트워크를 통해 빠르게 확산되며, 사용자 간 상호작용을 기반으로 전파되는 특성을 가진다[10]. 이러한 특성 때문에 네트워크 분석은 필수적이지만, 바이럴의 주요 특성인 '확산'을 고려한 연구는 여전히 부족하다. 이에 본 연구는 사용자 간 상호작용을 파악할 수 있는 네트워크 분석을 활용하여 보다 효과적으로 바이럴을 탐지하고, 단순 바이럴이 아닌 불법 바이럴을 탐지하는 데 중점을 둔다.

따라서 본 연구에서는 사용자 간의 언어·의미적 특성과 네트워크 특성을 통합적으로 반영할 수 있는 모델을 설계하였다. 이 모델은 세 가지 특성을 동시에 학습하도록 개발되었으며, 성능 평가 시 단순히 정확도뿐만 아니라 실제 불법 바이럴 리뷰를 놓치지 않는 것이 중요하다고 판단하였다. 이에 따라, 불법 바이럴 리뷰를 올바르게 탐지한 비율을 나타내는 Recall을 주요 성능 평가 지표로 설정하였다.

본 논문은 다음과 같이 구성된다. 제2장에서는 본 연구에서 설계한 불법 바이럴 리뷰 탐지 프레임워크와 방법론에 대해 설명하며, 제3장에서는 본 연구에 사용한 데이터와 모델을 설명하고, 이를 토대로 진행한 실험의 결과와 성능을 비교 분석한다. 마지막으로 제4장에서는 본 연구의 내용을 요약하고, 이를 통해 시사할 수 있는 내용으로 결론을 짓는다.

II. 방법론

2.1 Overview

본 연구에서는 불법 바이럴 리뷰와 일반 리뷰는 다음 기준으로 정의하였다. 불법 바이럴 리뷰는 특정 상품에 대해 리뷰를 해주는 조건으로 특정한 혜택을 받았음에도

불구하고, 이에 대한 경제적 이해관계를 밝히지 않는 리뷰로 정의하였다[6]. 일반 리뷰는 '#광고', '#제품제공' 등 광고를 받았음을 명시한 합법 바이럴 리뷰와 경험적 사실에 근거한 리뷰로 정의하였다[5].

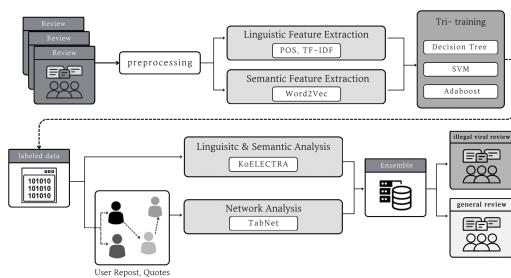


그림 1. 불법 바이럴 리뷰 탐지 과정이 담긴 프레임워크.

Fig. 1. Illegal Viral Reviews Detection Framework.

그림 1은 불법 바이럴 리뷰 탐지를 도출하는 프레임워크를 나타낸다. 먼저, X의 리뷰 데이터를 수집하고 전처리를 통해 분석 가능한 형태로 변환하였다. 이후 POS, TF-IDF, Word2Vec을 통해 언어·의미적 특성을 추출한 후 Tri-training을 진행하여 라벨링 된 데이터셋을 구축하였다. 구축된 데이터셋은 두 가지 방식으로 분석되었다. 첫째, 텍스트 분석을 위해 KoELECTRA 모델에 활용하였다. 둘째, 네트워크 분석을 위해 불법 바이럴 리뷰를 리포스트 혹은 인용한 계정을 활용하여 네트워크 피쳐를 추출한 다음 TabNet 모델에 활용하였다. 마지막으로, 각 단일 모델을 소프트 보팅 양상을 방법으로 결합하여 불법 바이럴 리뷰를 정확하게 탐지할 수 있는 최종 모델을 구축하였다.

2.2 세부 단계

2.2.1 데이터 라벨링

본 논문에서는 Tri-training 알고리즘을 활용한 반지도 학습을 통해 라벨을 붙이는 과정을 진행하였다. 반지도 학습은 라벨이 없는 데이터를 통합하는데 이점이 있으며,

라벨이 없는 데이터를 소수의 라벨이 있는 데이터와 함께 사용하는 것은 효과적으로 나타났다[11].

Tri-training은 세 개의 분류기를 사용하며 완전히 중복된 데이터셋이 필요하지 않고, 다른 지도 학습 알고리즘이 필요하지 않아 광범위하게 적용할 수 있다. Tri-training의 이러한 점은 알고리즘 효율성에 크게 기여하며 더 나은 일반화 능력을 얻을 수 있다[12].

그림 1과 같이 라벨링 된 세 개의 훈련 데이터셋을 각각의 분류기에 준비하고 세 분류기의 ‘다수결 방식’을 통해 라벨링이 되어있지 않은 나머지 데이터에 대한 라벨링을 진행한다. 예를 들어, 세 분류기 중 두 개의 분류기가 라벨링이 되어 있지 않은 데이터를 불법 바이럴 리뷰 데이터로 예측했을 경우, 다수결의 원칙에 따라 불법 바이럴 데이터로 분류된다. 이 데이터는 불법 바이럴로 예측하지 않았던 분류기의 훈련 데이터로 들어가 모델이 재훈련되도록 한다.

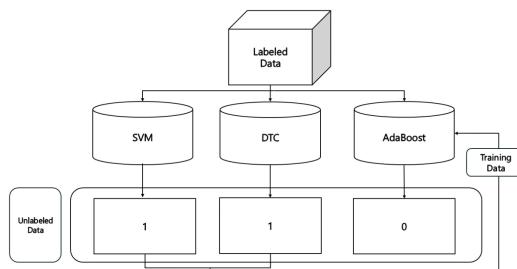


그림 2. Tri-training 알고리즘 과정을 시각화한 다이어그램.

Fig. 2. Diagram of the Tri-training Algorithm.

2.2.2 텍스트 및 네트워크 기반 모델

구축된 데이터셋은 텍스트 분석 방식과 네트워크 분석 두 가지 방식으로 분석되었다. 먼저, 언어·의미적 특성을 반영한 Korean ELECTRA(KoELECTRA) 모델을 사용하여 X의 불법 바이럴 리뷰 텍스트 분석을 통해 불법 여부를 판단하였다. KoELECTRA는

문장을 양방향 방식으로 문장의 전체 맥락을 이해할 수 있게 사전 훈련된 한국어 언어 모델인 KoBERT에 교체 토큰 탐지 방법을 결합한 모델이다[13][14].

또한, 사용자 간 네트워크 분석을 통해 추출한 다중 피처를 활용하여 불법 바이럴 포스트를 탐지하기 위해 Tabular Network(TabNet) 모델을 기반으로 진행하였다. TabNet은 정형 데이터에 적합하도록 설계된 딥러닝 모델로, Feature Transformer에서 각 스텝마다 중요한 피처를 선별하여 피처들의 중요성을 동적으로 평가가 가능한 모델이다[15].

네트워크 분석 모델에 사용된 다중 피처는 크게 두 가지로 분류하였다. 첫째, 명제를 수치화한 피처로 ‘Community’와 ‘AdRetweetCount(ARC)’를 설정하였다. Community는 네트워크 내 노드가 속한 그룹을 나타내는 범주형 변수이며, 노드 간의 연관성을 기반으로 계산된다. ARC는 각 계정이 불법 바이럴 포스트를 리포스트한 횟수를 의미한다.

둘째, 네트워크 구조적 특성을 기반으로 노드의 중심성과 영향력을 측정하는 피처로, ‘Betweenness Centrality(BC)’와 ‘Centrality Rank(CR)’를 설정하였다. BC는 노드가 다른 노드 쌍 간 최단 경로에서 차지하는 비율을 나타내며, 네트워크 내에서의 중개 역할이나 연결의 허브로서의 중요도를 측정한다. CR은 노드의 PageRank(PR) 값과 Degree Centrality(DC)를 결합하여 네트워크 내에서의 종합적인 영향력을 나타낸 지표이다[16][17].

2.2.3 소프트 보팅 기반 양상블

본 연구는 KoELECTRA와 TabNet을 소프트 보팅 기법을 활용하여 양상블하였다. 양상블 학습은 두 개 이상의 모델을 결합하여 보다 효과적인 모델을 만드는 학습 기법으로,

개별 모델의 강점을 활용하고 약점을 완화하여 다양한 작업에서 성능을 개선할 수 있다[18]. 양상블에는 보팅, 배깅, 부스팅 등의 다양한 기법이 존재하며 그 중 보팅 기법은 여러 모델을 통해 얻은 예측값들을 대상으로 다수결 투표를 하여 최종 클래스를 예측한다. 그중 소프트 보팅 기법은 각 모델이 예측한 확률과 가중치를 이용해 클래스별 최종 확률을 계산하며, 각 모델의 결과들을 유연하게 적용할 수 있다[19].

III. 실험

3.1 실험 데이터 구축

본 연구에서는 X의 리뷰 데이터와 네트워크 데이터를 수집하였다. 수집한 데이터는 총 9,014개로 계정ID, 포스트 내용, 리포스트 수, 리포스트 계정, 인용 계정의 데이터를 각각 수집하였다. 먼저, 불법 바이럴로 확인된 계정에서 800개의 라벨링 된 불법 바이럴 리뷰 데이터를 수집하였으며 광고 대가를 명시한 계정에서 800개의 라벨링 된 일반 리뷰 데이터를 수집하였다[20].

본 연구에서 진행한 데이터 전처리 과정은 다음과 같다. 먼저, 수집한 데이터에서 리뷰가 아닌 사람이 들어간 내용들은 제거하였다. 다음으로 불필요한 링크, 앤서인(@) 등을 삭제했고, X의 특성을 반영하는 해시태그는 남겨두었다. 또한 리뷰 텍스트를 토큰화하여 품사를 원형으로 변형한 뒤, 전처리된 단어 리스트를 하나의 문자열로 변환하여 최종적인 전처리 결과를 도출하였다. 최종 데이터셋은 1,600개의 라벨링 된 데이터와 7,414개의 라벨 없는 데이터로 구성되었다.

수집한 데이터 중 라벨이 없는 7,414개의 데이터는 Tri-training 알고리즘을 활용한

반지도 학습을 통해 라벨을 붙이는 과정을 진행하였다. 선행연구를 토대로 성능이 높은 분류기를 실험하여 Decision Tree Classifier(DTC), SVM, AdaBoost 3개의 분류기의 조합을 선정하였다[21][22][23]. 그 결과 각각의 성능은 88%, 87%, 80%로 나타났고 전체 성능이 87%로 나타났다.

3.2 모델링 및 분석

구축된 데이터셋은 언어·의미적 특성을 반영한 자연어 모델과 네트워크 특성을 반영한 모델 두 가지 방식으로 실험을 진행하였다.

첫째, KoBERT 기반 모델인 KoELECT-RA, 외의 KoRoBERTa, DistilKobert 모델에 대해 데이터셋을 훈련시킨 결과, KoELECTRA가 정확도와 Recall에서 가장 우수한 성능을 보여 KoELECTRA로 텍스트 분석을 진행하였다[24][25].

둘째, 불법 바이럴 포스트의 전파 패턴을 보기 위해 불법 바이럴 포스트를 전파한 계정들의 네트워크를 분석을 진행하였다. 선행연구를 기반으로 연구에 맞게 수정하여 “한 명의 X 유저가 다수의 광고성 트윗을 지속적으로 리포스트하며 공유한다.”, “X 유저들이 그룹을 만들어 광고성 포스트를 서로 리포스트하며 공유한다.”라는 두 가지

명제를 설정하였다[26]. 설정된 명제를 기반으로 노드는 계정, 엣지는 불법 바이럴 트윗을 리포스트 및 인용을 함으로써 형성된 관계로 설정하였으며 이러한 방식으로 구성된 네트워크는 그림 3과 같이 시각화된다.

그림 3에서는 한 명의 X 유저(노드)가 여러개의 엣지로 연결되어 있으며 같은 커뮤니티(그룹)를 나타내는 동일 색상의 노드가 클러스터를 형성하고 있다. 이를 통해 앞서 설정한 두 가지 명제를 증명하였다. 또한, 네트워크 분석의 결과 (네트워크 분석을 통해 증명된 명제를 바탕으로 추출된 다중 피처를 활용하였으며, TabNet 모델링을 통해 본 연구의 피처들의 중요성을 파악한 결과 Community, ARC, BC, CR 순으로 나타났다.

최종적으로는 KoELECTRA와 TabNet 모델을 소프트 보팅 기법을 활용하여 앙상블을 진행하였다. 일반적으로 두 모델의 평균을 계산하여 임계값을 0.5 기준으로 분류하지만, 본 연구에서는 임계값을 0.365로 수정하여 불법 바이럴 Re-call값을 올리고자 하였다.

3.3 실험 결과

모델 학습 결과, 언어적 및 의미적 특성을 분석하는 모델인 KoELECTRA의 경우 0.86의 정확도와 0.86의 Recall값을 가졌다. 네트워크를 분석하는 모델인 TabNet은 0.74의 정확도와 0.89의 Re-call값을 가졌다. 앙상블 모델의 경우 0.82의 정확도와 0.94의 Recall값을 가졌다. 표 1은 각 모델의 성능 지표를 정리한 것이다.

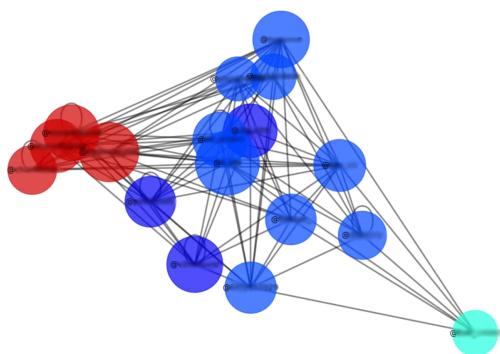


그림 3. 불법 바이럴 리뷰의 사용자 네트워크와 핵심 노드.

Fig 3. Network of Users and Influential Nodes in Illegal Viral Reviews.

표 1. 불법 바이럴 리뷰 판별 모델 성능

Table 1. Performance of the Illegal Viral Reviews Detection Model

Feature	Model	Accurancy	Precision	Recall	F-1 score
Linguistic + Semantic	KoELECTRA	0.86	0.86	0.86	0.86
Network	Tabnet	0.74	0.69	0.89	0.78
Linguistic + Semantic + Network	KoELECTRA + Tabnet	0.82	0.76	0.94	0.84

KoELECTRA는 가장 높은 정확도를 가졌으며, Recall값의 경우 KoELECTRA, TabNet을 소프트 보팅 앙상블로 결합한 모델이 가장 높은 값을 가졌다. 본 연구는 세 가지 모델의 성능 결과를 통해, 본 연구가 중점을 두었던 Recall이 가장 높은 앙상블 모델을 가장 성능이 좋은 모델로 판단하였다.

IV. 결론

본 연구는 언어적 및 의미적 특성 분석과 네트워크 특성 분석을 결합하여 불법 바이럴 리뷰를 효과적으로 탐지할 수 있는 모델을 구축하였으며, 실제 불법 바이럴 리뷰를 정확하게 탐지하기 위해 Recall 값이 가장 높은 모델을 찾는 것을 최우선적인 과제로 삼았다. 앙상블 모델은 리뷰 데이터에서 추출된 언어적 및 의미적 특성뿐만 아니라, 계정 간의 관계와 활동 패턴을 포함한 네트워크 특성을 모델에 통합함으로써 기존 선행연구의 탐지 방식의 한계를 극복하였다.

이러한 결과는 단순히 현재의 탐지 모델 개발에 활용될 뿐만 아니라, 향후 불법 바이럴 마케팅 진화에 대응할 수 있는 유연하고 강력한 탐지 시스템 구축의 기반을 마련하였으며, 소비자 보호와 공정한 온라인 리뷰 생태계 조성에 기여할 수 있는 중요한

학술적, 실용적 가치를 지니게 되었다.

V. 감사의 말

본 연구는 교육부 및 한국연구재단 대학혁신지원사업으로부터 지원받은 연구(2024)임을 알린다.

참고문헌

- [1] Liu X, Ren P, Xu Z, Xie W, Evolutive multi-attrib ute decision making with online consumer revie ws, Omega, 131, 103225, 2025.
- [2] Wang L, Che G, Hu J, Chen L, Online review helpfulness and information overload: The role s of text, image, and video elements, Journal of Theoretical and Applied Electronic Commerc e Research, 19, 2, 1243-1266, 2024.
- [3] W. Hasanat, A. Hoque, F. A. Shikha, M. Anwar, A. B. A. Hamid, and H. H. Tat, "The Impact of Coronavirus (COVID-19) on E-Business in Malaysia", AJMS, vol. 3, no. 1, pp. 85–90, Apr. 2020.
- [4] 김성용, 김동수, 외식 프랜차이즈 바이럴 마케팅이 소비자의 소비 행동에 관한 주관성 연구, 주관성연구, 61, 45-61, 2022.
- [5] 이선희, 인터넷 광고에 있어서 기만적 광고의 규제, 저스티스, 190, 255-285, 2022.
- [6] 이선희, 박희민, 머신러닝을 이용한 바이럴 원고 검출, 한국정보과학회 학술발표논문집, 2023, 1, 874-876, 2023.
- [7] 신동훈, 신우식, 김희웅, 작성자 언어적 특성 기반 가짜 리뷰 탐지 딥러닝 모델 개발, 정보시스템연구, 31, 4, 1-23, 2022.
- [8] Cano-Marin E, Mora-Cantallops M, Sanchez-A lonso S, The power of big data analytics over fake news: A scientometric review of Twitter as a predictive system in healthcare, Technologi cal Forecasting and Social Change, 190, 12238 6, 2023.
- [9] Mohawesh R., Bany Salameh H., Jararweh Y., Alkhalaileh M., Maqsood S., Fake review detecti

- on using transformer-based enhanced LSTM and RoBERTa, International Journal of Cognitive Computing in Engineering, 5, 250-258, 2024.
- [10] kavin B. P., Karki S., Hemalatha S., Singh D., Vijayalakshmi R., Thangamani M., Haleem S. L A, Jose D., Tirth V., Kshirsagar P. R., Adigo A. G., Machine learning-based secure data acquisition for fake accounts detection in future mobile communication networks, Wireless Communications and Mobile Computing, 635 6152, 2022.
- [11] Vale K.M.O., Gorgônio A C, Gorgônio F.D.L.E, Canuto A.M.D.P., An efficient approach to select instances in self-training and co-training semi-supervised methods, IEEE Access, 10, 7254-7276, 2022.
- [12] Meng F., Cheng W., Wang J., Semi-supervised software defect prediction model based on tri-training, KSII Transactions on Internet & Information Systems, 15, 11, 2021.
- [13] 강주영, 송민, 한국어 가짜 구매후기 생성과 탐지 성능 평가, 지능정보연구, 30, 2, 313-328, 2024.
- [14] Jang J. M., Min J. O., Noh H. S., KorPatELECTRA: A Pre-trained Language Model for Korean Patent Literature to improve performance in the field of natural language processing (Korean Patent ELECTRA), Journal of the Korea Society of Computer and Information, 27, 2, 15-23, 2022.
- [15] Arik S. Ö., Pfister T., TabNet: Attentive Interpretable Tabular Learning, Proceedings of the AAAI Conference on Artificial Intelligence, 35, 8, 6679-6687, 2021.
- [16] Kim H. J., Lee J. W., Two-layer PageRank centrality for dynamic networks with data integration, Journal of Network Theory, 18, 4, 355-370, 2024.
- [17] Zheng Y., Liu M., An enhanced measure of degree centrality in weighted complex networks, Complex Systems and Applications, 45, 1, 75-89, 2022.
- [18] Obaido G., An Improved Framework for Detecting Thyroid Disease Using Filter-Based Feature Selection and Stacking Ensemble, IEEE Access, vol 12, 89098-89112, 2024.
- [19] 전병욱, 강지수, 정경용, 도로교통 이머징 리스크 탐지를 위한 AutoML과 CNN 기반 소프트 보팅 양상을 분류 모델, 융합정보논문지, 11, 7, 14-20, 2021.
- [20] 백현미, 이새롬, 백혜진, 광고성 트윗과 비광고성 트윗의 확산 특성 비교 연구: 전자담배 트윗을 중심으로, 정보사회와 미디어, 20, 1, 45-67, 2019.
- [21] Mir A.Q., Khan F.Y., Chishti M.A., Online Fake Review Detection Using Supervised Machine Learning And BERT Model, ArXiv, 2301, 03225, 2023.
- [22] Ekramipooya, A., Boroushaki, M., & Rashtchian, D., Application of natural language processing and machine learning in prediction of deviations in the HAZOP study worksheet: A comparison of classifiers, Process Safety and Environmental Protection, 176, 65-73, 2023.
- [23] Akhtari C. G., Vyhnavi B., Deekshitha D., Advancement in NLP with Decision Tree: The Impact of social media on Enhancing Women's Safety in Indian Cities, Journal of Science & Technology, 8, 12, 195-207, 2023.
- [24] Kim T. Y., Baek S. U., Lim M. H., Yun B., Paek D., Zoh K. E., Yoon J. H., Occupation classification model based on DistilKoBERT: using the 5th and 6th Korean Working Condition Survey s, Annals of Occupational and Environmental Medicine, 36, 2024.
- [25] Singla A., Roberta and BERT: Revolutionizing Mental Healthcare through Natural Language, Shodh Sagar Journal of Artificial Intelligence and Machine Learning, 1, 1, 10-27, 2024.
- [26] Javed D., Jhanjhi N. Z., Khan N. A., Ray S. K., Al Mazroa A., Ashfaq F., Das S. R., Towards the future of bot detection: A comprehensive taxonomical review and challenges on Twitter /X, Computer Networks, 254, 110808, 2024.