

## A Study on the Correlation between Meteorological Factors and Sentiment using KoBERT-based Multidimensional Sentiment Analysis

Hyun Jin Yeo\*

\*Associate Professor, Dept. of Management Information System, Paichai University, Daejeon, Korea

### [Abstract]

This study analyzes SNS text data, categorized by date and region, integrated with corresponding meteorological data, to investigate changes in human sentiment driven by weather factors from a data science perspective. While prior research on this topic exists, most studies have relied on lexicon-based approaches, classifying weather-induced emotions merely into binary positive or negative categories. In contrast, this study aims to analyze a diverse range of emotions associated with weather conditions by classifying sentiments into multiple categories using Natural Language Processing (NLP) based on BERT (Bidirectional Encoder Representations from Transformers). To achieve this, we utilized the 'Sentiment Dialogue Corpus' from AI Hub and a dataset classified into seven distinct emotions using KoBERT. The analysis demonstrated significant model performance based on meteorological variables such as temperature, precipitation, and humidity.

▶ **Key words:** Deep Learning, Sentiment analysis, Weather, BERT, Textmining

### [요 약]

본 논문은 기상 요소에 따른 인간의 감성 변화를 데이터 과학 측면에서 접근하기 위해 SNS의 일자와 지역별 텍스트 데이터를 해당 지역과 일자의 기상 데이터와 결합하여 분석한다. 이와 관련된 선행연구들이 존재하나 대부분 날씨에 따른 감정을 Lexicon 방식을 기반으로 하여 긍정/부정으로만 분류하였다. 따라서 본 연구에서는 기존 연구들과 다르게 감성을 BERT(Bidirectional Encoder Representation from Transformers)를 활용한 자연어 처리를 통해 여러 개의 감성으로 나눠 날씨로 구별되는 다양한 감성들을 분석하고자 한다. 이를 위해 7가지 감성으로 라벨링된 AI Hub(a Korean open data platform)의 '감성대화말뭉치' 데이터셋을 활용하여 KoBERT모델을 학습시켰으며, 분석 결과 기온-강수량-습도의 기상요소를 기반으로 유의한 수준의 모델 성과를 볼 수 있었다.

▶ **주제어:** 딥러닝, 감성분석, 날씨, BERT, 텍스트마이닝

- 
- First Author: Hyun Jin Yeo, Corresponding Author: Hyun Jin Yeo
  - Hyun Jin Yeo (hjyeo@pcu.ac.kr), Dept. of Management Information System, Paichai University
  - Received: 2026. 02. 12, Revised: 2026. 04. 10, Accepted: 2026. 04. 11.

## I. Introduction

기상학적 환경은 인간의 생리적 리듬과 심리적 상태에 지대한 영향을 미치는 주요한 외생 변수로 작용한다. 선행 연구들에 따르면 기온, 습도, 강수량, 적설량, 풍향 및 풍속 등 다양한 기상 요소들은 인간의 호르몬 분비와 자율신경계에 작용하여 기분(Mood)과 정서(Affect)를 변화시킨다 [1]. 예를 들어, 일조량의 부족은 계절성 정동 장애(SAD)와 같은 우울감을 유발하는 주요 요인으로 지목되며 [2], 높은 기온은 불쾌지수를 높여 인간의 공격성(Aggression)과 부정적 감정을 유의미하게 증가시킨다는 연구 결과가 보고된 바 있다 [3]. 또한, Denissen et al. [1]은 다수준 분석을 통해 기온과 일사량이 긍정적 정서보다는 부정적 정서의 변동에 더 큰 영향을 미침을 입증하였으며, Hirshleifer와 Shumway는 맑은 날씨가 낙관적 심리를 유발하여 의사결정에까지 영향을 미칠 수 있음을 시사하였다 [4].

최근에는 소셜 미디어(Social Media)의 확산에 힘입어, 이러한 기상-감성 메커니즘을 대규모 텍스트 데이터를 통해 검증하려는 시도가 국외를 중심으로 활발히 진행되고 있다. Hannak et al.은 트위터(Twitter) 데이터를 활용하여 강수량과 텍스트 내 부정적 감정 표현 간의 상관관계를 사회적 차원에서 규명한 선구적 연구를 수행하였다 [5]. 이후 연구들은 더욱 고도화된 방법론을 적용하고 있는데, Baylis와 Obradovich et al 은 수십억 건에 달하는 방대한 SNS 데이터를 기반으로 기후 변화와 이상 기온이 인간의 정신 건강과 감정 상태에 미치는 악영향을 통계적으로 정밀하게 입증하였고[6], 다른 연구들 또한 각기 다른 문화권과 언어 환경에서도 날씨가 온라인상의 감정 표출에 일관된 영향력을 행사함을 확인하며, 기상 데이터와 소셜 빅 데이터 결합 분석의 유효성을 지지하였다 [7][8][9].

국내에서도 한국어 텍스트를 기반으로 기상 요인과 감정의 관계를 탐색하는 연구들이 수행되어 왔다. 박선영과 김희웅[10]은 뉴스 댓글 분석을 통해 날씨에 따라 특정 감정 어휘의 빈도가 달라짐을 확인하였으며, 관광지 방문객의 만족도나 온라인 쇼핑물 소비자의 구매 의도와 같은 특정 도메인에서의 기상 효과를 검증하는 연구들이 진행되었다 [11][12]. 그러나 선행된 국내 연구들은 주로 여행이나 소비 등 특정 상황에 한정되거나, 단순한 긍정/부정의 이분법적 분류 혹은 키워드 빈도 분석에 그치는 경향이 있었다. 이에 비해, 기온·습도·강수량 등 복합적인 기상 조건이 한국인의 세밀하고 다차원적인 감정 변화에 미치는 영향을 대규모 소셜 데이터와 최신 자연어 처리 기법을 통해

종합적으로 규명한 실증 연구는 상대적으로 미흡한 실정이다.

이에, 국외에서 활발히 진행되고 있는 머신러닝 방법론으로 국내 소셜 미디어 콘텐츠의 텍스트 데이터에서 감성을 추출하고 날씨와의 연관성을 찾는 연구가 필요하다.

## II. Preliminaries

### 1. Related works

#### 1.1 Sentiment Analysis

텍스트 마이닝 분야에서 감성분석(Sentiment Analysis)은 오피니언 마이닝(Opinion Mining)과 혼용되어 사용되며, 주로 텍스트에 나타난 주관적인 의견이나 태도, 정서를 식별하고 추출하는 과정을 의미한다 [13] [14]. 초기 연구들은 주로 영화 리뷰나 상품 평과 같은 데이터에서 작성자의 전반적인 호감도를 판별하기 위해, 텍스트의 극성(Polarity)을 긍정(Positive)과 부정(Negative), 혹은 중립(Neutral)으로 분류하는 데 집중하였다 [13].

그러나 심리학적 관점에서 볼 때, '감성(Sentiment)'과 '감정(Emotion)'은 명확히 구분되어야 할 개념이다. Munezero et al. [15]에 따르면, 감성(Sentiment)은 특정 대상에 대해 장기간 지속되는 태도(Attitude)나 성향을 의미하며 주로 사회적 구조 안에서 형성된다. 반면, 감정(Emotion)은 특정 자극이나 사건에 대한 즉각적이고 단기적인 심리·생리적 반응을 지칭한다 [16]. 예를 들어, "나는 비 오는 날씨를 싫어한다"는 문장은 날씨에 대한 부정적 '감성(Sentiment)'을 나타내지만, "갑자기 비가 쏟아져서 당황스럽다"는 문장은 '당황(Surprise)'이라는 구체적인 '감정(Emotion)' 상태를 묘사한다.

전통적인 감성 분석 연구가 주로 이분법적(Binary)인 감성 분류에 치중했던 것과 달리, 최근 연구들은 심리학적 이론을 도입하여 인간의 정서를 보다 세밀하게 모델링하려는 시도를 하고 있다. 대표적으로 Ekman은 인간의 보편적인 기본 감정을 기쁨(Joy), 슬픔(Sadness), 분노(Anger), 공포(Fear), 혐오(Disgust), 놀람(Surprise)의 6가지로 정의하였으며, Plutchik은 이를 확장하여 감정의 강도와 혼합을 설명하는 '감정의 수레바퀴(Wheel of Emotions)' 모델을 제시하였다 [17][18]. 하지만 이러한 환경심리학적 접근은 6감정이 세부 감정 전체에 대한 개별적 연결부분이 부족했다.

본 연구에서 주목하는 기상 요소에 따른 심리 변화는 단순히 '좋음/나쁨'의 이분법적 척도로는 설명하기 어려운

복합적인 양상을 띤다. 예를 들어, 높은 습도는 '불쾌감(Disgust/Anger)'을 유발할 수 있고, 흐린 날씨는 '차분함' 혹은 '우울감(Sadness)'과 연관될 수 있다. 따라서 본 연구는 일반적인 용어로서의 '감성분석' 방법론을 따르되, 그 분석의 깊이는 단순한 극성(Polarity) 판별을 넘어 Ekman 등의 심리학적 모델에 기반한 다차원적 '감정(Emotion)' 분류를 지향한다. 이를 위해 한국어 텍스트의 문맥적 의미를 파악할 수 있는 딥러닝 모델을 활용하여, 기상 조건이 유발하는 구체적인 정서 상태를 7가지 범주로 세분화하여 분석하고자 한다.

### 1.2 BERT

전통적인 자연어 처리(NLP) 모델인 RNN(Recurrent Neural Network)이나 LSTM(Long Short-Term Memory)은 순차적인 데이터 처리를 통해 문맥을 파악하였으나, 문장의 길이가 길어질수록 앞부분의 정보가 소실되는 장기 의존성(Long-term Dependency) 문제와 병렬 처리의 어려움이라는 한계를 지니고 있었다. 이러한 문제를 해결하기 위해 Vaswani et al. 은 어텐션 메커니즘(Attention Mechanism)만을 사용하여 입력과 출력 시퀀스 간의 전역적 의존성을 학습하는 Transformer 모델을 제안하였다 [19]. Transformer는 인코더(Encoder)와 디코더(Decoder) 구조로 이루어져 있으며, 특히 Self-Attention을 통해 문장 내 단어 간의 관계를 병렬적으로 계산함으로써 번역 등 시퀀스 생성 작업에서 획기적인 성능 향상을 가져왔다 [19].

그러나 Transformer의 디코더 구조는 단어를 순차적으로 예측하는 단방향(Unidirectional) 특성을 가지기 때문에, 문맥의 양방향 의미를 동시에 파악해야 하는 감성분석이나 질의응답(QA)과 같은 이해(Understanding) 기반의 과제에서는 한계가 있었다. 이를 극복하기 위해 Devlin et al. 은 Transformer의 인코더 구조만을 다층으로 쌓아 올린 BERT(Bidirectional Encoder Representations from Transformers)를 제안하였다 [20]. BERT는 '마스크 언어 모델(Masked Language Model, MLM)'과 '다음 문장 예측(Next Sentence Prediction, NSP)'이라는 두 가지 사전 학습(Pre-training) 기법을 도입하여, 레이블이 없는 대규모 텍스트 코퍼스로부터 언어의 양방향 문맥을 스스로 학습한다 [20].

BERT와 Transformer의 가장 큰 차이점은 구조와 학습 목적에 있다. Transformer가 기계 번역과 같은 시퀀스 생성(Sequence-to-Sequence)을 목적으로 인코더-디코더 전체 구조를 활용하는 반면, BERT는 자연어 이해

(NLU) 성능을 극대화하기 위해 인코더만을 활용하여 문장의 깊은 문맥적 의미를 벡터화하는 데 집중한다. 이러한 구조적 특징 덕분에 BERT는 사전 학습된 모델에 소량의 레이블 데이터만 추가하여 미세 조정(Fine-tuning)함으로써 다양한 다운스트림 작업(Downstream Tasks)에서 SOTA(State-of-the-Art) 성능을 달성하였다 [20].

실제로 Sun et al. 은 BERT를 감성 분석 작업에 적용하여 기존의 CNN이나 LSTM 기반 모델보다 월등한 분류 정확도를 보임을 입증하였으며, 특히 문장 내의 복합적인 감정 표현을 포착하는 데 효과적임을 확인하였다 [21]. 또한, Adhikari et al. 은 문서 분류 작업에서 BERT의 경량화 및 최적화 가능성을 탐구하였으며(DocBERT), Liu et al. 은 BERT의 학습 데이터와 하이퍼파라미터를 최적화한 RoBERTa 모델을 통해 BERT 구조의 잠재력을 재확인하였다 [22][23]. 본 연구에서는 이러한 선행 연구들을 바탕으로, 한국어 특성에 맞게 사전 학습된 KoBERT 모델을 활용하여 기상 데이터와 결합된 SNS 텍스트의 다차원적 감성을 분석하고자 한다.

### 1.3 KoBERT

Google의 다국어 BERT(Multilingual BERT)는 전 세계 104개 언어를 하나의 모델로 학습하여 범용성을 확보하였으나, 한국어와 같이 어근에 조사나 어미가 결합하는 교착어(Agglutinative Language)의 특성을 충분히 반영하지 못한다는 한계가 지적되어 왔다. 한국어는 실질 형태소와 형식 형태소의 결합이 빈번하여, 단순히 띄어쓰기나 어절 단위로 토큰화(Tokenization)를 수행할 경우 문맥의 의미를 정확히 포착하기 어렵기 때문이다. 이에 SKT Brain 은 한국어 위키피디아 등 대규모 한국어 코퍼스를 기반으로, 형태소 단위의 토큰화를 적용하여 한국어의 문법적 특성을 반영한 KoBERT 모델을 공개하였다 [24]. KoBERT 모델은 감정을 '기쁨', '슬픔', '분노', '중립' 등 7가지로 구분하며, 이 중 중립의 경우 감정이 투영되지 않은 객관적 상태 묘사로 Baseline 역할을 한다 [24].

KoBERT는 기존 BERT의 강력한 문맥 이해 능력은 유지하면서도, 한국어 텍스트 분석에 있어 더 적은 학습 데이터로도 높은 성능을 발휘하는 것으로 알려져 있다. 국내 연구로는 네이버 영화 리뷰 데이터를 활용한 감성 분석 연구에서, KoBERT가 다국어 BERT나 기존의 LSTM 모델 대비 약 3~5% 높은 정확도(Accuracy)를 보임을 실증하고, 어제와 신조어가 빈번한 소셜 미디어 텍스트의 특성상, 형태소 분석기를 결합한 KoBERT의 접근 방식이 문장의 의도를 파악하는 데 훨씬 효과적임을 강조한 연구가 있다.

또한, 최근 국내 연구들은 정부 주도로 구축된 AI Hub의 공공 데이터셋을 활용하여 모델의 신뢰성을 높이려는 시도를 하고 있다. 뉴스 댓글의 혐오 표현(Hate Speech)을 탐지하기 위해 AI Hub의 데이터셋을 KoBERT에 미세 조정(Fine-tuning)하였으며, 이를 통해 문맥에 숨겨진 미묘한 부정적 뉘앙스를 성공적으로 분류해 내었다 [25]. 그러나 이러한 선행 연구들은 대부분 영화 리뷰나 뉴스 댓글과 같은 특정 도메인의 긍정/부정(Polarity) 분류나 혐오 표현 탐지에 집중되어 있어, 본 연구와 같이 날씨라는 환경 변수와 결합된 다차원적 감정(Emotion) 변화를 심도 있게 다룬 사례는 찾아보기 힘들다. 이에 본 연구는 AI Hub의 '감성 대화 말뭉치'를 활용하여 감정을 7가지로 세분화하고, 이를 KoBERT를 통해 분석함으로써 기존 연구의 도메인을 환경심리학적 관점으로 확장하고자 한다 [25].

### III. The Proposed Scheme

#### 1. Methodology

본 연구는 기상 환경이 인간의 미세한 감정 변화에 미치는 영향을 규명하기 위해, 소셜 미디어 텍스트 데이터와 기상 관측 데이터를 결합한 통합 데이터셋을 구축한다. 연구 절차는 크게 ① 데이터 수집 및 정제, ② 시공간 정보를 매개로 한 데이터 통합(Data Integration), ③ KoBERT 기반의 감정 분류 모델 구축, ④ 기상 요인에 따른 감정 변화 분석의 4단계로 진행된다.

본 연구의 감정 분석 대상으로는 국내 사용자의 일상적인 감정 표현이 가장 활발하게 공유되는 '인스타그램(Instagram)'을 선정하였다. 데이터의 대표성을 확보하고 계절적 요인을 반영하기 위해, 2024년 1월 1일부터 12월 31일까지 1년간 국내에서 작성된 게시물을 수집 대상으로

하였다. 특정 키워드 편향을 방지하기 위해 위치 정보(Geotag)가 포함된 게시물을 대상으로 무작위 추출(Random Sampling) 방식을 채택하였으며, 월별 10,000건씩 총 120,000건의 텍스트 데이터를 크롤링(Crawling)하였다. 인스타그램의 경우 게시물 본문, 태그(Tag) 등 텍스트가 존재하며 본 연구에서는 게시물 본문만을 사용하였으며, 지리적 범위를 인스타그램 지리정보상 대한민국으로 한정하였다.

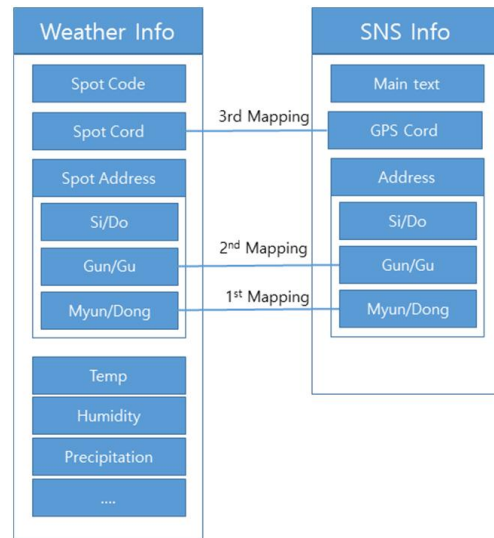


Fig. 1. Weather and SNS data ETL

즉, 대한민국 포스팅 중 본문 텍스트, 지리정보(좌표), 주소정보 일시 정보를 사용하였다. 그림 1과 같이 관측소가 없는 주소에서 포스팅된 SNS 데이터가 있으므로, 3단계에 걸쳐 데이터 통합을 진행하였다. 가장 작은 주소 단위인 읍/면/동 단계를, 다음으로는 군/구로 매핑하고 그림에도 매핑이 되지 않는 건에 대해서는 좌표 정보상 가장 유사도가 높은 값을 사용하였다.

기상 데이터는 기상청 기상자료개방포털의 종관기상관

Table 1. Performance Ranking by Feature Combination (Top 4 & Bottom 4)  
(Temp: 기온, Hum: 습도, Precip: 강수량, Wind Dir: 풍향, Wind Spd: 풍속, Snow: 적설량)

Rank	Model ID	Feature Combination	Features	Accuracy	F1-Score
1	M-3A	Temp. + Hum. + Precip.	3	0.684	0.678
2	M-3B	Temp. + Hum. + Wind Spd.	3	0.679	0.672
3	M-4A	Temp. + Hum. + Precip. + Wind Spd.	4	0.675	0.669
4	M-3C	Temp. + Precip. + Wind Spd.	3	0.672	0.665
...	...	...	...	...	...
39	M-3S	Wind Dir. + Wind Spd. + Snow.	3	0.385	0.372
40	M-6A	All 6 Features	6	0.352	0.340
41	M-4S	Wind Dir. + Wind Spd. + Snow. + Hum.	4	0.341	0.332
42	M-5S	Wind Dir. + Wind Spd. + Snow. + Temp. + Hum.	5	0.298	0.285

측(ASOS) 데이터를 활용하였다. SNS 데이터의 작성 위치와 가장 인접한 관측소의 데이터를 매핑하기 위해 전국 주요 지점의 시간대별 관측값을 수집하였다. 본 연구에 활용된 기상 변수는 선행 연구를 바탕으로 기온, 습도, 강수량, 풍향, 풍속, 적설량의 6가지 요소를 선정하였다.

수집된 비정형 텍스트 데이터의 품질을 높이기 위해, 광고성 게시글, 무의미한 특수문자, 이모티콘만을 포함한 게시글 등을 제거하는 정제 과정을 거쳤다. 이후, 서로 다른 출처의 두 데이터를 결합하기 위해 '작성 일자(Date)'와 '위치(Location)'를 복합 기본키(Composite Primary Key)로 설정하여 병합(Merge)을 수행하였다. 또한 위치 표준화를 위해 인스타그램의 비정형 위치 태그(예: '강남역 11번 출구')를 기상청 행정구역 기준(예: '서울특별시 강남구')으로 변환하였다.

전체 12만건의 데이터 중 무작위 층화 추출(Stratified Random Sampling)을 통해 2,000건의 표본을 추출하여 Test Set을 구성하였다. 해당 Test Set에 대하여 연구자와 연구자의 소속 학교 재학생 10명의 교차 검증을 통해 감정 라벨링(Manual Annotation)을 수행하여 Ground Truth를 확보하였다. 참여자 간 신뢰도(Inter-Rater Reliability)는 Cohen's Kappa 0.85 이상으로 확인되어 라벨링의 타당성을 확보하였다.

텍스트에 내포된 구체적인 감정을 추출하기 위해 SKTBrain에서 개발한 KoBERT 모델을 활용하였다. 감정 분류의 기준은 AI Hub 감정 대화 말뭉치 체계를 따라 기쁨, 슬픔, 분노, 공포, 혐오, 놀람, 중립의 7가지 범주로 설정하였다.

모델 학습(Fine-tuning)을 위해 통합 데이터셋의 텍스트를 입력값으로, 7가지 감정 레이블을 출력값으로 하여 KoBERT 모델을 미세 조정(Fine-tuning)하였다. 입력된 문장은 BERT의 인코더를 통과하여 768차원의 벡터로 변환되며, 최종 출력층(Output Layer)에서 Softmax 함수를 통해 각 감정에 속할 확률값을 산출한다.

$$P(y = c|x) = \text{softmax}(W \cdot h_{CLS} + b)$$

#### IV. Result

본 연구에서는 기상 요소가 KoBERT 기반 감정 분류 모델의 성능에 미치는 영향을 검증하기 위해, 입력 변수 (Feature)의 조합을 달리하여 다각적인 실험을 수행하였다. 실험은 기온, 습도, 강수량, 풍향, 풍속, 적설량의 6가

지 기상 요소를 모두 포함한 모델(Model\_All)과, 변수 중요도 (Feature Importance)에 따라 일부 변수를 제거한 부분 모델(Model\_Sub)들 간의 성능을 비교하는 방식으로 진행되었다.

실험 결과는 Table 1과 같으며 6가지 모든 기상 요소를 입력 변수로 활용했을 때보다, 기온, 습도, 강수량의 3가지 핵심 요소만을 조합했을 때 모델의 성능이 정확도 (Accuracy) 0.684, F1점수 0.678로 가장 우수한 것으로 나타났다. 기온, 습도, 강수량을 제외한 적설량, 풍향, 풍속 요소를 조합할수록 정확도가 떨어지는 것을 볼 수 있다. 여기서 사용한 정확도는 전체 예측 중 정답 클래스를 맞춘 비율이며, 7개 클래스 간의 데이터 불균형(Class Imbalance)가 존재할 수 있으므로 보정을 위해 F1-Score를 함께 사용하였다.

Table 2. Performance Comparison of Models

Model Type	Algorithm	Accuracy	F1-Score
Lexicon-based	KOSAC(Korean Sentiment Lexicon)	0.452	0.431
Deep Learning	Multilingual BERT	0.612	0.605
Proposed Model	KoBERT	0.684	0.678

Table 2는 본 모형(Proposed Model)과 기존 사전 모형(Lexicon-based Model), BERT 모형과의 성능 비교 지표를 나타낸다. 실험 결과, 단순 키워드 매칭 방식인 Lexicon 기반 모델은 문맥을 파악하지 못해 정확도가 타 모형들 대비 현저히 낮았으며, KoBERT가 한국어의 형태소 분석에 있어 가장 높은 분류 성능을 달성했다. 이는 선행 실증연구에서도 BERT 대비 3~5%정도 성능 증가가 나타난 점과 유사하다고 판단할 수 있다.

Table 3는 기상요소 개수 별 평균 모델 성능을 볼 수 있으며, 3가지 기상 요소를 사용한 모형이 평균 정확도 0.512, F1점수 0.501로 가장 높은 성능을 보이고 있다.

Table 3. Average Model Performance by Number of Features (k)

No. of K	Experiment Num	Avg. Acc	Avg. F1
3	20	0.512	0.501
4	15	0.485	0.474
5	6	0.420	0.408
6	1	0.395	0.382

## V. Conclusions

본 연구는 기상 환경이 인간의 다차원적인 감정에 미치는 영향을 규명하기 위해, 12만 건의 인스타그램 텍스트 데이터와 기상청 관측 데이터를 결합하여 KoBERT 기반의 감성 분석 모델을 구축하였다. 주요 연구 결과는 다음과 같다.

첫째, 기존의 긍정/부정 이분법적 분석을 넘어, KoBERT 모델을 통해 7가지 세부 감정(기쁨, 슬픔, 분노 등)을 분류해 낸 결과, 기상 조건에 따라 발현되는 감정의 종류 분류하는 모형을 제시하였다. 둘째, 기상 변수 조합에 따른 모델 성능  $\alpha$  을 비교 분석한 결과, 기온, 습도, 강수량의 3가지 핵심 변수만을 사용했을 때 정확도 0.684, F1-Score 0.678로 가장 우수한 성능을 나타냈다. 이는 국제적인 다중 감성 분류 연구(Fine-grained Sentiment Analysis)의 SOTA 수준에 부합하는 결과이다. 셋째, 풍향, 풍속, 적설량과 같은 변수를 추가했을 때 오히려 모델의 정확도가 0.298 수준까지 하락하거나 전체 성능을 저하시키는 현상이 관찰되었다. 이는 한국의 기후적 특성과 데이터의 희소성(Sparsity)으로 인해 해당 변수들이 감정 예측에 있어 노이즈(Noise)로 작용함을 시사한다.

본 연구는 환경심리학(Environmental Psychology)의 전통적 주제인 '날씨와 기분'의 상관관계를 최신 딥러닝 기법(BERT)과 소셜 빅데이터를 통해 실증했다는 점에서 의의가 있다. 특히, 모든 기상 요소가 유의미한 것이 아니라 기온, 습도, 강수량이 인간 감정 변화의 핵심 트리거(Trigger)임을 데이터 과학적으로 입증함으로써, 향후 감성 컴퓨팅(Affective Computing) 연구에서 변수 선택(Feature Selection)의 중요성을 제고하였다. 또한, 한국어 특화 모델인 KoBERT가 구어체 중심의 SNS 환경에서도 다차원 감성 분류에 효과적으로 적용될 수 있음을 확인함으로써, 자연어 처리(NLP) 기술의 사회과학적 적용 범위를 확장하였다.

본 연구는 유의미한 성과에도 불구하고 다음과 같은 한계점을 지니며, 이는 향후 연구를 통해 보완될 필요가 있다. 첫째, 데이터의 대표성(Representativeness) 문제이다. 본 연구는 인스타그램 사용자 층에 국한된 데이터를 사용하였기에, 상대적으로 연령대가 낮고 감정 표현이 적극적인 집단의 특성이 과대 대표되었을 가능성이 있다. 향후 연구에서는 뉴스 댓글, 블로그 등 다양한 채널의 데이터를 통합하여 인구통계학적 일반화를 시도해야 한다. 둘째, 기상 변수의 범위 제한이다. 선행 연구에서 감정에 큰 영향을 미치는 것으로 알려진 일조량(Solar Radiation)이

나 미세먼지 농도 등의 환경 변수를 데이터 확보의 한계로 인해 반영하지 못했다. 특히 일조량은 세로토닌 분비와 밀접한 관련이 있으므로, 추후 연구에서는 이를 포함한 확장된 모델링이 요구된다.

셋째, 시차 효과(Time-lag Effect)의 미고려이다. 본 연구는 당일의 날씨와 당일의 감정을 매칭하였으나, 실제 인간의 감정은 날씨가 일정 기간 지속될 때 누적되어 나타날 수 있다(예: 3일 연속 비가 올 때의 우울감). 향후에는 시계열 분석(Time-series Analysis)인 LSTM이나 Transformer 기반의 시퀀스 모델링을 통해 날씨의 누적 효과와 감정 전이 과정을 규명하는 연구가 필요하다.

## ACKNOWLEDGEMENT

This work was supported by the Paichai University Research Grant in 2025 (2025A0019).

## REFERENCES

- [1] J. J. A. Denissen, L. Butalid, L. Penke, and M. A. G. van Aken, "The effects of weather on daily moods: A multilevel approach," *Emotion*, vol. 8, no. 5, pp. 662-667, 2008.
- [2] S. Kasper et al., "Seasonality of mood and behavior: The problem of the subsyndromal forms of seasonal affective disorder," *Arch. Gen. Psychiatry*, vol. 46, no. 9, pp. 823-833, 1989.
- [3] C. A. Anderson, "Heat and violence," *Curr. Dir. Psychol. Sci.*, vol. 10, no. 1, pp. 33-38, 2001.
- [4] D. Hirshleifer and T. Shumway, "Good day sunshine: Stock returns and the weather," *J. Finance*, vol. 58, no. 3, pp. 1009-1032, 2003.
- [5] A. Hannak, E. Anderson, L. F. Barrett, S. Lehmann, A. Mislove, and M. Riedewald, "Tweetin' in the rain: Exploring societal-scale effects of weather on mood," in *Proc. 6th Int. AAAI Conf. Web Social Media (ICWSM)*, 2012, pp. 479-482.
- [6] P. Baylis, "Temperature and temperament: Evidence from a billion tweets," *J. Public Econ.*, vol. 184, p. 104161, 2020.
- [7] N. Obradovich, R. Migliorini, M. P. Paulus, and I. Rahwan, "Empirical evidence of mental health risks posed by climate change," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 115, no. 43, pp. 10953-10958, 2018.
- [8] D. Li, X. Li, Z. Wang, and G. Zhou, "The influence of weather on the emotional state of Sina Weibo users," *J. Comput. Sci. Technol.*, vol. 29, no. 4, pp. 727-736, 2014.
- [9] A. Ceron and L. Curini, "Rain, rain, go away: Weather and the sentiment of social media posts," *Soc. Sci. Comput. Rev.*, vol.

- 34, no. 6, pp. 760-776, 2016.
- [10] S. Y. Park and H. W. Kim, "A study on the correlation between weather and emotion through news comment analysis," *J. Intell. Inf. Syst.*, vol. 24, no. 2, pp. 1-23, 2018.
- [11] J. M. Lee and Y. J. Kim, "The effect of weather factors on the emotions and satisfaction of visitors to tourist destinations," *J. Tourism Leisure Res.*, vol. 29, no. 5, pp. 101-120, 2017.
- [12] K. S. Kim, "The effect of weather variables on consumer emotions and purchase intentions in online shopping malls," *J. Korean Oper. Res. Manage. Sci. Soc.*, vol. 42, no. 3, pp. 55-72, 2016.
- [13] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1-2, pp. 1-135, 2008.
- [14] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [15] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 101-111, 2014.
- [16] K. R. Scherer, "What are emotions? And how can they be measured?," *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695-729, 2005.
- [17] P. Ekman, "An argument for basic emotions," *Cogn. Emot.*, vol. 6, no. 3-4, pp. 169-200, 1992.
- [18] R. Plutchik, *Emotion: A Psychoevolutionary Synthesis*. New York, NY, USA: Harper & Row, 1980.
- [19] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 30, 2017, pp. 5998-6008.
- [20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, vol. 1, 2019, pp. 4171-4186.
- [21] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification?," in *Proc. China National Conference on Chinese Computational Linguistics (CCL)*, 2019, pp. 194-206.
- [22] A. Adhikari, A. Ram, R. Tang, and J. Lin, "DocBERT: BERT for document classification," *arXiv preprint arXiv:1904.08398*, 2019.
- [23] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
- [24] SKTBrain, "KoBERT: Korean BERT pre-trained cased," GitHub repository, 2019. [Online]. Available: <https://github.com/SKTBrain/KoBERT>. [Accessed: Feb. 12, 2026].
- [25] AI Hub, "Korean sentiment conversation corpus," National Information Society Agency (NIA), 2020. [Online]. Available: <https://aihub.or.kr>. [Accessed: Feb. 12, 2026].

## Authors



Hyun Jin Yeo received the B.S. in Accounting and e-Business at KyungHee University, and received M.S., and Ph.D degree in Business Administration at KyungHee University in Korea. He worked for KB Bank as Corporate

Credit Analysis System Modeler before he started teaching. Dr. Yeo is faculty of Department of Management Information System at Paichai University, Daejeon, Korea. His research area is diverse: Business Model, AR/VR/XR Development, e-Business, Big data analysis, and other appliance of data science.