

---

# COSE474-2024F: Final Project Proposal

## “Sentiment analysis using text”

---

2022320149 Subeen Ho

### 1. Introduction

In recent times, many individuals struggle to understand their own emotions, which often leads to mental health issues. A sentiment analysis model could help people become more aware of their emotional states, thereby preventing the onset of mental health problems.

According to a report by the World Health Organization (WHO), mental health conditions have been rising globally, with one of the contributing factors being an individual's inability to recognize and manage their emotions. Studies have also shown that self-awareness of emotional states is linked to better mental well-being, and tools that assist in this can have a preventive effect on emotional distress (WHO, 2022).

### 2. Problem definition & challenges

In this sentiment analysis project, the goal is to input text, determine whether the sentiment is positive or negative, and based on this, assess if a person's depressive level is high if negative texts persist for several days. By doing so, the project aims to provide useful insights for healthcare by categorizing positive or negative feedback regarding people's emotions.

#### 1. Interpretation by Domain

Words or phrases that convey sentiment can be interpreted differently depending on the domain. For example, the word “crash” may have a negative connotation in the stock market, but in video game communities, it can be used in a neutral or even positive context. These domain-specific characteristics make it challenging to generalize models across different fields.

#### 2. Handling Negations and Contextual Information

Handling negations and contextual information is another important challenge. Negation words like “not” or “never” can completely reverse the sentiment of a sentence. For instance, “This product is not bad” conveys a positive sentiment, even though it contains the word “bad.” Capturing such nuances requires sophisticated language understanding.

I aim to use the KoBERT model to infer the meaning of words within their contextual usage.

### 3. Related Works

#### 1. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis

This paper explores the application of BERT for review reading comprehension and aspect-based sentiment analysis. The authors employ post-training on the BERT model to adapt it for domain-specific data, such as product reviews, which improves the performance of identifying sentiments towards specific aspects (e.g., product quality, price). This post-training technique demonstrates significant improvements in sentiment classification in domain-specific tasks.

#### 2. Attention Is All You Need

This paper introduces the Transformer model, which has become the foundation for many state-of-the-art models like BERT and GPT. The Transformer model replaces sequential RNN or LSTM architectures with an attention mechanism, enabling parallel processing and more efficient training on large datasets. This model achieves significant improvements in natural language processing tasks, including sentiment analysis, by handling long-range dependencies and focusing on relevant parts of the input.

### 4. Datasets

In this project, I utilize the **Korean Emotion Lexicon** dataset provided by Hugging Face. This dataset is a well-structured resource for the classification of emotions in Korean text, providing labels for various emotions such as happiness, anger, and sadness. It has been widely used in emotion classification tasks and serves as a suitable benchmark dataset for this project's goal of binary sentiment analysis.

Notably, this dataset includes not only positive and negative sentiments but also the level of emotional activation, enabling an analysis that takes into account the overall intensity and activation of emotions. This feature is expected to allow for a clearer determination of whether the sentiment is positive or negative.

## 5. Brief/tentative schedule

1. First Week of November to Second Week of November  
Data collection + Preprocessing and model setup
2. Third Week of November  
Model training and evaluation
3. Fourth Week of November  
Result visualization

## 6. State-of-the-art methods and baselines

In the field of sentiment analysis, recent advancements in natural language processing (NLP) have been driven by deep learning models, particularly transformer-based architectures. These models outperform traditional machine learning methods by effectively capturing complex language structures and contextual relationships.

### 1. KoBERT

a variant of BERT optimized for the Korean language, has been shown to be particularly effective in Korean sentiment analysis tasks. Leveraging a pre-trained transformer on Korean text data allows it to capture unique linguistic features of the language, making it a powerful tool for understanding nuanced expressions of sentiment in Korean.

### 2. BERT

BERT has revolutionized sentiment analysis by using a bidirectional transformer architecture to understand the context of a word based on both its left and right surroundings. BERT-based models, especially pre-trained and fine-tuned on specific tasks (like KoBERT for Korean sentiment analysis), have achieved remarkable results in various NLP tasks, including sentiment classification. For sentiment analysis, BERT fine-tuned on sentiment datasets consistently outperforms traditional models, thanks to its ability to capture deeper contextual information.

### 3. GPT-based Models

While less commonly used specifically for sentiment analysis, generative models like GPT-3 have also shown potential in text classification tasks. These models generate context-aware representations of text and can be fine-tuned for sentiment classification. However, they are generally outperformed by BERT and its variants when it comes to pure classification tasks due to their generation-focused architecture.

## 7. references

Sun, Chi, et al. "BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis." Proceedings of NAACL-HLT 2019.

Vaswani, Ashish, et al. "Attention Is All You Need." Advances in Neural Information Processing Systems (NeurIPS), 2017.