MindCompass: NLP-based Psychological Diagnostic Diary

Yehdarm Kim, Nayeong Kim, Junhyun Park, and Giho Bang

Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu,
Suwon-si, Gyeonggi-do, Republic of Korea
{yesnica, ny8236, jenta1208, zzwhos}@g.skku.edu

Abstract. Mental health concerns are increasingly significant in today's society with South Korea reporting one of the highest suicide rates among OECD countries. Traditional mental health care methods often pose challenges due to their high costs, time commitments and the stigma associated with seeking help. To address these barriers, this study introduces a psychological diagnosis diary service powered by Natural Language Processing (NLP). By analyzing users' diary entries, the system identifies emotional states and provides personalized suggestions for self-care or professional assistance. This approach prioritizes user privacy, improves accessibility and promotes early detection, aiming to enhance individual well-being and reduce societal challenges associated with mental health.

Keywords: Natural Language Processing, Mental Health, KoBERT, KoELECTRA, Sentiment Analysis

1 Introduction

Mental health issues have grown to become a pressing concern worldwide, affecting individuals' quality of life and imposing a significant economic and social burden. South Korea, with a suicide rate of 26.6 per 100,000 individuals—the highest among OECD nations—highlights the critical need for effective mental health interventions. While traditional methods such as psychological counseling and therapy have proven effective, they are often inaccessible due to financial constraints, extended waiting times and the societal stigma surrounding mental health discussions.

Recent advancements in Natural Language Processing (NLP) provide a promising avenue to address these limitations. Through the analysis of textual data, NLP enables the identification of mental health conditions, such as depression and anxiety in a non-invasive manner. By examining personal diary entries, this technology can assess emotional states and offer timely interventions. Importantly, the anonymity of such systems helps reduce the stigma associated with seeking mental health support, making it an appealing and viable alternative.

This study presents an NLP-based diary service designed to aid in the management of mental health through sentiment analysis. The proposed system employs advanced Korean language models such as KoBERT to ensure accurate emotional evaluation. By delivering personalized feedback and suggestions, the service seeks to improve accessibility to mental health resources, alleviate pressure on professional care systems, and enhance national mental health outcomes.

The subsequent sections delve into the rationale behind the project, the underlying NLP techniques used for emotional assessment and the design and implementation details of the proposed system. This innovative approach aspires to make mental health care more inclusive, efficient and responsive to the needs of diverse populations.

2 Design for the Proposed System

2.1 Overall Architecture

The architecture of this project is designed to process emotion analysis efficiently by integrating an AI model, back-end data management, and a user-friendly front-end interface. The role and main features of each component are described as follows.

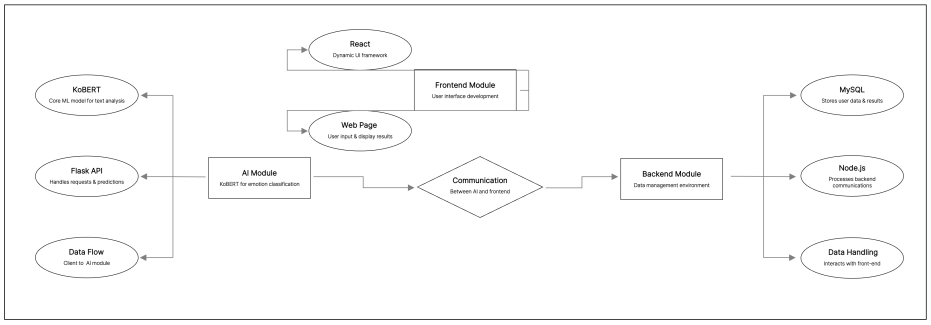


Fig. 1: The architecture of the emotion classification system, illustrating the interaction between the AI module, back-end module , and front-end module. Communication channels enable seamless data flow from user inputs to emotion predictions and storage, ensuring an efficient and interpretable process.

AI Module: The AI module serves as the core of the system, classifying user emotions. The centerpiece is the KoBERT model, which is fine-tuned using data augmentation techniques to provide robust contextual understanding and high classification accuracy. The AI module interacts with other components through a lightweight Flask API. It preprocesses text data received from the front-end, predicts emotions using KoBERT and returns the results.

The preprocessing phase prepares input text for the KoBERT model by converting it into a structured format for training and prediction. First, the text is tokenized using the KoBERT tokenizer, splitting it into word units and adding [CLS] and [SEP] tokens to mark sentence boundaries. The text is then adjusted to the maximum input length: longer sentences are truncated, while shorter ones are padded with [PAD] tokens. An attention mask is generated to distinguish real words from padding tokens. Finally, token IDs, attention masks and segment IDs are converted into tensors, ensuring the data is ready for input into the KoBERT model.

Back-end Module: The back-end handles data management and server-side logic. Node.js processes the server logic, communicates with the MySQL database, and manages user requests. MySQL stores user data and prediction results persistently. The backend also coordinates the data flow between the AI module and the front end, ensuring the stability of the system.

Front-end Module: The front-end, developed with the React framework, is responsible for the user interface. It provides real-time, responsive web experiences, collects text input from users and clearly displays emotion analysis results.

Data Flow: The system processes user inputs in a structured flow. First, users enter text data through the front-end interface. This data is sent to the back-end, where it undergoes preprocessing before being forwarded to the AI module. The AI module analyzes the text using the KoBERT model and generates predictions, which are returned to the back-end. The back-end processes these results and sends them back to the front-end, which visually presents them to the user. This flow ensures efficient handling of user requests and intuitive presentation of analysis results.

2.2 Core Skills/Techniques/Models

KoBERT: Our decision to use KoBERT (**K**orean **B**idirectional **E**ncoder **R**epresentations from **T**ransformers) stems from the foundational principles of BERT (**B**idirectional **E**ncoder **R**epresentations from **T**ransformers). BERT's bidirectional architecture enables it to process text in both forward and backward directions, capturing deeper contextual relationships within sentences compared to unidirectional models.[2] This characteristic makes BERT particularly effective for emotion analysis, where understanding subtle contextual nuances is essential.

BERT operates in two stages: pre-training and fine-tuning. During pre-training, it learns from large, unlabeled datasets to understand language structures, while fine-tuning involves tailoring the model for specific tasks with smaller, labeled datasets. This flexibility allows BERT to achieve high performance across various NLP tasks, even with limited labeled data.[2] These features provided the foundational basis for this project.

KoBERT builds upon BERT's architecture but is specifically optimized for Korean language processing.[4] It addresses the unique linguistic features of Korean, such as flexible word order and semantic shifts caused by particles and verb

endings, which general models trained on English often struggle with. KoBERT utilizes a SentencePiece tokenizer and has been pre-trained on 54 million words from Korean Wikipedia, enabling it to process text effectively at the subword level and capture Korean lexical structures more accurately.

KoBERT's bidirectional architecture further enhances its ability to analyze both forward and backward contextual relationships in sentences, making it particularly suited for tasks like emotion analysis, where a holistic understanding of sentence context is critical. It has outperformed other models, such as BERT Base Multilingual Cased and KoGPT2, in Korean sentiment analysis tasks, demonstrating its effectiveness.[4]

In our project, we aimed to provide users with an emotion classification tool capable of understanding nuanced diary entries. KoBERT was fine-tuned to classify emotions with high accuracy, leveraging its Korean-specific pre-training to capture subtle emotional cues. Its ability to balance accuracy and stability made it the ideal choice for this project, allowing users to document their emotions effectively and gain deeper insights into their mental health.

2.3 Challenges

The primary goal of this project was to develop a robust emotion classification model for diary entries. However, two significant challenges emerged during development: data imbalance and the black-box nature of the model.

Data Imbalance The dataset showed a skewed distribution of emotion labels, with some labels like "nervous" and "sadness" having over 2,000 instances, while others like "loss of confidence" and "sense of loss" had fewer than 500. This imbalance risked biasing the model toward frequent labels and underperforming on rarer labels.

To address this, we applied data augmentation techniques using KoELECTRA, a model tailored for Korean language processing. KoELECTRA's Replaced Token Detection (RTD) mechanism efficiently generates contextually appropriate text.[1] We implemented two augmentation techniques:

- **Random Masking Replacement:** This technique involves replacing a random selection of words in a sentence with mask tokens, which are restored by KoELECTRA's unmasker to generate contextually appropriate replacements. This technique avoids critical positions to maintain semantic integrity, introducing variability while preserving meaning.
- **Random Masking Insertion:** This method inserts mask tokens at random positions in a sentence, later replaced by KoELECTRA's unmasker with contextually fitting words. By exposing the model to diverse sentence structures, this approach enhances its ability to understand nuanced contexts. Both methods ensure augmented data remains coherent and diverse, leveraging KoELECTRA's contextual accuracy.

These techniques increased data diversity and effectively addressed label imbalance, ensuring more balanced training. KoELECTRA's ability to model Korean grammar and context enhanced the augmentation process, especially for smaller datasets, improving overall model performance.

Using KoELECTRA's generator as an unmasker offers several advantages. First, KoELECTRA is a pre-trained model specifically designed for the Korean language, with a deep understanding of Korean grammatical features and morphological structures. In the ELECTRA generator training process, some tokens are masked, and the model learns to predict the original identities of the masked-out tokens.[1] Additionally, KoELECTRA is a computationally efficient model, allowing it to operate quickly and effectively during the data augmentation process. Unlike traditional models using MLM (Masked Language Modeling), which learn from only a subset of the input tokens, KoELECTRA utilizes all input tokens during training, enabling it to learn more comprehensive information. This results in faster generation of augmented data and reduced computational costs for augmentation tasks. Therefore, we consider KoELECTRA as a robust approach to systematically improving model performance when used as an unmasker for augmentation.

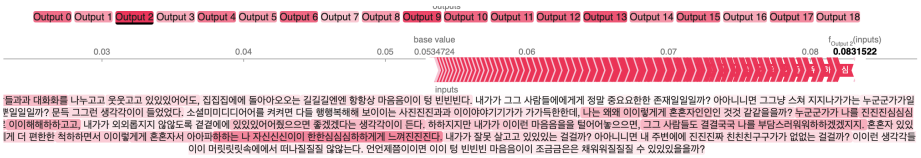


Fig. 2: When classifying a diary entry written to express the emotion "loneliness," the output successfully predicted "loneliness" as the label (Output2). LIME visualization revealed the input elements that contributed to this prediction, clearly showing which words and phrases the model focused on. For example, phrases like "feeling empty," "seems like I'm alone," and "so as not to feel lonely" were identified as key contributors to the model's decision.

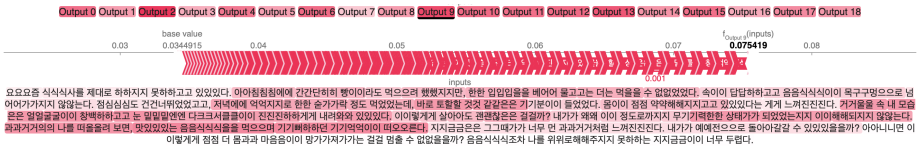


Fig. 3: Similarly, when classifying an entry expressing "loss of appetite," the output correctly predicted this emotion as "loss of appetite" (Output9). LIME visualization highlighted phrases like "I couldn't eat anymore" and "feeling like I might vomit immediately" as significant factors influencing the model's prediction.

Lack of Interpretability in the Black-Box Model While KoBERT excelled in classification performance, its decision-making process remained opaque, limiting trust and understanding of its predictions.

To address the lack of interpretability, we adopted LIME (**L**ocal **I**nterpretable **M**odel-Agnostic **E**xplanations), which provides visual explanations for model predictions by highlighting the contribution of individual words in the input text. LIME uses surrogate analysis, employing a simpler model to approximate and explain the behavior of the complex KoBERT model. LIME also propose a method to explain models by presenting representative individual predictions and their explanations in a non-redundant way, framing the ask as a submodular optimization problem.[3] This approach enhances transparency by revealing the key features influencing predictions, addressing the black-box nature of KoBERT.

The specific advantages of using LIME to interpret KoBERT's classification decisions are as follows. First, LIME alleviates the black-box nature of the KoBERT model by providing an interpretation of its predictions. While KoBERT performs high-accuracy classification by considering the context and structure of sentences, it does not directly reveal the reasoning behind its predictions. LIME visually represents how much specific words or phrases in the input sentence contribute to the model's predictions, making it possible to understand the rationale behind the outcomes. This is highly beneficial during the improvement and debugging process, as analyzing which factors the model focuses on allows for addressing its weaknesses.

Second, LIME is useful for verifying the reliability of the model. By identifying the elements with high contributions in KoBERT's predictions, LIME helps detect whether the model is overly reliant on specific words or has learned illogical patterns. This allows for a visual inspection of whether the training data and classification mechanisms are appropriate, further enhancing the model's trustworthiness.

LIME not only helps understand the reasoning behind predictions but also verifies the model's reliability by identifying potential over-reliance on specific words or illogical patterns. By ensuring that the decision-making process is logical and trustworthy, LIME improves both interpretability and performance. This process provided valuable insights for enhancing the model and communicating results effectively, as illustrated in Figures 2 and 3.

3 Machine Learning Model and Techniques

This project developed an NLP model for emotion classification using KoBERT and KoELECTRA, emphasizing data augmentation and fine-tuning for performance optimization.

3.1 Data Augmentation

To address data imbalance, KoELECTRA was employed to augment the dataset by masking and replacing words in sentences with contextually appropriate al-

ternatives. This process doubled the dataset and focused on enhancing underrepresented labels.

3.2 Model Fine-Tuning

The augmented dataset was used to fine-tune the KoBERT model, enabling it to adapt pre-trained knowledge to the specific task of emotion classification. Hyperparameters such as learning rate, batch size, and warmup ratio were optimized through iterative experiments to maximize performance. Fine-tuning allowed the model to discern nuanced differences between emotion labels, achieving high precision and generalization.

3.3 Evaluation and Interpretability

Model performance was evaluated using metrics like accuracy, precision, recall, and F1 scores. To ensure transparency, LIME was applied to visualize the contribution of specific words to predictions, addressing the black-box nature of the model. Cosine similarity validated that augmented data preserved the context of the original dataset, further improving reliability.

The fine-tuned KoBERT model achieved strong performance in emotion classification, effectively handling challenges like data imbalance and interpretability while demonstrating its suitability for real-world applications.

4 Implementation

4.1 UX/UI

The UX/UI design of this project focuses on simplicity, ease of use, and practical insights to support emotional well-being.

Main Page The main page features a clean design with a central prompt, encouraging users to start writing. Clear buttons like "Start New Entry" and "Continue Diary" are prominently placed for easy navigation.

Diary Writing Interface The writing interface provides a distraction-free environment, allowing users to focus on their entries without interruptions. The layout is simple and functional, ensuring a smooth writing experience.

Diary Analysis Page The analysis page presents emotional patterns through interactive graphs and charts, helping users understand trends over time. It also offers practical feedback, such as self-care suggestions or recommendations for further support, making the tool useful for personal growth.

4.2 Dataset for Machine Learning

The "Wellness Conversation Script Dataset," used in this project, consists of 5,232 sentences extracted from 16,000 counseling records from Kangnam Severance Hospital. It includes 19 emotion labels and features real patient utterances, enabling accurate and contextually rich emotion classification. The dataset's diversity allows for detailed analysis of psychological and mental health contexts, playing a crucial role in enhancing the KoBERT model's performance for emotion classification.

4.3 Frontend

In the frontend, React was used to build the web pages. The `Axios` module facilitates communication with the backend, and `Recharts` is utilized to display the emotions analyzed by the backend in a bar chart format. The `DateRangePicker` allows users to retrieve diary data written within a specific period. Routes were implemented to enable navigation between pages. A `Modal` component for writing diary entries, a `Tab` component and a `Nav` component for navigation were created and utilized within the page.

4.4 Backend

In the backend, Node.js and MySQL were used for development. The `Node.js` framework handles the server-side logic, while `MySQL` is used to store user data and emotion prediction results persistently. The backend provides a RESTful API, which facilitates communication with the frontend. This API receives the text data entered by the user on the webpage, processes it and passes it to the AI model for emotion prediction. The model's output is then sent back to the frontend, allowing users to view the results in real-time.

Furthermore, to manage user diary data efficiently, various tables were designed in the `MySQL` database, establishing relationships to optimize data processing. `MySQL` offers fast and reliable data handling, enabling the system to process large volumes of data and handle high traffic efficiently.

The backend also implements security measures, including data encryption, authentication and authorization management, ensuring that user information is protected at all times.

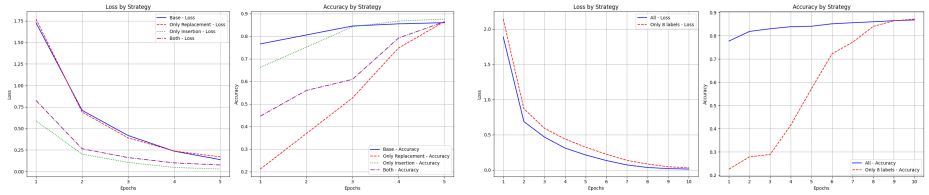
5 Evaluation

5.1 Experiments

In this project, two experiments were conducted. First, data augmentation techniques were applied to enhance data diversity and improve generalization performance. Second, the impact of augmenting specific data subsets on model performance was analyzed.

Augmentation Techniques The Base strategy uses the original data without augmentation, Only Replacement replaces specific words, Only Insertion adds new words, and Both combines the two. Figure 4 (a) and Table 1 compare these strategies. Only Insertion achieved the best results across metrics, including a Cosine Similarity of 0.973, outperforming Base, Only Replacement, and Both. The Base strategy showed the slowest improvement in training loss and test accuracy, while Only Insertion demonstrated the fastest and most consistent performance.

These findings highlight the effectiveness of insertion-based augmentation in enhancing data diversity and generalization, making it the preferred method for this project.



(a) Illustration of training loss and test accuracy trends across different data augmentation strategies.

(b) Shows the changes in training loss and test accuracy depending on the augmentation target.

Fig. 4: (a) Illustration of training loss and test accuracy trends across different data augmentation strategies. (b) Shows the changes in training loss and test accuracy depending on the augmentation target.

Strategy	Accuracy	F1-Score	Precision	Recall	Similarity
Base	0.840	0.871	0.861	0.863	-
Replacement	0.865	0.855	0.856	0.856	0.951
Insertion	**0.876**	**0.867**	**0.881**	**0.885**	**0.973**
Both	0.865	0.864	0.864	0.865	0.964

Table 1: Performance comparison of data augmentation strategies. Cosine similarity was used for similarity calculation.

Augmentation Targets Experiments were conducted to compare the performance of augmenting the entire dataset versus focusing on eight labels with fewer samples (fewer than 1,000).

Figure 4 (b) compares the training loss and accuracy progression between the full dataset and the eight key labels dataset. The eight-label dataset demon-

Strategy	Test Accuracy	F1-Score	Precision	Recall
Entire Dataset	0.868	0.868	0.868	0.868
Only 8 Labels	**0.872**	**0.871**	**0.872**	**0.872**

Table 2: Performance comparison between entire dataset augmentation and 8-label augmentation.

strated lower initial loss and faster reduction, stabilizing after the second epoch, while the full dataset showed higher initial loss with gradual improvement. In terms of accuracy, the eight-label dataset rose sharply after the third epoch, surpassing the full dataset at convergence.

Table 2 shows that the eight-label dataset achieved higher metrics (Test Accuracy: 0.872, F1-Score: 0.871) compared to the full dataset (0.868 across all metrics). This indicates that focusing on specific labels can enhance model performance, highlighting the importance of tailoring datasets to specific objectives.

5.2 Hyperparameter Tuning

Hyperparameter tuning was performed to optimize the performance of the KoBERT-based emotion analysis model. The tuning process systematically explored various combinations of key hyperparameters, including learning rate, batch size, warm-up ratio, dropout rate, maximum gradient norm, and input length, to identify the optimal configuration. The final selected hyperparameters are as follows: a learning rate of 5e-5, batch size of 16, warm-up ratio of 0.2, dropout rate of 0.1, 10 epochs, a maximum gradient norm of 1, log interval of 200, and maximum input length of 512.

This configuration achieved an accuracy of 87.24

5.3 Objective Achievement

The main goal of this project was to create a robust emotion classification model using user-generated text. Through experiments and hyperparameter tuning, the model achieved an accuracy of 87.24%, meeting the project's objectives. Efficient data augmentation techniques and tailored dataset composition were key to enhancing performance and ensuring reliable predictions.

The project highlights potential applications in areas like mental health management, customer service, and personalized user experiences. The use of targeted augmentation for specific labels underlines the importance of customized dataset design, which improves service accuracy and user satisfaction. This work demonstrates the practical relevance of emotion analysis and its versatility across various fields.

6 Limitations and Discussions

1. **Targeting Only People Who Write Diaries**
   The service is limited to individuals who write diaries and requires them to upload their entries online, which can reduce accessibility. There is also the question of how many people in need of help such as those struggling with depression would actually write diaries. People experiencing severe depression often become passive and if they do not engage in diary writing, the core purpose of our service—to send alerts to those in distress—becomes invalid.
2. **The Necessity of Professional Help**
   Our service analyzes only the content of the diary entries, but proper judgment and treatment require consulting a professional. For instance, if someone experiencing severe depression and stress writes only objective details about their day in the diary, our service would fail to detect their mental state.
3. **Security Issues**
   Our service is still in its early stages, and we have not given significant consideration to security. If the system were attacked, vulnerabilities could be exposed, potentially leading to the leakage of personal information. Diaries often contain deeply personal content, making this a critical issue, but our current level of development is still immature.
4. **Using Diary Data to Improve the AI Model**
   If our service were to evolve by allowing the AI to learn from diary data, it could raise concerns about privacy. To address this, it would be essential to obtain explicit user consent before using their data for AI training.
5. **Insufficient Testing**
   Due to the lack of individuals who have uploaded their diaries online, we have not been able to conduct adequate testing. Addressing this issue would require training the AI with user diary data, which connects to the privacy concerns mentioned in point 4.

7 Related Work

7.1 Dapda

Dapda is an application developed by LG U+ that analyzes emotions based on diary entries and provides encouragement tailored to the diary's content, similar to our service. The key differences are that Dapda is an app that cannot be used without downloading it, and it does not display emotional states but focuses solely on providing encouragement and advice.

7.2 Seamspace

Seamspace is a service that analyzes emotions based on diary entries, providing feedback and encouragement tailored to the emotions expressed in the diary. It

appears to be available on both mobile applications and web platforms. Unlike our service, Seamspace focuses not only on negative emotions but also includes positive emotions such as joy and happiness. The negative emotions it addresses are limited to four categories: stress, inferiority, depression, and anxiety.

7.3 Kids Diary

Kids Diary is an AI-based diary analysis service currently available in Korea. Specifically, it analyzes video diaries, providing insights into a child's psychological state when a video diary longer than one minute is uploaded. Our service differs in that it analyzes psychological states based on text-based diary uploads instead of videos and caters to all age groups, not just children.

8 Conclusion

This study introduced a novel NLP-based psychological diagnostic diary service designed to enhance mental health management by analyzing users' diary entries. By leveraging advanced Korean NLP models like KoBERT and implementing techniques such as data augmentation and LIME for interpretability, the proposed system achieved significant accuracy in emotion classification. While the service demonstrates great potential in promoting mental health accessibility and early intervention, several limitations, including privacy concerns and the need for professional consultation, must be addressed in future iterations. Overall, this work highlights the importance of integrating AI-driven technologies in mental health care to foster a more inclusive and effective support system.

References

[1] Kevin Clark et al. "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators". In: *arXiv preprint arXiv:2003.10555* (2020). URL: https://arxiv.org/abs/2003.10555.

[2] Jacob Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *arXiv preprint arXiv:1810.04805* (2018). URL: https://arxiv.org/abs/1810.04805.

[3] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ""Why should I trust you?" Explaining the Predictions of Any Classifier". In: *arXiv preprint arXiv:1602.04938* (2016). URL: https://arxiv.org/abs/1602.04938.

[4] SKTBrain. *KoBERT: Pretrained Language Models for Korean Text Understanding*. https://github.com/SKTBrain/KoBERT. 2019.