Knowledge-Based Systems 310 (2025) 112972

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

An intelligent marketing platform with influencer classification in social networking services

Xiaohong Yu, Jinyong Kim, Yoseop Ahn, Mose Gu, Jaehoon (Paul) Jeong *, JinYeong Bak, Jaemin Jo

Department of Computer Science and Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon, Gyeonggi-do, 16419, Republic of Korea

ARTICLE INFO

Keywords:
Social media
SNS
Influencer marketing
Korean text classification
KoELECTRA

ABSTRACT

Collaborations with influencers in social networking services (SNS) for online commercial advertisement has gained significant momentum in recent years. However, the existing approaches suffer from a lack of precision in matching advertisements to specific target user groups (e.g., influencers). To address the issue, in this paper we present an intelligent marketing platform that classifies influencers based on a newly built SNS post text dataset and the comprehensive performance factors. The proposed platform fine-tunes a pretrained language model with the SNS post text dataset of influencers, which can achieve a better classification accuracy. To further improve the matching accuracy for marketing partners, we integrate multiple performance features of influencers into the proposed platform. Our results show that the proposed model can provide more than 90% classification accuracy for business partners in their product advertisements. Furthermore, we used a real-world business requirement to implement and evaluate the proposed platform with a web-based system, which provides an easy way for online commercial advertisement.

1. Introduction

In our contemporary society, social networking services (SNS) have become a prominent platform for individuals, especially younger generations, to forge and sustain interpersonal relationships on a daily basis. Within these SNS channels such as Instagram, YouTube, and TikTok, a unique group of users known as influencers has emerged. These individuals consistently share various aspects of their personal lives and experiences across diverse scenarios, generating data that holds a significant commercial value. According to the influencer marketing report [1], the influencer marketing industry is expected to grow more than 20% yearly, resulting in 24 billion dollars by the end of 2024. In the growing process, more than half of the fashion brands and business companies leverage artificial intelligence (AI) technologies to discern influencers and execute marketing campaigns featuring these influencers to promote their products. Influencer marketing is widely regarded as an effective and comparatively cost-efficient means of product promotion in contrast to conventional methods such as printed, television, and outdoor advertisements [2]. Conventional marketing strategies tend to be broad, general, and costly, posing significant challenges for small businesses. Furthermore, traditional marketing techniques lack the capacity to accurately measure the efficiency of advertising campaigns.

Influencers' contents resonate with the Internet users who share similar experiences, potentially shaping trends for specific topics [3]. For instance, an influencer's movie review can spark discussions and heighten interest in viewing the film in a cinematic setting. This phenomenon presents an opportunity for companies in diverse sectors, such as fashion, cosmetics, and traveling, to leverage influencers' influence to promote their products or services. Advertising opportunities that were previously exclusive to prominent brands and large-scale enterprises have now become accessible to small businesses, particularly those operating at a local level, through the utilization of influencer marketing. Correspondingly, certain influencers with sizable followers can also derive financial remuneration by partnering with these small ventures and promoting their products or services. While influencer marketing presents a mutually advantageous opportunity for both enterprises and influencers, the process of identifying and selecting appropriate influencers can be time-consuming for enterprises. To address this challenge, a platform is required to facilitate the matching and collaboration between enterprises and influencers.

In light of the exponential proliferation of influencer marketing, there has been an emergence of scholarly endeavors dedicated to its various facets, including influencer identification [4], influence maximization [5–8], user incentivization [9] and modeling the influencer

* Corresponding author.
E-mail address: pauljeong@skku.edu (J.P. Jeong).

https://doi.org/10.1016/j.knosys.2025.112972
Received 20 July 2024; Received in revised form 27 December 2024; Accepted 4 January 2025
Available online 11 January 2025
0950-7051/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

marketing activities [10,11]. These empirical investigations demonstrate that collaborating with appropriate influencers can optimize the efficiency of advertising campaigns. However, the selection of suitable influencers remains a critical challenge. Notably, current research predominantly concentrates on analysis based on network topology and complex user attributes, neglecting the diverse thematic characteristics of the Internet celebrities, who are widely followed and loved on the Internet due to their special talents. Specifically, the cross-influences and comprehensive strengths of multiple topics are often overlooked [12]. A good influencer for one campaign should both have similar product features and good social network topology connectivity.

High quality influencers possess exemplary leadership abilities [13]. In order to ascertain the level of influence possessed by an influencer, it is insufficient to solely evaluate the numerical value of their followers. A comprehensive analysis encompassing various factors is essential, including the rate of follower growth, the frequency of media updates, the engagement of followers, and the distribution of media topics. Additionally, influencer post data encompasses a range of formats including images, videos, and text [14], all of which can be analyzed and categorized into distinct groupings. This process may result in a misalignment with the intended user demographic. Current research primarily classifies influencers on social networking sites based on textual analysis [15,16]. Nevertheless, the majority of SNS analytical models categorize influencers into broad groupings, which may not align effectively with the demands of corporate clients. Additionally, a substantial portion of the research examines English-based text generated by SNS influencers, neglecting non-English text analysis, such as Korean [17].

To address the challenges, we propose an intelligent social networking service (SNS) influencer classification and marketing system (hereinafter referred to as the SNS influencer marketing platform). This system facilitates collaboration between business clients and influencers, with a particular focus on small business owners. Unlike other research that primarily concentrates on strategies for identifying influencers, maximizing influence, and analyzing social media data, our work aims at providing a web-based, end-to-end service platform [18] within the influencer marketing domain. This system encompasses business clients releasing campaigns, influencers applying for campaigns, clients matching and selecting suitable influencers, influencer posting campaigns, and post-activity analysis.

Targeting the Korean influencer market, in the process of matching and selecting suitable influencers, the proposed platform classifies influencers based on the real-world category requirements from a business client, and also an integrated influencer analysis approach by analyzing the rate of follower growth, the frequency of media updates, the engagement of followers and so on. In the process of matching and selecting appropriate influencers to clients, the platform employs a deep learning-based text classification model to classify the media topics and analyzes multiple indicators such as follower growth rate, media update frequency, media content quality, and follower engagement. The comprehensive analysis result enables enterprises to discern the influence and advertising potential of influencers, providing a holistic view of their capabilities. Consequently, merchants (i.e., business clients) across diverse fields are facilitated in identifying suitable influencers for their online marketing campaigns.

The main contributions of this paper are as follows:

- **A post text dataset in Korean from influencers**: We collected and created a new text dataset from the Instagram platform via the official APIs for our influencer classification task, which consisted of eleven classes by a real-world business demand (see Section 4).
- **A fine-tuned text classification model**: We designed a fine-tuned text classification model based on the collected dataset for our influencer classification task in Korean language, which is used to detail the influencer active area features (see Section 5.1).
- **An integrated analysis for influencer classification**: We suggested an integrated influencer analysis approach that better shows the influence level, growth, content quality, follower status, likeability, and active topic features. The integrated approach used the proposed classification model and a comprehensive influence indicators supported by multiple performance features of influencers (see Section 5.5).
- **A real-world influencer marketing platform**: Our paper aims to provide a web-based, end-to-end service platform for the field of influencer marketing in the real world. We implemented and evaluated the proposed influencer marketing platform via a web-based system to verify the workflow for matching business campaigns with influencers (see Sections 6.4 and 6.5).

The rest of this paper is organized as follows: Section 2 summarizes the previous work related to our study. Section 3 explains the problem formulation of this paper by describing the architecture of our influencer marketing platform and its components, along with an implementation. Section 4 explains the newly built dataset used for our influencer marketing platform. Section 5 details the design of the classification for influencers according to their expertise fields with the comprehensive influence indicators. Section 6 describes the strategies to train the text classification model, shows the results of the performance evaluation of the influencer classification model, and demonstrates the implementation of our platform. Finally, we conclude this paper along with future work in Section 7.

2. Related work

The increasing prevalence of social media has generated a vast amount of data, which has greatly advanced scientific research. Numerous individuals have recognized the potential value of this data and have initiated efforts to extract meaningful information from it. Consequently, various social media analyses have emerged, including sentiment analysis [19–23], social media text classification [15,17,24–26], social network analysis and mining [16,27–29], and social recommender system [30,31]. Younger generations often turn to social media to follow current fashion trends, discover popular tourist destinations, find top-rated restaurants, or choose films based on online reviews. Merchants have recognized the advertising potential embedded within this vast amount of data.

This has driven the rapid growth of the influencer marketing industry, resulting in extensive research on topics such as influencer marketing campaign modeling [10,32], influencer identification [4,33], influence maximization [5–8,12,34], and user incentivization [9]. Influencer marketing campaign modeling employs graph-based models to simulate campaign spread, incorporating real-world factors such as customer behavior, willingness to pay, customer interest, and influencer engagement [10,32]. Influencer identification focuses on selecting the right influential users to maximize social media reach and enhance advertising returns by leveraging social network topology [4,9]. Influence maximization aims to improve the spread of influence within both known and unknown social network topologies through the use of greedy and heuristic algorithms [5–7]. User incentivization seeks to encourage product or service purchases within budget constraints, with approaches such as the reinforcement learning-based framework for optimizing incentive allocation [9]. The overarching goal of these studies is to identify influential users capable of generating cascading effects, thereby achieving sustained marketing impact.

The identification of suitable influencers is paramount for successful advertising campaigns. Previous research seeks influential users by utilizing the node connectivity of social network topology. Significant research has also been dedicated to identifying influencers through social media data [14,35–37]. Seungbae et al. [14] proposes a deep learning model that analyzes both textual and visual elements of influencer profiles, employing an attention mechanism to

prioritize specific posts for influencer classification. Farnadi et al. [37] presents a deep learning framework that merges features from text, images, and relationships, integrating user profiling information at each network decision. Despite the promise of multimodal profiling, several challenges remain. Features extracted from image data can be homogeneous and may lack diversity. While images often convey descriptive information, they may fail to capture the intended message as effectively as text, which can deliver more precise details. When influencers frequently post similar images with minimal text, multimodal profiling can prioritize visual information, potentially neglecting valuable textual insights. Furthermore, most existing studies primarily classify influencers into one single category, neglecting the diverse topic features of the information, particularly the cross-impact of multiple topics on information propagation. Addressing these challenges is crucial for future research to develop more robust and accurate influencer profiling methods, empowering brands to make informed decisions about influencer partnerships.

There are many research works performing the influencer social media text classification task. Ramadhani et al. [25] uses a machine learning model called support vector machine (SVM) to classify Instagram captions. Text classification classifies every Instagram caption into organized groups. But while machine learning-based models like SVM and term frequency-inverse document frequency (TF-IDF) can only perform syntactic parsing, they cannot perform semantic parsing for the text. There has been much improvement from only tracking the syntactic parsing to tracking the semantic context dependencies.

The transformer architecture [38] revolutionized natural language processing (NLP) by utilizing attention mechanisms to process and understand text. Bidirectional Encoder Representations from Transformers (BERT) [39] employs a masked language model mechanism to learn bidirectional representations from large corpora during pre-training, enabling natural language understanding that can be fine-tuned for various downstream tasks, such as text classification. To mitigate the computational demands of large language models, the DistilBERT [40] model was developed, applying knowledge distillation in the pretraining stage to reduce BERT's size by 40% while maintaining 97% of its performance. In contrast to BERT, which predicts masked tokens based on context alone, the ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately) model [41] utilizes a generator discriminator framework to achieve greater efficiency in natural language tasks. With the advancement of NLP and the continuous update of large language models, we can attain strong performance on downstream tasks by leveraging pretrained language models and fine-tuning them on smaller datasets [26,39]. Transformer-based models have paved the way for various NLP tasks, leading to the development of numerous models addressing different language challenges, including BERT-Base for 104 languages, KoBERT [42], distilKobert [43] and KoELECTRA [44] for Korean text analysis.

While much research focuses on English-based data analysis, there is a lack of studies analyzing non-English data. Our objective is targeting the Korean influencer marketing market by analyzing social media text data from Korean influencers, offering a comprehensive approach to gain in-depth insights of influencers and enhancing collaboration between business clients and influencers on an SNS marketing platform. In this paper, we fine-tune a deep learning-based model to perform the text classification of Korean-language Instagram influencers. In order to more comprehensively show the characteristics and potential of influencers, we categorize every text caption from all social media data of every influencer into eleven categories, and aggregate all the classification results to give a detailed classification distribution. Meanwhile, we also evaluate the influencers with the corresponding influence level, influencer growth, content quality [45], follower situation, and likability to give a comprehensive evaluation result of an influencer, which is very meaningful for matching influencers.

3. Problem formulation

An SNS influencer marketing platform is proposed to facilitate both the clients, especially small-sized business runners, and the influencers for the influencer marketing activities. The platform is a web-based, end-to-end service system in the influencer marketing domain. The system encompasses business clients releasing campaigns, influencers applying for campaigns, clients matching and selecting suitable influencers, influencer posting campaigns, and post-activity analysis.

For the process of matching and selecting suitable influencer, this paper proposed the classification model and comprehensive evaluation metrics of influencers to provide business clients with detailed information about them. Additionally, it applies the classification model and evaluation metrics to the platform to facilitate implementation and assessment of the system.

First, we provide an introduction to the main components of the platform as follows:

- **Clients**: Clients are companies or agencies seeking to advertise their products or services.
- **Influencers**: Influencers are individuals who actively engage in social networking platforms and have amassed a significant following. As consumers, they possess the ability to shape the behavioral patterns of their followers, thereby influencing trends and preferences within their respective communities.
- **Campaigns**: Campaigns are activities that clients post on the platform. The campaigns aim to recruit one or more influencers to advertise products or services.
- **Front-end**: Front-end is a web framework that displays information in a web browser. The information shown in the web browser can be configured via the graphical user interface according to different requirements. The front-end allows clients and influencers to send requests such as posting campaigns, seeking influencers, influencers applying for campaigns, and checking influencers' detailed information.
- **Back-end**: Back-end is a data processing bridge that processes the requests from the front-end, generates and sends the responses to the front-end. It typically includes a server, an App, and a database. Here, the server is simply an always-online process that listens to the information requests. The App is the influencer marketing user application where the data is processed and analyzed. The database is where platform data is stored for processing the requests from the App.

The architecture of the influencer classification system on the influencer marketing platform is shown in Fig. 1. The platform is designed as a real world system. The platform's architecture encompasses both a front-end and a back-end. The front-end includes a web browser functionality that displays user data written in hypertext markup language, designed specifically for influencer and client registration, campaign releasing, influencer matching, and post-activity reporting. The back-end is a data processing bridge that includes a server, an App and a database. The server receives front-end requests from the App manipulates data by communicating with the database that stores user data and text classification model parameters. The communication between the front-end and back-end employs the REST-API over HTTP or HTTPS [46]. The back-end receives REST-API requests from the front-end. Users can seamlessly deliver their intentions to the database via the front-end and back-end. Through this process, the back-end transmits the corresponding responses to the front-end.

Both influencers and business clients register with the platform. When the business clients seek suitable influencers for their marketing campaigns, they must evaluate which influencers can maximize advertising effectiveness. Our influencer classification system analyzes and classifies the text data of each influencer's posts, aggregating the classification results to provide a detailed expertise domain distribution. It

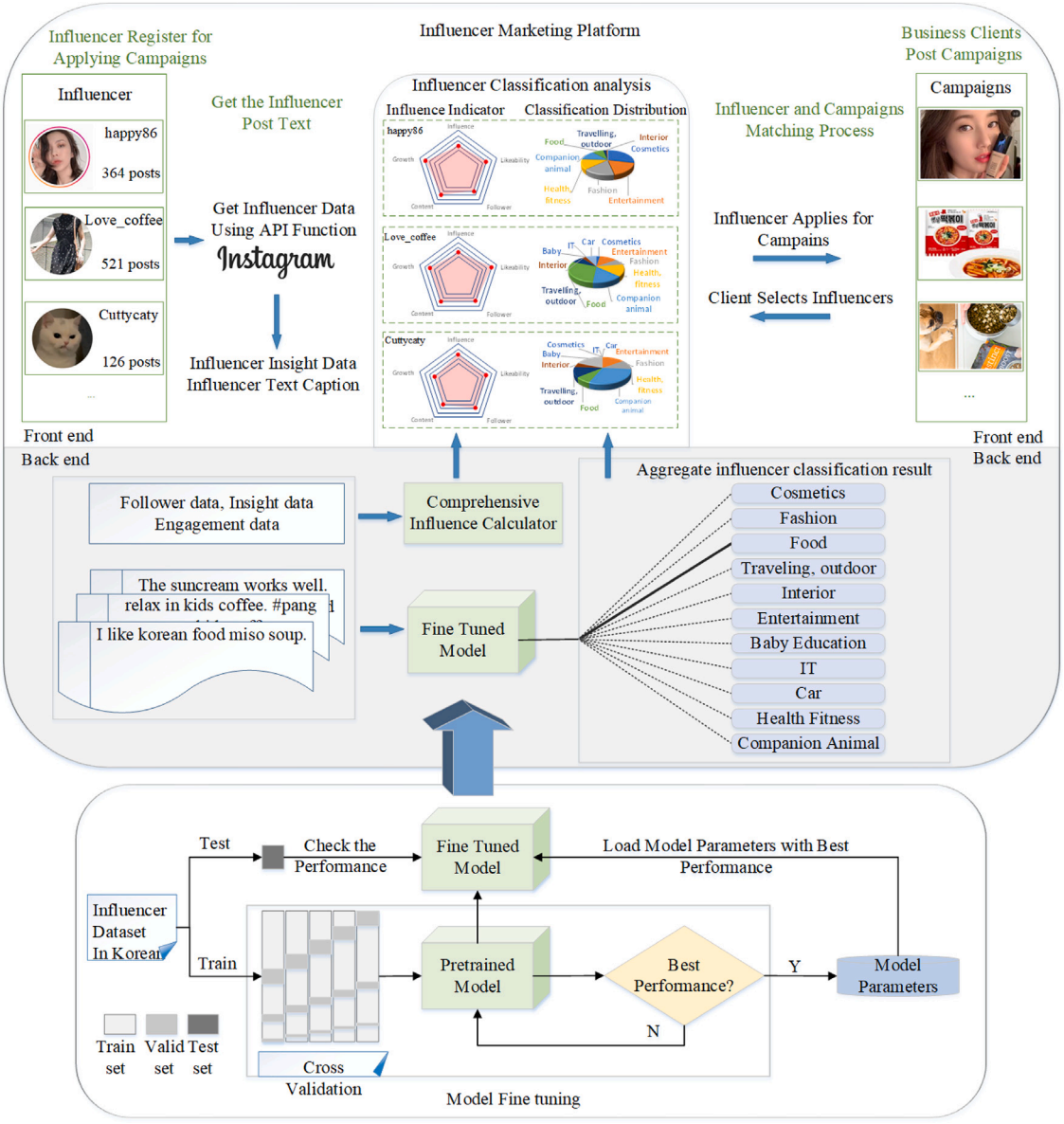

Fig. 1. The architecture of an influencer marketing platform with influencer classification services.

then integrates comprehensive decision factors, such as influence level, influencer growth, content quality, follower engagement, and likability, to deliver a thorough evaluation of each influencer.

When selecting an influencer for a campaign, a client reviews detailed information about each influencer, including their topics of posts and comprehensive influence indicators. To ensure an optimal matching, the client must understand the potential value which each influencer offers. Since engagement levels vary across different follower stages and topics of posts, the client can select influencers based on their specific field and potential impact.

To achieve the optimal matching between influencers and campaigns, this paper analyzes influencers based on the text captions from their social media posts. It presents the results of a classification analysis and calculates metrics such as influence level, growth, content quality, follower situation, and likability. These indicators are derived using insight data obtained through the Instagram API functions [47]; for the detailed API functions, refer to Appendix A. They can be used for the client to understand the potential of every influencer for maximizing commercial benefits for advertising campaigns.

4. Dataset

The dataset includes eleven categories: cosmetics, fashion, food, traveling/outdoor, interior, entertainment, baby education, IT, cars, health/fitness, and pet companions. Each category includes several subcategories to further specify its domain. The subcategories of each category are described in Table 1.

These categories are tailored to the platform's business characteristics and different from existing datasets. We need to create a new dataset for this classification work in Korean. According to the catalog details, we collected the dataset from Instagram using the Instagram API functions [47] while filtering out the movie stars' posts. We selected the subcategories as hashtags to search for if the hashtag exists or not on Instagram using hashtag search function, and then collected media data using the hashtag top media function.

Most of the raw data collected (e.g., post captions) was not associated with hashtags, and the subjects of some collected post captions were not related to the eleven categories. The data was rescreened and labeled manually. There are five people involved in manually labeling,

Table 1
SNS post text dataset description.

Category	Subclass scope
Cosmetics	Base skin care, anti-aging, body care, color makeup, hair and scalp, hair salon, shampoo, and conditioner products.
Fashion	Office, party, date, health, yoga and home clothing, swimwear, beachwear, women, men, unisex, shool style, and wedding clothing, suits, sportswear, men, women, children, daily fashion, leggings, and jewelry accessories.
Food	Korean, Japanese, Chinese, western food, fusion food, homemade, baking, healthy, party, camping food, popular restaurants, vegan restaurants, cafes, brunch desserts, rooftops, cocktails, wine, whiskey, bars, pubs, and clubs.
Traveling/outdoor	Domestic, abroad travel, camping, backpacking, mountaineer, cruise, group travel, USA, Japan, China, Southeast Asia, European, South America and Africa travel, and hotels.
Interior	Self-interior, nursery design, living room, kitchen, bedroom interior, one-room setups, plant arrangement, DIY (do it yourself) interior, office design, and interior tips.
Entertainment	Movies, music, performance shows, musicals, concerts, related activities, and celebrity peripheral products.
Baby education	Baby education, early education, English, Korean, Japanese education, Chinese characters, Chinese, foreign language education, prior learning, kids cafes, kindergartens, day care centers.
IT (information technology)	Personal computers, mobile phones, smartphones, tablets, game machine, game console, audio equipment, laptops, keyboards, monitors, and software.
Car	SUV (sport utility vehicle), sedan, domestic, imported, used car, off-road vehicle, sports, racing car, car modification, decoration, electric car, hybrid car, eco friendly car.
Health fitness	Yoga, stretching, pole dancing, weight training, Pilates, body profiles, fitness coaching, and fitness activities.
Companion animal	Dogs, cats, dog supplies, cat supplies, pet food, pet care services, pet pension, and dog training.

Table 2
SNS post text dataset in Korean.

Category	Sample count
Cosmetics	211
Fashion	205
Food	376
Traveling, outdoor	254
Interior	204
Entertainment	134
Baby education	204
IT	188
Car	200
Health fitness	201
Companion animal	205
Total	2382

two people in the field of influencer marketing, and three people in the project developing, a total of four native Korean speakers and one Chinese. Five people worked together on labeling and re-examined it at the end to ensure that there were no label conflicts. Totally, we collected 2,382 samples after re-screening and labeling. The dataset is unbalanced, as shown in Table 2. Some posts use an irrelevant hashtag, and some posts just have a hashtag without any other text. We selected samples with a text length of at least 10 words to provide an efficient learning model. The collected data combined into one CSV file with two columns: label and post caption. We named the dataset SNS post text dataset in Korean.

5. Method

5.1. Text classification model

In this paper, we present a text classification model (as shown in Fig. 2) specifically designed for influencer classification. The proposed model is a deep learning approach capable of categorizing text data into eleven classes, as outlined in Section 4. The text classification model utilizes prominent pretrained language models (PTMs) to comprehend the context of the input text data. These PTMs, including BERT-m, KoBERT, distilKoBERT, and KoELECTRA, have been trained on extensive Korean text corpora and have developed the ability to recognize significant features and patterns within text. We leverage these pretrained models as the foundation of our text classification model to extract contextualized representations. These representations encapsulate the semantic meaning and interconnections among words, enabling the model to grasp the overall context of the text.

To enhance the robustness of the model and mitigate over-fitting, a dropout layer is incorporated following the PTM. This layer helps improve the generalization capabilities of the model. Finally, a linear layer serves as the output layer of the model. The linear layer transforms the contextualized embeddings into a set of class scores, with each class represented by a distinct score. The class with the highest score is predicted as the model's output.

5.2. Data preprocessing

The raw data we collected cannot be used directly, as social media users often strive for brevity and employ neologisms (i.e., new words) [48] in their posts. This results in a variety of unique sentences, including informal language, nonstandard spellings, hashtags, emoticons, unnormalized text, slang, acronyms, and transliterations. These characteristics make social media text data distinct compared to data found in news articles, blogs, emails, and books. To facilitate effective tokenization, it is necessary to clean the data. Various text preprocessing techniques are typically employed, such as handling slang, abbreviations, emoticons, normalization, stemming, and lemmatization. To address issues related to Korean text appearing in Korean letters, we utilize the "Normalization Form Composition (NFC" function from the Unicode data library to normalize the social media text. Subsequently, we apply a regularization formula to select Korean words for training and testing in text classification.

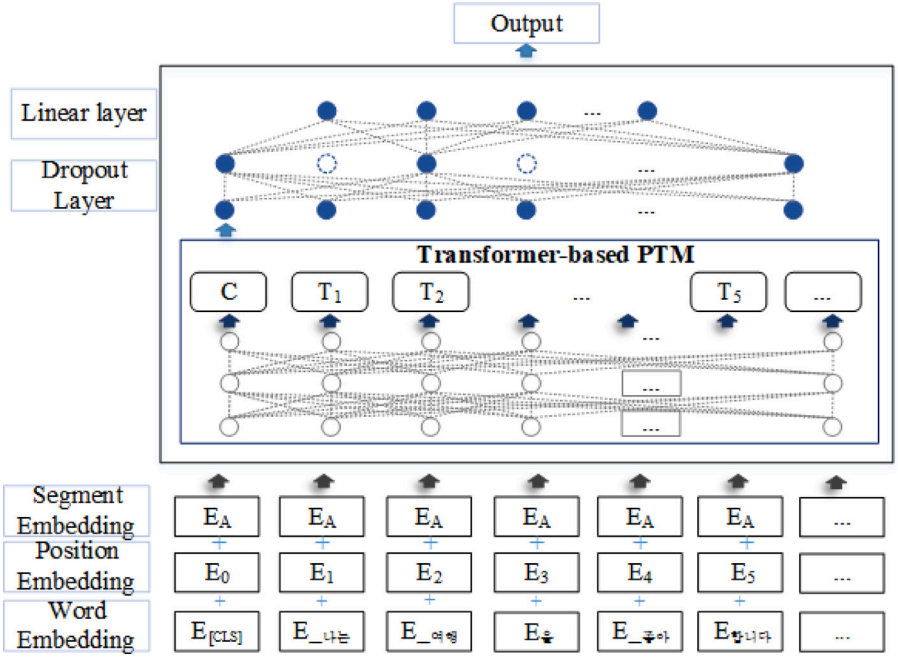

Fig. 2. Text classification model.

5.3. Input embedding

In the transformer-based text classification model, the input includes three parts, word embedding, position embedding and segment embedding. The word embedding was generated by the tokenizer function from different PTMs. To facilitate the learning and processing of text, tokenization must be employed to convert text data into a word embedding suitable for input into the model. During the tokenization process, Korean text is tokenized into tokens $\{x_1, x_2, x_3, \ldots, x_n\}$ using various tokenizers from different PTMs. These tokens are then converted to token indices in a vocabulary after adding several special tokens, such as $[cls]$, $[sep]$, and $[unk]$. Finally, a vector representation of word embedding is generated. Notably, the $[cls]$ token is added solely at the beginning of the first sentence to store the aggregated representation of the sentences. Conversely, the $[sep]$ token is added at the end of each sentence to signify its conclusion. Lastly, the $[unk]$ token is utilized when a word is absent from the vocabulary.

The transformer-based models process all input sequences in parallel, which causes the word position problem. To solve the problem, the position embedding was added to input. A sinusoidal function is used to indicate the word position sequence [38]. The Positional formula is as follows:

$$P_{pos}(pos, 2i) = sin(\frac{pos}{1000^{2i/d_{model}}}). \tag{1}$$

$$P_{pos}(pos, 2i+1) = cos(\frac{pos}{1000^{2i/d_{model}}}). \tag{2}$$

Here pos implies the position of the word in the sentence, and i implies the position of the embedding. In odd and even positions, the sin and cos functions are used separately. d_{model} represents the embedding dimension, which means the dimension of every word embedding.

The segment embedding was added to distinguish the first sentence from the other. We also call it token_type_id. When the downstream task is next sentence prediction and question answering, the input has two sentences; the first sentence is tagged with segment A with an E_A label that uses the value of 0, and the second sentence is tagged with segment B with an E_B label that uses the value of 1 to distinguish the two parts of the input. And also, at the end of the first sentence, the token $[sep]$ is added to separate the two sentences. In this paper, we perform text classification. All inputs are regarded as one sentence with segment encoding E_A using the value 0.

5.4. Loss function

A loss function (a.k.a., error function) is to calculate the difference between the predicted output and the actual target value. For our text classification task, to train a better model, a cross entropy function as a loss function is used to measure the probability of an example belonging to one class compared to another. The cross entropy function is shown below:

$$L_{ce} = -\omega_{y_n} log \frac{exp(x_{n,y_n})}{\sum_{c=1}^{C} exp(x_{n,c})}, \tag{3}$$

where x_n is the input, y_n is the target label, ω is the weight, and C is the number of classes. C equals eleven because the dataset has eleven classes. n indicates the batch size during the data training process.

The cross entropy function in (3) is calculated based on the softmax activation function [49]. The softmax activation function $\frac{exp(x_{n,y_n})}{\sum_{c=1}^{C} exp(x_{n,c})}$ converts the unnormalized probabilities output into normalized probabilities for determining one class as a target expertise domain. The cross-entropy function takes the output probabilities and measure the distance from the truth values (in an output vector). The objective is to decrease the loss function to make the output of the model as close as possible to the truth values.

5.5. Comprehensive influence indicators

The comprehensive influence indicators mean the five indicators that help to better understand the value of influencers, such as influence level, influencer growth, content quality, follower situation, and likability. The influence level is evaluated by how many followers the influencer has, and how many posts the influencer updates every month. Influencer growth is evaluated based on the number of followers and comments that the influencer gained in the past month, the number of posts that the influencer made in the past month. Content quality is evaluated based on the average number of views, likes, and shares for all the posts that the influencer has made. The follower situation is evaluated based on whole average engagement

Table 3
Model hyperparameters.

Hyperparameter	Value
Epoch	20
Learning Rate	5e−6
Optimizer	AdamW
Warm-up ratio	0.1
Scheduler	Cosine schedule with warm-up
Loss function	Cross entropy
Cross validation	StratifiedKFold
KFold	5
Train/test ratio	8 : 2

rate, which is the ratio of the number of the interactions that viewed, liked, commented, shared or saved the post of the influencer to the total number of followers. Likeability is evaluated based on the number of accounts that viewed the influencer's profile and posts.

During the campaign implementation phase on the influencer marketing platform, business clients initiate campaigns, and influencers on the platform subsequently apply, and then the business clients select the most suitable influencers from the applicant pool. In the selection process, in addition to the influencer's active topic, businesses utilize comprehensive influence indicators to identify the most valuable influencers for their campaigns. These indicators include the influencer's level of influence, growth, content quality, follower demographics, and likability. The influence level can be used to check how many followers the influencer has. The influencer growth can be used to detect whether the influencer is still active in the social media in recent months or not. The content quality can be used to evaluate whether the influencer consistently publishes quality posts that can have amazing attraction or not. The following situation can be used to check whether the influencer interacts frequently and closely with followers or not. The likeability can be used to check how many accounts the influencer is connecting to. From these five aspects, the potential of the influencer can be fully understood. The indicators contribute equally to the influencer identification.

6. Experiment

6.1. Experiment setup

We tried various strategies to obtain the best model performance by fine-tuning the text classification model on the SNS post text dataset in Korean. As the dataset is unbalanced, the minor classes may be ignored when performing the prediction. To avoid this bias problem, we used a random under-sampling technique to generate a new balanced dataset for fine-tuning the text classification model. Every category includes 134 data samples in the new balanced dataset. A cross validation technique was also applied to solve the insufficient data problem and to check if the performance was stable or not when training the model on different data. Different input lengths were set in order to search for an optimal parameter for the influencer marketing platform.

The dataset was first split into a training set and a test set, and then in the training set, StratifiedKFold function from scikit learn library was used to split the training set into a training subset and a validation subset that use different portions of the data to train and test a model on multiple iterations. After the cross validation process, the model with the best performance in the fine-tuning stage was evaluated on the test set. The hyperparameters of the text classification model are shown in Table 3.

After achieving optimal performance in the experiment, the text classification model with fine-tuned model parameters was utilized on the platform to perform the influencer classification task.

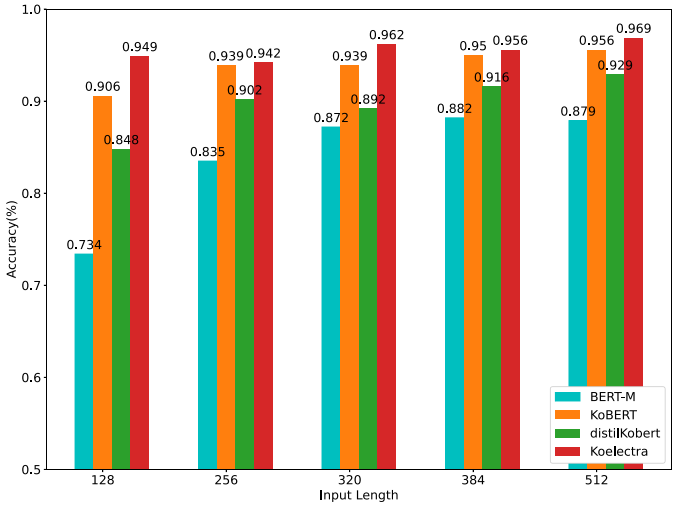

Fig. 3. Performance in different input lengths.

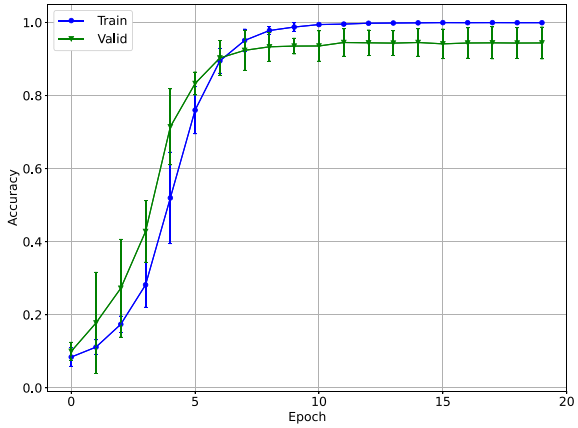

Fig. 4. Performance stability in cross validation.

6.2. Evaluation metrics

There are several evaluation metrics for the text classification task. In this paper, we use accuracy, precision, recall, and F1_score to evaluate the performance of the text classification model [50]. The equations of the metrics are shown as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%. \tag{4}$$

$$Precision = \frac{TP}{TP + FP} * 100\%. \tag{5}$$

$$Recall = \frac{TP}{TP + FN} * 100\%. \tag{6}$$

$$F1_score = 2 * \frac{Precision * Recall}{Precision + Recall} * 100\%. \tag{7}$$

For example, in the case of a binary classification task, TP means true positive, where a sample is correctly classified to positive. TN means true negative, where a sample is correctly classified to negative. FP means false positive, where a sample is incorrectly classified to positive. FN means false negative, where a sample is incorrectly classified to negative.

On the other hand, in the case of a multi-class classification task (e.g., 11 classes), let the 1st class be positive, and the other classes be

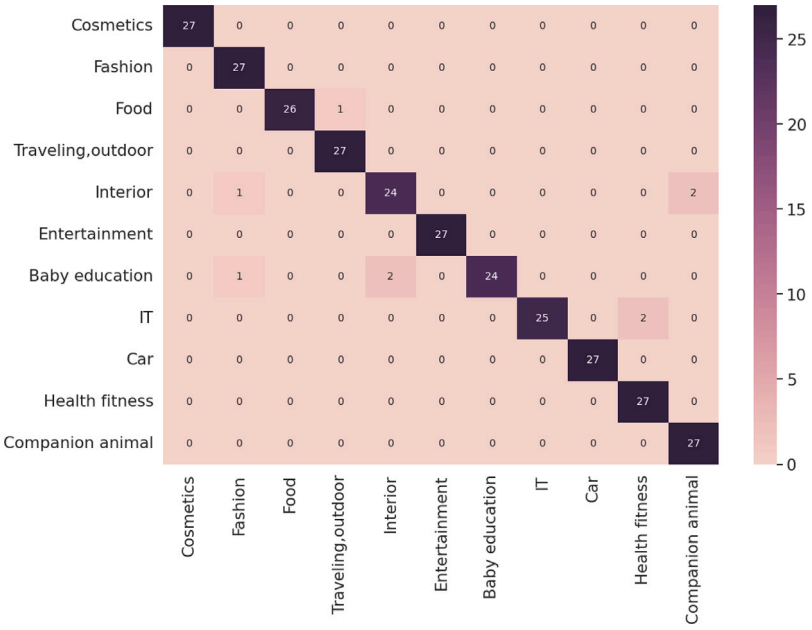

Fig. 5. Confusion matrix.

Table 4
The performance of four PTMs.

Model PTM	Accuracy	Precision	Recall	F1-score
BERT-M	0.7340	0.7392	0.7340	0.7332
KoBERT	0.9057	0.9081	0.9257	0.9058
DistilKoBERT	0.8484	0.8540	0.8484	0.8489
KoELECTRA	**0.9494**	**0.9518**	**0.9495**	**0.9498**

different negatives. TP means true positive, where a sample belonging to the 1st class is correctly classified to positive. TN means true negative, where a sample belonging to one of the other classes (except the 1st class) is correctly classified to its negative. FP means false positive, where a sample belonging to one of the other classes (except the 1st class) is incorrectly classified to positive (as the 1st class). FN means false negative, where a sample belonging to the 1st class is incorrectly classified to one of negatives (as one of the other classes except the 1st class).

The accuracy metric is the most widely used in classification tasks. The F1-score has become popular when it comes to the unbalanced dataset. The F1-score combines the precision and recall metrics into a single metric. The F1-score is designed to perform well on unbalanced data.

6.3. Performance

We carried out the text classification experiment based on different PTMs to ascertain which model is optimal for the platform. The performance of the text classification model corresponding to four metrics is shown in Table 4. During the experimentation, the input length parameter was set to 128. The findings indicate that the text classification model that incorporates the KoELECTRA as a pretrained model outperforms other pretrained models in terms of four metrics.

Additionally, we adjust the length of the input parameter to identify the optimal solution for the text classification task. The performance under different input lengths is depicted in Fig. 3. The findings consistently demonstrate that the text classification model utilizing KoELECTRA yields superior results in most cases. Therefore, for the purpose of the text classification task in this paper, we select the KoELECTRA as a PTM, and the input length is established at 512.

In order to assess the stability and robustness of the performance, cross validation was employed to determine whether the model remained robust or not when trained with different data sets. Fig. 4 presents the accuracy with error bars for each epoch. The curves in the figure represent the mean performance value for each epoch during different iterations of cross validation. The error bars depict the performance fluctuations of each epoch in terms of standard deviations during different iterations of cross validation. The performance demonstrated the robustness of the model in the downstream text classification task. Notably, the accuracy achieved 96.6% on the validation set.

The training subset and validation subset are employed to obtain model parameters that yield optimal performance. Subsequently, the model is further assessed using a test set. This evaluation provides a more objective assessment of the model's ability to generalize to unseen data. The accuracy on the test set achieved 96.9%. The evaluations that were performed on the test set and validation set yield similar results, indicating that the text classification model is well-tuned and exhibits stable performance. The confusion matrix is utilized to examine the classification details (as shown in Fig. 5). The confusion matrix can be defined as $M = [a_{ij}]_{m*n}$, while m and n both equal eleven for eleven categories. In the confusion matrix, the row categories correspond to the actual sample labels, which is the ground truth, the column categories correspond to the predicted sample labels. The numbers along the diagonal line represent the samples of either true positive (TP) or true negative (TN). On the other hand, the numbers in the other positions represent the samples of either false negative (FN) or false positive (FP). According to the accuracy formula in (4), the accuracy in the confusion matrix can be calculated in (8) as follows:

$$Accuracy_{cm} = \frac{\sum_{i=1}^{n} a_{ii}}{\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}} * 100\%, \tag{8}$$

where a_{ii} represents the number of samples well-classified on the diagonal of the confusion matrix, a_{ij} means the number of all samples in the confusion matrix, that is, the number of samples in the test set.

Based on the confusion matrix, the test set data was classified effectively, as the majority of the samples were correctly classified. However, one sample in the food category was misclassified as belonging to the traveling or outdoor category. Three samples in the interior category were misclassified as belonging to the fashion or pet companion categories. Three samples in baby education were misclassified as belonging to the fashion or interior categories. Lastly, two samples in IT were misclassified as belonging to the health fitness category.

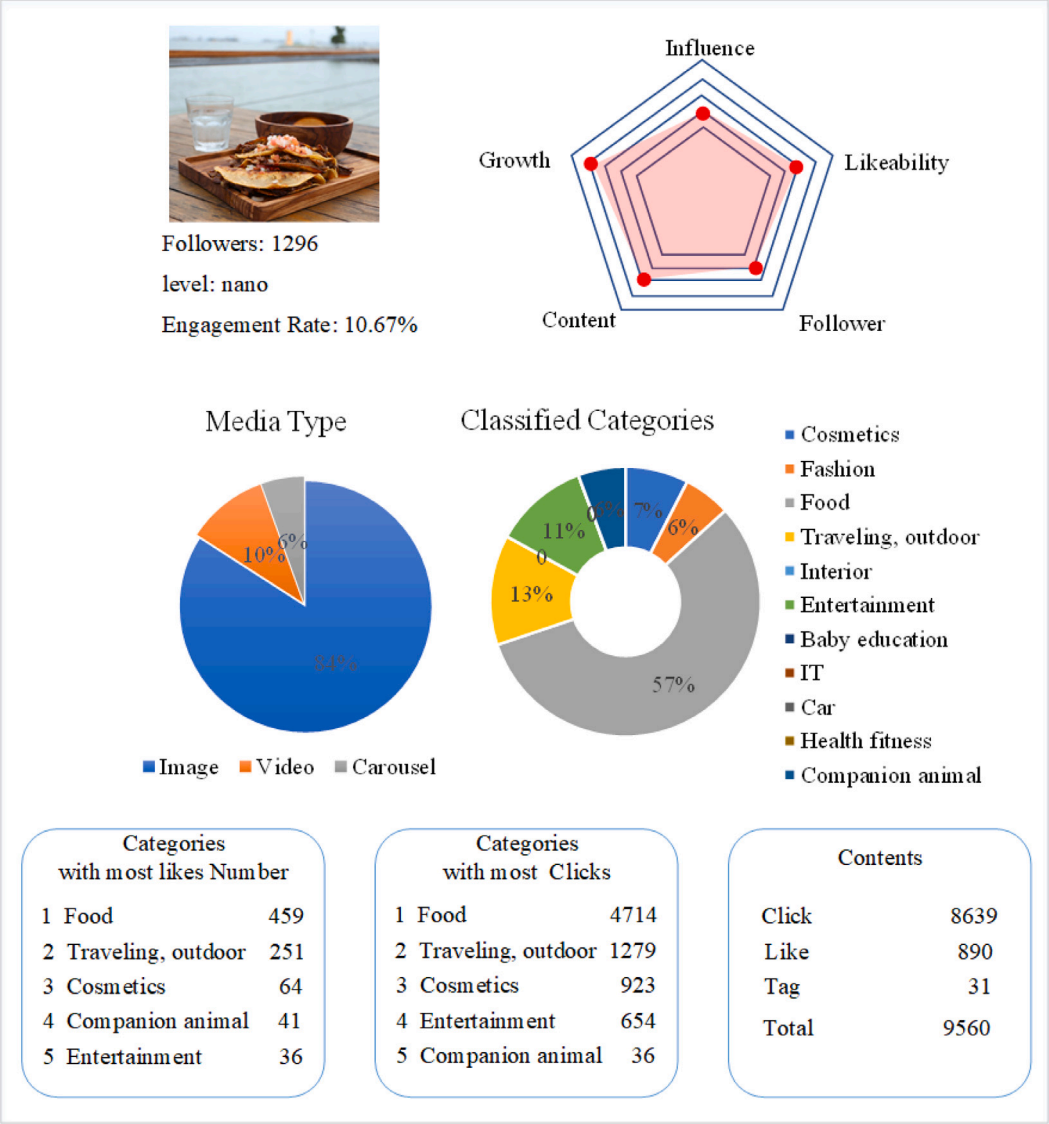

Fig. 6. Influencer classification analysis on the influencer marketing platform.

6.4. Implementation of the influencer marketing platform

After fine-tuning the text classification model, we further implement the activities on the influencer marketing platform. The platform is developed as a real-world solution and is applied to a startup company in the field of influencer marketing. The platform comprises a front-end and a back-end. The front-end component includes a web page that facilitates the registration of clients and influencers, the initiation of campaigns, influencer applying campaigns, clients selecting influencers and so on. Additionally, the back-end component manages requests from the front-end, manipulates data by interacting with a database. For the detailed description of our platform, refer to our previous work [18].

On the platform, both influencers and business clients register their participation using the web page browser with the database. Clients initiate campaigns, and influencers submit applications to those campaigns. The clients can then select suitable influencers accordingly. When the clients review an influencer's information, the platform initially requests and acquires all the media and insight associated with that influencer. Subsequently, every post text is categorized into a category, and a detailed classification distribution is accumulated utilizing a text classification model stored within the database.

Table 5
Performance on the influencer marketing platform.

Influencer	Accuracy	Top-3 accuracy
Influencer ID1	83.3%	87.5%
Influencer ID2	96.4%	96.4%
Influencer ID3	36.8%	42.1%
Total	76.1%	78.87%

Simultaneously, comprehensive influence indicators are computed using the influencer insight data. Upon the completion of this process, the platform conducts a thorough examination of each influencer, generating a visual analysis report that clearly presents the influencer's information facilitating the matching of influencers, as shown in Fig. 6. This figure shows an example of an influencer analysis report. The influencer information includes the classification result distribution shown in the categories, the comprehensive influence indicators including influence level, growth, content, follower, likeability, the media type distribution, and the detailed engagement situation. From the analysis report, the client can clearly understand the advertising potential of each influencer. After the selection, the selected influencer will create new posts on his (or her) social media App like Instagram, to promote

Table A.6
Instagram graph API functions.

Function name	API function	Description
Hashtag Top Media	GET/{ig_hashtag_id}/ top_media? & user_id= {user_id}fields={fields}	Return most popular media objects that have been tagged with the hashtag.
Hashtag Search	GET/ig_hashtag_search? user_id={user-id}&q={q}	Return the IG hashtag ID if the hashtag exists on the Instagram.
Business Discovery	GET/ig-user-id/ business_discovery	Return basic data about another Instagram business user such as follower number and media objects.
Insights	GET /{ig-user-id}/insights? metric=reach, profile_views &period=day	Return insights of another Instagram user including follower information, website clicks, reach, etc.

the campaign. Finally, each client receives the campaign promotion report to check the benefits of the advertising campaign. Therefor, our influencer marketing platform works well with our text classification model.

6.5. Evaluation of the influencer marketing platform

We also evaluated the text classification model's performance on the platform. Unlike testing the model in the dataset, the data on the platform is real world data. As the platform was in the development stage, one client and three influencers registered themselves with it. We only evaluated based on the data of three influencers. The data is open domain data, so we provide accuracy and top 3 accuracy as metrics. The result is shown in Table 5.

For the first two influencers, the influencer marketing platform demonstrates a high level accuracy comparable to the dataset. However, the third influencer exhibits lower accuracy. Upon reviewing the text content, it was observed that the text of the third influencer primarily consists of short content with a length of fewer than 10 words. This may limit the information available for effective text classification. The text classification model effectively classifies the text of influencer ID1 and influencer ID2, even when they incorporate numerous sentences containing hashtags. This performance can be attributed to the pretrained model of KoELECTRA, which utilizes a small generator network to offer alternative sampled tokens rather than masking the input tokens. Additionally, KoELECTRA identifies the authenticity among these tokens, further enhancing performance. Therefore, it is concluded that our influencer marketing platform runs well such that the text classification model has considerable performance in the real world data.

7. Conclusion

This paper introduces an intelligent marketing platform with influencer classification in social networking services (called SNS). This paper provide a web-based, end-to-end service platform for the real world influencer marketing industry. For facilitating the matching between the business clients and influencers, the dataset was collected to match the category requirements of the platform. The text classification model, which was based on PTMs, was designed and fine-tuned by the dataset for the platform influencer classification task. This paper also calculates the comprehensive influence indicators of the influencers to help understand their potential. With the influencer classification result and comprehensive influence indicators, the platform can generate a detailed analysis report for every influencer in order to enable business clients to choose the right influencers for their advertising campaigns. Our influencer marketing platform works well with our text classification model. The performance on the marketing platform is also considerable, as the data on the platform are open-domain data, not limited to the dataset. Also, the dataset size for fine-tuning the model is still small due to the limitation of data collection time. For future work, we will keep collecting datasets to improve performance and make better generalization on the platform. We will also pursue a better methodology for the influencer classification analysis on our influencer marketing platform along with images and video clips of influencers.

CRediT authorship contribution statement

Xiaohong Yu: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jinyong Kim:** Validation, Software, Investigation, Data curation, Conceptualization. **Yoseop Ahn:** Validation, Software, Formal analysis, Data curation. **Mose Gu:** Writing – original draft, Visualization, Software. **Jaehoon (Paul) Jeong:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **JinYeong Bak:** Writing – review & editing, Supervision, Investigation, Formal analysis. **Jaemin Jo:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jaehoon (Paul) Jeong reports financial support was provided by Institute of Information & Communications Technology Planning & Evaluation (IITP). Jaehoon (Paul) Jeong reports a relationship with Sungkyunkwan University - Natural Sciences Campus that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grants funded by the Korea government (MSIT) (No. RS-2022-II221015, Development of Candidate Element Technology for Intelligent 6G Mobile Core Network and No. RS-2022-II221199: Regional Strategic Industry Convergence Security Core Talent Training Business). This work was supported in part by the grant of IPCX (Innovative Plan Creative eXecution), Inc. and the IPCX's help for dataset collection.

Appendix A. Instagram graph API functions

The dataset was collected using the Instagram Graph API functions [47]. We used several API functions (as shown in Table A.6) in this paper.

We selected the subcategories as hashtags to search if the hashtag exists or not on Instagram using the hashtag search function, and then collected media data using the hashtag top media function. Here, the fields in the function include post caption, post id, like_count,

comment _count, media_type, media_url, permalink, and timestamp. We use post_captions since we perform text classification. like_count, and comment_count are used to calculate the comprehensive influence indicators. The business discovery API and Insights API functions were used in the influencer analysis stage. The business discovery API function was used to access all media data of one influencer based on the user_id. The Insights API function was used to access the influencer's Insight which includes the number of followers, followers distribution, impressions, reach, website clicks, and profile views, which are used for calculating the comprehensive influence of the influencer measured by five indicators.

Data availability

Data will be made available on request.

References

[1] W. Geyser, The state of influencer marketing 2024: Benchmark report, 2024, https://influencermarketinghub.com/influencer-marketing-benchmark-report/.

[2] L. Wang, Z. Yu, B. Guo, D. Yang, L. Ma, Z. Liu, F. Xiong, Data-driven targeted advertising recommendation system for outdoor billboard, ACM Trans. Intell. Syst. Technol. 13 (2) (2022) 23, http://dx.doi.org/10.1145/3495159.

[3] Y. Bu, J. Parkinson, P. Thaichon, Influencer marketing: Homophily, customer value co-creation behaviour and purchase intention, J. Retail. Consumer Serv. 66 (2022) 102904, http://dx.doi.org/10.1016/j.jretconser.2021.102904.

[4] S. Wu, W. Li, H. Shen, Q. Bai, Identifying influential users in unknown social networks for adaptive incentive allocation under budget restriction, Inform. Sci. 624 (2023) 128–146, http://dx.doi.org/10.1016/j.ins.2022.12.071.

[5] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09, Association for Computing Machinery, New York, NY, USA, 2009, pp. 199–208, http://dx.doi.org/10.1145/1557019.1557047.

[6] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, Association for Computing Machinery, New York, NY, USA, 2003, pp. 137–146, http://dx.doi.org/10.1145/956750.956769.

[7] S. Mihara, S. Tsugawa, H. Ohsaki, Influence maximization problem for unknown social networks, in: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, 2015, pp. 1539–1546, http://dx.doi.org/10.1145/2808797.2808885.

[8] B. Wilder, L. Onasch-Vera, J. Hudson, J. Luna, N. Wilson, R. Petering, D. Woo, M. Tambe, E. Rice, End-to-end influence maximization in the field, in: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2018, pp. 1414–1422.

[9] S. Wu, W. Li, Q. Bai, GAC: A deep reinforcement learning model toward user incentivization in unknown social networks, Knowl.-Based Syst. 259 (2023) 110060, http://dx.doi.org/10.1016/j.knosys.2022.110060.

[10] R. Doshi, A.R. Ranganathan, S. Rao, Modeling influencer marketing campaigns in social networks, 2021, CoRR abs/2106.01750. arXiv:2106.01750.

[11] L. Zhang, J. Shen, J. Zhang, J. Xu, Z. Li, Y. Yao, L. Yu, Multimodal marketing intent analysis for effective targeted advertising, IEEE Trans. Multimed. 24 (2022) 1830–1843, http://dx.doi.org/10.1109/TMM.2021.3073267.

[12] C. Jiang, W. Li, S. Wu, Q. Bai, OMT: An operate-based approach for modelling multi-topic influence diffusion in online social networks, in: W. Zhang, L. Zou, Z. Maamar, L. Chen (Eds.), Web Information Systems Engineering – WISE 2021, Springer International Publishing, Cham, 2021, pp. 542–556.

[13] M.D. Veirman, V. Cauberghe, L. Hudders, Marketing through instagram influencers: the impact of number of followers and product divergence on brand attitude, Int. J. Advert. 36 (5) (2017) 798–828, http://dx.doi.org/10.1080/02650487.2017.1348035.

[14] S. Kim, J.-Y. Jiang, M. Nakada, J. Han, W. Wang, Multimodal post attentive profiling for influencer marketing, in: Proceedings of The Web Conference 2020, WWW '20, 2020, http://dx.doi.org/10.1145/3366423.3380052.

[15] M. Pennacchiotti, A.-M. Popescu, A machine learning approach to Twitter user classification, Proc. Int. AAAI Conf. Web Social Media 5 (1) (2011) http://dx.doi.org/10.1609/icwsm.v5i1.14139.

[16] M.S. Rahman, H. Reza, A systematic review towards big data analytics in social media, Big Data Min. Anal. 5 (3) (2022) 228–244, http://dx.doi.org/10.26599/BDMA.2022.9020009.

[17] D. Rogers, A. Preece, M. Innes, I. Spasić, Real-time text classification of user-generated content on social media: Systematic review, IEEE Trans. Comput. Soc. Syst. 9 (4) (2022) 1154–1166, http://dx.doi.org/10.1109/TCSS.2021.3120138.

[18] M. Gu, J. Kim, Y. Shim, J.P. Jeong, SNS influencer marketing platform for market advertisement of small-sized business runners, in: 24th International Conference on Advanced Communication Technology, ICACT, 2022, pp. 513–517, http://dx.doi.org/10.23919/ICACT53585.2022.9728965.

[19] K. Chakraborty, S. Bhattacharyya, R. Bag, A survey of sentiment analysis from social media data, IEEE Trans. Comput. Soc. Syst. 7 (2) (2020) 450–464, http://dx.doi.org/10.1109/TCSS.2019.2956957.

[20] H.-C. Soong, N.B.A. Jalil, R. Kumar Ayyasamy, R. Akbar, The essential of sentiment analysis and opinion mining in social media : Introduction and survey of the recent approaches and techniques, in: 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics, ISCAIE, 2019, pp. 272–277, http://dx.doi.org/10.1109/ISCAIE.2019.8743799.

[21] T. Wang, K. Lu, K.P. Chow, Q. Zhu, COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model, IEEE Access 8 (2020) 138162–138169, http://dx.doi.org/10.1109/ACCESS.2020.3012595.

[22] U. Naseem, I. Razzak, M. Khushi, P.W. Eklund, J. Kim, COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis, IEEE Trans. Comput. Soc. Syst. 8 (4) (2021) 1003–1015, http://dx.doi.org/10.1109/TCSS.2021.3051189.

[23] F.K. Sufi, I. Khalil, Automated disaster monitoring from social media posts using AI-based location intelligence and sentiment analysis, IEEE Trans. Comput. Soc. Syst. (2022) 1–11, http://dx.doi.org/10.1109/TCSS.2022.3157142.

[24] J. Wen, P. Liu, A classification method for english texts based on hybrid recurrent neural network and graph construction in social recommendation systems, IEEE Syst. J. 17 (4) (2023) 5272–5279, http://dx.doi.org/10.1109/JSYST.2023.3283147.

[25] P.P. Ramadhani, S. Hadi, Text classification on the instagram caption using support vector machine, J. Phys. Conf. Ser. 1722 (1) (2021) 012023, http://dx.doi.org/10.1088/1742-6596/1722/1/012023.

[26] D. Kim, J. Koo, U.-M. Kim, EnvBERT: Multi-label text classification for imbalanced, noisy environmental news data, in: 2021 15th International Conference on Ubiquitous Information Management and Communication, IMCOM, 2021, pp. 1–8, http://dx.doi.org/10.1109/IMCOM51814.2021.9377411.

[27] F. Bonchi, C. Castillo, A. Gionis, A. Jaimes, Social network analysis and mining for business applications, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) http://dx.doi.org/10.1145/1961189.1961194.

[28] J. Wang, Y. Yang, Q. Liu, Z. Fang, S. Sun, Y. Xu, An empirical study of user engagement in influencer marketing on weibo and WeChat, IEEE Trans. Comput. Soc. Syst. 10 (6) (2023) 3228–3240, http://dx.doi.org/10.1109/TCSS.2022.3204177.

[29] Y. Wang, Z. He, J. Hu, Traffic information mining from social media based on the MC-LSTM-conv model, IEEE Trans. Intell. Transp. Syst. 23 (2) (2022) 1132–1144, http://dx.doi.org/10.1109/TITS.2020.3021096.

[30] P. Kazienko, K. Musial, T. Kajdanowicz, Multidimensional social network in the social recommender system, IEEE Trans. Syst., Man, Cybern. - Part A: Syst. Humans 41 (4) (2011) 746–759, http://dx.doi.org/10.1109/TSMCA.2011.2132707.

[31] S. Dhelim, H. Ning, N. Aung, R. Huang, J. Ma, Personality-aware product recommendation system based on user interests mining and metapath discovery, IEEE Trans. Comput. Soc. Syst. 8 (1) (2021) 86–98, http://dx.doi.org/10.1109/TCSS.2020.3037040.

[32] Y. Xiao, Y. Zhu, W. He, M. Huang, Influence prediction model for marketing campaigns on e-commerce platforms, Expert Syst. Appl. 211 (2023) 118575, http://dx.doi.org/10.1016/j.eswa.2022.118575.

[33] Y. Hu, J. Li, Y. Ruan, Finding influencers in complex networks: A novel method based on information theory, IEEE Syst. J. 16 (2) (2022) 3372–3380, http://dx.doi.org/10.1109/JSYST.2021.3119081.

[34] W. Li, Y. Hu, C. Jiang, S. Wu, Q. Bai, E. Lai, ABEM: An adaptive agent-based evolutionary approach for influence maximization in dynamic social networks, Appl. Soft Comput. 136 (2023) 110062, http://dx.doi.org/10.1016/j.asoc.2023.110062.

[35] C.-C. Hung, Y.-C. Huang, J.Y. jen Hsu, D.K.-C. Wu, Tag-based user profiling for social media recommendation, 2008, https://api.semanticscholar.org/CorpusID:15474630.

[36] Q. You, S. Bhatia, J. Luo, A picture tells a thousand words – About you! User interest profiling from user generated visual content, 2015, arXiv:1504.04558.

[37] G. Farnadi, J. Tang, M. De Cock, M.-F. Moens, User profiling through deep multimodal fusion, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 171–179, http://dx.doi.org/10.1145/3159652.3159691.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2017, CoRR abs/1706.03762. arXiv:1706.03762.

[39] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, 2018, CoRR abs/1810.04805. arXiv:1810.04805.

[40] V. Sanh, L. Debut, J. Chaumond, T. Wolf, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, 2020, arXiv:1910.01108.
[41] K. Clark, M.-T. Luong, Q.V. Le, C.D. Manning, ELECTRA: Pre-training text encoders as discriminators rather than generators, 2020, arXiv:2003.10555.
[42] SKTBrain, SKTBrain/Kobert: Korean bert pre-trained cased (Kobert), 2021, https://github.com/monologg/KoBERT-Transformers. (Accessed: 21 June 2021).
[43] J. Park, DistilKoBERT: Distillation of kobert, 2019, GitHub. https://github.com/monologg/DistilKoBERT.
[44] J. Park, KoELECTRA: Pretrained ELECTRA model for Korean, 2020, GitHub. https://github.com/monologg/KoELECTRA.
[45] S. Barta, D. Belanche, A. Fernández, M. Flavián, Influencer marketing on TikTok: The effectiveness of humor and followers' hedonic experience, J. Retail. Consumer Serv. 70 (2023) 103149, http://dx.doi.org/10.1016/j.jretconser.2022.103149.
[46] L. Li, W. Chou, W. Zhou, M. Luo, Design patterns and extensibility of REST API for networking applications, IEEE Trans. Netw. Serv. Manag. 13 (1) (2016) 154–167, http://dx.doi.org/10.1109/TNSM.2016.2516946.
[47] Instagram, Instagram graph API:Developer documentation, 2024, https://developers.facebook.com/docs/instagram-api. (Accessed: 21 June 2024).
[48] M. Asif, D. Zhiyong, A. Iram, M. Nisar, Linguistic analysis of neologism related to coronavirus (COVID-19), Soc. Sci. Humanit. Open 4 (1) (2021) 100201, http://dx.doi.org/10.1016/j.ssaho.2021.100201.
[49] J. Bridle, Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters, in: D. Touretzky (Ed.), Advances in Neural Information Processing Systems, Vol. 2, NIPS '89, Morgan-Kaufmann, 1989, pp. 211–217.
[50] M. Sokolova, N. Japkowicz, S. Szpakowicz, Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation, in: A. Sattar, B.-h. Kang (Eds.), AI 2006: Advances in Artificial Intelligence, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 1015–1021.