Design and Build an Instagram Content Sentiment Analysis Application Using the *Bidirectional Encoder Representation From Transformer Model*

Fadila *, Yeka Hendriyani , Zulhendra , Khairi Budayawan

Department of Electronics Engineering, Universitas Negeri Padang, Padang, Indonesia

ORCID: https://orcid.org/0009-0000-6506-7212

Abstract **- The high activity of Instagram users in Indonesia, with the number of users reaching 89.15 million in January 2023, reflects the great potential for sentiment analysis on social media. This research aims to develop a web-based application that can analyze the sentiment of Instagram comments in real-time regarding government policies, using the BERT (*Bidirectional Encoder Representation from Transformer*) model. BERT is a deep learning model developed by Google in 2018, which has the ability to understand word context bidirectionally. This research uses a method of processing comment text from Instagram with NLP techniques and trains the IndoBERT model, which is specially adapted for Indonesian. The dataset used includes labeled data from Instagram and Kaggle *scraping*, as well as real-time data captured using Instaloader. The results showed that the BERT model achieved 91.57% accuracy in classifying sentiment as positive or negative. In conclusion, the developed application successfully provides an effective platform for sentiment analysis of Instagram content, providing valuable insights into public opinion towards government policies. This research recommends further exploration of various BERT models in Indonesian language social media sentiment analysis.**

Keywords **—BERT, NLP, sentiment analysis, machine learning, government policy.**

I. INTRODUCTION

Social media as information and communication technology facilitates interaction without time and space limitations, one platform that allows free expression and opinion is Instagram [1]. Based on the *We Are Social* report in January 2023, the number of Instagram users in Indonesia reached 89.15 million, making it the country with the 4th largest number of users in the world [2]. This phenomenon reflects the high activity of Indonesian people on social media platforms, especially Instagram. The active involvement of Instagram users in sharing information, interacting, collaborating, and conveying ideas through posts opens up opportunities to carry out sentiment analysis [3].

DOI: 10.18782/ijsehtxx-xx
https://doi.org/10.18782/ijeclxx-xx

Corresponding author: Fadila ,
affiliation
Email: *kabar5627@gmail.com,*

Received: August 2024.
Revised: September 2024.
Accepted: October 2024.
Published: October 2024.

Sentiment analysis assesses the opinions or emotions contained in the text, making it possible to understand how users feel about a particular topic or event which can produce valuable data and information, including comments from the Community [1]. The data generated from these activities provides valuable insights into societal trends, preferences and views, and helps in understanding social dynamics, culture and political trends.

Government policies are often the subject of debate and public attention, especially when they involve significant changes in people's daily lives. Openness in policy making is a necessity for local governments to implement good public services [4]. This policy can influence government decisions and trigger various comments on social media. Therefore, sentiment analysis of public comments regarding government policies on social media can provide a clear picture of public perceptions and help policymakers respond more effectively.

To understand these reactions more deeply and objectively, sentiment analysis technology based on *Natural Language Processing* (NLP) can be used. Sentiment analysis is closely related to NLP, namely the ability of machines to understand human language through the use of computer algorithms, mathematics, and computational linguistics [5]. NLP, which focuses on training computers to understand, process, and produce language, is a branch of *artificial intelligence* (AI) [6].

In the context of sentiment analysis, NLP is used to extract meaning from text produced by individuals, including netizen comments on social media, news articles, and others. In the context of sentiment analysis, NLP is used to extract meaning from text produced by individuals, including comments on social media and news articles. Algorithms and NLP techniques make it possible to extract valuable information from the large and diverse text data on Instagram. One very effective NLP

technique is BERT (*Bidirectional Encoder Representations from Transformers*).

BERT replaces recurrent networks with attention mechanisms *and* understands contextual relationships between distant words in NLP. Developed based on the *Transformer architecture* , the BERT model has been proven to be more effective in handling NLP tasks compared to traditional approaches [5]. BERT has been a pioneer in *transfer learning* in the field of NLP. This model allows the use of general representations for specific tasks after the training process. Thus, BERT models have changed the paradigm of natural language processing and paved the way for more sophisticated applications, such as text classification, sentiment analysis, and deeper language understanding [7].

Based on research results on reviews of the Gundala film on YouTube, it shows that BERT achieved an average accuracy of 66.7% higher than *Naïve Bayes* and *Support Vector Machine* (SVM) for Indonesian sentiment analysis. These findings confirm BERT's superiority in understanding context and semantics in depth, which cannot be achieved by traditional algorithms such as Naïve Bayes and SVM [8].

As the need for *real-time data analysis* and data-driven decisions increases, the development of web-based applications for sentiment analysis is becoming a potential solution. This application provides efficient and intuitive access to analyze Instagram comments in real-time, making it easy to access and use, and ensuring analysis results can be accessed quickly and securely. This application aims to process comment text from Instagram posts using the BERT model. The main focus of the research is real-time sentiment analysis of Instagram content, with the aim of classifying opinions as positive or negative, displaying post images, captions, and sentiment analysis results, including percentages and details of comments by sentiment category. The results of this analysis are expected to provide valuable insight into public sentiment towards applicable government policies

II. METHODS

A. Research Data

1) Labeled Dataset

The dataset dataset_ta.csv was used in this research, which was obtained from *scraping* Instagram and Kaggle data. This dataset has been labeled with positive (label 1) and negative (label 0) sentiment, consisting of two main columns: *comment* (comment text) and label (sentiment value). This dataset is used to train the BERT model to recognize and classify comment sentiment in *real-time data*

2) Real-Time Comment Data from Instagram

This research also uses real-time comment data from Instagram taken with Instaloader. The data taken includes *usernames* and *comments* from Instagram users. By inputting their Instagram *username* and *password*, users can authenticate and retrieve comment data for *real-time sentiment analysis*

B. Bert Models

Bidirectional Encoder Representation from Transformer (BERT) is a language model developed by Google in 2018 and is an example of a *deep learning model*. BERT has the ability to understand the context of words in sentences bidirectionally (two-way) [9].

There are two steps in the performance of the BERT model, namely *pre-training* and *fine-tuning* . The *pre-training* process is carried out using a very large dataset and Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) techniques, the *fine-tuning process* is carried out using a dataset that is relevant to the task to be completed [10].

C. System Development Process

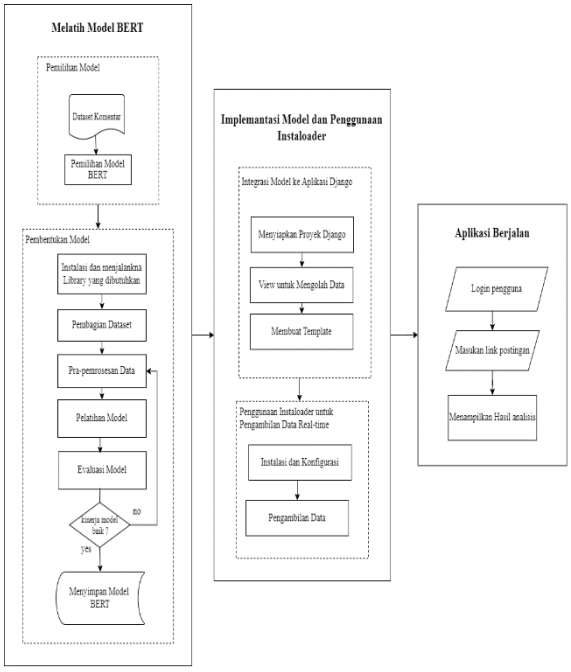


Figure 1. System development process

1) Training a BERT Model

The BERT model used in this research is a BERT model trained for Indonesian. This model was then retrained (fine-tuned) using the dataset dataset_ta.csv to classify comment sentiment.

a. Model Selection

The BERT model used is a BERT model that has been trained for Indonesian (IndoBERT). The selection of this model was based on its ability to understand and process text in Indonesian with a high level of accuracy.

b. Model Formation

The BERT model used is a BERT model that has been trained for Indonesian (IndoBERT). The selection of this model is based on its ability to understand and process text in Indonesian with a high level of accuracy.

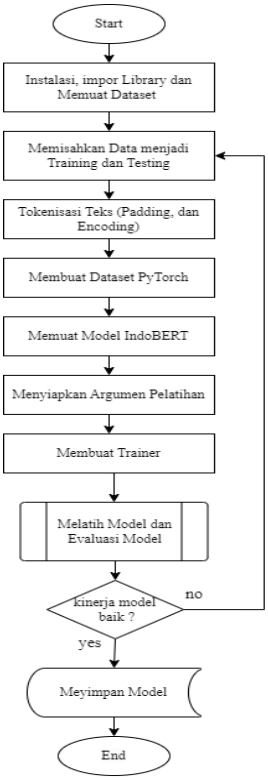


Figure 2. BERT model

2) Model Integration and Use of Instaloader

a. Create a Django project

framework is the main framework that has significant advantages in web development compared to other frameworks based on *object-oriented programming* [11]. To start the application, make sure Python and Django are installed. Create a new project and then create a virtual environment to isolate dependencies. Install libraries such as Transformers, Torch, and Instaloader. After that, create an application using the python manage.py startapp command. Prepare templates for input and web display, then integrate BERT models for sentiment analysis in the Django *framework*.

b. Integrating BERT Models

The BERT model shown in the figure has gone through a training process and is saved along with various important configurations, including config.json, model.safetensors, special_tokens_map.json, tokenizer_config.json, and vocab.txt. To integrate this model into a Django application.

c. Implementation in views.py

The code that will be implemented in the Django framework in the view.py file aims to carry out sentiment analysis on Instagram comments using the IndoBERT model and instaloader is used to retrieve data from Instagram. This application is designed to allow users *to log* into an Instagram account, retrieve comments from a post and analyze the sentiment of those comments using a pre-trained machine learning model.

3) Running Application

The application starts with the user *logging in*, then they enter the link of the Instagram post they want to analyze. The application takes comment data and analyzes sentiment using BERT model.

D. Design Process

1) Use Case Diagrams

Use Case Diagram is a type of diagram in *the Unified Modeling Language* (UML), used to model interactions between external actors and the system [12]. This diagram helps in identifying the main functions of the system and the relationships between the actors and *use cases* involved. In this system, flowcharts are used for *real-time sentiment analysis of Instagram content*.

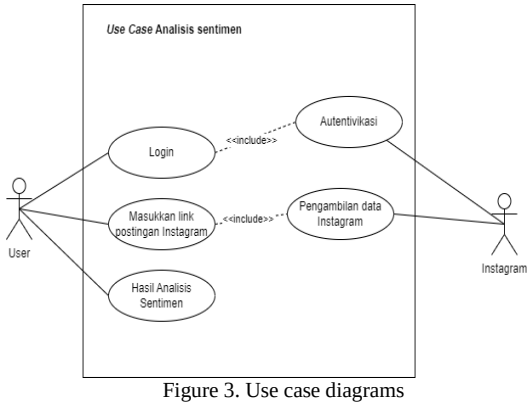


Figure 3. Use case diagrams

2) Activity Diagrams

Activity Diagram, a type of diagram in *the Unified Modeling Language (UML)*, is used to model workflows or activities in a system [12]. In this system, the sentiment analysis system consists of two main parts, namely user interaction and system flow.

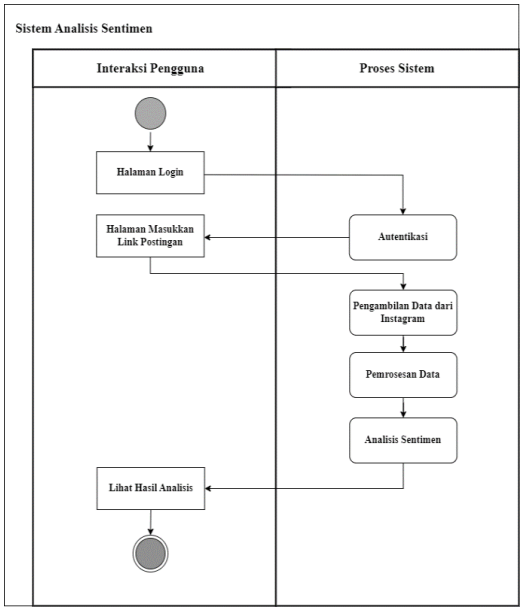


Figure 4. Activity diagrams

3) Design the Interface

a. Login page design

This page is designed so that users can enter their Instagram *username* and *password* . The page layout will display columns for entering *the username* and *password* , as well as a button for *logging in.*

Figure 5. *Login* page design

b. Instagram post link input page design

This page is provided for users to enter post links from the Instagram account they want to analyze. This page displays clear placement for each input field, complete with labels that ensure users know what information needs to be entered.

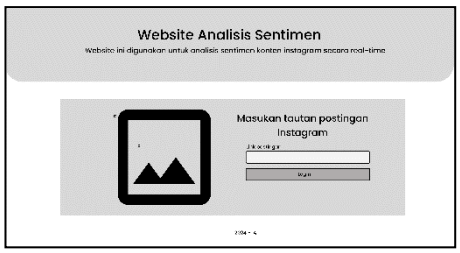


Figure 6. Post link input page design

c. Design the analysis results page

This page displays images of posts as well as descriptions and sentiment analysis results, including the percentage of positive and negative sentiment as well as detailed comments based on sentiment categories.

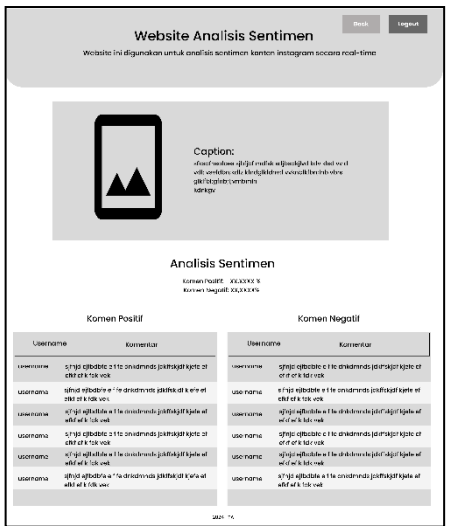


Figure 7. Design the analysis results page

III. RESULTS

A. *Results of BERT Model Establishment*

The results of the training and model building process in Google Colab produce a model and tokenizer which are saved in the local directory (./indobert_sentiment_model).

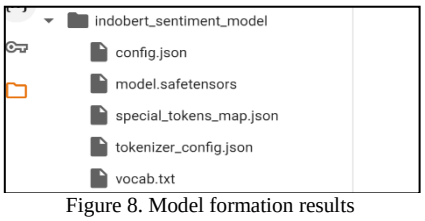
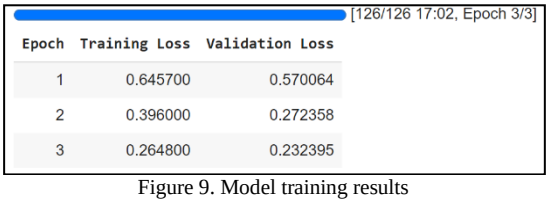


Figure 8. Model formation results

B. *Model Testing*

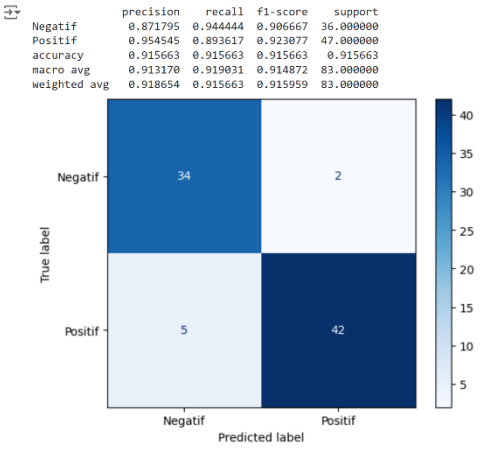
The model training results show a significant reduction in training loss and validation loss, which indicates that the BERT model successfully learns from the data and improves its ability to classify sentiment. This decrease indicates that the model not only memorizes the training data but can also generalize well to data it has never seen before.



Epoch	Training Loss	Validation Loss
1	0.645700	0.570064
2	0.396000	0.272358
3	0.264800	0.232395

Figure 9. Model training results

Accuracy testing using evaluation metrics such as precision, recall, and F1-score ensures that the model provides reliable results in sentiment classification. Overall, this application is able to provide effective and useful sentiment analysis for users.



	precision	recall	f1-score	support
Negatif	0.871795	0.944444	0.906667	36.000000
Positif	0.954545	0.893617	0.923077	47.000000
accuracy	0.915663	0.915663	0.915663	0.915663
macro avg	0.913170	0.919031	0.914872	83.000000
weighted avg	0.918654	0.915663	0.915959	83.000000

Figure 10. Accuracy testing

The model shows good performance with an accuracy of 91.57%, successfully classifying positive and negative sentiments equally, with high precision and recall in both classes, indicating that this model is effective in detecting and classifying sentiments on the given data.

C. *Application Interface*

1) Login Page

login page allows users to enter authentication data, namely the Instagram *username* and *password*, to

access the application. Users are expected to fill in the fields provided with valid information to continue the authentication process.

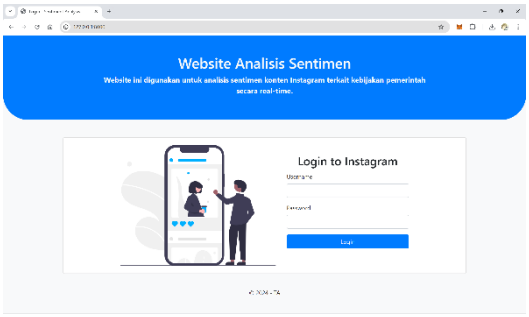


Figure 11. *Login* page

2) *Post Link Input Page*

This page allows users to enter the link of the Instagram post they want to analyze. Users are expected to enter the relevant URL into the column provided, so that the application can process and analyze the post.

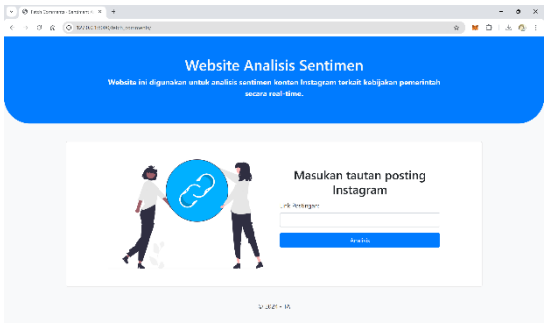


Figure 12. Post *link* input page

3) *Analysis Results Page*

This page displays the analysis results of the Instagram posts that have been entered, including images and descriptions of the posts. Users can see a summary of sentiment analysis results, which includes the percentage of positive and negative sentiment, as well as detailed comments grouped by sentiment category. This information is presented to provide a comprehensive overview of the sentiment contained in the post.

Figure 13. analysis results page

D. System Testing

TABLE I
Functional Testing

No.	Test	Expected results	Test Results	Status
1	Test the login page	Users with valid authentication information can proceed to the next page.	The system successfully authenticates the user and displays appropriate error messages if it fails.	As expected
2	Test instagram link login page	The system processes valid links and rejects invalid links with an error message.	The system successfully processes valid links and displays an error message for invalid links.	As expected
3	Testing analysis results	The system displays sentiment analysis results accurately.	The system successfully displays the sentiment percentage and comment classification.	As expected

E. Discussion

In a new Criminal Code sentiment analysis study on Twitter data using the BERT model, it was found that the BERT model achieved 81% accuracy on test data, outperforming the Support Vector Machine (SVM) with an accuracy difference of 6% [13]. This research also implements the BERT (*Bidirectional Encoder Representations from Transformers*) model with a model accuracy of 91.57% for real-time sentiment analysis of Instagram comments, using IndoBERT, which is a special adaptation of BERT for Indonesian. The research process begins with installing the necessary libraries and loading the dataset dataset_ta.csv, which contains comments labeled with positive and negative sentiments. This dataset is divided into training and test data to train and evaluate the model. Text tokenization is carried out to convert comments into tokens that can be processed by the BERT model, and then converted into numeric vectors.

The model training results show a significant reduction in *training loss* and *validation loss*, which indicates that the BERT model successfully learns from the data and improves its ability to classify sentiment. This decrease indicates that the model not only memorizes the training data but can also generalize well to data it has never seen before. Accuracy testing using evaluation metrics such as precision, recall, and F1-score ensures that the model provides reliable results in sentiment classification. Overall, this application is able to provide effective and useful sentiment analysis for users.

The integration of BERT models into Django-based applications enables efficient sentiment analysis of Instagram comments. The application interface includes a *login page*, Instagram link input, and analysis results page. Functional testing confirmed that the app can authenticate users, process

Instagram links, and display sentiment analysis results accurately.

IV. Conclusions

This project succeeded in designing and developing an effective web-based application for sentiment analysis of Instagram content using the Django framework. This application is not only able to retrieve comment data in real-time but also uses the BERT model to perform sentiment analysis efficiently. By providing an effective platform for interpreting public opinion, this application fulfills the main objective of the project.

Future research is recommended to explore and compare the performance of various BERT models in the context of Instagram content sentiment analysis. This comparison will provide in-depth insight into the strengths and weaknesses of each model in social media sentiment analysis, especially in Indonesian.

References

[1] Permatasari, P. A., Linawati, L., & Jasa, L. (2021). Survei tentang analisis sentimen pada media sosial. *Maj. Ilm. Teknol. Elektro*, 20(2), 177.

[2] Annur, C. M. (2023, April 5). Jumlah Pengguna Instagram Indonesia Terbanyak ke-4 di Dunia. *Databoks*. Diakses dari https://databoks.katadata.co.id/datapublish/2023/05/04/jumlah-pengguna-instagram-indonesia-terbanyak-ke-4-di-dunia

[3] Anisah, N., Sartika, M., & Kurniawan, H. (2021). Penggunaan media sosial Instagram dalam meningkatkan literasi kesehatan pada mahasiswa. *Jurnal Peurawi: Media Kajian Komunikasi Islam*, 4(2), 94-112.

[4] Anggraini, N., Taufiqurokhman, T., Fauzi, D. M., & Watriningsih, W. (2020). Kebijakan Pelayanan Publik Pemerintah Daerah di Era Disrupsi. *Petanda: Jurnal Ilmu Komunikasi dan Humaniora*, *3*(1), 24-39.

[5] Chandradev, V., Suarjaya, I. M. A. D., & Bayupati, I. P. A. (2023). Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT. *Jurnal Buana Informatika*, 14(02), 107-116.

[6] Rumaisa, F., Puspitarani, Y., Rosita, A., Zakiah, A., & Violina, S. (2021). Penerapan Natural Language Processing (NLP) di bidang pendidikan. *Jurnal Inovasi Masyarakat*, 1(3), 232-235.

[7] Mudding, A. A. (2024). Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film. *Journal of System and Computer Engineering (JSCE)*, 5(1), 36-43.

[8] Fimoza, D. (2021). Analisis Sentimen Terhadap Film Indonesia dengan Pendekatan Bert (*Doctoral dissertation*, Universitas Sumatera Utara).

[9] Asri, Y., & Kuswardani, D. (2024). MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi. Uwais Inspirasi Indonesia.

[10] Tandijaya, J. H., Liliana, L., & Sugiarto, I. (2021). Klasifikasi dalam Pembuatan Portal Berita Online dengan Menggunakan Metode BERT. *Jurnal Infra*, *9*(2), 320-325.

[11] Sabita, H., Herwanto, R., Syafitri, Y., & Prasetyo, B. D. (2022). Pengembangan Aplikasi Akreditasi Program Studi Berbasis Framework Django. *Jurnal Informatika*, 22(1), 33-37

[12] Arianti, T., Fa'izi, A., Adam, S., & Wulandari, M. (2022). Perancangan Sistem Informasi Perpustakaan Menggunakan Diagram UML (Unified Modelling Language). *Jurnal Ilmiah Komputer Terapan dan Informasi*, *1*(1), 19-25.

[13] Abdurrazzaq, M. A., & Tjiong, E. L. (2022). Analisis Sentimen KUHP Baru Pada Data Twitter Menggunakan BERT. *Jurnal Komunikasi, Sains dan Teknologi*, 1(2), 83-88.