



## **Implementation of Deep Learning Method Using BERT Model in Career Choice Analysis of Gen Z**

**Silvia Ramadhani\*, Abdul Halim Hasugian**

Computer Science, State Islamic University of North Sumatra, Medan, Indonesia

Email: <sup>1</sup>\*slvrmdhni11@gmail.com, <sup>2</sup>abdulhalimhasugian@uinsu.ac.id

(\* : slvrmdhni11@gmail.com)

Submitted: 22/07/2025; Accepted: 30/07/2025; Published: 31/07/2025

**Abstract**—The development of digital technology has significantly influenced how individuals, particularly Generation Z (born between 1997 and 2012), make career decisions. Faced with an abundance of digital information, many individuals in this cohort experience difficulties in selecting career paths that align with their interests, abilities, and labor market demands. This study analyzes the career preferences of Generation Z using a deep learning approach through the Bidirectional Encoder Representations from Transformers (BERT) model, specifically the IndoBERT variant, which is pre-trained on Indonesian-language data. The research data were collected from textual responses to Google Form questionnaires, focusing on four digital career paths: Software Engineer, Content Creator, Digital Marketing, and Entrepreneur. From 601 data samples, sentiment analysis revealed that 57.85% of the responses were positive, while 42.15% were negative. Classification results indicated that Content Creator was the most preferred career, followed by Entrepreneur, Digital Marketer, and Software Engineer. Model evaluation showed a test accuracy of 51.24%, with better performance in categories that had larger data volumes. These findings demonstrate that IndoBERT is effective in capturing opinions and career tendencies from unstructured text and provides a scientific basis for educational institutions, industries, and policymakers to design more relevant career development strategies in the digital era.

**Keywords:** Generation Z; career choice; BERT; sentiment analysis; deep learning

### **1. INTRODUCTION**

The rapid development of digital technology has influenced various aspects of life, including how individuals especially Generation Z determine their career paths[1]. Born between 1997 and 2012, this generation has grown up in an environment saturated with the internet, social media, and mobile technology[2]. Despite having broad access to information, Gen Z often faces difficulties in choosing careers that align with their interests, abilities, and labor market demands. In today's digital era, they tend to be drawn to professions related to technology and creativity, such as Software Engineer, Content Creator, Digital Marketing, and Entrepreneur. These fields are considered attractive due to their flexibility, high income potential, and strong connection to the rapidly evolving digital ecosystem.

There are four main factors that influence the career decisions of Generation Z: work conditions and environment, job prospects and duration, intrinsic values and contribution to work, and work schedule flexibility. However, previous research remains limited to conventional quantitative survey approaches[3]. Generation Z is a productive age group that is beginning to enter the workforce and is expected to become the main driving force of future labor markets. They tend to prefer careers that offer flexibility, creativity, and strong connections to technology and social media. However, the abundance of digital information often leads to confusion in choosing careers that align with their interests and market demands[4]. Therefore, a systematic and data-driven analysis is needed to understand Gen Z's career preferences. The results of such analysis can serve as a foundation for educational institutions, companies, and policymakers to design more targeted career development strategies and to bridge the gap between current and future workforce needs.

This study aims to enhance that approach by applying the BERT deep learning model, which is capable of analyzing unstructured textual data in a deep and contextual manner, particularly from social media. This method is expected to provide a more comprehensive understanding of Generation Z's career preferences within the digital ecosystem.

Sentiment analysis is a method used to identify and categorize opinions or emotions expressed in digital text, aiming to determine the author's attitude toward a particular topic whether it is positive, negative, or neutral. This method is commonly applied to product reviews, news articles, or social media content to understand public opinion trends and consumer perceptions, which are then used as a basis for decision-making[5]. Sentiment analysis, also known as opinion mining, is a field of study in computing that aims to identify and represent opinions, sentiments, evaluations, emotions, subjectivity, attitudes, and viewpoints contained within a text. This method helps in understanding how an individual or a group responds to a particular topic through the language they use[6].

Machine learning is a branch of artificial intelligence that can be defined as a method that enables systems to learn from experience, adapt to new data, perform actions automatically, and make accurate and relevant decisions. This technology plays a role in supporting decision-making processes, where systems can learn from available data, recognize patterns, and automatically generate decisions without direct human intervention[7]. Machine learning is a branch of computer science that develops algorithms capable of recognizing patterns in data and learning automatically without being explicitly programmed. This technology allows systems to become more adaptive and efficient[8].

Deep learning is a branch of machine learning that relies on artificial neural networks with multiple layers (deep neural networks) to process and understand complex data. Each layer plays a role in extracting and interpreting information from the data at different levels[9]. Deep learning utilizes complex models for various applications such as



speech recognition, natural language processing, and image recognition, supported by large amounts of data and high computational power. Its main advantage is the ability to automatically learn from unstructured or unlabeled data, making it highly effective for solving complex tasks that are difficult to address using traditional learning methods. As a result, deep learning has become a crucial technology in fields such as artificial intelligence, big data analytics, and the development of advanced technologies[10].

BERT (Bidirectional Encoder Representations from Transformers) is an encoder model based on the Transformer architecture that uses attention mechanisms to understand the contextual relationships between words, even those that are far apart, without relying on recurrent networks. BERT generates contextualized representations for each word in a sentence and has been shown to significantly improve model performance on various complex and sequential NLP tasks[11]. Natural Language Processing (NLP) is a branch of artificial intelligence in which computers are designed to communicate with humans using natural language, such as Indonesian[12].

This study aims to analyze the sentiments and perceptions of Generation Z toward four major digital career paths: Software Engineer, Content Creator, Digital Marketing, and Entrepreneur by utilizing the BERT algorithm on questionnaire data collected via Google Forms. Through this approach, the research seeks to identify career preferences among Generation Z and the factors influencing their choices. The results of this analysis can serve as a valuable reference for educational institutions, companies, and career counselors in designing more targeted and relevant career development programs suited to the needs of the younger generation in the digital era.

Previous research utilized the BERT model to analyze sentiment in hotel reviews, primarily focusing on classifying customer opinions into categories such as positive and negative. In contrast, this study applies BERT to analyze career choices among Generation Z, involving more complex and diverse data related to individual perceptions, motivations, and preferences regarding careers. The analytical objective also differs; rather than simply performing classification, it aims to understand the context and patterns behind career decision-making. This highlights a clear research gap, as there is limited exploration of BERT's application in the field of career development, particularly among Indonesian youth. By addressing this gap, this study contributes both methodologically and practically: methodologically by extending the use of NLP-based sentiment analysis into sociological and psychological domains, and practically by providing insights that can inform educational institutions and policymakers in shaping future career guidance strategies. Thus, this study expands the application of BERT beyond commercial use to include generational studies and decision-making processes.

Previous research utilized BERT Uncased and IndoBERT for emotion classification in Indonesian text using a dataset of 2,515 entries with nine emotion labels. The results showed that BERT Uncased achieved 90% accuracy, while IndoBERT reached 81%, with the best performance observed at the 10th epoch. IndoBERT proved effective in identifying emotions, achieving up to 97.6% accuracy on opinion sentences[13]. Another study titled "Hoax News Detection through Website Comparison Using a Deep Learning Approach with the BERT Algorithm" compared the performance of three algorithms: BERT, SVM, and Random Forest in classifying hoax news. The results showed that BERT significantly outperformed the others, achieving both an accuracy and ROC-AUC score of 0.99. This advantage is attributed to BERT's ability to deeply process contextual information through its multi-layered architecture. Additionally, the text analysis revealed that hoax news is often accompanied by manipulated visuals, such as images or videos that do not match the actual content[14].

This research is important because Generation Z is poised to dominate the workforce, and their career preferences will directly shape future labor market trends. Moreover, by leveraging advanced NLP techniques, this study bridges the methodological gap between computational approaches and career development research in Indonesia.

## **2. RESEARCH METHODOLOGY**

### **2.1 Research Framework**

Based on this study, there are several research stages. The research was conducted using data from social media. The image below illustrates the systematic stages of the research process. It begins with the Planning phase, followed by Data Collection (gathering data from social media). Next, the Method Analysis stage is carried out to determine the appropriate approach, which is then followed by the Application of Method (implementation of the IndoBERT method). The Testing Method phase is conducted to evaluate the model using the processed data. The outcomes are presented in the Results stage, and finally, an Evaluation is performed to assess the overall performance of the model. This visualization provides a clear and structured overview of the research workflow implemented in this study.

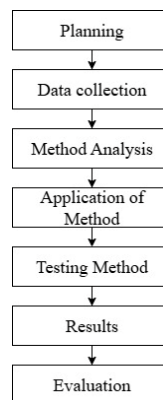


Figure 1. Research Framework

### 2.1.1 Planning

This study begins with data collection through open-ended questionnaires distributed online to Generation Z to explore their career preferences and the factors influencing their career choices in the digital era. The respondents' textual answers go through preprocessing stages such as data cleaning, normalization, stopword removal, and stemming. The processed data is then analyzed using the IndoBERT model to identify patterns or trends in Gen Z's career choices more deeply with the help of NLP technology[15].

### 2.1.2 Data Collection

Data collection in this study was carried out through the design and online distribution of open-ended questionnaires to Generation Z, aiming to explore their career choices and the influence of the digital environment. The free-text responses were gathered in digital format and initially reviewed to ensure completeness and relevance before being analyzed using the BERT model. This approach aims to directly capture Gen Z's career preferences in the digital era.

### 2.1.3 Method Analysis

The method analysis section focuses on processing text data from questionnaires using the BERT deep learning model, specifically IndoBERT, which is tailored for the Indonesian language[16]. The data is first processed through several preprocessing stages, including data cleaning, normalization, stopword removal, and stemming, followed by tokenization using the IndoBERT tokenizer. After preprocessing, the model is trained to identify patterns in Gen Z's career preferences and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The ultimate goal of this method is to produce a deeper understanding and a text-based predictive model to support a career recommendation system. These steps are illustrated in the flowchart below.

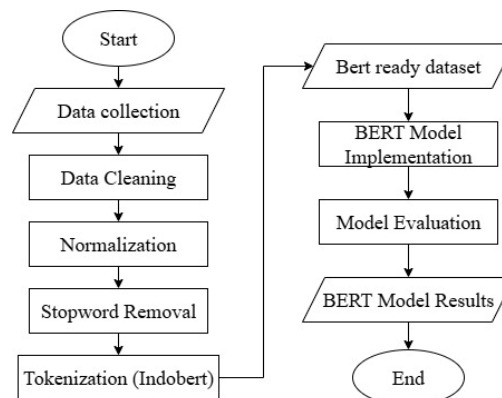


Figure 2. Flowchart of the BERT model system

### 2.1.4 Application of Method

This study employs IndoBERT to classify Gen Z's career preferences based on narrative texts from questionnaires. IndoBERT was chosen because it is specifically designed for the Indonesian language and trained on a large Indonesian language corpus, allowing it to understand the structure, context, and local vocabulary more accurately than the original BERT model, which is based on English[17]. After going through preprocessing stages such as data cleaning, normalization, stopword removal, and stemming, the data is trained and evaluated using metrics such as accuracy,



precision, recall, and F1-score. This method demonstrates the effectiveness of natural language processing (NLP) in deeply and contextually understanding and mapping Gen Z's career preferences.

### 2.1.5 Testing Method

The method testing section aims to evaluate the performance of the IndoBERT model after being trained with preprocessed questionnaire data. The dataset is divided into training and testing sets with a ratio of 80:20[18]. The model is trained to recognize language patterns in each career category, then tested with new data to measure its ability to generalize.

### 2.1.6 Results

The results section shows that the IndoBERT model successfully classified Gen Z's career preferences based on questionnaire responses into four main categories: software engineer, content creator, digital marketing, and entrepreneur. These results illustrate the distribution of career interests among respondents, such as the high level of interest in becoming a content creator.

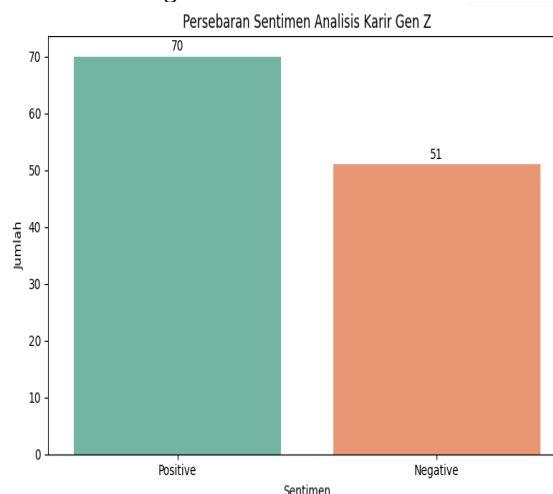
### 2.1.7 Evaluation

The evaluation stage aims to measure the performance of the IndoBERT model in classifying questionnaire text responses based on the correct career labels. The evaluation uses metrics such as accuracy, precision, recall, and F1-score to assess the model's correctness, sensitivity, and balanced performance in recognizing patterns in the test data. This evaluation ensures the model's reliability in understanding Gen Z's career preferences from narrative texts.

## 3. RESULT AND DISCUSSION

### 3.1 Data Collection

The data for this study was obtained through open-ended questionnaires distributed between May 27 and June 18, 2025, with a total of 602 respondents. The responses were converted into .csv format and processed using Python. Preprocessing steps included normalization (removing punctuation, converting all text to lowercase, and removing stopwords), followed by tokenization using the IndoBERT tokenizer. The labeled data was then split using the K-Fold Cross Validation technique to ensure fair model evaluation. Sentiment classification was performed using the IndoBERT model into two categories: positive and negative. Additionally, respondents' career choices were analyzed to identify career trends among Generation Z. The following is the sentiment label distribution of the entire dataset used.



**Figure 3.** Data distribution

### 3.2 Preprocessing Data

The data collected from respondents must go through preprocessing stages before it can be analyzed. The purpose of preprocessing is to transform the initially unstructured and noisy dataset into clean data that is ready for processing. The following is a flow that outlines the system stages during the data preprocessing process, which includes Data Cleaning, Normalization, Stopword Removal, and Tokenization (IndoBERT).

#### a. Data Cleaning





The cleaning stage aims to remove unnecessary elements from the data, such as numbers, symbols, emojis, links, and punctuation marks. This process involves detecting these characters in each line of the data and removing them to produce cleaner, more analyzable text. The table below shows a Sample of the Data Cleaning Process.

**Table 1.** Sample of Data Cleaning Process

| Initial Sentiment                                                                                                                                                                                                                                                                                                                                                                                                                                            | Data Cleaning Text                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sangat menarik, zaman sekarang semuanya serba digital, jadi apa apa menggunakan AI, AI termasuk digital yang bisa membantu kita mencari informasi yang lebih dalam tentang pekerjaan mungkin mengikuti zaman karena dengan menjadi content creator saya bisa menghasilkan uang dan menjadi seleb asekðŸ’f banyaknya syarat" loker pada jaman sekarang sehingga sulit untuk mendapatkan pekerjaan media sosial kerjanya fleksibel tidak di bawah tekanan bos  | sangat menarik, zaman sekarang semuanya serba digital, jadi apa apa menggunakan ai, ai termasuk digital yang bisa membantu kita mencari informasi yang lebih dalam tentang pekerjaan mungkin mengikuti zaman karena dengan menjadi content creator saya bisa menghasilkan uang dan menjadi seleb asek banyaknya syarat" loker pada jaman sekarang sehingga sulit untuk mendapatkan pekerjaan media sosial kerjanya fleksibel tidak di bawah tekanan bos      |
| Wah keren, sangat cerdas Karena fomo, gen z ini banyak kemauan nya jadi pasti mereka milih karir yg kerja santai tapi penghasilan nya banyak Karena ini serba digital, sudah dipastikan posisi ini cocok buat saya perkembangan dunia kerja digital sekarang sangat pesat, lebih fleksibel, dan berbasis teknologi gaji dan hari libur bisa kerja dr rumah Sangat baik, mempermudah pencari kerja dan penyedia Sesuai minat dan hobi Pekerjaan yg saya sukai | wah keren, sangat cerdas karena fomo, gen z ini banyak kemauan nya jadi pasti mereka milih karir yg kerja santai tapi penghasilan nya banyak karena ini serba digital, sudah dipastikan posisi ini cocok buat saya perkembangan dunia kerja digital sekarang sangat pesat, lebih fleksibel, dan berbasis teknologi gaji dan hari libur bisa kerja dr rumah sangat baik, mempermudah pencari kerja dan penyedia sesuai minat dan hobi pekerjaan yg saya sukai |

#### b. Normalization

After the data cleaning process, the next step is normalization, which involves removing symbols, punctuation marks, and standardizing words. The table below shows a Sample of Normalization Process

**Table 2.** Sample of Normalization Process

| Data Cleaning Text                                                                                                                                                                                                  | Normalized Text                                                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sangat menarik, zaman sekarang semuanya serba digital, jadi apa apa menggunakan ai, ai termasuk digital yang bisa membantu kita mencari informasi yang lebih dalam tentang pekerjaan mungkin mengikuti zaman karena | sangat menarik zaman sekarang semuanya serba digital jadi apa apa menggunakan ai ai termasuk digital yang bisa membantu kita mencari informasi yang lebih dalam tentang pekerjaan mungkin mengikuti zaman karena dengan |



|                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| dengan menjadi content creator saya bisa menghasilkan uang dan menjadi seleb aseks                                                                                                                                                        | menjadi content creator saya bisa menghasilkan uang dan menjadi seleb aseks                                                                                                                                                            |
| banyaknya syarat" loker pada jaman sekarang sehingga sulit untuk mendapatkan pekerjaan media sosial kerjanya fleksibel tidak di bawah tekanan bos                                                                                         | banyaknya syarat loker pada jaman sekarang sehingga sulit untuk mendapatkan pekerjaan media sosial kerjanya fleksibel tidak di bawah tekanan bos                                                                                       |
| wah keren, sangat cerdas karena fomo, gen z ini banyak kemauan nya jadi pasti mereka milih karir yg kerja santai tapi penghasilan nya banyak karena ini serba digital, sudah dipastikan posisi ini cocok buat saya                        | wah keren sangat cerdas karena fomo gen z ini banyak kemauan nya jadi pasti mereka milih karir yg kerja santai tapi penghasilan nya banyak karena ini serba digital sudah dipastikan posisi ini cocok buat saya                        |
| perkembangan dunia kerja digital sekarang sangat pesat, lebih fleksibel, dan berbasis teknologi gaji dan hari libur bisa kerja dr rumah sangat baik, mempermudah pencari kerja dan penyedia sesuai minat dan hobi pekerjaan yg saya sukai | perkembangan dunia kerja digital sekarang sangat pesat lebih fleksibel dan berbasis teknologi gaji dan hari libur bisa kerja dr rumah sangat baik mempermudah pencari kerja dan penyedia sesuai minat dan hobi pekerjaan yg saya sukai |

c. Stopword removal

Stopwords are common words such as "and," "or," "that," and similar terms that frequently appear in text but carry little meaning in sentiment analysis. Therefore, they are typically removed during data preprocessing to improve efficiency and focus on more meaningful words. The table below shows Sample of Stopword Removal Process

**Table 3.** Sample of Stopword Removal Process

| Normalized Text                                                                                                                                                                                                                                                                                     | Stopword Removal Text                                                                                                                                                                                                                                 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sangat menarik zaman sekarang semuanya serba digital jadi apa apa menggunakan ai ai termasuk digital yang bisa membantu kita mencari informasi yang lebih dalam tentang pekerjaan mungkin mengikuti zaman karena dengan menjadi content creator saya bisa menghasilkan uang dan menjadi seleb aseks | sangat menarik zaman sekarang semuanya serba digital jadi apa apa menggunakan ai ai termasuk digital bisa membantu mencari informasi lebih dalam pekerjaan mungkin mengikuti zaman menjadi content creator bisa menghasilkan uang menjadi seleb aseks |
| banyaknya syarat loker pada jaman sekarang sehingga sulit untuk mendapatkan pekerjaan media sosial kerjanya fleksibel tidak di bawah tekanan bos                                                                                                                                                    | banyaknya syarat loker jaman sekarang sulit mendapatkan pekerjaan media sosial kerjanya fleksibel tidak bawah tekanan bos                                                                                                                             |
| wah keren sangat cerdas karena fomo gen z ini banyak kemauan nya jadi pasti mereka milih karir yg kerja santai tapi penghasilan nya banyak karena ini serba digital sudah dipastikan posisi ini cocok buat saya                                                                                     | wah keren sangat cerdas fomo gen z banyak kemauan jadi pasti milih karir yg kerja santai tapi penghasilan banyak serba digital dipastikan posisi cocok buat                                                                                           |







|                 |                                           |                     |                 |
|-----------------|-------------------------------------------|---------------------|-----------------|
|                 | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0     |                 |
|                 | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0           |                 |
|                 | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       |                     |                 |
| persaingan      | [CLS] persaingan kerja makin ketat tapi   | 2 6283 494 2855     | 1 1 1 1 1 1 1 1 |
| kerja makin     | makin banyak peluang buat punya skill     | 5983 469 2855       | 1 1 1 1 1 1 1 1 |
| ketat tapi      | terutama skill bidang it kerja bisa       | 271 2882 968        | 1 1 1 1 1 1 1 1 |
| makin banyak    | dikerjakan mana aja bidang teknologi      | 1121 7963 1580      | 1 1 1 1 1 1 1 1 |
| peluang buat    | punya banyak peluang pilih software       | 7963 1254 1701      | 1 1 1 1 1 1 1 1 |
| punya skill     | engineer suka logika cod ##ing rasanya    | 494 166 6565        | 1 1 1 0 0 0 0 0 |
| terutama skill  | senang kalau bisa buat program bisa bantu | 1152 1465 1254      | 0 0 0 0 0 0 0 0 |
| bidang it kerja | orang [SEP] [PAD] [PAD] [PAD] [PAD] [PAD] | 1429 1121 271       | 0 0 0 0 0 0 0 0 |
| bisa            | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 2882 959 3389       | 0 0 0 0 0 0 0 0 |
| dikerjakan      | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 21058 1506 7858     | 0 0 0 0 0 0 0 0 |
| mana aja        | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 11114 55 3514       | 0 0 0 0 0 0 0 0 |
| bidang          | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 3000 599 166 968    | 0 0 0 0 0 0 0 0 |
| teknologi       | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 986 166 1621 232    | 0 0 0 0 0 0 0 0 |
| punya banyak    | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 3 0 0 0 0 0 0 0 0 0 | 0 0 0 0 0 0 0 0 |
| peluang pilih   | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0 0 0 | 0 0 0 0 0 0 0 0 |
| software        | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0 0 0 | 0 0 0 0 0 0 0 0 |
| engineer suka   | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0 0 0 |                 |
| logika coding   | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0 0 0 |                 |
| rasanya         | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0 0 0 |                 |
| senang kalau    | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0 0 0 |                 |
| bisa buat       | [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]       | 0 0 0 0 0 0 0 0 0 0 |                 |
| program bisa    | [PAD] [PAD] [PAD]                         | 0 0 0 0 0 0         |                 |
| bantu orang     |                                           |                     |                 |

### 3.3 Implementation of the BERT Model

In this study, the IndoBERT model (indobenchmark/indobert-base-p1) was used for sentiment classification of Indonesian-language text. After preprocessing steps such as normalization, stopword removal, and tokenization, the data were converted into token IDs and attention masks to be fed into the BERT for Sequence Classification model. The model was trained for three epochs using the AdamW optimizer with a learning rate of 2e-5. Evaluation was performed using K-Fold Cross Validation with metrics including accuracy, precision, recall, and F1-score.

### 3.4 Model Evaluation

The horizontal bar chart visualizes the career preferences of Gen Z respondents. The graph displays the number of respondents on the horizontal axis and the types of careers on the vertical axis. The results show that Content Creator is the most preferred career choice (183 respondents), followed by Entrepreneur (157), Digital Marketing (146), and Software Engineer as the least chosen (115). These findings indicate that Gen Z tends to be more interested in careers that support creativity and self-expression rather than technical professions.



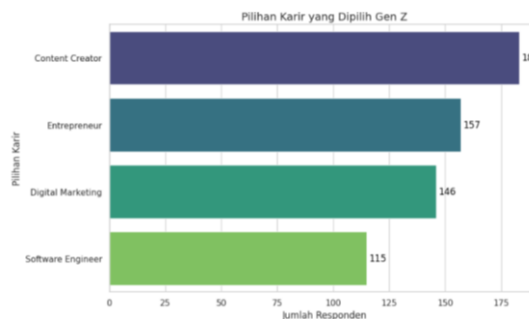


Figure 5. Career Choices Preferred by Gen Z

a. Confusion matrix

The confusion matrix shows the number of correct and incorrect predictions for both positive and negative classes, allowing us to understand how well the model can distinguish between the two. This evaluation result provides a comprehensive overview of the model's ability to generalize to new data.

The following figure presents the results of the confusion matrix, which illustrates the number of correct and incorrect predictions made by the model for each class, both positive and negative. This visualization helps evaluate how accurately the model classifies the data.

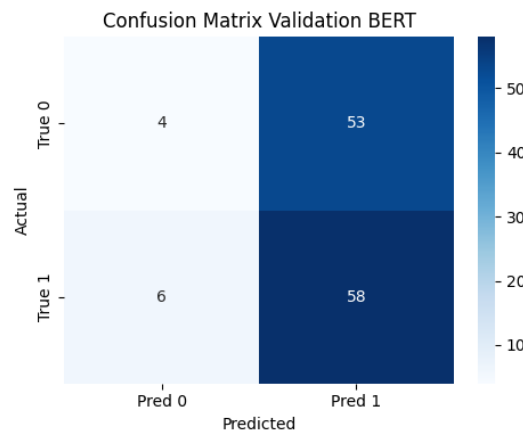


Figure 6. Confusion Matrix

b. Accuracy

The accuracy score measures the proportion of correct predictions overall. It is calculated by dividing the number of correct predictions by the total number of predictions. A high accuracy score indicates that the model performs well overall in correctly classifying sentiments.

The image below presents the accuracy results achieved by the IndoBERT model in classifying the data. This visualization provides an overview of how effectively the model can recognize patterns and make accurate predictions based on the data that has undergone the training process.

```

=== Classification Report ===

```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.4000    | 0.0702 | 0.1194   | 57      |
| 1            | 0.5225    | 0.9062 | 0.6629   | 64      |
| accuracy     |           |        | 0.5124   | 121     |
| macro avg    | 0.4613    | 0.4882 | 0.3911   | 121     |
| weighted avg | 0.4648    | 0.5124 | 0.4068   | 121     |

Figure 7. Accuracy

$$Accuracy = \frac{58 + 4}{58 + 4 + 53 + 6} = \frac{62}{121} \times 100\% = 51,24\%$$

c. Precision

1. For the negative class



$$Precision = \frac{4}{4 + 6} = \frac{4}{10} \times 100\% = 40\%$$

2. For the Positive class

$$Precision = \frac{58}{58 + 53} = \frac{58}{111} \times 100\% = 52,25\%$$

- d. Recall

1. For the negative class

$$Recall = \frac{4}{4 + 53} = \frac{4}{57} \times 100\% = 7,02\%$$

2. For the Positive class

$$Recall = \frac{58}{58 + 6} = \frac{58}{64} \times 100\% = 90,62\%$$

- e. F1 - Score

1. For the negative class

$$F1 - Score = 2 \times \frac{0,4 \times 0,0702}{0,4 + 0,0702} = 2 \times \frac{0,0281}{0,4702} \times 100\% = 11,97\%$$

2. For the Positive class

$$F1 - Score = 2 \times \frac{0,5225 \times 0,9062}{0,5225 + 0,9062} = 2 \times \frac{0,4739}{1,4287} \times 100\% = 66,34\%$$

### 3.5 Implementation

This study utilizes the BERT (Bidirectional Encoder Representations from Transformers) model to analyze textual data from questionnaires regarding Generation Z's career preferences in the digital era. After undergoing normalization, text cleaning, and stopword removal, the data is tokenized using the IndoBERT tokenizer and then split into training and validation sets. The model is trained over several epochs while monitoring training loss and validation accuracy. Evaluation is conducted using accuracy metrics, a confusion matrix, and performance visualizations. The results are used to identify sentiment patterns and career trends among Gen Z, providing deeper insights into the factors influencing their career decisions in the digital age.

## 4. CONCLUSION

Based on the final results of the problem-solving process in this study, several conclusions can be drawn as follows. First, from a total of 601 questionnaire responses regarding Gen Z's career preferences, it was found that 57.85% expressed positive sentiment, while 42.15% reflected negative sentiment. Second, the implementation of the BERT model to classify Gen Z's career choices successfully mapped respondents into four predefined categories: content creator, entrepreneur, digital marketing, and software engineering. The prediction results showed that the majority of respondents preferred careers as content creators, followed by entrepreneurs, then digital marketers, with software engineering being the least chosen. These findings indicate a strong interest among Generation Z in creative, independent, and digital fields.

Third, the evaluation of the IndoBERT model during training, visualized through graphs and epoch-based metrics, showed a decreasing trend in training loss from epoch 1 to 3 indicating improved learning on the training data. However, validation accuracy remained stagnant between 47% and 49%, suggesting that the model struggled to generalize to the validation set despite a sharp drop in training loss from 0.32 to 0.11. Lastly, the BERT model used in this study achieved a fairly good accuracy rate of 51.24% in classifying career choices based on opinion texts. The evaluation results showed varied precision, recall, and F1-scores across categories, with generally satisfactory performance. The model performed well on categories with larger datasets, while its performance was relatively lower for categories with fewer data samples.

## REFERENCES

- [1] S. D. Saraswati, N. Alifia, and G. B. Setyawan, "YUME : Journal of Management Pengaruh Persepsi Generasi Z Terhadap Pilihan Karir Pada Mahasiswa Universitas Karya Husada," vol. 8, no. 2, pp. 1691–1699, 2025.
- [2] M. Alimuddin and P. Poddala, "Prospek Digital Marketing Untuk Generasi Muda Dalam Berwirausaha," *J. Career Dev.*, vol. 1, no. 1, pp. 54–70, 2023.
- [3] Eka Nurkhasannah and Shafa Anggita Putri, "Pengaruh Persepsi Generasi Z Terhadap Pilihan Karir pada Mahasiswa Program Studi Manajemen Universitas Pamulang," *J. Manag. Creat. Bus.*, vol. 3, no. 1, pp. 251–260, 2025, doi: 10.30640/jmcbus.v3i1.3703.





- [4] D. Mawabagja, A. Kurniawan, R. Claudia, S. Islami, and R. Fadjri, "Strategi Adaptasi Manajemen Puncak dalam Menghadapi Transformasi Tenaga Kerja Generasi Z Pendahuluan Transformasi tenaga kerja yang terjadi saat ini dipengaruhi oleh masuknya generasi Z," vol. 03, no. 03, pp. 132–142, 2025.
- [5] Z. W. Farida and N. Rochmawati, "Analisis Sentimen Masyarakat terhadap Fenomena Childfree Menggunakan Metode Long Short Term Memory dan Bidirectional Encoder Representations from Transformers di Twitter," *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 369–376, 2024, doi: 10.26740/jinacs.v5n03.p369-376.
- [6] R. M. R. W. P. K. Atmaja and W. Yustanti, "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *Jeisbi*, vol. 02, no. 03, p. 2021, 2021.
- [7] S. Prasetyo and T. Dewayanto, "Penerapan Machine Learning, Deep Learning, Dan Data Mining Dalam Deteksi Kecurangan Laporan Keuangan-a Systematic Literature Review," *Diponegoro J. Account.*, vol. 13, no. 3, pp. 1–12, 2024, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/accounting>
- [8] R. Kurniawan, P. B. Wintoro, Y. Mulyani, and M. Komarudin, "Implementasi Arsitektur Xception Pada Model Machine Learning Klasifikasi Sampah Anorganik," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 2, pp. 233–236, 2023, doi: 10.23960/jitet.v11i2.3034.
- [9] I. C. Azhari and T. Haryanto, "Modeling Of Hyperparameter Tuned RNN-LSTM and Deep Learning For Garlic Price Forecasting In Indonesia," *J. Informatics Telecommun. Eng.*, vol. 7, no. 2, pp. 502–513, 2024, doi: 10.31289/jite.v7i2.10714.
- [10] P. Hafidzah, S. Maryani, B. Y. Ihsani, E. Erwin, and A. K. Niswariyana, "Penerapan Deep Learning dalam Menganalisis Sentimen di Media Sosial," *Semin. Nas. Paedagoria*, vol. 4, pp. 328–339, 2024.
- [11] Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *J. Buana Inform.*, vol. 14, no. 02, pp. 107–116, 2023, doi: 10.24002/jbi.v14i02.7244.
- [12] M. Furqan, S. Sriani, and M. N. Shidqi, "Chatbot Telegram Menggunakan Natural Language Processing," *Walisongo J. Inf. Technol.*, vol. 5, no. 1, pp. 15–26, 2023, doi: 10.21580/wjit.2023.5.1.14793.
- [13] W. Alan, "Klasifikasi Emosi pada Teks Menggunakan Deep Learning," *Tugas Akhir di Univ. Islam Indones.*, vol. 6, no. 1, 2021.
- [14] A. Ripa'i, F. Santoso, and F. Lazim, "Deteksi Berita Hoax dengan Perbandingan Website Menggunakan Pendekatan Deep Learning Algoritma BERT," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1749–1758, 2024, doi: 10.33379/gtech.v8i3.4541.
- [15] Y. A. Singgalen, "IndoBERT-Based Sentiment Analysis for Understanding Hotel Guests ' IndoBERT-Based Sentiment Analysis for Understanding Hotel Guests ' Preferences," no. March, 2025, doi: 10.47065/josyc.v6i2.6864.
- [16] E. Mundika, G. H. Martono, and R. Rismayati, "Sentiment Analysis of Service and Facility Satisfaction at Computer Lab of Universitas Bumigora Using Indobert," vol. 5, no. 2, pp. 475–484, 2025, doi: 10.30811/jaise.v5i2.6798.
- [17] G. T. Fadilah, L. Muflikhah, and R. S. Perdana, "Analisis Sentimen Produk Hijab Pada E-Commerce Tokopedia Menggunakan Algoritma Support Vector Machine Dan Indobert Embedding," vol. 9, no. 2, pp. 2548–964, 2025, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [18] A. S. Rizky and E. Y. Hidayat, "Emotion Classification in Indonesian Text Using IndoBERT," *Comput. Eng. Appl.*, vol. 14, no. 1, pp. 2252–4274, 2025.