Journal of Applied Informatics and Computing (JAIC)
Vol.8, No.1, July 2024, pp. 7~12
e-ISSN: 2548-6861

IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election

Dwi Ismiyana Putri [1]*, Ari Nurul Alfian [2], Mardi Yudhi Putra [3]*, Putro Dwi Mulyo [4]*****
* Rekayasa Perangkat Lunak, Universitas Bina Insani
** Manajemen Informatika, Universitas Bina Insani
*** Teknik Informatika, Universitas Bina Insani
dwiismiyana@binainsani.ac.id [1], arin@binainsani.ac.id [2], mardi@binainsani.ac.id [3], putrodwi31@gmail.com [4]

Article Info

Article history:

Received 2024-04-02
Revised 2024-04-23
Accepted 2024-05-07

Keyword:

Election,
IndoBERT,
Sentiment Analysis,
Twitter.

ABSTRACT

Elections are one of the key moments in a country's democracy. Indonesian elections have a significant impact on regional and global politics. Twitter being one of the popular social media platforms becomes a powerful tool for political campaigns. This makes it an ideal source to analyze public opinion during the 2024 general election, particularly the upcoming Presidential Election (Pilpres). IndoBERT is the model chosen to analyse the sentiment from the dataset in this study using a zero-shot learning approach. Based on the evaluation results, the accuracy value of the 2024 presidential election classification is 0.60 (60%), tends to predict with a good value in the positive label of 0.74 (74%) for F1-Score. This model is considered quite good at predicting negative labels but the results are not too optimal with a value of 0.49 (49%). Confusion Matrix in this IndoBERT model is more likely to label tweets with positive things, by detecting negative labels quite well.

I. INTRODUCTION

The development of technology today continues to grow without limits. What we want to do can happen with just one or two clicks using the internet. This also applies when we communicate or share something on a platform on the internet, social media. Twitter is one of the most popular social media in the world. According to [1], the advertising company pointed out that marketers could reach 372.9 million Twitter users as early as April 2023. Meanwhile, Indonesia itself shows that Twitter has 24 million active users in 2023, which is equivalent to 8.7% of the total population today.

Indonesia, the third largest democracy in the world, will hold general elections in 2024 [2]. Elections are one of the important moments in a country's democratic life [3]. As one of the largest countries in the world [4], Indonesian elections have a significant impact on regional and global politics. Twitter is one of the most popular social media platforms and is often used for political discussion and communication [5].

This makes it an ideal source for analysing public opinion during the 2024 general election, particularly the upcoming Presidential Election. Twitter has become a powerful tool for political campaigns [6] to reach voters and a tool for organising protests, for citizens to express their opinions and engage in political discourse. By analysing the sentiment of tweets related to the presidential election, political analysts can gain insight into public opinion and attitudes towards candidates and issues. This information can later be used to inform political campaigns, media coverage, and policy decisions. Therefore, Twitter has become an important platform for understanding public sentiment [7] during the election season in Indonesia. Sentiment analysis is a method used to identify and understand the positive, negative, or neutral sentiment [8] behind analysed text.

The 2024 elections will simultaneously determine the composition of parliament, the president, and various local leaders [9] who will play an important role in policy-making and the direction of the country for years to come. With a large electorate, cultural diversity and complex issues, the 2024 elections in Indonesia represent a major challenge in organising, facilitating and monitoring a fair and transparent democratisation process [10]. The political dynamics, citizen participation, and debates around key issues in this election will play a crucial role in shaping Indonesia's future. Therefore, a deep understanding of voter preferences,

changing opinions and political dynamics [11] that evolve along the journey towards the 2024 General Election is of great relevance.

One of the newly developed methods that can be used to identify and understand the political sentiment of the 2024 General Election is IndoBERT [12]. IndoBERT is a pre-trained BERT for Indonesian that uses a large Indonesian language corpus of four billion words in the pre-training [13]. The BERT (Bidirectional Encoder Representations from Transformers) method is one of the effective techniques in performing sentiment analysis [14] on text data. BERT is a language model that is trained unsupervised using millions of texts from the internet, so it has a better understanding of the context and meaning of words [15].

In the context of the 2024 General Election in Indonesia, sentiment analysis using Twitter data with the BERT method can provide valuable insights into people's views on candidates, political parties, political issues, and the election process itself. By understanding people's sentiments, relevant parties, such as candidates, political parties, and electoral institutions, can take better actions to fulfil people's expectations and needs in the 2024 Election process. In addition, sentiment analysis can also help the general public to understand the various perspectives that exist and make more informed decisions in choosing leaders and political parties that match their values and aspirations.

II. METHOD

The method in this study uses data collection as follows.

A. Data Collection Methods

Literature Study

In the literature study method, researchers collect data by reading and understanding the opinions of experts related to the problem being examined. Sources taken from the literature method are books, papers, national and international journals, and online media.

Twitter Data Collection

This method involves using the Twitter Streaming API to collect tweets in real-time that include keywords related to the 2024 presidential election. Monitoring Twitter for tweets containing keywords of presidential candidate names on 13 and 14 February 2024. Using web scraping techniques to retrieve tweets from relevant Twitter accounts, such as those of candidates, political parties, or public figures related to the 2024 presidential election. This makes it possible to focus on tweets from significant sources.

B. Research Stages

The flow of methods in this research can be seen in Figure 1. This research contains steps including data collection, preprocessing, data labelling, modelling, testing, and evaluation.

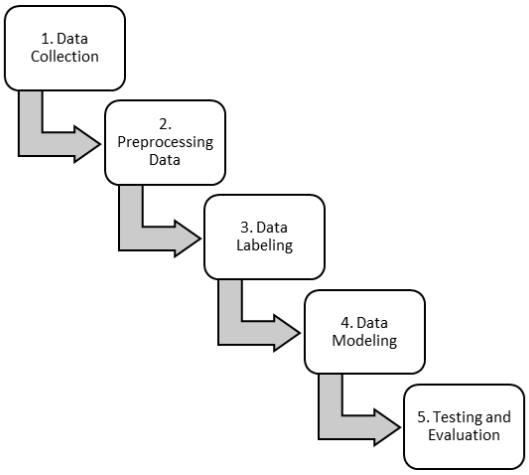

Figure 1. Research Stages

The steps in the study are outlined as follows:

1) *Data Collection.* Data collection or data collection is obtained from Crawling Data on Twitter Social Media, which starts by first obtaining Consumer Key, Authentication Token and Auth_Token by entering the page https://developer.twitter.com/en. Then the tokens that have been obtained are imported using the Python library using Google Colab. Obtained as much as 9,136 raw data based on the keyword name of the presidential candidate.

2) Data Preprocessing. After the data is successfully collected, the data is then processed first by cleaning the text which consists of removing irrelevant special characters, punctuation marks, or symbols that do not contribute to the sentiment analysis of this presidential election.

3) *Data Labelling.* The data is then reprocessed by labelling (positive, neutral, negative) and filtered based on 3 (three) criteria, namely the name of each presidential candidate. Furthermore, the data that has been filtered according to the criteria is randomly selected again. The data selected was 300 data consisting of 100 data for each category. This was done to speed up processing time and to obtain a more representative sample.

4) Data Modelling. Data modelling is an important stage in the development of artificial intelligence models, especially in the context of natural language processing (NLP). In this research, IndoBERT, the Indonesian variant of the pre-trained BERT model, is used to perform sentiment analysis on Indonesian text datasets.

5) *Test and Evaluation*. After conducting the training process on the indoBERT model using the processed data, the next step is the testing and evaluation stage. At this stage, we will test the trained model using data that has never been seen before, and measure its performance in predicting text sentiment using Google Colab as an interactive notebook to run Python code.

III. RESULTS AND DISCUSSION

This research covers the entire sentiment analysis process of data from Twitter, starting from data collection using Data Crawling to the model evaluation stage, the following are the results of each stage carried out.

A. Data Crawling

Data Crawling is a technique in collecting data from various sources on the internet. In this research, data crawling targets twitter social media as a platform that is considered good enough in collecting datasets in the form of text related to public opinion about the 2024 presidential election. After collecting data, 9,136 raw data were obtained using the keyword based on name of the presidential candidate. The following is an example of the results of data collected at an early stage or raw data that has not been cleaned, found in Table 1.

TABLE I
RAW DATA CRAWLING RESULTS

Gara garaDIRTY VOTE" NEGRI KONOHA RAME. Perkiraan kecurangan pemilu 2024 terbongkar dan menurut saya itu sesuai DATA DAN FAKTA dilapangan. #Pemilu2024 #PemiluDamai2024 #AniesMuhaimin2024 #PrabowoGibran2024 #GanjarMahfud #bawaslu #konoha #jokowi #mahkamahkonstitusi"
Assalamualaikum, salam pagi sehat utk smua Ingat besok tggl 14 Februari dbilik suara pda kertas suara pilpres warna abu2 **Coblos No. 1** @aniesbaswedan @cakimiNOW " , jngan lupa juga tlong djaga situng suara "AMIN" dtps2 msing !!! 💪👍👇 #AMINAjaDulu #AniesMuhaimin2024 https://t.co/D35elrmz5y"
@data_retra @DirtyVote @aniesbaswedan @cakimiNOW Seorang Presiden bagi² bansos dipinggir jalan tak ubahnya seorang RT. 🤣🤣🤣🤣presiden kurang kerjaan #AniesMuhaimin2024 Harus ada perubahan
Gak perlu ragu lagi sih buat milih Pak @ganjarpranowo sebagai presiden selanjutnya. Beliau paling paham masyarakatnya dan tulus banget! Yukk ikut menangkan Pak Ganjar! #GanjarMahfud2024 #ganjarpranowo #mahfudMD #SatSetLebihBaik #pastilebihbaik #capres2024 #reels #Pilpres2024 https://t.co/2V4DYx1D07
Ganjar Pranowo Ngaku Tak Risau Baca Survei Pilpres 2024, Di Langit Sudah Digariskan Pemenangnya https://t.co/8widfgcRgB lewat @tribunnews #GanjarMahfud #ganjarpranowo #pilpres2024 #capres2024
Paslon nomor urut 2, Prabowo Subianto-Gibran Rakabuming Raka, diprediksi memenangkan Pilpres 2024. https://t.co/6itHUNTnOQ #PrabowoGibran #capres2024 #pilpres2024 https://t.co/iQBM84OJWv
Sampurasun Kota Bandung. Eh Tunggu Ada yang Naik Motor Sambil Menyapa Warga Jawa Barat, Siapa Tuh? 🤔 #GanjarPranowo #LanjutGan #Capres2024 #Gandara #GANjarDAnRAkyat #GandaraIndonesia #GenZdukungGanjar Towel Sejarah Baru Joel Kojo Australia #STYSTAY Malaysia Round of 16 Korsel https://t.co/qm44woREN5
Prabowo gibran untuk indonesia emas□ #Capres2024 #prabowo https://t.co/vFah035WF5

B. Preprocessing Data

The next process is pre-processing. After the data is collected, it turns out that there are still some symbols that are less meaningful such as (@, ^, $, %, &, !). This makes the sentence untidy and difficult to read. Therefore, the results of the crawling must go through a data cleaning process. Then, in an opinion sentence there are words that are less meaningful and non-standard language, therefore the word must be removed and converted into a more standard form of language. To implement it, researchers need to adjust the lexicon into Indonesian. This stage also changes the original twitter account name previously found in the raw data to the default name as "@user" and removes the link contained in the text which can be seen in Figure 2.

```
import re

def preprocess(text):
    new_text = []
    #ganti semua mention menjadi @user
    #ganti semua link menjadi http
    for t in text.split(" "):
        t = '@user' if t.startswith('@') and len(t) > 1 else t
        t = 'http' if t.startswith('http') else t
        new_text.append(t)
    new_string =  " ".join(new_text)

    #hapus karakter non ascii
    new_string  = re.sub(r'[^\x00-\x7F]+',' ', new_string)
    return new_string
```

Figure 2. Preprocessing Function

The following are samples of dataset results that have passed the preprocessing stage can be seen in Table 2.

TABLE II
PREPROCESSING RESULTS DATASET

Kalau no 1 menang, gue ga nonton konseran kpop selama anies menjabat . Bismillah menang
@user pak anies saya sebel bgt sm bapak, jd td saya tusuk jidatnya
@user Tadi ke bilik langsung buka surat suara abu-abu, baca bismillah sama shalawat, tak coblos senyumnya Pak Anies, pengin banget punya presiden pak @user #sayacoblosno1 http
@user Pilih siapa sum, anies, prabowo, ganjar, atau sumarna?
@user LO MAU BELAIN PRABOWO SAMPAI BERBUSA SILAHKAN! TP GAK BAKALAN HAPUS KLO DIA PENJAHAT HAM !! Ttd NdrewsTjan !! Trims ! http
@user @user @user yaampuun wong ini
mostly orang yang komen soal dirty vote, mk, anies gimana, prabowo gimana, ganjar gimana, semua punya kepentingan. dan cara pandang mereka didasari oleh apa yang mereka ingin lihat dan keinginan mereka didasari oleh apa yang mereka telah rasakan.
Komentar Menohok Dokter Tirta Soal TKN Prabowo-Gibran yang Kebakaran Jenggot Gara-gara Film Dirty Vote, katanya... - http
@user PSI dulu Ganyang Prabowo sekarang jd Penyembah nya mana ada orang beradab ga punya MALU begini ampas apa yg bisa dijual dari parte recehan gini kl ga dilirik Jokowers jd gelandangan udah ini mah..
@user @user @user Kkir ini tidak solit , gimana mau solit orang sudah deklarasi prabowo persiden aja cak imin masih manuver ke ganjar bertemu dengan ganjar yg jelas" capres dari koalisi lain"
@user Bapak gak tau kalo pak Ganjar dapat penghargaan lagi?
Ganjar Pranowo dan Siti Atikoh Ganjar menyalami warga yang berekerumun dan menunggu kehadiran mereka, saat mendampingi Presiden Jokowi meninjau Pasar Grogolan, Pekalongan Warga Berteriak Pak Ganjar ay lop yu, Pak Ganjar Presiden http http http

C. Data Labelling

The 9,136 data that has been collected is then separated according to the emotion category, namely positive, neutral, and negative. The data labelling process carried out by researchers uses a manual method, namely by reading one by one and analysing the dataset to find out the emotions contained in each opinion sentence. Table 3 shows the number of results from data labelling.

TABLE III
NUMBER OF LABELLED DATA

Label	Data
Positif	3.572
Netral	1.893
Negatif	4.149

The process of determining emotion using the manual method results in data labels as in Table 3. In determining this emotion, the research was assisted by the team to minimise errors in determining the emotion in the text. The results of data labelling can be seen in Table 4 below.

TABLE IV
EXAMPLE OF DATA LABELLING RESULT

No.	Data	Label
1.	@user Klo Ganjar Mahfud bakalan bnrmembuat Indonesia semakin berjaya	Positif
2.	AHY ajak warga Semarang menangkan Prabowo-Gibran agar Indonesia semakin maju. dzp PDemokrat AgusYudhoyono Demokrat Bersama Prabowo http	Netral
3.	@user Si emak hayangna anies tp di tunjuken nu lain pan goblog Tlng di tindak lanjut dong @user	Negatif

To identify emotions in the text, the researcher used the following manual method:

1) In the Positive Label, in the sentence, "@user Klo Ganjar Mahfud bakalan bnrmembuat Indonesia semakin berjaya dimana dapat dilihat pada kalimat tersebut terdapat kata "membuat Indonesia semakin berjaya", which is a positive word for the progress of the Indonesian nation.
2) In the Neutral Label, in the sentence "AHY ajak warga Semarang menangkan Prabowo-Gibran agar Indonesia semakin maju. dzp PDemokrat AgusYudhoyono Demokrat Bersama Prabowo", this sentence contains information for a person or group. Researchers considered the sentence to be neutral.
3) On Negative Labels, in the sentence "@user Si emak hayangna anies tp di tunjuken nu lain pan goblog Tlng di tindak lanjut dong @user", there is a negative emotion which is clarified by the word "goblog" in the sentence expressed.

After the data was identified and labelled, 300 data were selected for processing. Consisting of 100 data each for each keyword, namely the name of the presidential candidate. This was done to speed up processing time and to obtain a more representative sample.

D. Data Modelling Using IndoBERT

The next stage after pre-processing the data and labelling the data is to determine the model to be used. The model used in this research is IndoBERT using the Zero-Shot Learning approach. With the zero-shot learning approach, the model is given a general understanding of the various sentiment classes that may be related to the presidential election, such as positive, negative, or neutral, without requiring specific examples of each presidential candidate. As such, the model can analyse sentiments related to the presidential election more broadly and can identify sentiments related to new presidential candidates that have not been seen before. This provides great flexibility in sentiment analysis, especially when there is a change in context or new information emerges.

E. Test and Evaluation

The final stage is to test and evaluate the results of the test dataset. In the context of zero-shot learning, the model is given a general understanding of various classes or concepts without the need to see specific training examples for each class. Even though the model does not see training examples for each class, accuracy evaluation can be done using test data or validation data that has been labelled. This test data can be used to test the model's ability to recognise classes that have been taught zero-shot. By evaluating accuracy using test data, researchers can assess how well the model is able to predict the classes that have been given to the model zero-shot. Before testing, the process of installing the Library Transformer that will be used is found in Figure 3 below.

```
Requirement already satisfied: transformers in
/usr/local/lib/python3.10/dist-packages (4.38.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from transformers) (3.13.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.19.3 in
/usr/local/lib/python3.10/dist-packages (from transformers) (0.20.3)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.10/dist-packages (from transformers) (1.25.2)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from transformers) (24.0)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.10/dist-packages (from transformers) (6.0.1)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.10/dist-packages (from transformers) (2023.12.25)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from transformers) (2.31.0)
Requirement already satisfied: tokenizers<0.19,>=0.14 in
/usr/local/lib/python3.10/dist-packages (from transformers) (0.15.2)
Requirement already satisfied: safetensors>=0.4.1 in
/usr/local/lib/python3.10/dist-packages (from transformers) (0.4.2)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.10/dist-packages (from transformers) (4.66.2)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub<1.0,>=0.19.3->transformers) (2023.6.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub<1.0,>=0.19.3->transformers) (4.10.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers)
(3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers)
(2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers)
(2024.2.2)
```

Figure 3. Library Transformer

Next, use the Pipline transformer with Custom Model. This method is used if you want to compare the most accurate model. Furthermore, Figure 4 shows the model that will be tested is the IndoBERT Model.

```
model = [
    "taufiqdp/indonesian-sentiment",
    "akahana/indonesia-sentiment-roberta",
    "cardiffnlp/xlm-twitter-politics-sentiment",
    "cardiffnlp/xlm-roberta-base-sentiment-multilingual",
    "candra/indobertweet-sentiment2",
    "intanm/indonesian_financial_sentiment_analysis_10",
    "lxyuan/distilbert-base-multilingual-cased-sentiments-student",
    "indolem/indobert-base-uncased",
    "cahya/bert-base-indonesian-1.5G"
]
```

Figure 4. Pipline Transformer with Custom Model

```
from transformers import pipeline
from transformers import AutoTokenizer, AutoModelForSequenceClassification


pretrained= model[4]
model = AutoModelForSequenceClassification.from_pretrained(pretrained)
tokenizer = AutoTokenizer.from_pretrained(pretrained)

sentiment_analysis_custom = pipeline("sentiment-analysis", model=model,
tokenizer=tokenizer)
```

Figure 5. Pipline IndoBERT Model

Furthermore, sentence inference, whole data inference, and label standardisation are performed. The following is the Visualisation of Candidate Data Resulting from Labelling and Prediction in Figure 6.

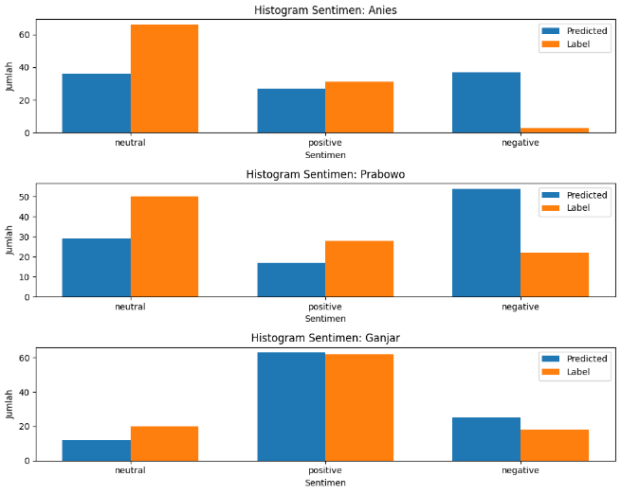

Figure 6. Visualisation of Candidate Data Resulting from Labelling and Prediction

F. *Evaluation Results*

Evaluation was carried out on the IndoBERT model. Testing is done so that researchers know whether the model is running properly and accordingly.

From the evaluation of the model, the measure of accuracy results for the IndoBERT Model is 0.60 or 60%, tends to predict with a good value in positive labels 0.74 or 74% for the balance measure (f1-score). This model can predict negative labels but the results are not very optimal with a value of 0.49 or 49%. Table V below shows the classification report.

TABLE V
CLASSIFICATION REPORT

	Precision	Recall	F1-Score	Support
Negative	0.34	0.91	0.49	43
Neutral	0.74	0.42	0.54	136
Positive	0.79	0.69	0.74	121
Accuracy			0.60	300
Macro Avg	0.62	0.67	0.59	300
Weigted Avg	0.70	0.60	0.61	300

Figure 7 shows a 3x3 matrix related to the test results on validation data. The total data in the validation data is 300 data. The True Positive (TP) value on neutral labels is 39 data, False Positive (FP) is 58 + 19 = 77 data. Then True Negative (TN) is 57 + 84 = 141 data and False Negative (FN) is 2 + 18 + 2 +21 = 43 data. In the Confusion Matrix, the IndoBERT model is more likely to label tweets with positive things. However, to detect negative labels is quite good.

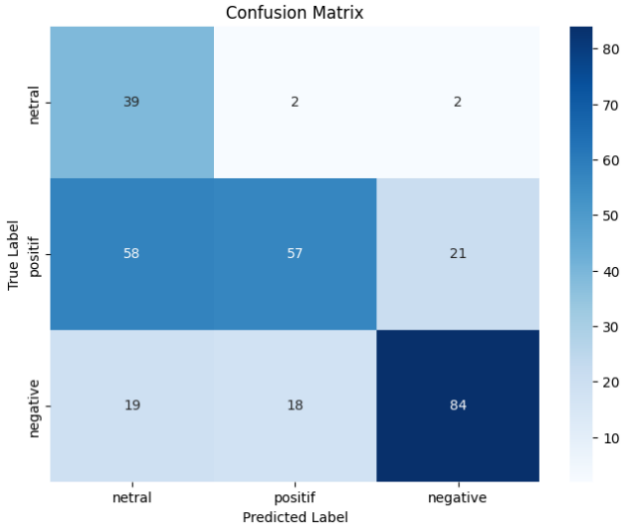

Figure 7. Confusion Matrix Visualisation

G. Word Cloud

The visual representation of the text between the presidential candidates is represented with a word cloud, where the frequency of occurrence of words in the text is depicted with different sizes and colours. Words that appear more frequently in the text will be displayed with a larger size and sometimes with a more striking colour, while words that appear less frequently will be displayed with a smaller size. Figure 8 shows the Word Cloud of the Three Presidential Candidates.

Figure 8. Word Cloud of the Three Presidential Candidates

IV. CONCLUSIONS

This research describes the sentiment analysis process of data from Twitter Social Media related to the Presidential Election using the IndoBERT model. The process starts with data collection through Data Crawling, followed by preprocessing, labelling, data modelling, and evaluation using Google Colab. A total of 9,136 data were obtained from Twitter with the keywords of the presidential candidates' names, each data had three emotion labels (positive, neutral, and negative), which were processed and filtered into 300 representative samples. The IndoBERT model in this study uses a zero-shot learning approach allowing the model to recognise sentiment without the need to see specific training examples. In the evaluation stage, it obtained an accuracy value of 0.60 or 60%, tending to predict with a good value in positive labels of 0.74 or 74% for the balance measure (f1-score). This model can predict negative labels but the results are not very optimal with a value of 0.49 or 49%. On the Confusion Matrix this IndoBERT model is more likely to label tweets with positive things. However, to detect negative labels is quite good.

ACKNOWLEDGEMENTS

Gratitude is extended to Universitas Bina Insani for funding this research.

BIBLIOGRAPHY

[1] DATAREPORTAL, "TWITTER USERS, STATS, DATA & TRENDS," *We Are Social*, 2023. https://datareportal.com/essential-twitter-stats?utm_source=DataReportal&utm_medium=Country_Article_Hyperlink&utm_campaign=Digital_2023&utm_term=Indonesia&utm_content=Facebook_Stats_Link (accessed Oct. 25, 2023).

[2] I. Krismantari and N. Fitri Ramadhani, "Pemilu 2024 akan jadi pesta demokrasi terbesar di dunia: 5 hal yang perlu kamu ketahui," *The Conversation*, 2023. https://theconversation.com/pemilu-2024-akan-jadi-pesta-demokrasi-terbesar-di-dunia-5-hal-yang-perlu-kamu-ketahui-211069 (accessed Oct. 25, 2023).

[3] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.

[4] T. Patrio Sorongan, "Daftar 10 Negara Terluas di Dunia, Indonesia Nomor Berapa?," *CNBC Indonesia*, 2023. https://cnbc.id/NlOCBq (accessed Oct. 24, 2023).

[5] F. Zahria Emeraldien, R. Jefri Sunarsono, and R. Alit, "Twitter Sebagai Platform Komunikasi Politik Di Indonesia," *J. Teknol. dan Inf.*, vol. 14, no. 1, pp. 21–30, 2019, [Online]. Available: www.statisticbrain.com.

[6] M. A. Muqsith, *Pesan Politik di Media Sosial Twitter*. Surabaya: Jakad Media Publishing, 2022.

[7] A. Perdana, A. Hermawan, and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 2, pp. 195–200, 2022, doi: 10.32736/sisfokom.v11i2.1412.

[8] S. M. P. Tyas, R. Sarno, A. T. Haryono, and K. R. Sungkono, "A Robustly Optimized BERT using Random Oversampling for Analyzing Imbalanced Stock News Sentiment Data," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, 2023, pp. 897–902, doi: 10.1109/ICCoSITE57641.2023.10127725.

[9] Biro Humas Lemhannas RI, "Pemilu 2024 Waktunya Demokrasi Gagasan, Bukan Demokrasi Pengkultusan," *Biro Humas Lemhannas RI*, 2022. https://www.lemhannas.go.id/index.php/publikasi/press-release/1670-pemilu-2024-waktunya-demokrasi-gagasan-bukan-demokrasi-pengkultusan (accessed Oct. 24, 2023).

[10] P. F. Sarumaha, "TANTANGAN DAN STRATEGI DALAM MEWUJUDKAN KESUKSESAN PEMILU 2024 DI NIAS SELATAN," *Bawaslu Nias Selatan, Kordiv PHL dan Humas*, 2021. https://niasselatan.bawaslu.go.id/tantangan-dan-strategi-dalam-mewujudkan-kesuksesan-pemilu-2024-di-nias-selatan/ (accessed Oct. 24, 2023).

[11] BBC NEWS INDONESIA, "Pemilu 2024: Pemilih muda, dinamika politik penuh kejutan, dan potensi polarisasi - Sejumlah hal yang perlu Anda ketahui," 2023. https://www.bbc.com/indonesia/indonesia-66531834 (accessed Oct. 24, 2023).

[12] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 746–757, 2023, doi: 10.26555/jiteki.v9i3.26490.

[13] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10thInternational Joint Conference on Natural Language Processing*, 2020, pp. 843–857, doi: https://doi.org/10.48550/arXiv.2009.05387.

[14] G. Shetty, A. Nair, P. Vishwanath, and A. Stuti, "Sentiment Analysis and Classification on Twitter Spam Account Dataset," *Proc. - 2020 Adv. Comput. Commun. Technol. High Perform. Appl. ACCTHPA 2020*, pp. 111–114, 2020, doi: 10.1109/ACCTHPA49271.2020.9213206.

[15] S. Tabinda Kokab, S. Asghar, and N. Shehneela, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, p. 100157, 2022, doi: https://doi.org/10.1016/j.array.2022.100157.