OPEN ACCESS

ISSN 2460-9056
http://socj.telkomuniversity.ac.id/ijoict/



Intl. Journal on ICT
Vol. 9, No. 1, Jun 2023. pp. 44-56
doi: doi.org/10.21108/ijoict.v9i1.718

# Multi Aspect Sentiment Analysis of Mutual Funds Investment App Bibit Using BERT Method

Serly Setyani *1, Sri Suryani Prasetiyowati 2, Yuliant Sibaroni 3

1,2,3 *School of Computing, Telkom University*
*Jl. Telekomunikasi No. 1, Bandung, Jawa Barat, Indonesia*

1 serlysetyani@student.telkomuniversity.ac.id
2 srisuryani@telkomuniversity.ac.id
3 yuliant@telkomuniversity.ac.id

**Abstract**
Rapid technological developments and emerging applications can facilitate and provide solutions to problems. One of them is an application that can facilitate investment transactions. An investor no longer needs to visit the location of an investment company to make an investment. Investors can carry out all investment transactions through the smartphone screen. Bibit is an investment application that can help investors invest in mutual funds. There are many reviews submitted by users every day, therefore an aspect-based sentiment analysis is needed to identify the aspects and user sentiments of each review. Sentiment analysis based on a review of aspects of the Bibit application will be carried out in this study using the BERT method with IndoBERT who has been previously trained. The results of the classification of multi-aspect sentiment analysis showed that the highest accuracy was achieved by the service aspect at 92%, the user satisfaction aspect was 87%, and the lowest accuracy was achieved by the system aspect at 75%. Based on the sentiment analysis results, companies can improve aspects of the Bibit application system and services to provide better services & functionality.

**Keywords:** investment, Bibit, IndoBERT, aspect-based sentiment analysis

## I. INTRODUCTION

Before the emergence of applications related to investment transactions, an investor is required to visit the investment company concerned to carry out investment transactions actively. With the rapid development of technology, investment transaction activities can be carried out and completed through smartphone screens. Seeds are applications that take advantage of technological developments in the financial sector [1]. This application provides a variety of mutual fund products that users can buy according to their risk profile. Apart from buying mutual fund products, users can also create portfolios, switch mutual fund products, and sell mutual fund products.

Application users usually provide input through reviews and ratings on the application. Reviews and ratings play a very important role in decision-making when developing applications. However, the large number of user reviews makes it difficult for developers to make decisions in application development, so sentiment analysis is needed to identify the aspects discussed. This problem is called Aspect Based Sentiment Analysis (ABSA). ABSA can analyze the entire text to identify aspects, both attributes and components, that will be

Received on 7 Jun 2023. Revised on 17 June 2023. Accepted and Published on 30 June 2023.

detected by sentiment [2]. Reviews explain the overall sentiment towards the application and various other aspects such as system, service, and user satisfaction. An example of a review from the dataset that the writer took from the Bibit application: "Bibit sangat bagus, selama menggunakan aplikasi ini saya sangat puas, cs nya juga fast respon. penjelasanya juga lumayan mudah di pahami, pembayaran dan pelayanan nya mudah dan cepat". The review contains several general aspects, such as system aspects with positive sentiments, user satisfaction with positive sentiments, and service with positive sentiments.

Several studies in the field of sentiment analysis have been carried out. Research [3], uses the deep learning method which is an alternative to conducting aspect-based sentiment analysis, because of its ability to capture grammatical and semantic features of the text. Meanwhile [4], conducted a comparative aspect-based sentiment analysis on Amazon electronic product reviews using the Naive Bayes, SVM, LSTM, and BERT methods. The results obtained stated that BERT had a superior value compared to other methods, with an accuracy value of 88.48% and an f-1 score of 89.41%. Sentiment analysis using Bidirectional Encoder Representations from Transformers (BERT) which reviewed the Google Play Store application obtained an accuracy value of 84% [5], However, in this study, sentiment analysis has not been carried out at the aspect level.

IndoBERT is a variation of the BERT model developed by Google researchers [6]. IndoBERT has delivered an excellent performance in sentiment analysis. Meanwhile, Santiago [7], conducted a sentiment classification using 50,000 film review data sets taken from the IMDB page. In this study, a comparison was made between using BERT and supervised learning models such as SVC and logistic regression. The results obtained state that BERT gets the highest accuracy value of 93%.

In this study, a multi-aspect sentiment analysis was carried out on the review of the Seed application taken from the Google Play Store using the BERT method with the pre-trained IndoBERT model to determine sentiment. Selection of BERT with IndoBERT pre-training as a model considering that the methods can work well in sentiment analysis.

## II. Literature Review

BERT is a model based on the Transformer architecture and has the unique ability to attend to both preceding and succeeding contexts within a sentence [8]. Through a two-stage pre-training process, BERT can learn rich language representations that can be fine-tuned for specific language tasks. Evaluations demonstrate that BERT performs well, surpassing other methods in some language processing tasks. Thus, BERT represents a significant innovation in the field of natural language understanding and makes a substantial contribution to improving performance across various language processing tasks.

Several studies were conducted in the aspect-based sentiment analysis of user reviews using BERT. In research [9], Dimas Samodra Bimaputra conducted an aspect-based sentiment analysis on reviews of several hotels in Bali from the Tripadvisor page using BERT. By using a batch size of 16, epoch 2, and Adam's learning rate, this study achieves a very high level of accuracy, with 95% accuracy in the architectural aspect, 100% in the security aspect, 95% in the service aspect, 99% in the cleanliness aspect, and 99% in the aspect of convenience.

In another study, Priyan Fadhil Supriyadi [10] conducted a multi-aspect sentiment analysis on Xiaomi smartphone reviews with datasets taken from social media Twitter using BERT and IndoBERT. Researchers conducted several scenarios on BERT and IndoBERT to determine their performance. From the scenarios carried out, with a batch size of 16, an epoch of 4, and a learning rate of 3e-5, BERT and IndoBERT provided the best accuracy values. The accuracy performance of the BERT in this study was 91% on the battery aspect, 95% on the camera aspect, and 94% on the screen aspect. Then IndoBERT produces an accuracy of 97% on the battery aspect, 96% on the camera aspect, and 97% on the screen aspect. From the performance results, it was found that IndoBERT has a superior accuracy value than BERT.

In further research, Emerson Lopes [11] conducted research related to aspect-based sentiment analysis on Portuguese language hotel reviews taken from the TripAdvisor page using the BERT method and pre-trained

BERTimbauto. Each aspect is given three sentiments, namely negative (-1), neutral (0), and positive (1). In this study, ABSA was carried out using the largest Portuguese language corpus, and then the model received two inputs written in different formats. Even though the BERT model has been experimented with in NLP case studies several times, it raises biases during training. To overcome this, post-training was carried out on the model. The results of the experiment showed that the f1-score was 70% without post-training and 77% with post-training.

Based on several studies related to these aspect-based sentiment analyses, this research will also implement aspect-based sentiment analysis of user reviews using BERT. This study uses IndoBERT p-2 as a pre-train because the pre-train is trained with larger data and newer data and will fetch Bibit user review data from the google play store.

## III. RESEARCH METHOD

In Fig. 1, there is an overview of the stages of building a multi-aspect sentiment analysis system used in this study. The system starts by crawling Bibit user review data. Then proceed with determining aspects, data labelling, preprocessing, and k-fold cross-validation by dividing the k dataset and k experimental iterations. Furthermore, the classification is carried out using a data train using BERT with pre-trained IndoBERT on each aspect. Then in the final stage, an evaluation is carried out on the model that has been built.

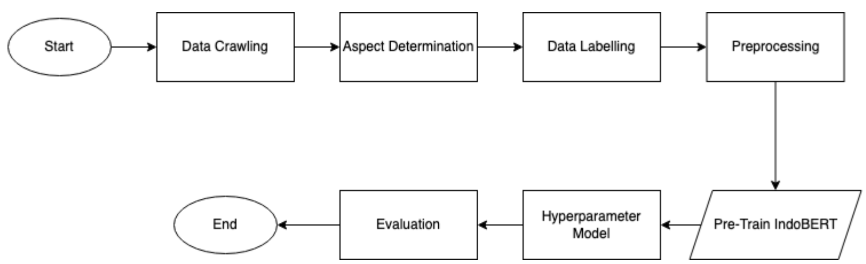


Fig. 1. Sentiment Analysis System Design

### A. Data Crawling

The data used in this study are user reviews of the Bibit application taken from the Google Play Store from January 2019 to December 2022, as many as 3004 data. The review data collection method is web scrapping with the help of a Python library called Google-Play-Scrapper. The reviews taken for this research are content and user rating scores for the Bibit application and the collection of reviews are then stored in comma-separated values (CSV) format. Table 1 is an example of the results of reviewing the Bibit application in this research.

TABLE I
EXAMPLE OF CRAWLING DATA RESULTS BIBIT APP REVIEWS

| ***Review*** |
|---|
| *Sudah saatnya Bibit memiliki mode gelap (dark mode), belum mendukung sistem navigasi android 12 sehingga ketika saat melakukan pembelian masih mudah untuk keluar aplikasi. padahal hanya sekedar melakukan gestur kembali untuk membatalkan pada opsi pembelian. Tampilan portofolio butuh dipercantik agar bisa senada dengan Android 12.* |
| *Saya gak bisa buka aplikasi ini kenapa ya ? Ada apa? Apa lagi gangguan di aplikasinya. Padahal sebelumnya lancar aja, uang investasi saya ada disitu ya.. Tapi gak bisa di buka terus* |
| *Selama pemakaian disini saya selalu puas, trading lancar, dari segi grafiknya bagus, TP dan SL nya kena terus, jadi ga sia sia kan kalo kita setting sedikian rupa bisa kita tinggal atau close aplikasinya, ga melulu harus dipantau di tungguin setiap saat juga. Top banget deh.* |

B. Delone Mclean Model

In multi-aspect sentiment analysis, the Delone McLean model can be used to understand the factors influencing user sentiment towards an information system or application. Delone McLean's model evaluates information systems implemented by a company to find parts of the system that need to be repaired, enhanced, or maintained [12]. This method was proposed by William Delone and Ephraim R. McLean in 1987, they stated that six dimensions influence the success of an information system. The six dimensions are:

1) Information quality

Information quality discusses the output produced by an information system.

2) System quality

System quality discusses the characteristics of information systems such as the ease of learning the system, system flexibility, and system reliability.

3) Service quality

Service quality discusses the services received by users from the information system used.

4) Usage intentions

Usage intentions discuss the level and ways in which users take advantage of the capabilities of an information system.

5) User satisfaction

User satisfaction discusses user satisfaction regarding responses or impressions of information system services.

6) System benefit

System benefit discusses the system's impact, results and benefits to user needs and company success.

C. Aspect Determination

In this study, the authors used three aspects of the Delone - McLean model, namely system, service, and user satisfaction. The choice of using this aspect is because, in the Delone McLean Model, it is an aspect that is often discussed by application users in the success of an information system [13]. Each review will be categorised into aspects and labelled with the appropriate sentiment when the review talks about any of the following:

1) System

The system aspect of the Bibit application covers performance, reliability, security, and user convenience. The positive sentiment reflects user satisfaction, the negative sentiment reflects dissatisfaction, and the neutral sentiment indicates neither pleasure nor displeasure.

2) Service

The service aspect of the Bibit application focuses on the quality of services provided to users. It includes the system's willingness and responsiveness to address user issues, as well as its ability to meet user needs. The positive sentiment reflects user satisfaction, the negative sentiment reflects dissatisfaction, and the neutral sentiment indicates neither pleasure nor displeasure.

3) User satisfaction

User satisfaction refers to the level of satisfaction experienced by users when using the Bibit application. It includes factors like user experience, ease of use, and effectiveness. Positive sentiment indicates satisfied user reviews, while negative sentiment reflects dissatisfaction or discomfort. Neutral sentiments refer to reviews without clear expressions of pleasure or displeasure.

## D. Data Labelling

After collecting the data, the next step is to label the data. In this process, each retrieved review will be labelled. The label to be used is in the form of numbers, where the value "1" represents positive sentiment where reviews contain positive or good words for using the Bibit application, then the value "2" represents negative sentiment where reviews contain words that describe dissatisfaction with the use of the application Bibit, and a value of "0" represents neutral sentiment where the review contains words that do not contain words that contain positive and negative words that are too strong and aspects that are not being discussed in a review.

Labelling of 3004 data was carried out by 3 people by dividing by the number of datasets equally and providing a brief containing guidelines to assist in the process of labelling sentiments and relevant aspects in each review. The system aspect has 875 data with negative labels, 1364 data with positive labels, and 765 data with neutral labels. The service aspect has 111 data with negative labels, 230 data with positive labels, and 2663 data with neutral labels. Meanwhile, in the aspect of user satisfaction, there are 1051 data with negative labels, 1718 data with positive labels, and 233 data with neutral labels. An example of data labelling in the Bibit application review is in Table 3.

TABLE II
EXAMPLE OF DATA LABELLING

| Reviews | *System* | *Service* | *User Satisfaction* |
|---|---|---|---|
| Aplikasi ini bagus banget buat pemula yang belajar menabung dan berinvestasi. CS nya juga fast respon. Penjelasan nya lumayan mudah di pahami. Pembayaran dan pelayanan nya mudah dan cepat. | 1 | 1 | 1 |
| aplikasi nya bagus kok.. cuman utuk waktu kerja nya sangat lama. mebutuhkan peroses perhari.. spai 3 hri lebih...semoga kedepan nya bisa lebih membaik. utuk jual beli nya gk membutuhkan peroses yng lama | 1 | 2 | 1 |
| Sejauh ini aplikasi untuk investasi tapi nilainya bisa kecil yg terbaik masih bibit menurutku, dan sangat cocok untuk pemula yg baru belajar investasi. | 1 | 0 | 1 |

## E. Preprocessing data

After labelling the data, the next step is to preprocess the data. The purpose of doing data preprocessing is to make the data more structured by not changing the meaning of the data and simplifying the model in classification. The following are several stages of data preprocessing:

1) Case Folding

In the case folding stage, the process is carried out to convert the text into a unified form, namely making all review data into lowercase letters [14].

2) Data Cleaning

At the data cleansing stage, the process is carried out to remove symbols, and numbers (0-9), delete links with http/https and www patterns, HTML tags, and remove characters other than the alphabet, and punctuation marks in the dataset. The deleted punctuation marks can be seen in Table 3.

TABLE III
REMOVED SYMBOLS AND PUNCTUATION

| Description | Punctuation & Symbols |
|---|---|
| Hashtags | # |
| At | @ |
| Percent | % |
| Point | . |
| Coma | , |
| Question mark | ? |
| Exclamation mark | ! |
| Quotation mark | " ' |

3) Tokenization

In the next preprocessing stage, the tokenization process will be carried out. Tokenization is the process of breaking sentences into smaller units of words [15].

4) Stopword Removal

In this process, the removal of common words that often appear in a sentence that has no significant meaning is carried out. The purpose of doing stopword removal is to reduce unnecessary words in the dataset and help the model increase efficiency and accuracy in carrying out sentiment analysis later. The stopword removal process is assisted by a Python library called Sastrawi. The Sastrawi version used in this study is version 1.0.1. List of stopword used in this studty can be seen in Table IV.

TABLE IV
LIST OF STOPWORDS

| ***Stopwords*** | | | | | | | |
|---|---|---|---|---|---|---|---|
| ada | apabila | beberapa | buat | dekat | entah | ingin | kapan |
| adalah | apakah | begini | bukan | demi | terhadap | ini | karena |
| agak | atau | begitu | cuma | dengan | hal | itu | ke |
| agar | bagai | belum | percuma | depan | hampir | jangan | kemudian |
| akan | sebagai | sebelum | dahulu | dia | hanya | jika | kenapa |
| aku | bagaimana | berapa | dalam | dialah | harus | juga | kepada |
| anda | bagi | biasa | dan | dong | hendak | justru | ketika |
| antara | bahwa | bila | dapat | dulu | hingga | kalau | nah |
| apa | bahwasanya | bisa | dari | enggal | ia | kalian | pada |
| mengapa | sebanyak | boleh | daripada | entah | ialah | kami | paling |
| per | saja | saling | sama | sana | sangat | sedang | sekali |
| sementara | semakin | sering | tadi | tak | tidak | wah | yang |

5) Nonstandard Word Normalization

Furthermore, changing words that are not standard or follow the rules that apply in Indonesian to become standard words or follow the Kamus Besar Bahasa Indonesia (KBBI) [16]. For non-standard dictionaries, the author uses a dictionary that has been provided by Kurniadi[1] on his GitHub pages. The examples of non-standard words normalization to standard word can be seen in Table V.:

TABLE V
EXAMPLE OF NONSTANDARD WORD NORMALIZATION

| Nonstandard Word | Standard Word |
|---|---|
| jan | jangan |
| mmg | memang |
| y | ya |
| dripada | dari pada |
| yg | yang |
| ngak | tidak |
| sy | saya |
| bs | bisa |
| riweh | ribet |

## F. Pre-Train IndoBERT

IndoBERT generally has the same architecture as BERT. However, IndoBERT was trained using large Indonesian text data with 180 epochs so that it can understand Indonesian representations and phrases well. In this study, the BERT pre-train used was IndoBERT p-2.

## G. Hyperparameter Model

After implementing the IndoBERT pre-train on the model, it is necessary to adjust the hyperparameters on the model to produce the best output. As for the parameters that can be adjusted and the range of parameter values that work well in almost any task is [8]:

1) Batch Size

Batch size is the number of data samples that are processed in one gradient calculation during model training. In general, the batch sizes used are 16 and 32.

2) Epoch

Epoch is one training cycle in the deep learning model, this parameter determines how many times the algorithm has worked through the entire dataset forward and backwards. The epoch values that are often used are 2, 3, and 4.

3) Learning Rate

The learning rate is a parameter that determines how many steps the optimization algorithm takes in finding the minimum cost function or loss function. The learning rate values that are often used are 5e-5, 4e-5, 3e-5, and 2e-5.

[1] https://raw.githubusercontent.com/ShinyQ/One-Click-Sentiment_BE/main/app/dataset/Kamu-Alay.csv

At this stage, experiments were carried out using the parameter values previously mentioned. The experimental results show that using epoch 3, learning rate of 5e-5, and batch size of 32, managed to achieve the best accuracy and f1-score. The parameter values were chosen in this study can be seen in Table VI.

TABLE VI
HYPERPARAMETER MODEL

| Epoch | Batch Size | Learning Rate | System F1 | System ACC | Service F1 | Service ACC | User Satisfaction F1 | User Satisfaction ACC |
|---|---|---|---|---|---|---|---|---|
| 2 | 16 | 2e-5 | 0.71 | 0.76 | 0.71 | 0.92 | 0.71 | 0.87 |
| | | 3e-5 | 0.71 | 0.76 | 0.68 | 0.92 | 0.72 | 0.87 |
| | | 4e-5 | 0.71 | 0.76 | 0.64 | 0.92 | 0.73 | 0.87 |
| | | 5e-5 | 0.71 | 0.76 | 0.70 | 0.92 | 0.73 | 0.87 |
| | 32 | 2e-5 | 0.71 | 0.75 | 0.69 | 0.93 | 0.69 | 0.87 |
| | | 3e-5 | 0.71 | 0.75 | 0.73 | 0.92 | 0.70 | 0.87 |
| | | 4e-5 | 0.71 | 0.76 | 0.73 | 0.93 | 0.71 | 0.87 |
| | | 5e-5 | 0.72 | 0.76 | 0.72 | 0.93 | 0.71 | 0.87 |
| **3** | 16 | 2e-5 | 0.72 | 0.76 | 0.73 | 0.92 | 0.73 | 0.87 |
| | | 3e-5 | 0.72 | 0.76 | 0.71 | 0.93 | 0.74 | 0.87 |
| | | 4e-5 | 0.71 | 0.74 | 0.71 | 0.93 | 0.73 | 0.87 |
| | | 5e-5 | 0.70 | 0.74 | 0.71 | 0.92 | 0.72 | 0.88 |
| | **32** | 2e-5 | 0.73 | 0.75 | 0.63 | 0.93 | 0.72 | 0.87 |
| | | 3e-5 | 0.73 | 0.75 | 0.68 | 0.93 | 0.73 | 0.87 |
| | | 4e-5 | 0.73 | 0.76 | 0.68 | 0.93 | 0.73 | 0.87 |
| | | **5e-5** | **0.73** | **0.76** | **0.73** | **0.92** | **0.73** | **0.87** |
| 4 | 16 | 2e-5 | 0.73 | 0.76 | 0.65 | 0.93 | 0.75 | 0.87 |
| | | 3e-5 | 0.73 | 0.76 | 0.68 | 0.92 | 0.75 | 0.86 |
| | | 4e-5 | 0.71 | 0.75 | 0.68 | 0.93 | 0.74 | 0.86 |
| | | 5e-5 | 0.73 | 0.71 | 0.65 | 0.93 | 0.75 | 0.86 |
| | 32 | 2e-5 | 0.73 | 0.76 | 0.59 | 0.92 | 0.73 | 0.87 |
| | | 3e-5 | 0.75 | 0.76 | 0.61 | 0.93 | 0.74 | 0.87 |
| | | 4e-5 | 0.73 | 0.76 | 0.69 | 0.92 | 0.75 | 0.87 |
| | | 5e-5 | 0.72 | 0.75 | 0.68 | 0.92 | 0.72 | 0.87 |

## H. K-fold Cross Validation

In the final stage, a model evaluation will be carried out using k-fold cross-validation, with the aim of evaluating the model by training and testing data on different subsets. In this study, the value of k used in the k-fold cross-validation is 10. This value is very commonly used because it can minimize bias [17].

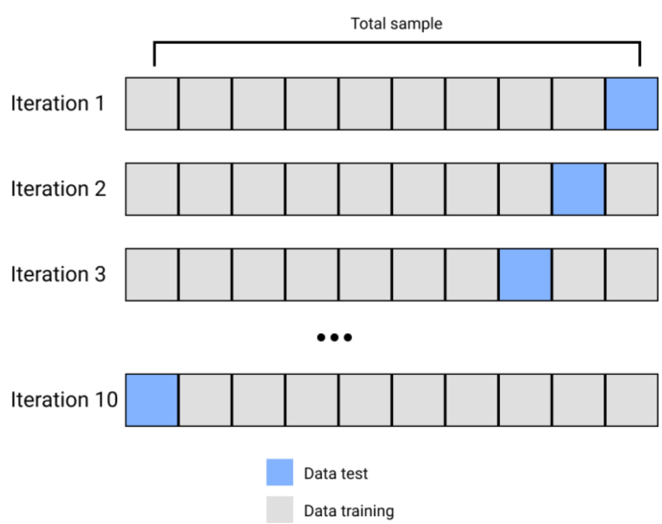


Fig. 2. 10 K-fold Cross Validation

Fig. 2. illustrates k-fold cross-validation, with a value of k = 10. In this process, k iterations will be carried out to calculate the evaluation value, so that k evaluation values will also be obtained. To get the accuracy value and f-1 score in each k-fold cross-validation iteration, where the formula for accuracy and f1-score can be seen in the:

1) Recall

Recall compares the True Positive (TP) value with lots of truly positive.

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{1}$$

2) Precision

Precision is a comparison between True Positive (TP) with lots of data that is predicted to be positive.

$$Precissions = \frac{TP}{TP + FP} \tag{2}$$

3) F1-score

F1-score is the ratio value of the average comparison of precisions with recall values.

$$F1 - score = \frac{2 \cdot Recall \cdot Precissions}{(Recall + Precessions)} \tag{3}$$

4) Accuracy

Accuracy is the ratio of the correct value predictions, both positive and negative, with all the existing data.

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \tag{4}$$

I. Confusion Matrix

Confussion matrix ( Table VII) is a matrix used to analyze the performance of the model in classifying. In this matrix, true positives (TP) denote the number of positive data that are correctly classified as positive, true negatives (TN) are the number of negative data that are correctly classified as negative, false positives (FP) are the number of negative data that are incorrectly classified as positive, and false negative (FN) is the number of positive data that is incorrectly classified as negative.

TABLE VII
CONFUSION MATRIX

| | **Predicted 0** | **Predicted 1** |
|---|---|---|
| **Actual 0** | TN | FP |
| **Actual 1** | FN | TP |

## IV. RESULTS AND DISCUSSION

In this research, we evaluate the model in conducting multi-aspect sentiment analysis on the Bibit application review dataset. This multi-aspect sentiment analysis involves identifying positive, negative, and neutral sentiments and identifying aspects commented on in a review text. To evaluate the model, k-fold cross-validation is used with a total of 10 folds. The metrics used in model evaluation are accuracy and f1-score. The model hyperparameters used in the model train process are batch size 32, epoch 3, and learning rates 5e-5.

TABLE VIII
F1-SCORE PERFORMANCE OF EACH ASPECTS

| K-fold | System | User Satisfaction | Service |
|---|---|---|---|
| 1 | 0.67 | 0.75 | 0.60 |
| 2 | 0.75 | 0.65 | 0.73 |
| 3 | 0.72 | 0.74 | 0.91 |
| 4 | 0.76 | 0.82 | 0.76 |
| 5 | 0.73 | 0.61 | 0.72 |
| 6 | 0.64 | 0.76 | 0.93 |
| 7 | 0.74 | 0.76 | 0.84 |
| 8 | 0.78 | 0.76 | 0.61 |
| 9 | 0.73 | 0.72 | 0.59 |
| 10 | 0.75 | 0.75 | 0.59 |
| **Average** | **0.73** | **0.73** | **0.73** |

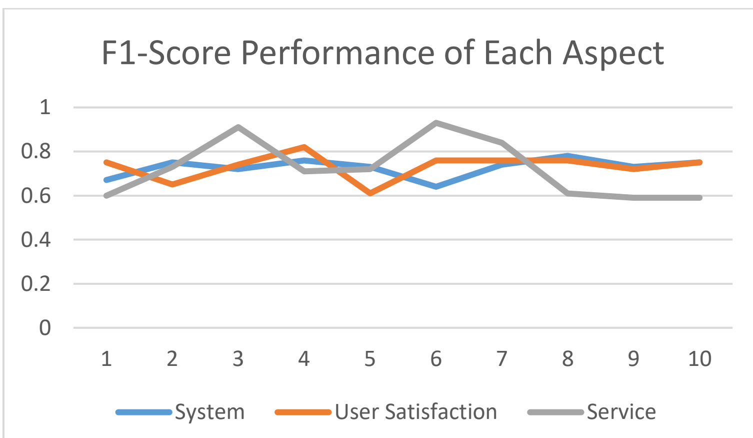


Fig. 3. F1-Score Performance of Each Aspect

Based on Table 9 and Fig. 3., the average f1-score for the system, user satisfaction, and service is 0.73. However, the f1-score values for user satisfaction and service vary across folds, unlike the stable values for aspect systems. The balanced f1-scores and accuracy of the system aspects indicate good performance in predicting classes accurately. However, the significant differences in f1-scores and accuracy between user satisfaction and service aspects are due to imbalanced data labels. Some k-fold cross-validation iterations lack data with neutral, positive, or negative labels, leading to decreased IndoBERT performance. Nevertheless, the average f1-score of 0.73 with 3 labels can still be considered good.

TABLE X
ACCURACY PERFORMANCE OF EACH ASPECT

| K-fold | System | User Satisfaction | Service |
|---|---|---|---|
| 1 | 0.74 | 0.90 | 0.89 |
| 2 | 0.72 | 0.85 | 0.93 |
| 3 | 0.77 | 0.90 | 0.92 |
| 4 | 0.78 | 0.87 | 0.92 |
| 5 | 0.72 | 0.88 | 0.92 |
| 6 | 0.76 | 0.86 | 0.92 |
| 7 | 0.74 | 0.86 | 0.94 |
| 8 | 0.77 | 0.88 | 0.93 |
| 9 | 0.74 | 0.86 | 0.96 |
| 10 | 0.77 | 0.87 | 0.91 |
| **Average** | **0.75** | **0.87** | **0.92** |

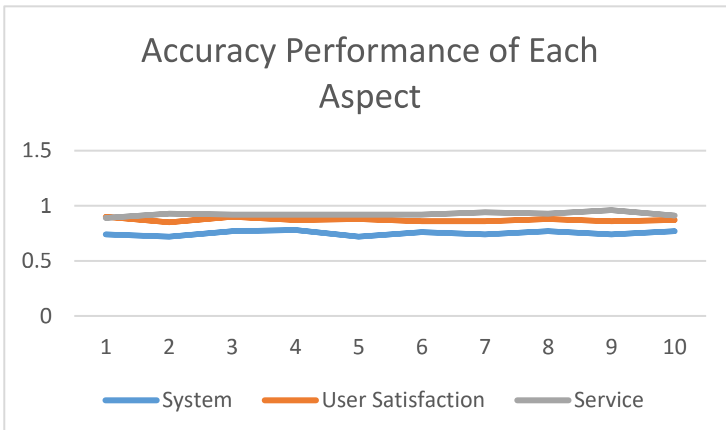


Fig. 4. Accuracy Performance of Each Aspect

Additionally, Table 10 and Fig. 4. show that accuracy performance shows that the system has a high level of accuracy, reliable service, and satisfactory user satisfaction. The average accuracy obtained on the system aspect reaches 0.92, the user satisfaction aspect reaches 0.87, and the service aspect reaches 0.75, demonstrating the ability of the model to perform a good classification of sentiment analysis. In addition, a stable level of accuracy at each validation fold shows the consistency of system performance in providing accurate prediction. A similar study using the same method and hyperparameter using a dataset of Xiaomi reviews was able to gain higher accuracy [10].

TABLE XI
SENTIMENT RESULT OF EACH ASPECT

| Label | System | User Satisfaction | Service |
|---|---|---|---|
| Negative | 3600 | 4800 | 200 |
| Positive | 7600 | 11400 | 400 |
| Neutral | 8800 | 3800 | 19400 |

In the previous stage, the model was trained using 3004 labelled data. Furthermore, to find out user sentiment towards the Bibit application, a sentiment analysis was carried out on 20,000 unlabeled Bibit application review data on each of its aspects. This data will be predicted using the best model from the experimental results that have been carried out.

On the system aspect, most reviews are neutral with 8800, followed by reviews with a positive sentiment of 7600, and reviews with a negative sentiment of only 3600. This shows that in general, Bibit users provide neutral reviews of the system aspects of the application. On service aspect, based on the results of sentiment analysis on unlabeled data from Bibit application reviews on the service aspect, it can be concluded that most reviews are neutral at 19400, while reviews with positive sentiment are only 400, and reviews with negative sentiment are only 200. On the aspect of user satisfaction, based on the results of sentiment analysis it can be concluded that as many as 11400 users give positive sentiments to the Bibit application in fulfilling user satisfaction. Meanwhile, 3800 users gave neutral sentiments, and 4800 users gave negative sentiments towards the Bibit application in fulfilling user satisfaction using the Bibit application.

## V. CONCLUSION

Multi-aspects sentiment analysis was carried out using IndoBERT using the Bibit application review dataset taken from the Google Play store, it can be concluded that using 32 batch size, 3 epoch, and 5e-5 learning rate can achieve the highest performance. The results of this study show that the service aspect has the highest average accuracy score of 0.92, the user satisfaction aspect has an average accuracy value of 0.87, and the system aspect has the lowest average score of 0.72. Judging from the accuracy results obtained, the model used can perform sentiment analysis well. From the results of the sentiment analysis, the company can consider making improvements to the system and service aspects to provide even better functionality and service to its users. However, this research still needs improvement to get even better model performance results. Because there are unbalanced data factors in the dataset used.

## REFERENCES

[1] H. Nurul, “Reksadana Bibit – Cara Beli, Jenis, dan Keunggulannya.” https://lifepal.co.id/media/bibit-reksadana/ (accessed Jun. 18, 2023).

[2] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, pp. 422–430, Apr. 2021, doi: 10.30865/MIB.V5I2.2845.

[3] H. H. Do, P. W. C. Prasad, A. Maag, and A. Alsadoon, “Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review,” *Expert Syst Appl*, vol. 118, pp. 272–299, Mar. 2019, doi: 10.1016/J.ESWA.2018.10.003.

[4] M. P. Geetha and D. Karthika Renuka, “Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model,” *International Journal of Intelligent Networks*, vol. 2, pp. 64–69, Jan. 2021, doi: 10.1016/J.IJIN.2021.06.005.

[5] K. S. Nugroho, A. Y. Sukmadewa, H. W. DW, F. A. Bachtiar, and N. Yudistira, “BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews,” *ACM International Conference Proceeding Series*, pp. 258–264, Jul. 2021, doi: 10.1145/3479645.3479679.

[6] M. F. Mubaraq and W. Maharani, “Sentiment Analysis on Twitter Social Media towards Climate Change on Indonesia Using IndoBERT Model,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, pp. 2426–2431, Oct. 2022, doi: 10.30865/MIB.V6I4.4570.

[7] S. González-Carvajal and E. C. Garrido-Merchán, “Comparing BERT against traditional machine learning text classification,” May 2020, Accessed: Mar. 27, 2023. [Online]. Available: https://arxiv.org/abs/2005.13012v2

[8] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North*, pp. 4171–4186, 2019, doi: 10.18653/V1/N19-1423.

[9] D. S. Bimaputra and E. Sutoyo, “Aspect-Based Sentiment Analysis of Hotels in Bali on Tripadvisor Using BERT Algorithm,” *International Journal of Advances in Data and Information Systems*, vol. 4, no. 2, Apr. 2023, doi: 10.25008/IJADIS.V4I2.1284.

[10] P. F. Supriyadi and Y. Sibaroni, “Xiaomi Smartphone Sentiment Analysis on Twitter Social Media Using IndoBERT,” *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 1, pp. 19–30, Feb. 2023, doi: 10.30865/jurikom.v10i1.5540.

[11] E. Lopes, L. Freitas, G. Gomes, G. Lemos, L. Hammes, and U. Corrêa, “Exploring BERT for Aspect-based Sentiment Analysis in Portuguese Language,” *Proceedings of the International Florida Artificial Intelligence Research Society Conference, FLAIRS*, vol. 35, 2022, doi: 10.32473/FLAIRS.V35I.130601.

[12] S. Mardiana, J. H. Tjakraatmadja, and A. Aprianingsih, “International Journal of Economics and Financial Issues DeLone-McLean Information System Success Model Revisited: The Separation of Intention to Use-Use and the Integration of Technology Acceptance Models,” *International Journal of Economics and Financial Issues*, no. 5, pp. 10–11, 2015, [Online]. Available: http:www.econjournals.com

[13] W. H. DeLone and E. R. McLean, “Information Systems Success: The Quest for the Dependent Variable,” *https://doi.org/10.1287/isre.3.1.60*, vol. 3, no. 1, pp. 60–95, Mar. 1992, doi: 10.1287/ISRE.3.1.60.

[14] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi,” *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, p. 012017, Jun. 2020, doi: 10.1088/1757-899X/874/1/012017.

[15] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Soc Netw Anal Min*, vol. 11, no. 1, pp. 1–19, Dec. 2021, doi: 10.1007/S13278-021-00776-6/FIGURES/5.

[16] M. Theo, A. Bangsa, S. Priyanta, and Y. Suyanto, “Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 2, pp. 123–134, Apr. 2020, doi: 10.22146/IJCCS.51646.

[17] M. Kuhn and K. Johnson, “Applied predictive modeling,” *Applied Predictive Modeling*, pp. 1–600, Jan. 2013, doi: 10.1007/978-1-4614-6849-3/COVER.