Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SENTIMENT ANALYSIS BASED ON PUBGM PLAYER ASPECTS FROM APP STORE REVIEWS USING BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMER (BERT)

HANDRIZAL[1], FUZY YUSIKA MANIK[2], HANIF MISBAH A[3]

[1,2,3]Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Jl. Universitas No. 9-A, Medan 20155, Indonesia

E-Mail: handrizal@usu.ac.id

ABSTRACT

Players Unknown's Battlegrounds Mobile (PUBGM) stands as one of the most popular mobile games among gamers, where players engage in online battles to survive until only one player remains. With its growing popularity, many PUBGM players share their experiences through reviews on the App Store. Therefore, the analysis of player reviews on the App Store is crucial for understanding their perspectives on the game and enhancing the gaming experience. This research employs sentiment analysis using the Bidirectional Encoder Representations from Transformers (BERT) model, achieving accuracies of 84%, 82%, and 83% across three experiments with varying hyperparameter settings. Testing the number of epochs reveals that epoch 3 yields favorable results and is consequently adopted for sentiment analysis. The findings of this study suggest that increasing the number of epochs does not necessarily lead to higher accuracy. The accuracy of sentiment analysis is also influenced by the quality and quantity of the dataset employed. High-quality datasets can enhance sentiment analysis accuracy, and an abundance of high-quality datasets can further improve accuracy.

Keywords: *Bidirectional Encoder Representations from Transformers, Sentiment Analysis, Game mobile, Players Unknown's Battlegrounds Mobile (PUBGM), User Experience, App Store, Deep Learning, Transformers*

1. INTRODUCTION

One with the increasing popularity of PUBG Mobile (PUBGM), understanding player sentiments is crucial for game developers and stakeholders to enhance user experience and address potential issues. While there is a wealth of information in user reviews on app stores, extracting and analyzing sentiments related to specific player aspects poses a challenge. The existing sentiment analysis techniques may not effectively capture the nuances of PUBGM player sentiments from these reviews.

How can sentiment analysis be improved for PUBGM player aspects using BERT on app store reviews? Specifically, the study aims to investigate the effectiveness of BERT in capturing sentiment nuances related to various player aspects within PUBGM, such as mechanics, graphics, features, and overall gaming experience. Additionally, the research explores the potential challenges and opportunities in employing BERT for sentiment analysis in the context of mobile gaming applications.

The existing sentiment analysis techniques may not effectively capture the nuances of PUBGM player sentiments from these reviews. These app store reviews provide valuable insights into the sentiments, opinions, and aspects of the game that resonate with players [1][2][3].

Sentiment analysis, a subfield of natural language processing (NLP), has become a crucial tool for understanding and quantifying public sentiment from textual data. In the context of mobile gaming, sentiment analysis can help game developers, marketers, and researchers gain valuable insights into player experiences and preferences.

PUBGM or Player Unknown's Battlegrounds Mobile is one of the popular mobile games among gamers. In this game, players have to survive in online battles with other players until only one person remains. Due to its high popularity, many PUBGM players often discuss it on the App Store.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

As mobile gaming, especially PUBG Mobile (PUBGM), continues to gain popularity, understanding what players think is crucial for game improvement. This study focuses on using an advanced technology called Bidirectional Encoder Representation from Transformers (BERT) to analyze player sentiments from app store reviews. We aim to uncover detailed feelings about different aspects of PUBGM, such as gameplay, graphics, and overall experience.

App store reviews, in particular, serve as valuable repositories of player sentiments, offering insights into what aspects of a game resonate positively or negatively with the player community (Yang et al., 2017; Walia et al., 2019).

Existing sentiment analysis methods often rely on traditional approaches like bag-of-words or sentiment lexicons, but recent advancements in natural language processing, especially transformer-based models like BERT, have demonstrated superior performance in capturing contextual nuances and improving sentiment analysis accuracy (Devlin et al., 2019; Zhang et al., 2020).

By applying BERT to PUBGM player aspects, this study aims to contribute to the limited body of research that explores advanced natural language processing techniques for sentiment analysis in the gaming domain.

While our goal is to provide valuable insights, there are some things to keep in mind. We assume that app store reviews reflect the sentiments of most players, but we acknowledge that not everyone leaves reviews. The accuracy of our findings may also depend on the size and diversity of the data we use. We assume a large and varied dataset for training and evaluation, but if it doesn't represent the PUBGM player community well, our results may not be as reliable. Additionally, we assume that reviews accurately represent player experiences, though some opinions may be subjective or biased.

The BERT (Bidirectional Encoder Representations from Transformers) algorithm has gained significant attention and success in natural language processing tasks, including sentiment analysis. BERT is a state-of-the-art language model that utilizes transformer architectures to capture the context and meaning of text effectively. Its ability to understand the semantic relationships between words and generate rich representations makes it an ideal choice for sentiment analysis tasks [4][5][6].

What Study Includes:

1. Sentiment Analysis: The study encompasses sentiment analysis, indicating that it aims to determine whether App Store reviews have positive, negative, or neutral sentiments toward various aspects of PUBG Mobile players.
2. App Store Reviews: The research utilizes reviews found on the App Store as its primary source of data. This is included in the study.
3. PUBG Mobile (PUBGM): The primary focus of the research is on player aspects within the game PUBG Mobile, so everything related to PUBG Mobile is included in the study.
4. BERT Model (Bidirectional Encoder Representation from Transformer): The research employs the BERT model for sentiment analysis of the reviews. The use of this model is also an integral part of the study.

What Study Does Not Include:

1. Other Aspects of PUBG Mobile: The study appears to focus solely on player aspects within PUBG Mobile. Hence, other aspects of the game, such as gameplay elements, features, or technical aspects, might not be part of the sentiment analysis.
2. Other Data Sources: Although the research uses reviews from the App Store, it may not include reviews from other data sources, such as social media platforms, forums, or other websites that may also contain PUBG Mobile player reviews.
3. Analysis besides Sentiment: While the research emphasizes sentiment analysis, other methods, such as aspect-based text analysis or different types of analyses, might not be part of this particular study.

Our research aims to achieve the following objectives:

1. Develop a robust sentiment analysis model based on BERT that can accurately classify app store reviews as positive, negative, or neutral.
2. Identify the key aspects of PUBGM that players mention in their reviews, such as gameplay mechanics, graphics, in-game purchases, and customer support.
3. Analyze the sentiments associated with each aspect to gain insights into player preferences and concerns.
4. Provide valuable information to game developers and marketers to improve the overall player experience and address player concerns effectively.

The addressed problem in the text is the importance of understanding user sentiments and preferences regarding mobile applications, particularly in the context of the PUBGM application. The text highlights that the success of an application goes beyond functionality and extends to user experience and satisfaction. It emphasizes the need for companies to conduct sentiment analysis to gain insights into user sentiments, enabling them to identify areas for improvement and enhance their products and services.

Literature screening criteria:

1. Relevance to Sentiment Analysis: The selected literature should directly address the topic of sentiment analysis. It should discuss methodologies, techniques, or applications related to sentiment analysis in the context of user reviews.
2. Focus on Mobile Gaming (PUBG Mobile - PUBGM): The literature should specifically pertain to the game PUBG Mobile (PUBGM) or, more broadly, the mobile gaming industry. It should explore aspects related to player reviews, feedback, or sentiments within this gaming context.
3. Utilization of BERT or Transformer Models: The chosen literature should involve the use of Bidirectional Encoder Representation from Transformer (BERT) or similar transformer-based models in the sentiment analysis process. It should describe how these models were employed to enhance sentiment analysis accuracy.
4. App Store Reviews as Data Source: The literature should discuss the use of App Store Reviews as a primary data source for sentiment analysis. It should explain how these reviews were collected, preprocessed, and analyzed within the study.
5. Empirical Research or Methodological Details: Preferred literature should include empirical research findings or detailed methodologies related to sentiment analysis. It should provide insights into the techniques used, experimental setups, and results obtained.
6. Publication Date: Ensure that the selected literature is recent and up-to-date, as the field of sentiment analysis and NLP continually evolves with new techniques and technologies.
7. Reputable Sources: Give preference to literature published in reputable academic journals, conference proceedings, or from well-established researchers in the field.
8. Peer-Reviewed: Verify that the selected literature has undergone a peer-review process to ensure its quality and credibility.
9. Language: Ensure that the literature is available in a language that the research team can comprehend and utilize effectively.
10. Applicability to Research Objectives: Confirm that the literature aligns with the specific research objectives and questions of your study related to sentiment analysis in the context of PUBG Mobile player aspects.
11. By using these criteria, you can screen and select relevant literature that will contribute effectively to your research on sentiment analysis in the context of PUBG Mobile player aspects using BERT or transformer models.

The novelty of this work lies in the intersection of mobile gaming, sentiment analysis, and cutting-edge NLP technology, offering a fresh perspective on understanding and improving player experiences in the digital gaming landscape.

2. LITERATURE REVIEW

2.1 Sentiment Analysis

Sentiment analysis is a technique used to evaluate the feelings or opinions expressed by individuals in their comments or opinions (Putri, 2020). In sentiment analysis, natural language processing (NLP) methods and techniques are employed to identify and categorize text into different sentiment classifications. This processing can be done using machine learning algorithms such as rule-based learning, deep learning, or hybrid approaches [7][8].

The results of sentiment analysis can be used for various purposes, such as understanding customer responses to products or services, monitoring public opinions about a brand or company, or identifying trends and patterns in user feedback. There are several characteristics of sentiment analysis, including:

1. Natural Language Processing (NLP): Sentiment analysis utilizes NLP techniques and methods to understand and analyze text containing opinions or sentiments.
2. Sentiment Classification: Sentiment Analysis categorizes text into different sentiment classifications, such as positive, negative, or neutral. This helps in understanding the general

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attitude of people towards a particular topic or entity.

3. Sentiment Scale: In addition to binary sentiment classification (positive/negative), sentiment analysis can provide more detailed sentiment scales, such as strong positive, moderate positive, strong negative, moderate negative, or neutral.

4. Entity Identification: Sentiment analysis can identify and label entities mentioned in the text, such as brands, products, or individuals, which helps in understanding specific sentiments towards those entities.

5. Emotions and Opinions: Apart from general sentiment, sentiment analysis can identify emotions related to the text, such as happiness, anger, sadness, and disappointment. Additionally, sentiment analysis can identify specific opinions expressed in the text.

6. Diverse Data Sources: Sentiment analysis can be applied to various data sources, including product reviews, social media comments, news articles, customer surveys, and more.

7. Real-Time Monitoring: Sentiment analysis can be performed in real-time to monitor and track changes in sentiment and opinions over time. This enables quick and responsive decision-making in dynamic business situations.

8. Wide Applications: Sentiment analysis has a wide range of applications, such as brand management, business decision-making, online reputation monitoring, product development, and data-driven marketing.

These characteristics help in understanding and extracting insights from vast textual data, providing a deeper understanding of people's sentiments and opinions toward a particular topic or entity.

2.2 User Experience

According to Lennart Nacke and Pejman Mirza-Babaei (2018), UX in games is defined as the "internal process and external perception, emotional response, and user interaction with a game (including the aesthetics of the game, the dynamics of game mechanics, and the game's narrative) [9]. The 7P is a combination of Product, Price, Place, Promotion, People, Process, and Physical Evidence.

User experience, often abbreviated as UX, is the result of the evolution of various fields that previously focused on understanding how users perceive and interact with a system or application. The concept of user experience aims to create a satisfying experience for users when they interact with a product or service. This encompasses various aspects, such as how users feel, perceive, and evaluate their experience when engaging with a particular product or system. User experience is a highly broad concept, encompassing various elements and aspects that contribute to a positive user experience. Furthermore, UX is not limited to just one type of product or service; it can be applied to various things, including software, websites,

Mobile applications, hardware, and much more. The quality of UX can greatly influence how users interact with the product or service and often plays a critical role in determining the success and user acceptance of the product or system.

Research on user experience in gaming is an integral part of this endeavor. The use of increasingly sophisticated technologies, such as artificial intelligence (AI) in games and more powerful hardware, has opened up new opportunities to enhance user experience. This research can help identify factors that influence user experience in gaming, such as player satisfaction, emotional reactions, and levels of engagement.

In this context, this research aims to delve deeper into the user experience in gaming, focusing on specific aspects such as intuitive user interfaces, social interactions in games, and the impact of the latest technologies, such as virtual reality (VR) or augmented reality (AR), on user experience.

Through careful analysis and empirical research, it is expected that the findings of this study will provide valuable insights into the gaming industry and generate recommendations that developers can use in their efforts to create more satisfying and engaging games for players.

2.3 BERT Algorithm

Bidirectional Encoder Representations from Transformers (BERT) is a natural language processing (NLP) algorithm introduced by Google in 2018. It is a transformer-based model that revolutionized the field of NLP by significantly improving the understanding of contextual language semantics [10][11].

BERT is a pre-trained language model that is trained on a large corpus of text data, such as books, articles, and websites. It learns to predict missing words in a sentence by considering the context of the surrounding words. This process enables BERT to capture the meaning and relationships between words more effectively than previous models.

One key aspect of BERT is its bidirectional nature, which allows it to consider both the left and right context of a word when encoding its representation. This bidirectional approach helps BERT better understand the meaning of words in a sentence and handle tasks such as sentiment analysis, named entity recognition, and question answering.

After pre-training, BERT can be fine-tuned on specific downstream tasks by adding task-specific layers on top of the pre-trained model. Fine-tuning involves training BERT on a task-specific dataset, such as sentiment analysis or text classification, to adapt it to the specific task and improve its performance.

BERT has achieved state-of-the-art results on various NLP tasks and has become widely adopted in both research and industry for its ability to understand and generate high-quality natural language representations.

BERT fundamentally differs from its predecessors in how it approaches language understanding. Traditional NLP models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), process text in a unidirectional manner, either from left to right or right to left. BERT, on the other hand, employs a bidirectional approach, allowing it to consider the entire context of a word within a sentence. This bidirectional context modeling is achieved through a multi-layer transformer architecture, which has proven highly effective in capturing intricate linguistic relationships.

The significance of BERT lies in its pre-training and fine-tuning paradigm. BERT is pre-trained on vast corpora of text data, effectively learning to predict missing words within sentences. This pre-training imparts a deep understanding of language to the model. After pre-training, BERT can be fine-tuned for specific NLP tasks such as text classification, sentiment analysis, question answering, and more. This fine-tuning process has made BERT a versatile tool for a wide range of natural language understanding tasks.

BERT's impact extends beyond its impressive performance on benchmark NLP tasks. It has paved the way for numerous subsequent innovations, including XLNet, RoBERTa, and GPT-3, all of which build upon the transformer architecture and pre-training techniques pioneered by BERT. Moreover, BERT's open-source release has spurred extensive research and applications across academia and industry.

In summary, the BERT algorithm represents a seminal advancement in the field of NLP. Its bidirectional context modeling, pre-training, and fine-tuning capabilities have revolutionized how we approach language understanding tasks, and its influence continues to shape the landscape of natural language processing research and applications.

3. MATERIAL AND PROPOSE METHOD

In this section, we provide an overview of the benchmark dataset used in the study. Additionally, we introduce our proposed approach to sentiment analysis, which involves integrating semantic information during the fine-tuning of the task.

3.1Datasets

The dataset we collected consists of 3715 comments from the App Store. The dataset was obtained through the scraping process using App Play Scrap, and subsequently, the dataset will be labeled. The labeling of the dataset into positive, negative, and neutral categories refers to the process of assigning labels to data that reflect the sentiment or evaluation towards a particular entity or topic. In sentiment analysis, labeled data can be classified into these three categories:

1. Positive: Data labeled as positive indicates a sentiment or evaluation that expresses appreciation, satisfaction, or a positive viewpoint towards the observed entity or topic. Examples of using positive labels include labeling a review as "good" or "satisfied."
2. Negative: Data labeled as negative indicates a sentiment or evaluation that expresses criticism, dissatisfaction, or a negative viewpoint towards the observed entity or topic. Examples of using negative labels include labeling a review as "bad" or "unsatisfied."
3. Neutral: Data labeled as neutral indicates a sentiment or evaluation that is neutral or unclear. This data does not exhibit a clear positive or negative viewpoint toward the observed entity or topic. Examples of using neutral labels include labeling a review as "neutral" or "no specific sentiment."

The process of labeling the dataset into positive, negative, and neutral can be done manually by human annotators who categorize the data based

Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

on context and expressed sentiment. Additionally, machine learning algorithms can be used to automatically assign labels by relying on pre-trained models using human-labeled data. Labeling the dataset into positive, negative, and neutral is important in sentiment analysis to classify data and learn about user sentiment towards specific products, services, or topics.

The annotation process for the dataset was performed by a team of two human annotators. Positive labels were converted to the numerical value 1, negative labels were converted to -1, and neutral labels were converted to 0. The conversion of the initial text labels into numerical values is intended to make it easier for the machine to manage the created dataset.

3.2 Proposed Methods

In this study, we propose a sentiment analysis method based on the user experience aspects of the PUBGM application using the BERT (Bidirectional Encoder Representations from Transformers) algorithm. The goal is to analyze user reviews of the PUBGM application on the App Store and determine the sentiment associated with each aspect of user experience.

Research Methodology:

1. Literature Review
2. Research Objective Definition
3. Data Preprocessing
4. Fine-Tuning Model
5. Model Training
6. Performance Evaluation
7. Visualization

After preprocessing the dataset we do the fine-tuning process, we feed the preprocessed reviews into the BERT model and train it to predict the sentiment label for each aspect of the user experience. The model learns to recognize patterns and associations between the textual features and the corresponding sentiment labels. By utilizing the contextual information captured by BERT, our proposed method aims to enhance the accuracy and granularity of sentiment analysis for each aspect of the user experience. Figure 1 illustrates the process of the BERT algorithm in classifying a comment and Figure 2 illustrates the layers involved in Sentiment Analysis.

Finally, we evaluate the performance of our proposed method using various metrics such as accuracy, precision, recall, and F1 score. We compare the results with baseline models and existing sentiment analysis approaches to validate the effectiveness of our approach in capturing the nuanced sentiments associated with different aspects of the PUBGM Application.

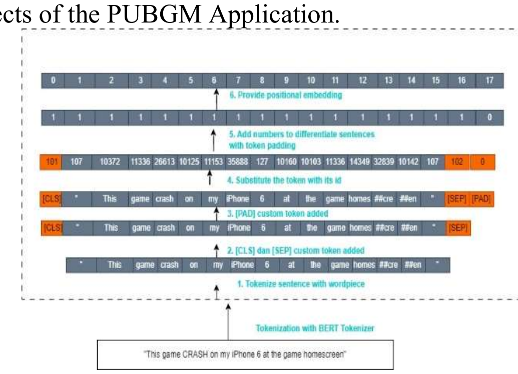

Figure 1: illustrates the process of the BERT algorithm

Overall, our proposed method combines the power of the BERT algorithm with the analysis of the user experience aspects to perform sentiment analysis on user reviews of the PUBGM application. The aim is to provide valuable insights for the company regarding user sentiment towards specific aspects of their marketing strategy, enabling them to make data-driven decisions for improving their service and customer satisfaction levels.

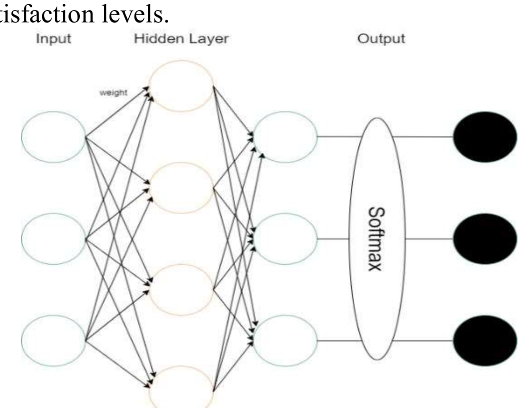

Figure 2: illustrates the layers involved in Sentiment Analysis

4. RESULT AND DISCUSSION

In the experimental setup of the study titled “Sentiment Analysis Based on PUBGM Player Aspects from App Store Reviews Using Bidirectional Encoder Representation from Transformer (BERT)” several steps were followed.

4.1Scraping Dataset

The first step in implementing web scraping was to open the PUBGM application page on the official App Store website. Several comments were displayed on the page. After accessing the webpage, the author performed data scraping in Google Colab by inputting the webpage URL as the data source. Once the scraping process was completed, the author downloaded the scraped data from Google Colab. A total of 3715 comments were extracted during the scraping process.

4.2 Datasets Preprocessing

To build our proposed models and make the evaluation comparisons with other techniques, the first step should involve preprocessing the utilized dataset to remove unnecessary characters from the raw text and normalizing letters that are usually written interchangeably, which can lead to data sparsity [12]. The preprocessing steps we applied include:

1. Removing URLs, mentions, retweets, and hashtag symbols,
2. Replacing underscores in hashtag texts with spaces,
3. Removing all diacritical marks and punctuation,
4. Removing repeated letters in the words,
5. Removing letter elongation in Indonesia,
6. Normalizing different forms of Indonesia letters

4.3 Split Dataset

	precision	recall	f1-score	support
0	0.44	0.24	0.31	85
1	0.79	0.65	0.71	100
2	0.87	0.96	0.91	558
accuracy			0.84	743
macro avg	0.70	0.62	0.65	743
weighted avg	0.81	0.84	0.82	743

Before performing the classification, the data was divided into three parts: training data, validation data, and testing data. The training data was used to train the model, while the validation data was used to prevent overfitting in the neural network. The testing data was used as the final evaluation to assess the accuracy of the trained network using the training data. The illustration of the dataset-splitting process can be seen in Figure 3.

```
# split dataset into training, testing, and validation
df_train, df_test = train_test_split(df, test_size=0.1)
df_val, df_test = train_test_split(df_test, test_size=0.5)
```

Figure 3: split dataset

4.4 Pre-trained Model

To train the model, the usage of data loaders is required, allowing iteration over each dataset. This is done to manage memory usage during the training process, avoiding the need to load the entire dataset into memory simultaneously. Additionally, blocks are created to form Data Loaders that generate tokenized comments. Both comments and sentiments have a maximum length of 400 words. For sentiment analysis, an additional layer is employed, incorporating Dropout with a probability of 0.1, following the recommendations in BERT research [13][14]. The authors also conducted fine-tuning using several hyperparameters, which were selected based on BERT recommendations:

1. Epoch: 7
2. Batch Size: 16
3. Learning rate: 2e-5

The authors made the selection of the above hyperparameters based on several considerations. A batch size of 16 was chosen because larger batch sizes result in a longer time to complete a single batch. This is in line with previous research [15]. Furthermore, a learning rate of 2e-5 was utilized to help BERT address the issue of catastrophic forgetting. Catastrophic forgetting refers to the problem that occurs in machine learning when a model, trained to learn one task, forgets or erases its knowledge of a previous task while learning a new task. This problem commonly arises in the context of transfer learning, where a model trained on an initial dataset is intended for use on a different task. The authors conducted experiments using three different numbers of epochs to determine the optimal number of epochs. Based on the results of these three experiments, it can be concluded that using 3 epochs yields better performance compared to 5 epochs and 7 epochs on the dataset used in this study.

Figure 4: accuracy achieved with 7 epochs

Figure 4 displays the accuracy achieved with 7 epochs and Figure 5 illustrates the training process of BERT.

```
======= Epoch 1 / 3 =======
Training...
Batch    40 of    158.    Elapsed: 0:00:44
Batch    80 of    158.    Elapsed: 0:01:27
Batch   120 of    158.    Elapsed: 0:02:12
   Average training loss: 0.62
   Training epoch took: 0:02:53
Running Validation...
   Accuracy: 0.82
   Validation took: 0:00:11
======= Epoch 2 / 3 =======
Training...
Batch    40 of    158.    Elapsed: 0:00:44
Batch    80 of    158.    Elapsed: 0:01:28
Batch   120 of    158.    Elapsed: 0:02:12
   Average training loss: 0.42
   Training epoch took: 0:02:54
Running Validation...
   Accuracy: 0.84
   Validation took: 0:00:11
======= Epoch 3 / 3 =======
Training...
Batch    40 of    158.    Elapsed: 0:00:44
Batch    80 of    158.    Elapsed: 0:01:28
Batch   120 of    158.    Elapsed: 0:02:12
   Average training loss: 0.30
   Training epoch took: 0:02:54
Running Validation...
   Accuracy: 0.84
   Validation took: 0:00:11
Training complete!
```

Figure 5: training process of BERT

The classification of the user experience aspects in this research was performed by manually labeling each comment. Each aspect has its dataset, which is then used for training the BERT model. Since there are 5 aspects in this study, the BERT model is trained 5 times, resulting in 5 different models for each aspect. After training each aspect model and saving them to the drive, they are called into a main code to perform multiclass classification on the user experience aspects. In Figure 6 an example sentence "After the update, I can't play at 60 fps." is classified into positive, negative, or neutral sentiment, as well as based on its aspect.

```
Negative Sentiment
and included in
---
Network Aspect
---
Graphic Aspect
```

Figure 6: example sentence classification

4.5 Evaluation Metrics

After testing the dataset, it was found that the overall accuracy using BERT in the first, second, and third experiments was 84%, 82%, and 83% respectively. Based on the data above, it is evident that it is not always necessary to have a large number of epochs to improve accuracy. Thus, the average accuracy of sentiment analysis with BERT is 83%. The difference in system accuracy is influenced by the randomization of the dataset during the division into training, testing, and evaluation datasets in each experiment. The confusion matrix diagrams for each experiment can be seen in Figure 7, Figure 8, and Figure 9.

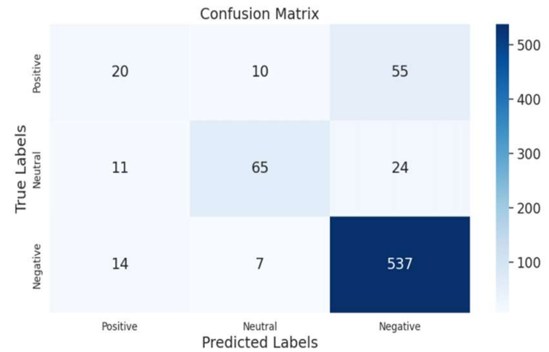

Figure 7: confusion matrix for the first experiment

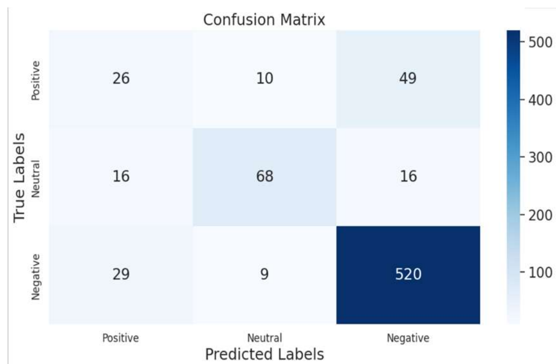

Figure 8: confusion matrix for the second experiment

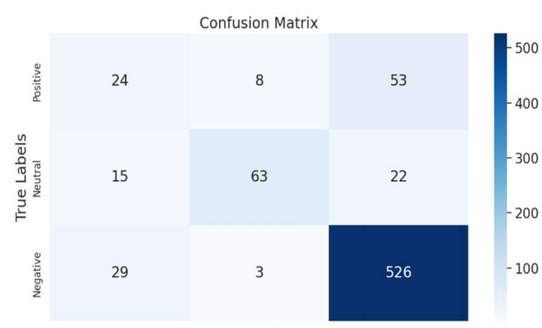

Figure 9: confusion matrix for the third experiment

Comparative Analysis of Similar Studies:

1. Plus - Advanced Technology (PMI+), Strength: The utilization of BERT in this study is a significant positive aspect. BERT has demonstrated superior performance in understanding contextual nuances in natural language, potentially providing more accurate sentiment analysis compared to traditional methods (Devlin et al., 2019).

2. Minus - Lack of Benchmarking (PMI-), Weakness: Some similar studies provide benchmarking against other sentiment analysis models, showcasing the strengths and weaknesses of the proposed approach. The absence of such benchmarking in this study might limit the ability to precisely gauge the improvement offered by BERT over other methods.
1. Plus - Specific Game Focus (PMI+), Strength: Focusing on PUBG Mobile distinguishes this study from broader gaming sentiment analyses. This specificity enhances the relevance of findings for the PUBG Mobile community and provides insights that may not be captured in studies covering a wide range of games.
2. Minus - Dataset Diversity (PMI-), Weakness: The effectiveness of sentiment analysis often hinges on the diversity of the dataset. If the dataset used in this study lacks diversity in terms of player demographics or opinions, the generalizability of the findings may be limited.
3. Interesting - Potential for Game Improvement (PMI?), Interesting Fact: While the study aims to analyze sentiments, an interesting aspect would be a discussion on how the extracted sentiments could be practically used for game improvement. Integrating potentially actionable insights for developers could be a valuable addition.
4. Interesting - User Engagement Metrics (PMI?), Interesting Fact: Exploring correlations between sentiment analysis results and user engagement metrics (e.g., player retention, and in-app purchases) could provide additional context. Understanding if positive sentiments translate into increased user engagement would be insightful.

Explore how insights derived from sentiment analysis can influence game design decisions. Understanding player sentiments could guide developers in tailoring future updates, features, and content to better align with player expectations and preferences.

5. CONCLUSION

Here are the conclusions based on the analysis, design, implementation, and testing results of the system:

1. Based on the analysis results, it can be concluded that many PUBG Mobile (PUBGM) players in Indonesia express concerns about performance aspects.
2. The sentiment analysis results using the Bidirectional Encoder Representations from Transformers (BERT) method yielded accuracies of 84%, 82%, and 83% in three different experiments. In these experiments, hyperparameters such as a batch size of 16, a learning rate of 2e-5, and epochs with variations of 3, 5, and 7 were used.
3. From the testing results with three different epoch settings, namely 3, 5, and 7, it was found that epoch 3 produced the best results in sentiment analysis. This indicates that having more epochs does not necessarily result in higher accuracy.
4. It can be concluded that the quality and quantity of data in the dataset significantly impact the accuracy achieved when implementing the BERT model. Better dataset quality tends to lead to higher accuracy levels. Additionally, the addition of quality data contributes to improved accuracy in sentiment analysis.

Overall, the experiments and evaluations demonstrate the effectiveness and superiority of the BERT algorithm in sentiment analysis compared to traditional machine learning models.

REFERENCES

[1]. Chen, Hao, et al. "Neural sentiment classification with user and product attention." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.
[2]. Haddi, Elham, et al. "Sentiment analysis techniques and applications: A survey." Journal of Information Science 42.5 (2016): 1-19.
[3]. Tang, Duyu, et al. "Aspect level sentiment classification with deep memory network." Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies. 2016.
[4]. Devlin, Jacob, et al. "BERT: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
[5]. Vaswani, Ashish, et al. "Attention is all you need." Advances in Neural Information Processing Systems. 2017.
[6]. Huang, Lihan, et al. "BERT-based models for sentiment analysis fine-tuning." arXiv preprint arXiv: 2010.16061 (2020).
[7]. Putri, I. S. (2020). Sentiment analysis: A systematic literature review. Journal of

Physics: Conference Series, 1567(4), 042032.
[8]. Kiritchenko, Svetlana, et al. "Sentiment analysis in multiple languages: A cross-lingual review." Journal of Artificial Intelligence Research 69 (2020): 675-722.
[9]. Nacke, L., & Mirza-Babaei, P. (2018). Games User Research: A Case Study Approach. CRC Press.
[10]. Devlin, Jacob, et al. "BERTology meets biology: Interpreting attention in protein language models." NeurIPS Workshop on Machine Learning for Molecules and Materials. 2019.
[11]. Clark, Kevin, et al. "Electra: Pre-training text encoders as discriminators rather than generators." arXiv preprint arXiv: 2003.10555 (2020).
[12]. Arifin, Azhar, et al. "Toward Indonesian text normalization: Building a dataset and neural model." Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC). 2020.
[13]. Wolf, Thomas, et al. "Transformers: State-of-the-art natural language processing." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020.
[14]. Brownlee, Jason. "How to develop word-based neural language models in Python." Machine Learning Mastery, 2019.
[15]. J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, & S. H. Deng. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. Journal of Electronic Science and Technology, 17(1), 26-40.