# Identifying clickbait in online news using deep learning

**Andry Chowanda[1], Nadia[2], Lie Maximilianus Maria Kolbe[1]**
[1]Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[2]Cyber Security Program, Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

## Article Info

## ABSTRACT

Several industries use clickbait techniques as their strategy to increase the number of readers for their news. Some news companies implement catchy headlines and images in their news article links, with the expectation that the readers will be interested in reading the news and click the provided link. The majority of the news is not hoax news. However, the content might not be as grand as the catchy headlines and images provided to the readers. This research aims to explore the classification model using machine learning to identify if the headlines are classified as clickbait in online news. This research explores several machine learning techniques to classify clickbait in online news and comprehensively explain the results. Several popular machine learning techniques were implemented and explored in this research. The results demonstrate that the model trained with fast large margin provides the best accuracy and classification error (90% and 10%, respectively). Moreover, to improve the performance, bidirectional encoder representations from transformers architecture was used to model clickbait in online news. The best BERT model achieved 98.86% in the test accuracy. BERT model requires more time to train (0.9 hour) compared to machine learning (0.4 hour).

*Corresponding Author:*

Andry Chowanda
Department of Computer Science, School of Computer Science, Bina Nusantara University
Jakarta 11480, Indonesia
Email: achowanda@binus.edu

## 1. INTRODUCTION

With the advancement of technology, the content of the internet is exploding from social media to news, from learning content to entertainment media, as well as from information to hoaxes. This digital transformation era and the fourth industrial revolution transformed the news technology from printed to online news. The digitisation of the news industry provides both positive and negative effects to the industries and societies. On the positive side, societies as readers can access the news from anywhere and anytime. The readers also are able to access breaking news immediately without waiting for the paperboy to arrive to hand in the newspaper to them. The readers also can share what they have been read with their friends and colleagues through social media or other media. This led to the negative effects of the digitisation of the news industries, where not every news the readers shared have been authenticated and validated their correctness and truth. Moreover, several industries use clickbait techniques to increase the number of readers for their news. Some of the news company implements catchy headlines and images in their news article links, with the expectation that the readers will be interested in reading the news and in clicking the provided link [1]. The majority of the news is not hoax news. However, the content might not be as grand as the catchy headlines and images provided to the readers.

This research aims to explore the classification model using machine and deep learning to identify if the headlines are classified as clickbait in online news. Several works have been done to classify clickbait from the headline for online news media [1], [2]. However, there are no comprehensive exploration and analysis on which machine and deep learning techniques provide the best results to classify clickbait from the headline for online news media. Hence, this research explores several machine and deep learning techniques to classify clickbait in online news and comprehensively explain the results. Several popular machine learning techniques were implemented and explored in this research: Naive Bayes, generalised linear model (GLM), logistic regression, fast large margin, artificial neural network (ANN), gradient boosted and support vector machine (SVM). The dataset used in this research was downloaded from Kagle, with 32,000 data where 16,000 data were labelled as click baits and 16,000 data labelled as non-clickbait. he results demonstrate that the model trained with fast large margin provides the best accuracy and classification error (90% and 10%, respectively). Moreover, the model was trained in a total of 0.4 hours in a CPU computer. Moreover, to improve the performance of the model, bidirectional encoder representations from transformers (BERT) was implemented to model clickbait classification from online news. The best BERT model achieved 98.85% in the test accuracy. BERT model requires more time to train (0.9 hour) compared to machine learning (0.4 hour). The rest of the paper is organised as follows: the related work regarding clickbait classification are presented in section 2. The methods proposed in this research is thoroughly demonstrated in the section 3 and the results are comprehensively discussed in section 4. Finally, the conclusion and future work direction are presented in section 5.

## 2.    RECENT WORK

Research for classification on clickbait in machine learning is not new topic research. On the contrary, it is has been going on for several years. Some of the researchers have researched the clickbait classification using several methods. According to Zou [3], used multi-classification by applying a self-attentive recurrent neural networks (RNN) mechanism on the hidden states of bidirectional gated recurrent unit (bi-GRU). In addition, the self-attentive applied a token level attention mechanism to infer token importance level in annotation distribution prediction. By applying a self-attentive neural network, manual feature engineering is not involved because it can be trained end-toend, and other external information is not needed. This method could achieve 0.033 mean squared error (MSE), with an F1 score (0.683), accuracy (0.856), and running time (00:03:27) with 80 percent of the labelled dataset sample, which is the best result at clickbait challenge 2017. Anand *et al* [2] mentions that neural network architecture based on RNN for detecting clickbait. They used distributed word representations learned and convolutional neural networks (CNN) for character embedding. So, the experiment evaluates the performance of character embedding, word embedding, and combination with the proposed combination embedding method, resulting in the best with bi-directional long short term memory (biLSTM). The accuracy, precision, and recall are 0.98 with an F1-score of 0.98 and receiver operating characteristics (ROC)-area under the curve (AUC) of 0.99. Although this [2] using RNN as Zhou [3], the result is better because of the proposed combination embedding method. Pujahari and Sisodia [4], tried to build up the detecting clickbait techniques by combining the categorization technique. The hybrid categorization separates clickbait by integrating different features using eleven features, recategorized sentence structure, and clustering using word vector similarity based on the t-stochastic neighbourhood embedding (tSNE) approach. The proposed combination of the technique is proven to be more sturdy with a better result. Pujahari and Sisodia [4] using decision tree (DT), SVM, and random forest (RF) as classification algorithm, and turns out all the accuracy, precision, and recall is better using proposed combination techniques. The best result is the combination of all categorization or clustering features with SVM as the classification algorithm with accuracy 0.97, 0.97 precision, and recall of 0.96.

Papadopoulou *et al.* [5] proposed two-level classification by combining combine the outputs of 65 first-level classifiers with the second-level feature vector in clickbait challenge. Unfortunately, this classification architecture does not perform very well, and it has only got 0.63 for the F-score, 0.91 for precision, and 0.49 for recall. Moreover, the evaluation showed that the experiment did not get a good result, either for text features alone or complex combinations. Omidvar *et al.* [6] tried to propose deep learning methods for clickbait detection. To find the best deep learning architectures, many different kinds of architectures were trained and implemented. The proposed model for clickbait detection [6] was consist of 5 steps, including bi-directional gated recurrent unit (GRU), biLSTM, word embedding vector using GloVe, the combination of forwarding GRU and backward GRU. This proposed model was the best model in terms of mean squared error with 0.03, although its precision was only 0.73. A deep learning framework using natural language cues as in [7] to improve clickbait detection as clickbait has done many bad effects. The trained framework is used for decision-making by classifying headlines for clickbait. Linguistic analysis is part of the speech analysis module, and the decision-making task of classification uses long short-term

memory (LSTM). The result of this framework's architecture is 0.97 of accuracy. Thakur does other research about clickbait detection by using deep learning [8], [9] which suggests the recurrent CNN overcomes the heavy feature engineering in clickbait detection. This method has been tested and turns out that the accuracy is better than the other clickbait-detection algorithms such as LSTM [10]–[12], CNN [8], [13], [14], BERT [15], [16], or conventional machine learning algorithms [17]–[21].

## 3. PROPOSED METHOD

Several popular machine learning techniques were explored in this research. They are: Naive Bayes, GLM, logistic regression, fast large margin, ANN, gradient boosted and SVM. The algorithms were used due to their popularity to deal with several classification problems including natural language processing problems (e.g. click bait classification) [22]–[25]. The dataset used in this research was downloaded from Kagle, with 32,000 data where 16,000 data labeled as click baits and 16,000 data labeled as non click baits. Table 1 and Figure 1 illustrate the profile of the dataset and Table 2 demonstrates the examples of the dataset contents.

Table 1. The database profile

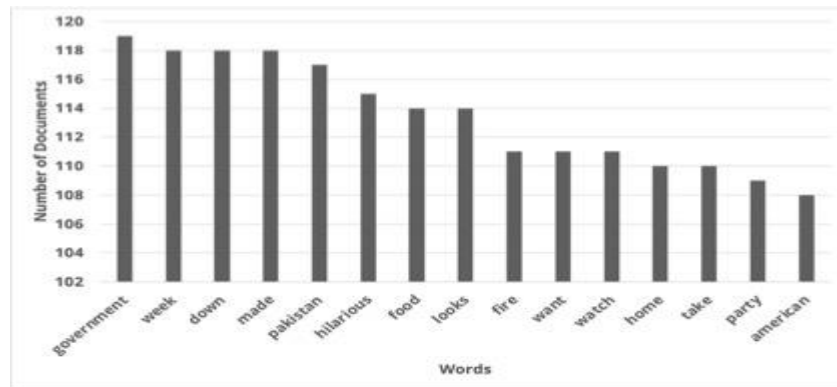| No | Class | Training | Testing | Total |
|---|---|---|---|---|
| 1 | Non-clickbait (0) | 12,800 (80%) | 3,200 (20%) | 16,000 (100%) |
| 2 | Clickbait (1) | 12,800 (80%) | 3,200 (20%) | 16,000 (100%) |



Figure 1. Number of words in documents

Table 2. The database examples

| Examples | Class |
|---|---|
| The new star war the force awakens trailer is here to give you chills | 1 |
| This vine of new york on celebrity big brother is fucking perfect | 1 |
| A couple did a stunning photo shoot with their baby after learning she had an inoperable brain tumor | 1 |
| Natalie dormer and sam claflin play a game to see how they'd actually last in the hunger games | 1 |
| The most Canadian groom ever left his wedding to plow out his guests in a snow storm | 1 |
| Bill changing credit card rules is sent to obama with gun measure included | 0 |
| New year introduces illinois texting while driving ban, among other laws | 0 |
| Tropical storm dolores now active | 0 |
| Invitational games for the deaf, taipei 2008' starts, new slogan for 2009 summer deaflympics unveil | 0 |
| Medtronic paid $788,000 to doctor accused of faking study | 0 |

The headline features were represented by using term frequency–inverse document frequency transform (TF-IDF) vector representation. The vector consists of 1,500 words for the representation. The most words across all the documents are: government, week, down, made, and Pakistan, detected in 119, 118, 118, 118, and 117 documents respectively. The extracted features then trained with naive bayes, GLM, logistic regression, fast large margin, ANN, gradient boosted and SVM. Several hyperparameters of each machine learning algorithms were also explored. In the fast-large margin, the hyperparameter explored was the hyperparameter C, where it indicates the error term's penalty parameter, was explored. The parameter C was explored in six different settings from 0.001 to 100 multiplied by 10 in every phase. In the gradient boosted tree, the hyperparameters explored were the combination of [30; 90; and 150], [2; 4; and 7], and

[0,001; 0,01; and 0,1] for number of trees, max depth of each tree, and the learning rate respectively. Moreover, for SVM, the hyperparameters explored were the combination of Gamma (γ) ([0,005; 0,05; 0,5; and 5]) and the cost parameter of C ([0,1; 01, 10, 100; and 1,000]).

Finally, seven layers of ANNs (one input layer, one output layer and five hidden layers) were explored in this research. The input layer consists of 1,500 words vector with rectifier activation units and 0.00001 L1 regularisation to prevent over fitting problem. The hidden layers have 200 nodes in each layer with rectifier activation units and 0.00001 L1 regularisation. The output layer consists of two nodes (0 and 1) with softmax activation units and 0.00001 L1 regularisation. Moreover, four different settings of BERT [15] was implemented to model clickbait classification from online news. The hyper-parameters of all BERT model are: batch size=32, maximum epochs=10, initial learning rate=0.001 with early stopping and learning rate reduction applied. The early stopping and learning rate reduction set Patience level of 10, minimum delta of 0.001 and factor for reduction=0.1. The first and second BERT model implements "BERT-base-uncased" pre-trained model. The first BERT model set the maximum of the input length to 128, while the second BERT model set the maximum of the input length to 64. The third and fourth BERT model implements "bert-large-uncased" pre-trained model. The third BERT model set the maximum of the input length to 128, while the fourth BERT model set the maximum of the input length to 64. The training time for all models trained with machine learning and deep learning were recorded to be evaluated.

## 4. RESULTS AND DISCUSSION

Nine machine learning algorithms and several hyperparameters settings combinations have been explored in this research to build the best classifier to detect clickbait from online news. The algorithms are naive bayes, GLM, logistic regression, fast large margin, ANN (vanilla deep learning), gradient boosted and SVM. Table 3 demonstrates the overview of results trained by the algorithms. The best accuracy and classification error was achieved by the model trained with fast large margin algorithms, resulting in 90%, 10%, ±0.2%, and 0.4 hours of testing accuracy, testing classification error, testing standard deviation and training time, respectively. The second-best result was achieved by the model trained with the GLM, resulting in 89%, 11%, ±0.4%, and 0.6 hours of testing accuracy, testing classification error, testing standard deviation and training time, respectively. Moreover, the third-best model was the one trained with seven layers ANNs with 87%, 13%, ±0.42% and 1.6 hours of testing accuracy, testing classification error, testing standard deviation and training time, respectively. The worst result was achieved by the model trained with DT, resulting in 50%, 50%, ±0.1% and 0.3 hours of testing accuracy, testing classification error, testing standard deviation, and training time, respectively. Despite being the worst model, the models trained with tree-based algorithms (i.e. DT, gradient boosted tree, and RF) provide the best standard deviation results with ±0.1 of standard deviation. Moreover, the fastest training time was achieved by the DT algorithm with only 0.3 hours of training time in a CPU powered computer.

Table 3. Machine learning results

| No | Algorithm | Accuracy (%) | Error (%) | SD (%) | Time (h) |
|----|-----------|--------------|-----------|--------|----------|
| 1 | Naive Bayes | 58 | 42 | ±0.4 | 0.5 |
| 2 | Generalised linear model | 89 | 11 | ±0.4 | 0.6 |
| 3 | Logistic regression | 82 | 18 | ±0.3 | 0.4 |
| 4 | Fast large margin | 90 | 10 | ±0.2 | 0.4 |
| 5 | Artificial neural network | 87 | 13 | ±0.2 | 1.6 |
| 6 | Random forest | 66 | 34 | ±0.1 | 0.8 |
| 7 | Gradient boosted trees | 78 | 22 | ±0.1 | 0.8 |
| 8 | Decision tree | 50 | 50 | ±0.1 | 0.3 |
| 9 | Support vector machine | 77 | 23 | ±0.2 | 0.6 |

Table 4 demonstrates the confusion matrix of the best model trained by fast large margin. The best model achieved the values of 90.5% and 89.6% of testing precision for class 0 (non-clickbait) and class 1 (clickbait), respectively. Moreover, the model also achieved the values of 89.4% and 90.7% of testing recall for class 0 (non-clickbait) and class 1 (clickbait), respectively. More detailed results are shown in Table 5. The best model achieved 90.05%, 9.95%, 96.5%, 90.05%, 90.05%, 90.05%, 85.69%, and 93.96% for the training accuracy, classification error, AUC, precision, recall, F1 score, sensitivity and specificity, respectively, with relatively low standard deviation values for each performance metric. Moreover, Table 6 illustrates the overview of the models trained by deep learning architecture (i.e. BERT). The model trained with bert-base-uncased pre-trained model and maximum of input length of 128 achieved the accuracy score of 98.82%, loss of 3.6789e-04, and require 0.9 hour of training time. The model trained with bert-base-

uncased pre-trained model and maximum of input length of 64 achieved the accuracy score of 98.86%, loss of 2.9029e-04 and require 0.9 hour of training time. This model is the best model trained compared to the other BERT models. Moreover, the model trained with bert-large-uncased pre-trained model and maximum of input length of 128 achieved the accuracy score of 98.62%, loss of 6.1851e-04, and require 1.3 hour of training time. The model trained with bert-large-uncased pre-trained model and maximum of input length of 64 achieved the accuracy score of 98.64%, loss of 5.2386e-04, and require 1.3 hour of training time. There are no significant differences between the models trained with maximum input of 64 and 128.

Table 4. Fast large margin confusion matrix

|        | True 0 | True 1 | Precision (%) |
|--------|--------|--------|---------------|
| Pred 0 | 2,862  | 299    | 90.5          |
| Pred 1 | 338    | 2,901  | 89.6          |
| Recall | 89.4%  | 90.7%  |               |

Table 5. Fast large margin detailed results

| Performance metrics  | Value/testing set (%) | Standard deviation |
|----------------------|-----------------------|--------------------|
| Accuracy             | 90.05                 | 0.002              |
| Classification error | 9.95                  | 0.002              |
| AUC                  | 96.5                  | 0.002              |
| Precision            | 90.05                 | 0.005              |
| Recall               | 90.05                 | 0.008              |
| F1 score             | 90.05                 | 0.003              |
| Sensitivity          | 85.69                 | 0.008              |
| Specificity          | 93.96                 | 0.006              |

Table 6. Deep learning results

| No | Algorithm     | Accuracy (%) | Loss       |
|----|---------------|--------------|------------|
| 1  | BERT base-128 | 98.82        | 3.6789e-04 |
| 2  | BERT base-64  | 98.86        | 2.9029e-04 |
| 3  | BERT large-128| 98.62        | 6.1851e-04 |
| 4  | BERT large-64 | 98.64        | 5.2386e-04 |

Figure 2 shows the details of training and validation accuracy as shown in Figure 2(a) and loss in Figure 2(b) of the model trained with BERT-base with maximum length of input of 64 (the best model). The result demonstrates that the accuracy and loss of the training and validation did not improved after the third epochs. The best training accuracy achieved by the model was 100% and the best validation accuracy achieved by the model was 98.96%. Table 7 demonstrates the details results of the model trained with BERT-base with maximum length of input of 64 (the best model). Both classes (clickbait and true) demonstrates balanced results of precision, recall, and F1-score. The macro and weighted average score of precision, recall and F1-score are all 98.86%. Figure 3 illustrates confusion matrix of the model trained with BERT-base with maximum length of input of 64 (the best model).



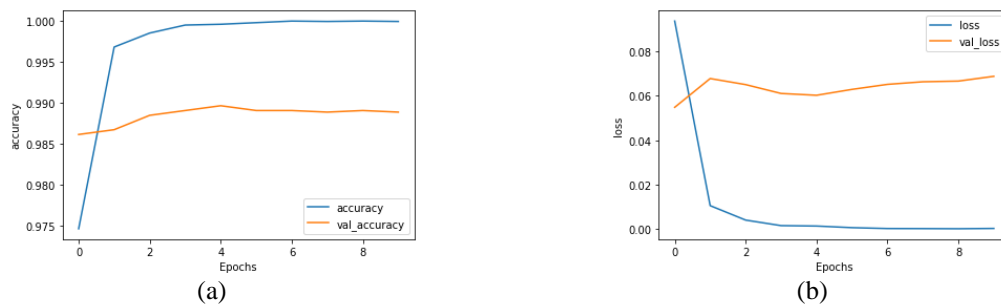(a)                              (b)

Figure 2. Best model using BERT (a) training and validation accuracy and (b) loss

Table 7. Deep learning results

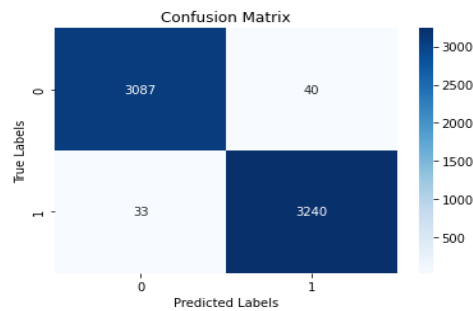|              | Precision | Recall | F1-Score |
|--------------|-----------|--------|----------|
| 0            | 0.9894    | 0.9872 | 0.9883   |
| 1            | 0.9889    | 0.9899 | 0.9889   |
| Accuracy     | -         | -      | 0.9886   |
| Macro AVG    | 0.9886    | 0.9886 | 0.9886   |
| Weighted AVG | 0.9886    | 0.9886 | 0.9886   |

Figure 3. Confusion matrix best model using BERT

## 5.    CONCLUSION AND FUTURE NETWORK

Several machine learning techniques have been deployed to build the best classifier to detect clickbait from online news. Several hyper-parameters settings were also explored to find the best model for the clickbait classifier. The results demonstrate that the model trained with fast large margin provides the best accuracy and classification error (90% and 10%, respectively). The model was trained for 0.4 hours using a CPU, resulting in ± 0.2% of standard deviation. The second best results were achieved by the model trained with the GLM with the accuracy of 89% and classification error of 11% and standard deviation of ±0.4%. The third best model was achieved by training seven layers of ANNs. It performs the accuracy of 87%, classification error of 13% and standard deviation of ±0.2%. Finally, the best standard deviation was achieved by the models trained by tree-based models (i.e. RF, gradient boosted trees, and DT) with ±0.1 of standard deviation. Moreover, the fastest model trained in this research with the selected dataset was the decision tree (DT) model with ±0.3 hours of training times. Moreover, to improve the performance, BERT architecture was used to model clickbait in online news. The best BERT model achieved 98.86% in the test accuracy. BERT model requires more time to train (0.9 hour) compared to machine learning (0.4 hour).

Other deep learning architectures can provide better results for future research directions. Some architectures such as CNN, RNN such as LSTM, or the combination of the architectures. Some attention models also can be explored in future work. Moreover, the transformer architecture can also be implemented to improve the results of classifying clickbait from online news. More datasets can be added or augmented from the newest online sources to provide more power to the deep learning architectures. More references can be added to give more variation to the data. Moreover, local contents can be added to the dataset. Finally, some combination of features representation (e.g. the embedding) can be explored to provide better representations for the clickbait classifier training.

## REFERENCES

[1]    A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly, "Stop clickbait: detecting and preventing clickbaits in online news media," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 9–16, doi: 10.1109/ASONAM.2016.7752207.

[2]    A. Anand, T. Chakraborty, and N. Park, "We used neural networks to detect clickbaits: you won't believe what happened next!," in *Advances in Information Retrieval*, Cham: Springer, 2017, pp. 541–547, doi: 10.1007/978-3-319-56608-5_46.

[3]    Y. Zhou, "Clickbait detection in tweets using self-attentive network," *Arxiv-Computer Science*, pp. 1–5, 2017, doi: 10.48550/arXiv.1710.05364.

[4]    A. Pujahari and D. S. Sisodia, "Clickbait detection using multiple categorisation techniques," *Journal of Information Science*, vol. 47, no. 1, pp. 118–128, 2021, doi: 10.1177/0165551519871822.

[5]    O. Papadopoulou, M. Zampoglou, S. Papadopoulos, and I. Kompatsiaris, "A two-level classification approach for detecting clickbait posts using text-based features," *Arxiv-Computer Science*, pp. 1–5, 2017, doi: 10.48550/arXiv.1710.08528.

[6]    A. Omidvar, H. Jiang, and A. An, "Using neural network for identifying clickbaits in online news media," in *Information Management and Big Data*, Cham: Springer, 2019, pp. 220–232, doi: 10.1007/978-3-030-11680-4_22.

[7]    B. Naeem, A. Khan, M. O. Beg, and H. Mujtaba, "A deep learning framework for clickbait detection on social area network using natural language cues," *Journal of Computational Social Science*, vol. 3, no. 1, pp. 231–243, 2020, doi: 10.1007/s42001-020-00063-y.

[8]    D. S. Thakur and S. Kurhade, "Context-based clickbait identification using deep learning," in *2021 International Conference on Communication information and Computing Technology (ICCICT)*, 2021, pp. 1–5, doi: 10.1109/ICCICT50803.2021.9510141.

[9]    P. Mowar, M. Jain, R. Goel, and D. K. Vishwakarma, "Clickbait in YouTube: prevention, detection and analysis of the bait using ensemble learning," *Arxiv-Computer Science*, pp. 1–26, 2021.

[10]   N. Bhoj, A. R. Dwivedi, A. Tripathi, and B. Pandey, "LSTM powered identification of clickbait content on entertainment and news websites," in *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2021, pp. 29–33, doi: 10.1109/CICN51697.2021.9574677.

[11]   M. Marreddy, S. R. Oota, L. S. Vakada, V. C. Chinni, and R. Mamidi, "Clickbait detection in telugu: overcoming NLP challenges in resource-poor languages using benchmarked techniques," in *2021 International Joint Conference on Neural Networks*

(IJCNN), 2021, pp. 1–8, doi: 10.1109/IJCNN52387.2021.9534382.

[12]  B. Siregar, I. Habibie, E. B. Nababan, and Fahmi, "Identification of Indonesian clickbait news headlines with long short-term memory recurrent neural network algorithm," *Journal of Physics: Conference Series*, vol. 1882, no. 1, pp. 1–8, 2021, doi: 10.1088/1742-6596/1882/1/012129.

[13]  N. Kaothanthong, S. Kongyoung, and T. Theeramunkong, "Headline2Vec: a CNN-based feature for thai clickbait headlines classification," *International Scientific Journal of Engineering and Technology (Isjet)*, vol. 5, no. 1, pp. 20–31, 2021.

[14]  B. Gamage, A. Labib, A. Joomun, C. H. Lim, and K. Wong, "Baitradar: a multi-model clickbait detection algorithm using deep learning," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2665–2669, doi: 10.1109/ICASSP39728.2021.9414424.

[15]  T. Liu, K. Yu, L. Wang, X. Zhang, H. Zhou, and X. Wu, "Clickbait detection on WeChat: a deep model integrating semantic and syntactic information," *Knowledge-Based Systems*, vol. 245, p. 108605, 2022, doi: 10.1016/j.knosys.2022.108605.

[16]  J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Arxiv-Computer Science*, pp. 1–16, 2019.

[17]  M. Jain, P. Mowar, R. Goel, and D. K. Vishwakarma, "Clickbait in social media: detection and analysis of the bait," in *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, 2021, pp. 1–6, doi: 10.1109/CISS50987.2021.9400293.

[18]  S. R. Dam, S. P. Panday, and T. B. Thapa, "Detecting clickbaits on nepali news using SVM and RF," in *Proceedings of 9th IOE Graduate Conferenc*, 2021, vol. 9, pp. 140–146.

[19]  Vincent, S. Regina, K. Purwandari and F. I. Kurniadi, "Clickbait Headline Detection Using Supervised Learning Method," *2022 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, 2022, pp. 408-412, doi: 10.1109/IoTaIS56727.2022.9975866.

[20]  R. Gothankar, F. D. Troia and M. Stamp, "Clickbait Detection for YouTube Videos," *In Cybersecurity for Artificial Intelligence. Springer*, Cham, pp. 261-284, 2021.

[21]  R. Agarwal, S. Gupta, and N. Chatterjee, "Profiling Fake News Spreaders on Twitter: A Clickbait and Linguistic Feature Based Scheme," *In International Conference on Applications of Natural Language to Information Systems, Springer*, Cham, pp. 345-357, 2022.

[22]  A. Chowanda, R. Sutoyo, and S. Tanachutiwat, "Exploring text-based emotions recognition machine learning techniques on social media conversation," *Procedia Computer Science*, vol. 179, pp. 821-828, 2021, doi: 10.1016/j.procs.2021.01.099.

[23]  A. Chowanda and Y. Muliono, "Emotions Classification from Speech with Deep Learning," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022.

[24]  K. Sailunaz, M. Dhaliwal, J. Rokne, R. Alhajj, "Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, vol. 8, no. 1, pp. 1-26, 2018, doi: 10.1007/s13278-018-0505-2.

[25]  N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowledge and Information Systems*, vol. 62, no. 8, pp. 2937-2987, doi: 10.1007/s10115-020-01449-0.

## BIOGRAPHIES OF AUTHORS

**Andry Chowanda** 🆔 📊 SC 🔵 is a Computer Science lecturer in Bina Nusantara University Indonesia. His research is in agent architecture and machine (and deep) learning. His work mainly on how to model an agent that has capability to sense and perceive the environment and react based on the perceived data in addition to the ability of building a social relationship with the user overtime. In addition, he is also interested in serious game and gamification design. He can be contacted at email: achowanda@binus.edu.

**Nadia** 🆔 📊 SC 🔵 is a cyber security lecturer in Bina Nusantara University. Her undergraduate and master degree from Bina Nusantara Universtity. Her main research is in the field of natural language processing and cyber security. She can be contacted at email: nadia002@binus.ac.id.

**Lie Maximilianus Maria Kolbe** 🆔 📊 SC 🔵 is a computer science lecturer in Bina Nusantara University. His undergraduate and master degree from Bina Nusantara Universtity. His main research is in the field of natural language processing, machine learning and competitive programming. He can be contacted at email: lie.kolbe@binus.ac.id.