International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

https://doi.org/10.47001/IRJIET/2024.811015

Sentiment Analysis of Imbalanced Sarcastic Flood Disaster Texts Using Deep Learning Models

[1]*Nur Khamidah, [2]Khairil Anwar Notodiputro, [3]Sachnaz Desta Oktarina

[1]Student, Department of Statistics and Data Science, IPB University, Bogor, Indonesia
[2]Lecturer, Department of Statistics and Data Science, IPB University, Bogor, Indonesia
[3]Professor Lecturer, Department of Statistics and Data Science, IPB University, Bogor, Indonesia
*Corresponding Author's E-mail: nur.khamidah@apps.ipb.ac.id

***Abstract* - Sentiment analysis often faces challenges like manual labeling, sarcasm detection, and imbalanced class labels. Using Twitter/X data for sentiment analysis is resource-intensive due to manual labeling. The BERT model is adequate for Indonesian sentiment analysis, but sarcasm remains challenging. This research evaluates the performance of BERT, LSTM, and BERT-LSTM models for classifying sarcastic text data, specifically in flood-related posts from Indonesia. We used Twitter/X data from December 19, 2023, to January 13, 2024, labeled by three annotators. We handle imbalanced data using techniques like Random Undersampling, SMOTE, and SMOTETomek. We assessed model performance with ANOVA based on balance-weighted accuracy. The BERT and BERT-LSTM models excelled, achieving balance-weighted accuracy values of 98.61% and 98.06%, respectively. This research advances sentiment analysis methods, particularly for natural disaster contexts in Indonesia.**

Keywords: sentiment, flood, sarcasm, imbalance data, deep learning model.

I. INTRODUCTION

Indonesia ranks 12th out of 35 countries at high risk of fatalities and economic losses due to various disasters [1]. Floods are the most common natural disaster, with a total of 788 reported cases in 2021 [2], 1,524 cases in 2022 [3], and 989 cases in 2023 [4]. When floods occur, the term ''flood'' (or mentioned as ''banjir'' in Bahasa Indonesia) often becomes the most discussed topic on social media platforms such as X or Twitter. These platforms have a ''trends for you'' feature that captures information about users' most frequently discussed topics or words within a certain period [5], [6].

The advancement of technology has fueled interest in developing natural language processing (NLP) techniques related to text data analysis. One application is sentiment analysis, which falls under supervised learning and requires labeled data [7]. Previous research using X data (formerly Twitter) for sentiment analysis still involved manual labeling in their method [8], [9]. The challenge of manual labeling is the time and resources it consumes, often resulting in class imbalance issues, where documents addressing specific issues/topics tend to dominate one class. Imbalanced data is a common problem in classification analysis, especially with large datasets. This issue is more pronounced in sentiment classification because it often involves data skewed towards one class [10]. If not adequately addressed, it can lead to inaccurate classification results, with the model tending to classify the minority class into the majority class, resulting in unstable classification outcomes [11]. Methods like the Synthetic Minority Over-sampling Technique (SMOTE) [11], [12], [13], its development SMOTEBoost[14], and ExtraTrees[15] provide more stable classification results for handling imbalanced data.

Long Short-Term Memory (LSTM) is an artificial neural network architecture developed to process and model data with temporal or sequential dependencies. In natural language processing and sentiment analysis, the LSTM architecture is advantageous because it can remember and process information related to previous contexts in a sentence or document and classify observations into specific categories or sentiments. LSTM can remember long-term information (long-term memory) and short-term information (short-term memory) simultaneously [16]. Previous algorithms, Recurrent Neural Networks (RNNs), had limitations in storing information with long-term dependencies, leading to the LSTM cells' proposal [17]. This method improves the memory capacity of the cells by repeatedly applying gates within the cells. Since its development, LSTM has been modified and popularized by many researchers [18], [19]. We use the LSTM algorithm for sentiment classification, and several studies have shown that LSTM performs better and can be combined with various other deep learning algorithms to improve classification accuracy [20], [21], [22], [23], [24].

Another challenge in semantic text analysis is sarcasm, which is difficult to analyze automatically, even for humans, and can potentially lead to misinterpretation of text sentiment.

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

https://doi.org/10.47001/IRJIET/2024.811015

Determining sentiment in sarcastic sentences remains a significant challenge in text processing [25], [26]. Yunitasari et al., in their research [27] found that sentiment analysis using the Naïve Bayes algorithm achieved an accuracy of 74.91% without sarcasm detection, which increased by 5.49% with sarcasm detection. Additionally, Muhaddisi et al. (2021) achieved sentiment analysis accuracy of 60% with Random Forest and 71% with Double Random Forest when detecting sarcasm in Indonesian text data.

Google's AI Language Lab introduced the BERT language model (Bidirectional Encoder Representations from Transformers) under J. Devlin's leadership in 2018 [26]. IndoBERT is a development of BERT created for the Indonesian language, pre-trained on more than 220 million words. BERT can understand natural language like humans using a bidirectional deep learning approach. Compared to Logistic Regression, Naïve Bayes, BiLSTM, and multilingual BERT models, IndoBERT achieves the highest accuracy of 84.13% when used for sentiment analysis in the Indonesian language [29].

Referring to the issues explained, this research aims to evaluate the performance of BERT, LSTM, and BERT-LSTM models in classifying text data containing sarcasm. Additionally, it seeks to compare and obtain the best model among the three for classifying sentiment in data with imbalanced class labels, explicitly focusing on tweets/posts related to flood disasters in Indonesia.

II. METHOD

2.1 Simulation Data Generation

The simulation study in this research aims to evaluate the performance of classification models in recognizing text data that contains sarcasm. As explained in the previous section, sarcastic text data consists of favorable terms but conveys an opposite sentiment. We suspect that sarcasm interferes with the performance of sentiment classification models, so it is necessary to understand how well the model can recognize sarcasm in text data before predicting the sentiment.

This research's mechanism for generating simulation data involves preprocessing and exploring the empirical text data obtained. Then, we generate new text data regarding flood disasters containing sarcasm using ChatGPT 3.5, with a corpus matched to the empirical data. This generation intends to produce simulation data with corpus distribution and characteristics like the empirical data under study. The detailed steps for simulation data generation appear in Figure 1.

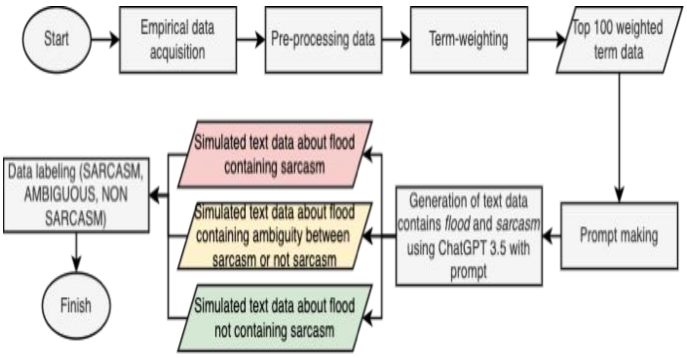


Figure 1: Simulation data generation procedure

Next, from the above steps, a total of 180 sentences were obtained with the following details in Figure 2.

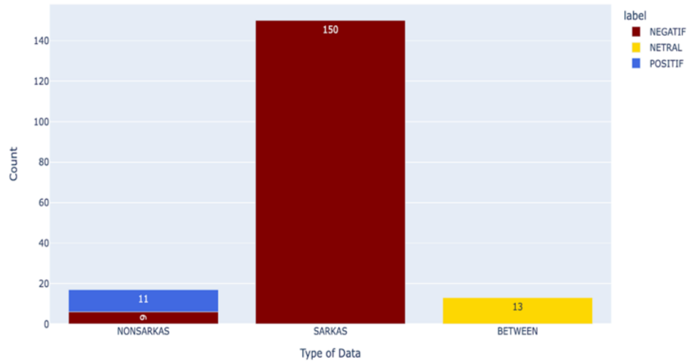


Figure 2: Distribution frequency of simulation text data

2.2 Empirical Data Acquisition

The empirical data in this study consists of data obtained by crawling from the X/Twitter social media platform using the "requests" package in Visual Studio Code version 1.85.1 with Python programming language version 3.10.6. We extract the tweets using facilities provided by the X API with a Basic package limited to 10,000 posts per subscription. The data consists of posts shared by X/Twitter social media users from December 19, 2023, to January 13, 2024, totaling 10,479 posts. The query used appears in Table 1 as follows:

Table 1: Parameters and values that are used to crawl the empirical data

Parameter	Value
Query	banjirlang:id -is:retweet -is:reply -has:mentions -has:links
Tweet Fields	created_at,lang,text,geo,author_id,id,public_metrics,referenced_tweets
Expansions	geo.place_id,author_id
Place Fields	contained_within,country,country_code,full_name,geo,id,name,place_type
User Fields	description,username,id
Start Time	2023-12-19T00:01:01.000Z
End Time	2024-01-13T23:59:59.000Z

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

https://doi.org/10.47001/IRJIET/2024.811015

This study focuses on supervised learning-based machine learning methods, so we manually label the obtained data. After collecting data, we get 1,582 tweets relevant to the previous research objectives. Subsequently, we performed labeling by three annotators/labelers from three different fields (Indonesian Language and Literature (labeled as BHS), Statistics (labeled as STA), and Psychology (labeled as PSI), totaling nine annotators. This process examines how different annotator fields could influence classification performance, which we will test in the results section.

2.3 Data Analysis Procedure

Data processing for analysis in this study was conducted through the Google Chrome browser using the cloud-based platform Google Colaboratory and the Python 3.10.12 programming language. We carry out the data analysis using three types of processors, including the CPU for data preparation, NVIDIA GPU T4 for lightweight analysis, as well as NVIDIA GPU A100 and NVIDIA GPU V100 for running deep learn-ing models that require high computation. Some modules used in this research include Transformers (4.40.0), Tensorflow (2.15.0), Keras (2.15.0), Imbalanced-learn (0.10.1), and Tf-keras (2.15.1). Further-more, the research procedure can be seen more clearly in Figure 3.

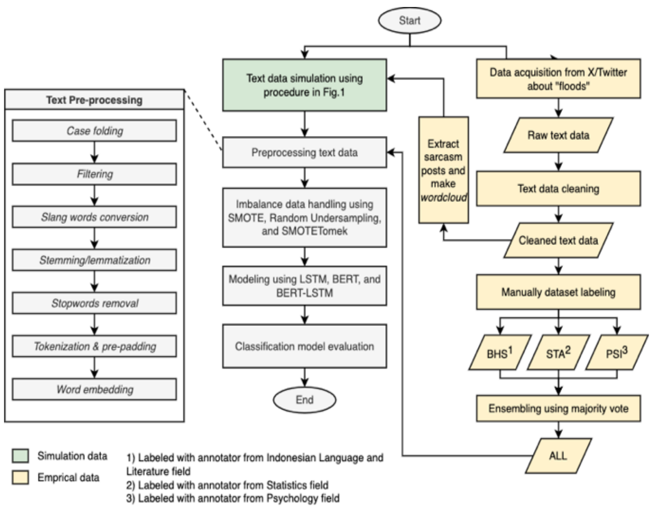


Figure 3: Flowchart of the research procedure

III. RESULTS AND DISCUSSION

3.1 Simulation Study

This simulation study tests the classification models' capability to predict the sentiment labels/classes from the generated simulation data. Previously, we generated the data based on a corpus containing sarcasm obtained from empirical data. Based on exploration using a word cloud, simulation data with word distribution and its comparison with empirical data were obtained as follows.

(i)

(ii)

Figure 4: Comparison of word cloud results from both data types: (i) simulation data and (ii) empirical data on posts containing sarcasm

Figure 4 shows that in the generated simulation data, the distribution of words used is like the words/terms selected by X/Twitter users in expressing sarcastic flood disasters. Furthermore, we perform the three models (BERT, LSTM, and BERT-LSTM) with and without handling imbalanced classes to observe how the models perform in predicting text data containing sarcasm. From the results of the simulation study conducted, the classification performance on the simulation data is as follows in Table 2.

Table 2: The classification result based on a balanced weighted accuracy score on simulation data

Imbalance	NONE		RUS		SMOTE		SMOTETomek	
Model	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
LSTM	0.8417	0.0525	0.4889	0.3495	0.6722	0.2644	0.6750	0.3181
BERT	**0.9861**	0.0270	0.4000	0.3394	0.9417	0.0607	0.9278	0.1097
BERT-LSTM	0.9778	0.0410	**0.8750**	0.0851	**0.9806**	0.0348	**0.9444**	0.0752

The results of modeling using three types of models (BERT, LSTM, and a hybrid BERT-LSTM model) and four types of imbalanced data handling scenarios (NONE/not handled, SMOTE, RUS/Random Undersampling, and

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

https://doi.org/10.47001/IRJIET/2024.811015

SMOTETomek/hybrid) in Table 2 above show that overall, the BERT-LSTM model produced the best performance in predicting sarcasm, with the highest accuracy achieved across all imbalanced data handling techniques. In the LSTM model, using imbalanced data handling techniques did not improve the model's performance in detecting sarcasm. In the BERT model, using imbalanced data handling techniques with Random Undersampling did not perform as well as the other techniques because the Random Undersampling technique involved data elimination, indicating that the lack of data affects the model's performance. According to research by Ezen-Can (2020), the BERT model used on small-sized datasets provides unstable and overfitted performance. Furthermore, the BERT-LSTM/hybrid model detects sarcasm well with all imbalanced data handling techniques. As shown in Table 2, we can conclude that the hybrid model performs excellently in detecting sarcasm across all created data scenarios. Additionally, based on execution time, the BERT-LSTM method with SMOTETomek averaged 31.808 seconds per epoch, which is shorter than SMOTE, which took 33.086 seconds per epoch. This result indicates that data generated using the SMOTETomek/hybrid method for handling imbalanced data undergoes synthesis followed by the elimination of duplicate data, resulting in a smaller dataset compared to data handled with SMOTE, which only involves the synthesis process without elimination.

Performance classification evaluations were conducted on each model using analysis of variance (ANOVA) and post hoc multiple comparisons using Tukey's Honestly Significant Difference (HSD) test. We assess the modeling results of the simulation data using a two-way analysis of variance (two-way ANOVA) with the obtained analysis of variance table as follows in Table 3.

Table 3: The classification result based on a balanced weighted accuracy score on simulation data

Source of variance	df	F score	p-value
Imbalance	3	31.388	$1.14 \times 10^{-14*}$
Model	2	62.342	$< 2 \times 10^{-16*}$
Imbalance: Model	6	4.847	0.000203^{*}

*) Significant at the 5% level of significance

From the analysis of the balance-weighted accuracy values in Table 3, the test obtained that based on a significance level of 5%, there is a significant difference in the balance-weighted accuracy values influenced by the differences in classification models, imbalanced data handling techniques, and their combinations. This result indicates that in the simulated data, the choice of model to predict sarcasm in each imbalanced data handling technique and vice versa significantly affects the balance-weighted accuracy values produced at a 5% significance level.

Next, we conduct a post hoc analysis using Tukey's Honestly Significant Difference (HSD) Test to compare the selection of imbalanced data handling techniques for each LSTM, BERT, and BERT-LSTM model based on the balance-weighted accuracy values obtained when predicting sarcasm. Based on the analysis of the variance table above, the post hoc analysis results of the balance weighted accuracy values for each imbalanced data handling technique and classification model appear in Figure 5.

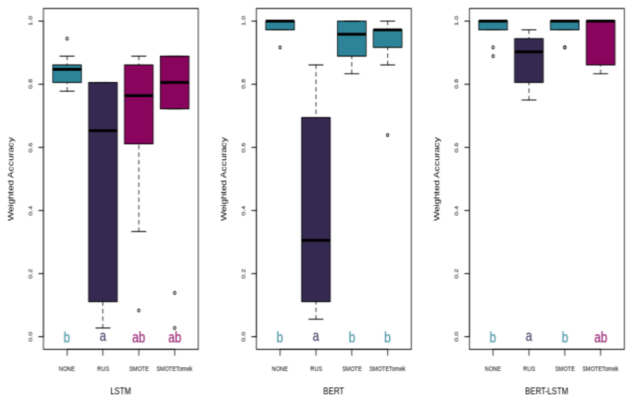


Figure 5: Post-hoc analysis result on simulation data from each model

The results in Figure 5 indicate that for the LSTM and BERT models, selecting the Random Undersampling technique for imbalanced data handling resulted in different balance-weighted accuracy values compared to not applying any handling, where no handling yielded better and more stable balance-weighted accuracy predictions. Meanwhile, there was no difference in the balance weighted accuracy values when no imbalanced data handling technique was applied compared to the SMOTE and SMOTETomek techniques, whether for the LSTM, BERT, or BERT-LSTM models in predicting sarcasm.

3.2 Empirical Study

3.2.1 Data Exploration

We conduct data exploration to examine the characteristics of the text data. In this study, we also generated a word cloud to visualize the distribution of words used by users in their uploaded posts and performed hashtag extraction. Hashtags consist of words or phrases used periodically in several uploaded posts as markers and are generally related to specific topics or identities.

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

https://doi.org/10.47001/IRJIET/2024.811015

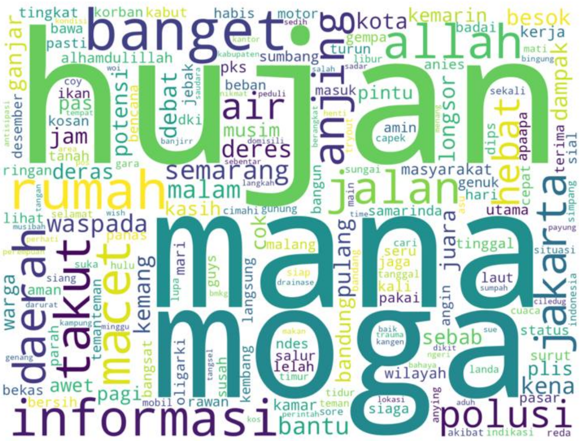


Figure 6: Word cloud of the empirical data

From the word cloud in Figure 6, it can present that some commonly used words by people to express floods on X/Twitter include “hujan” (rain), “moga” (hope), “takut” (fear), “informasi” (information), “rumah” (home), and others. Several geographical locations or city names frequently appear, including Kemang, Jakarta, Semarang, Bandung, Malang, and Tangsel (Tangerang Selatan). This result indicates that based on the empirical data obtained, the online community indicates and widely discusses floods occurring in these areas. Furthermore, some emotional terms expressed by people on X/Twitter regarding floods include “takut” (fear), “anjing” (dog, kind of curse way), “allah,” “alhamdulillah,” “anying” (kind of curse word), “cok” (damn), and others.

Next, we conduct the topic modeling using Latent Dirichlet Allocation (LDA), with an initial parameter tuning to find the optimal number of topics based on the log-likelihood score. The optimal number of topics (n) obtained was 2 with the highest log-likelihood value. From these results, we conduct further topic modeling, and the topics obtained appear in Table 4.

Table 4: Topics extracted and the corpus distribution of each topic

No	CorpusDistribution	Topic
1	‘‘banjir’’, ‘‘hujan’’, ‘‘ujan’’, ‘‘banget’’, ‘‘rumah’’, ‘‘tahun’’, ‘‘baru’’, ‘‘terus’’, ‘‘semoga’’, ‘‘jalan’’, ‘‘macet’’, ‘‘dimana’’, ‘‘jadi’’, ‘‘jakarta’’, ‘‘hari’’, ‘‘malam’’, ‘‘takut’’, ‘‘allah’’, ‘‘pulang’’, ‘‘kena’’	Traffic congestion due to floods d uring th enew year
2	‘‘banjir’’, ‘‘untuk’’, ‘‘pks’’, ‘‘kota’’, ‘‘deres’’, ‘‘warga’’, ‘‘air’’, ‘‘semoga’’, ‘‘bantuan’’, ‘‘bencana’’, ‘‘masyarakat’’, ‘‘banget’’, ‘‘terdampak’’, ‘‘alhamdulillah’’, ‘‘deh’’, ‘‘waspada’’, ‘‘memberikan’’, ‘‘kayak’’, ‘‘potensi’’, ‘‘hujannya’’	Community assistance for flood-affected areas

3.2.2 Imbalance Data Handling

This study applies three types of imbalanced data handling methods for comparison: oversampling with SMOTE, undersampling with Random Undersampling, and hybrid with SMOTETomek. Before this, visualization was conducted for each dataset to observe the class proportions differences in each dataset in Figure 7.

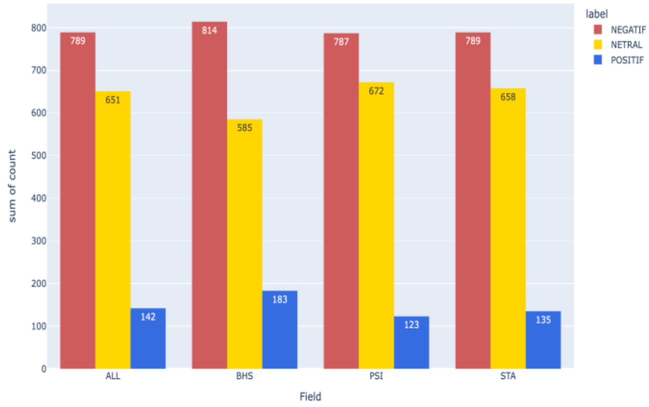


Figure 7: Distribution of the class proportion in each data

Each dataset exhibits class proportion differences, which trigger data imbalance issues. Therefore, in this study, data imbalance handling will be performed before modeling proceeds.

3.2.3 Model Evaluation

The modeling results using three types of models (LSTM, BERT, and BERT-LSTM) on four types of imbalanced data handling (no handling, SMOTE, Random Undersampling, and SMOTETomek), as well as on four types of datasets (BHS, STA, PSI, and ALL), yielded the following balance weighted accuracy values as shown in Table 5.

Table 5: The results of balanced accuracy of all models and datasets

Imbalance	Imbalance	ALL		BHS		STA		PSI	
		Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
NONE	LSTM	0.749	0.087	0.737	0.043	0.742	0.088	0.775	0.028
	BERT	**0.944**	0.049	**0.925**	0.076	**0.945**	0.058	**0.945**	0.050
	BERT-LSTM	0.918	0.056	0.917	0.076	0.938	0.056	0.922	0.072
SMOTE	LSTM	0.570	0.206	0.545	0.187	0.680	0.116	0.579	0.271
	BERT	**0.905**	0.073	**0.878**	0.082	**0.912**	0.063	0.891	0.112
	BERT-LSTM	0.893	0.079	0.854	0.096	0.871	0.073	**0.900**	0.104
RUS	LSTM	0.550	0.156	0.520	0.134	0.495	0.056	0.509	0.165
	BERT	0.826	0.059	**0.858**	0.072	**0.831**	0.077	**0.826**	0.065
	BERT-LSTM	**0.835**	0.072	0.809	0.058	0.795	0.083	0.810	0.058
SMOTE-Tomek	LSTM	0.535	0.274	0.501	0.280	0.260	0.290	0.579	0.271
	BERT	**0.889**	0.080	**0.875**	0.095	**0.927**	0.069	**0.903**	0.054
	BERT-LSTM	0.885	0.073	0.851	0.093	0.907	0.070	0.886	0.082

Based on Table 5, we can observe that there is relatively no difference in the classification performance results among the ensemble dataset ALL, BHS, STA, and PSI based on the balance-weighted accuracy values. Regarding the LSTM model, the best balance-weighted accuracy values were obtained across all datasets when no imbalanced data handling technique was applied, as it resulted in the most stable balance-weighted accuracy values. The most stable balance-

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

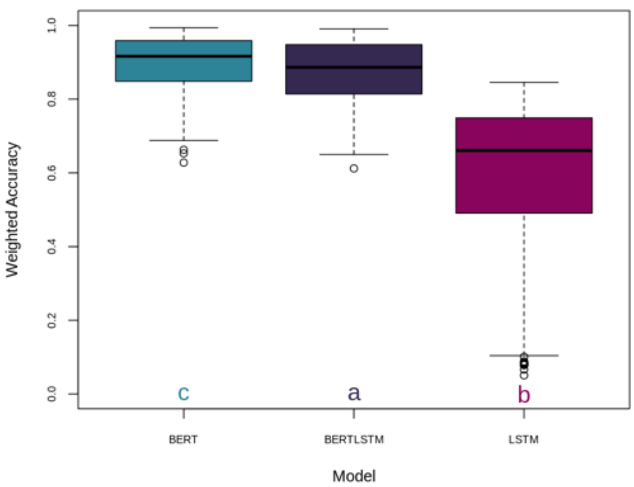
ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

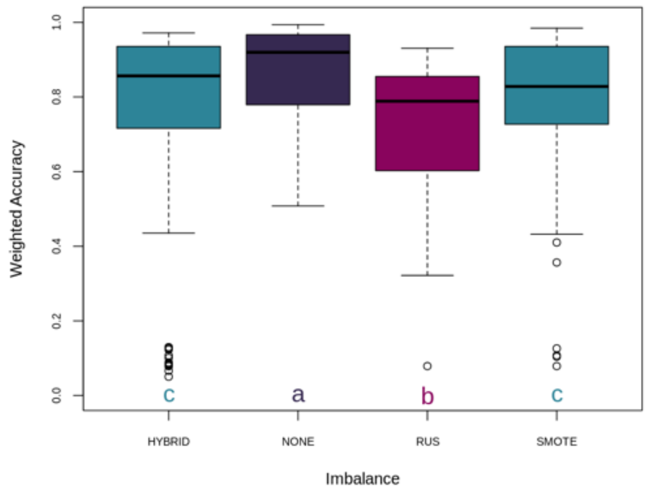
https://doi.org/10.47001/IRJIET/2024.811015

weighted accuracy values are found in the STA dataset when we apply the LSTM model without imbalanced data handling. Furthermore, when we apply the BERT model across all datasets, it is observed that not applying imbalanced data handling techniques and applying SMOTETomek yield the best and most stable balance weighted accuracy values. Additionally, the SMOTETomek technique on the BERT-LSTM model has a shorter runtime, averaging 65.121 seconds per epoch compared to SMOTE, which takes 111.4656 seconds per epoch. Thus, with the same model, SMOTETomek is nearly twice as fast to execute compared to SMOTE. This result is consistent with the simulation results. Meanwhile, for the BERT-LSTM model, on the ensemble (ALL) dataset, applying Random Undersampling, and on the PSI dataset, applying SMOTE results in the best balance weighted accuracy. When compared based on BERT and BERT-LSTM models, they generally yield relatively better and stable balance-weighted accuracy values compared to the LSTM model. This result indicates that on empirical data with more varied linguistic features, since this research uses data from X/Twitter posts, BERT and BERT-LSTM perform better in predicting the sentiment of text data due to BERT's performance in recognizing the semantic contexts [29].

We conducted a variance analysis of balance-weighted accuracy values on the empirical data result in Table 5, and we obtained that there is a significant difference in balance-weighted accuracy values influenced by the variance in classification models and the variance in imbalanced data handling techniques. However, balance-weighted accuracy values have no significant difference based on the interaction effect and annotator field. This result indicates that the selection of models and imbalanced data handling techniques significantly affect the balance-weighted accuracy values at a 5% significance level in empirical data. Based on the analysis of the variance table, the post hoc test results of balance-weighted accuracy values appear in Figure 8.



(i)



(ii)

Figure 8: Post-hoc testing results of empirical data based on the influence of (i) classification model selection and (ii) imbalanced data handling techniques

The results in Figure 8 indicate that statistically, there are differences in the weighted accuracy values obtained from the test data when applying each classification model. Each LSTM, BERT, and BERT-LSTM model provides different predictive weighted accuracy values, where the LSTM model yields the lowest weighted accuracy compared to the other two models. Furthermore, based on imbalanced data handling techniques, there is no difference in the predictive weighted accuracy values obtained when using the SMOTE and SMOTETomek imbalanced data handling techniques, but they both are different in execution time. These results are relatively consistent with the simulation study findings discussed earlier.

IV. CONCLUSION

The simulation study results indicate that the BERT and BERT-LSTM models better classify sarcastic text data with high and stable balance-weighted accuracy. Additionally, using SMOTE and SMOTETomek techniques in the simulation data can enhance model performance, considering the relatively small dataset size. Overall, the BERT-LSTM model applied to simulation data exhibits a more stable performance than the other two methods. Regarding imbalanced data handling techniques, we can observe that Random Undersampling does not perform better than no handling or applying SMOTE and SMOTETomek techniques. This result is because Random Undersampling eliminates data, resulting in a smaller training dataset and suboptimal deep-learning modeling in this study. However, based on the stability of the results and model execution time, the SMOTETomek technique is recommended over SMOTE, as outlined in the previous section.

© 2024-2017 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 8, Issue 11, pp 150-158, November-2024
https://doi.org/10.47001/IRJIET/2024.811015

The analysis of empirical data yields results like those of the simulation study. The empirical study finds that the BERT and BERT-LSTM models achieve better balanced weighted accuracy values even without handling imbalanced data. Furthermore, the BERT and BERT-LSTM models perform relatively better than the LSTM models. This study suggests that the presence of BERT and BERT-LSTM models can produce accurate classification even with imbalanced data.

Exploratively, X/Twitter users express opinions about flood disasters using words that match the conveyed sentiment. As expressions of negative sentiment, positive connotation words are also frequently used when expressing the opposite sentiment. Some geographical information is also commonly used by the public to express opinions about flood disasters occurring in specific areas.

Recommendations that we can suggest for this study include developing models using additional hyperparameter tuning, such as dropout proportions in the model, which can improve model performance on test data and address overfitting generated by the model. The data obtained in this study mainly consists of data that is not related to the discussed research context but contains keywords relevant to the research objectives. It is expected that future research will develop more advanced models related to detecting posts that are not relevant to the research context more effectively before conducting sentiment classification modeling.

ACKNOWLEDGEMENT

We would like to acknowledge the Indonesia Education Fund for Education (LPDP) from the Ministry of Finance for granting the scholarship and supporting this research.

REFERENCES

[1] World Bank, “Indonesia - Vulnerability | Climate Change Knowledge Portal.” Accessed: Jun. 10, 2023. [Online]. Available: https://climateknowledgeportal.worldbank.org/country/indonesia/

[2] Badan Pusat Statistik, “BanyaknyaDesa/KelurahanMenurutJenisBencana Alam dalamTigaTahunTerakhir (Desa), 2021.” Accessed: Jun. 10, 2023. [Online]. Available: https://www.bps.go.id/indicator/168/954/1/banyaknya-desa-kelurahan-menurut-jenis-bencana-alam-dalam-tiga-tahun-terakhir.html

[3] Pusdatinkom Badan Nasional PenanggulanganBencana (BNPB), “StatistikBencanaMenurutJenis.” Accessed: Jun. 10, 2023. [Online]. Available: https://dibi.bnpb.go.id/kbencana2

[4] Pusdatinkom Badan Nasional PenanggulanganBencana (BNPB), “StatistikBencanaMenurutJenis.”

[5] R. N. Amalia, K. Sadik, and K. A. Notodiputro, “A Preliminary Study of Sentiment Analysis on COVID-19 News: Lesson Learned from Data Acquisition, Pre-processing, and Descriptive Analytics,” *BAREKENG: JurnalIlmuMatematika dan Terapan*, vol. 17, no. 4, pp. 1901–1914, Dec. 2023, doi: 10.30598/barekengvol17iss4pp1901-1914.

[6] C. Oktarina, K. A. Notodiputro, and Indahwati, “Kajian PerbandinganMetode K-Means dan K-Medoids untukMenggerombolkan Data Twitter,” IPB University, 2020. Accessed: May 24, 2023. [Online]. Available: https://repository.ipb.ac.id/handle/123456789/102886

[7] A. E. S. Saputro, K. A. Notodiputro, and I. A, “Study of Sentiment of Governor’s Election Opinion in 2018,” *Int J Sci Res Sci Eng Technol*, pp. 231–238, Dec. 2018, doi: 10.32628/ijsrset21841124.

[8] A. Ghifari, K. A. Notodiputro, and B. Sartono, “AnalisisPopularitas IPB berdasarkan Data Twitter Menggunakan Awan Kata, Geovisualisasi, dan PemodelanKlasifikasi,” IPB University, 2018. Accessed: May 24, 2023. [Online]. Available: https://repository.ipb.ac.id/handle/123456789/94386

[9] F. T. Saputra, Y. Nurhadryani, S. H. Wijaya, and Defina, “AnalisisSentimen Bahasa Indonesia pada Twitter MenggunakanStruktur Tree BerbasisLeksikon,” IPB University, 2020. Accessed: May 24, 2023. [Online]. Available: https://repository.ipb.ac.id/handle/123456789/105583

[10] S. Li, G. Zhou, Z. Wang, S. Y. M. Lee, and R. Wang, “Imbalanced Sentiment Classification,” in *International Conference on Information and Knowledge Management.*, Association for Computing Machinery, 2011.

[11] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, “Classification of Imbalanced Data:Review of Methods and Applications,” *IOP Conf Ser Mater Sci Eng*, vol. 1099, no. 1, p. 012077, 2021, doi: 10.1088/1757-899x/1099/1/012077.

[12] S. D. Oktarina, H. Wijayanto, and H. R. Yarah, “Low Birth Weight Classification With Synthetic Minority Over Sampling Technique Random Forest,” *Jurnal Kesehatan Ibu dan Anak*, vol. 17, no. 1, pp. 46–56, Oct. 2023, doi: 10.29238/kia.v17i1.1802.

[13] N. E. Zendrato, B. Sartono, and U. D. Syafitri, “Identifikasi Karakteristik Rumah Tangga Penerima Bantuan Sosial Menggunakan Metode Feature Importance Model Berbasis Pohon Klasifikasi,” IPB University, 2022. Accessed: May 24, 2023. [Online].

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

https://doi.org/10.47001/IRJIET/2024.811015

Available: https://repository.ipb.ac.id/handle/123456789/113984

[14] R. Kurnia, B. Sartono, and I. M. Sumertajaya, "Classifier Learning For Imbalanced Dataset Using Modified SMOTEBoost Algorithm And Its Application On Credit Scorecard Modeling," IPB University, 2013. Accessed: May 24, 2023. [Online]. Available: https://repository.ipb.ac.id/handle/123456789/67070

[15] I. Tamara, B. Sartono, and A. Kurnia, "Kajian Kinerja AlgoritmeKlasifikasi Extra-Trees pada Permasalahan Data Kelas Tak Seimbang," IPB University, 2022. Accessed: May 24, 2023. [Online]. Available: https://repository.ipb.ac.id/handle/123456789/113230

[16] I. Augenstein, T. Rocktäschel, A. Vlachos, and K. Bontcheva, "Stance Detection with Bidirectional Conditional Encoding," in *Proceedings ofthe 2016 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2016, pp. 876–885.

[17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[18] F. Gers, "Long Short-Term Memory in Recurrent Neural Networks," 2001, doi: 10.5075/epfl-thesis-2366.

[19] F. A. Gers and J. Schmidhuber, "Recurrent Nets that Time and Count," *IEEE*, 2000.

[20] K. Barik, S. Misra, A. K. Ray, and A. Bokolo, "LSTM-DGWO-Based Sentiment Analysis Framework for Analyzing Online Customer Reviews," *ComputIntellNeurosci*, vol. 2023, pp. 1–19, 2023, doi: 10.1155/2023/6348831.

[21] M. N. Moghadasi, M. Sc, Y. Zhuang, V. Sheng, S. Yu, and M. Sheridan, "Sentimental Semantic Classification using Decomposed LSTM over Big Data," Texas Tech University, 2020.

[22] W. J. Murdoch, P. J. Liu, and B. Yu, "Beyond Word Importance: Contextual Decomposition to Extract Interactions from LSTMs," *ICLR*, 2018, [Online]. Available: http://arxiv.org/abs/1801.05453

[23] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis," *Multimed Tools Appl*, vol. 78, no. 18, pp. 26597–26613, 2019, doi: 10.1007/s11042-019-07788-7.

[24] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.

[25] L. Septiani and Y. Sibaroni, "Sentiment Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia," *Jurnal Linguistik Komputasional*, vol. 2, no. 2, pp. 62–67, 2019, [Online]. Available: https://twitter.com/hipwee/status/734249192273645568?y

[26] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings ofNAACL-HLT*, 2019, pp. 4171–4186. [Online]. Available: https://github.com/tensorflow/tensor2tensor

[27] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 1, p. 53, Jan. 2019, doi: 10.22146/ijccs.41136.

[28] A. Muhaddisi, B. N. Prastowo, and D. U. Kusumaning Putri, "Sentiment Analysis With Sarcasm Detection On Politician's Instagram," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 4, p. 349, Oct. 2021, doi: 10.22146/ijccs.66375.

[29] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *arXiv preprint*, 2020, [Online]. Available: http://arxiv.org/abs/2011.00677

AUTHORS BIOGRAPHY

Nur Khamidah completed his undergraduate studies in the Department of Statistics at Brawijaya University in 2021 with a thesis titled "Comparison of Classification Accuracy using CART (Classification and Regression Tree) and Logistic Regression on ILI (Influenza-Like Illness) Patients at Dr. Soetomo General Hospital, Surabaya". Before pursuing his master's degree, he worked as an IT Developer at a multinational private company for approximately six months. He is currently undertaking his master's studies in the Department of Statistics at IPB University, specializing in Statistics and Data Science, with a scholarship from the Ministry of Finance of the Republic of Indonesia. His areas of interest include big data analytics, computational statistics, and text analysis. She can be contacted at e-mail: nur.khamidah@apps.ipb.ac.id or nurkhamidah.sta@gmail.com

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 11, pp 150-158, November-2024

https://doi.org/10.47001/IRJIET/2024.811015

Khairil Anwar Notodiputro earned his bachelor's and master's degrees from IPB University. He obtained his Ph.D. from Macquarie University, Australia. Currently, he is a professor in the Department of Statistics at IPB University and serves as a visiting professor at Songkla University, Thailand. He holds a functional position as the head of the doctoral program in the Department of Statistics at IPB University. Throughout his academic career, he has held various positions in scientific organizations, mainly in the field of statistics. These roles include Head of the Department of Statistics, Director of the Preparatory Level Program, Head of the Information Resource Institute, and Dean of the Graduate School at IPB University. Outside of IPB University, he has served as Head of the Research and Development Agency of the Ministry of Education and Culture of the Republic of Indonesia, Chairman of the Indonesian Statistician Association (Ikatan Statistisi Indonesia/ISI), and Chairman of the Supervisory Board at Hasanuddin University. He can be contacted at e-mail: khairilnotodiputro@gmail.com

Sachnaz Desta Oktarina completed his undergraduate studies in the Department of Statistics at IPB University in 2010. He then pursued his master's and doctoral studies at the University of Tsukuba, Japan, specializing in Agrobio Resources Science & Technology for his master's and Appropriate Technology and Sciences for Sustainable Development for his doctorate. He completed his doctoral studies in 2017 with a dissertation titled “Community-Driven Development and Poverty Alleviation: An Economic Evaluation of the National Program for Community Empowerment in Indonesia.” Currently, he is a lecturer at the Department of Statistics at IPB University, teaching courses in statistics and data analysis, statistical optimization, statistical analysis for health sciences, data exploration and visualization, and statistical methods. She can be contacted at e-mail: sachnazdes@apps.ipb.ac.id

Citation of this Article:

Nur Khamidah, Khairil Anwar Notodiputro, & Sachnaz Desta Oktarina. (2024). Sentiment Analysis of Imbalanced Sarcastic Flood Disaster Texts Using Deep Learning Models. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(11), 150-158. Article DOI https://doi.org/10.47001/IRJIET/2024.811015

****