



ARTICLE

Feature-Based Augmentation in Sarcasm Detection Using Reverse Generative Adversarial Network

Derwin Suhartono^{1,*}, Alif Tri Handoyo¹ and Franz Adeta Junior²

¹Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

²Cyber Security Program, Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

*Corresponding Author: Derwin Suhartono. Email: dsuhartono@binus.edu

Received: 23 August 2023 Accepted: 27 November 2023 Published: 26 December 2023

ABSTRACT

Sarcasm detection in text data is an increasingly vital area of research due to the prevalence of sarcastic content in online communication. This study addresses challenges associated with small datasets and class imbalances in sarcasm detection by employing comprehensive data pre-processing and Generative Adversarial Network (GAN) based augmentation on diverse datasets, including iSarcasm, SemEval-18, and Ghosh. This research offers a novel pipeline for augmenting sarcasm data with Reverse Generative Adversarial Network (RGAN). The proposed RGAN method works by inverting labels between original and synthetic data during the training process. This inversion of labels provides feedback to the generator for generating high-quality data closely resembling the original distribution. Notably, the proposed RGAN model exhibits performance on par with standard GAN, showcasing its robust efficacy in augmenting text data. The exploration of various datasets highlights the nuanced impact of augmentation on model performance, with cautionary insights into maintaining a delicate balance between synthetic and original data. The methodological framework encompasses comprehensive data pre-processing and GAN-based augmentation, with a meticulous comparison against Natural Language Processing Augmentation (NLPAug) as an alternative augmentation technique. Overall, the F1-score of our proposed technique outperforms that of the synonym replacement augmentation technique using NLPAug. The increase in F1-score in experiments using RGAN ranged from 0.066% to 1.054%, and the use of standard GAN resulted in a 2.88% increase in F1-score. The proposed RGAN model outperformed the NLPAug method and demonstrated comparable performance to standard GAN, emphasizing its efficacy in text data augmentation.

KEYWORDS

Data augmentation; Generative Adversarial Network (GAN); Reverse GAN (RGAN); sarcasm detection

1 Introduction

Interpretation of a statement is crucial to determine the results of the analysis. Meanwhile, the results of the proper analysis based on data lead to the right action. Currently, there is an abundance of information being shared on social media platforms in the form of statements, thoughts, or comments. These expressions encompass both positive and negative sentiments. However, it is within this spectrum



of statements that negative sentiments are occasionally veiled through the use of sarcasm. Sarcastic remarks, by nature, contain an implied message, rendering them more challenging to decipher.

Sarcasm, as defined, is a form of negative sentiment concealed within seemingly pleasant sentences [1]. Recent studies have further categorized sarcasm as an aggressive variant of irony used to convey unfavorable messages [2]. It is often intertwined with various forms of irony [3]. Sarcasm can manifest through both verbal and textual communication. Verbal sarcasm carries distinct characteristics such as volume, speaking tempo, tone of voice, and accompanying gestures, making it relatively discernible [1]. Conversely, textual sarcasm, commonly encountered on social media and product/service reviews, presents a more formidable challenge due to the absence of these contextual cues [4].

Over the past five to ten years, the research landscape has witnessed a notable surge in studies pertaining to sarcasm detection [5]. This surge underscores the pivotal role sarcasm detection plays in facilitating well-informed decision-making through the interpretation of sarcastic expressions. Fig. 1 provides an overview of the trends in sarcasm detection research spanning from 2010 to 2022.

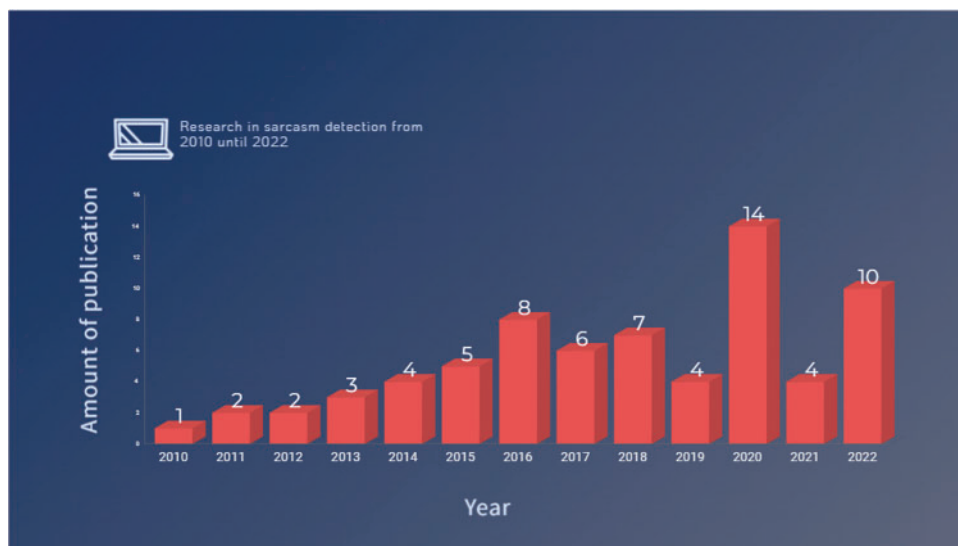


Figure 1: Sarcasm trend research from 2010 to 2022

While previous research efforts have employed a spectrum of methodologies, these endeavors predominantly fall within two overarching domains: machine learning and deep learning. Machine learning-based approaches have been explored utilizing techniques such as the Support Vector Machine (SVM) [6], Lexical influence [7], and the ensemble method of SVM, K-Nearest Neighbor (KNN), and decision tree [8].

However, traditional machine learning approaches have exhibited limitations when confronted with sarcastic statements carrying implicit messages, as they struggle to contextualize the sentence as a whole. This necessitated a transition towards deep learning methods. Subsequently, research has embraced a deep learning paradigm for sarcasm classification, incorporating techniques such as multi-layer perceptrons [9] and hybrid neural networks that combine Convolutional Neural Networks (CNN) and bidirectional Long Short Term Memory (LSTM) architectures [10]. While these endeavors have primarily focused on model development, this research endeavors to bridge the gap by exploring and developing augmentation techniques tailored specifically for sarcasm data.

Apart from advancing deep learning model methodologies, this research acknowledges the significance of data augmentation in enhancing a model's classification capability. Existing research has explored a range of data augmentation techniques to improve model performance in sarcasm detection. However, one relatively uncharted avenue within the realm of sarcasm text augmentation is the application of Generative Adversarial Networks (GANs). GAN-based augmentation has yielded satisfactory results in image-processing domains such as medical imaging [11], face detection [12], and agriculture [13]. Nevertheless, its potential in sarcasm text augmentation remains underexplored.

Inspired by the success of GANs in augmenting datasets, this research introduces a novel framework employing the Reverse Generative Adversarial Neural Network (RGAN) technique. This framework aims to enhance the accuracy of deep learning models in sarcasm detection. The fundamental premise of RGAN involves reversing the labels of genuine and synthetic data. This reversal encourages the generator to produce data closely resembling real data while challenging the discriminator to develop a more comprehensive understanding of subtle distinctions between authentic and synthetic data.

In summary, the contributions of this research encompass:

- The author's proposed framework introduces a novel approach for enhancing sarcastic data through the utilization of a Reverse Generative Adversarial Network (RGAN). The purpose of reversing the labels of actual and fake data is to encourage the generator to produce data that closely resembles real data while simultaneously pushing the discriminator to develop a more comprehensive understanding of the subtle differences between real and fake data.
- The research involved the execution of tests and subsequent analysis to provide evidence supporting the effectiveness of data augmentation through the use of RGAN in enhancing the model's ability to differentiate between sarcastic and non-sarcastic texts. This was compared to the alternative methods of synonym replacement in NLPAug and the traditional GAN method. Then, This research also analyses the distribution of data generated from GAN-based models.
- Performed RGAN testing on balanced and unbalanced datasets. Tests were conducted with 4 augmentation scenarios on each dataset based on percentages of 15%, 30%, and 45% and adjusting the number of data additions with the highest class. To analyze GAN's efficiency further, this research also analyses the distribution of data generated from GAN-based models.

The remainder of this paper is structured as follows: [Section 2](#) reviews previous research on sarcasm detection and augmentation techniques in sarcasm datasets. [Section 3](#) goes over the datasets used, pre-processing techniques, proposed models, and experimental methods. [Session 4](#) explains the data generated by GAN as well as the experimental results. Finally, in [Session 5](#), the conclusions of this research are discussed.

2 Related Works

In this session, we will discuss previous research in detecting sarcasm and augmentations used in text data. The summary of previous research shown in [Table 1](#) shows that sarcasm sentence research tends to explore less data augmentation.

Table 1: Summary of previous sarcasm detection research

Title	Approach	Algorithm	Accuracy	Dataset	Augmentation
Sarcasm detection in tweets: A feature-based approach using supervised machine learning models [14]	Machine learning	Decision tree	91.84% accuracy	Tweet that contain #sarcasm (37583 sarcastic sentence and 39216 non-sarcastic)	No
Ensemble classification approach for sarcasm detection [8]	Machine learning	Ensemble (SVM, LR, DT)	98.37% accuracy	Tweet that contain #sarcasm	No
Detection on sarcasm using machine learning classifiers and rule based approach [15]	Machine learning	Random forest	76% accuracy	SemEval 2018 T3	No
C-Net: Contextual network for sarcasm detection [16]	Deep learning	C-Net (Based on Bert Model)	75% F1-score	Twitter and reddit dataset	No
Sarcasm detection using multi-head attention based bidirectional LSTM [17]	Deep learning	LSTM with multihead attention	77.48% F1-score	Private dataset (57000 sarcasm and 57000 non-sarcastic)	No
Sarcasm detection using hybrid neural network [10]	Deep learning	LSTM-CNN combination	89.7% accuracy	Headline dataset (11725 sarcastic and 14984 non-sarcastic)	No
Plumeria at SemEval-2022 task 6: Robust approaches for sarcasm detection for English and Arabic using transformers and data augmentation [18]	Deep learning	BERT	87% highest accuracy using words repetition augmentation	Private dataset from Twitter (6570 for sarcastic in English language and 4502 from Arabic language)	External dataset, word embedding (token replacement), repeating words

(Continued)

Table 1 (continued)

Title	Approach	Algorithm	Accuracy	Dataset	Augmentation
Sarcasm detection in Twitter—Performance impact while using data augmentation: word embeddings [19]	Deep learning	RoBERTa	40.44% F1-score	iSarcasm, Ghosh, Ptacek, SemEval-18 dataset	NLPAug
GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples [20]	Deep learning	BERT	62.67% F1-score	Multi-Genre Natural Language Inference (MNLI) dataset	GAN based

2.1 Sarcasm Detection with Machine Learning

Previous studies have explored the detection of sarcasm through the utilization of multiple machine learning models, which are combined utilizing ensemble learning techniques [8]. The dataset utilized in this study was sourced from the Twitter social media platform, comprising instances that were classified as either sarcasm or non-sarcasm. Ensemble learning encompasses various combinations of models. In a general sense, an ensemble learning approach that incorporates Support Vector Machines (SVM), Logistic Regression (LR), and Decision Trees (DT), utilizing a voting system to determine class prediction, demonstrates superior average accuracy performance when evaluated on five distinct datasets, surpassing alternative ensemble models. The Principal Component Analysis (PCA) algorithm for dimension reduction is utilized in order to represent numerous features through a decreased feature dimension in the experiments. The ensemble learning of Support Vector Machines (SVM), Linear Discriminant Analysis (LD), and Decision Trees (DT) achieved an accuracy of 98.37% on the evaluated dataset. While Support Vector Machines (SVM), Logistic Regression (LR), and Decision Trees (DT) are capable of identifying the relationship between words in a phrase, it should be noted that typical LR models are not specifically designed to handle sequential data, such as text. Traditional machine learning (ML) methods also exhibit inadequate capability in comprehending context, resulting in a failure to grasp semantic links between words. This phenomenon has the potential to result in misclassification or the occurrence of false positives.

In the same year, a study conducted by Godara et al. [8] yielded findings that were consistent with the prior research. Nevertheless, the current study does not employ an ensemble learning methodology and conducts the classification procedure separately for each model [14]. The dataset was obtained by utilizing an Application Programming Interface (API) provided by Twitter. Specifically, comments containing the hashtag #sarcasm were selected, resulting in a dataset comprising a total of 76,799 tweets. The experimental findings indicate that the Decision Tree algorithm achieves the highest level

of accuracy, specifically 91.84%. This outcome is attributed to the utilization of the sarcastic feature set, which comprises various linguistic elements associated with sarcasm, including question marks, exclamation marks, and repeated ellipses. The feature set for sarcasm detection includes both positive and negative sentiment data as additional evidence for identifying sarcastic sentences.

Previous studies have employed a rule-based approach to identify ironic statements, a specific form of sarcasm [15]. The dataset employed in this study is derived from SemEval 2018-T3, which consists of phrases that exhibit irony. The SVM, Naive Bayes, Decision Tree, and Random Forest algorithms are enhanced by the utilization of rule-based lexical and semantic techniques. These techniques serve to eliminate irrelevant words and assess the level of sarcasm, thus improving the ability to recognize contextual information within a phrase. The Random Forest algorithm yields the most accurate results.

2.2 Sarcasm Detection with Deep Learning

There have been studies to detect sarcasm in texts. Recognizing the form of sarcasm in a text is very useful for analyzing customer satisfaction and providing the right steps for making business decisions. However, detecting sarcasm remains difficult, particularly in understanding the context of a sarcastic sentence. There are researchers who use multi-head attention on bidirectional LSTM to detect sarcasm [17]. The private dataset utilized is a collection of comments that include quotation marks, exclamation points, and a mix of question marks and exclamation points. These characteristics are thought to be able to aid the model in identifying the context of sarcasm. Meanwhile, the bidirectional LSTM has forward and backward modes to capture all of the information from a sentence. With multi-head attention, which gives each word a different weight, it is possible to understand the relationship between complex words [21]. Compared to SVM and bidirectional without multi-head attention in this study, multi-head attention mixed with bidirectional LSTM performs better because it can capture word context. Despite the use of an attention mechanism in this research, the model has an accuracy of less than 80%. Regardless of the fact that the dataset used has features such as the number of quotes, exclamation marks, question marks, ellipses, and interjections, this research does not investigate dataset augmentation.

Using C-Net, there is research on how to categorize sarcasm and non-sarcasm [16]. C-Net is composed of many Bidirectional Encoder Representation from Transformers (BERT) models that are trained independently on the response data, the last sentence of the context set, the second last sentence of context, and the first sentence of context, and are then integrated at the fusion layer. BERT [22] is a highly effective natural language understanding model. BERT can read sentences from left to right and vice versa in order to better understand the context of the sentence as a whole. Some words from the dataset will be masked during the tokenization process, and the model will make efforts to guess these words based on the unmasked words. Aside from that, BERT can predict subsequent sentences. The C-Net model experiment makes use of dialog-sentence-formatted datasets from Twitter and Reddit. Sentences in the text are marked sequentially using timestamping. Overall, this study contrasts traditional machine learning with a transformer-based approach. According to the results, the transformer model outperforms all traditional machine learning methods, with F1-scores of 75% and 66.3% on the Twitter and Reddit datasets, respectively. Similarly to previous studies by [17], the research did not investigate dataset augmentation. The dataset used is quite small, with less than 10,000 data points for Twitter and Reddit.

Research on sarcasm detection using a hybrid neural network consisting of CNN and Bidirectional LSTM with an attention module has been carried out [10]. CNN can benefit from input encoded from LSTM by spotting n-gram word patterns. Due to the weighting provided by the attention module, the

model can then better understand the context of a word. Using the attention module, CNN hybrid architecture and Bidirectional LSTM can detect incongruity in a sentence. The test accuracy obtained by comparing the baseline model and the intended method was 84.88% and 89.7%, respectively. Model development can significantly improve accuracy, but the quality of the model is not solely determined by the architecture. This research does not show the pre-processing side of the dataset used, and no dataset augmentation is explored.

2.3 Data Augmentation for Sarcasm Detection

Prior studies on the detection of sarcastic sentences have employed various techniques, such as augmenting existing sarcastic datasets by the incorporation of external datasets, utilizing word embedding methodologies, and employing sentence repetition [18]. The external dataset utilized is sourced from the Twitter social media network, as well as the SemEval-18 and ArSarcasm-v2 datasets. In contrast, the primary dataset included in this study is a proprietary dataset comprising sarcastic statements in both English and Arabic languages, with a total of 6570 and 4502 instances, respectively. The process of external dataset augmentation involves merging the original dataset with an external dataset in order to introduce a significant level of variability. Word embedding is a technique employed to substitute words within a sentence with synonymous alternatives. In order to get a balanced distribution of sardonic and non-sarcastic classes, the technique of repeating words is employed to replicate the same case. The accuracy of validation and tests has shown a significant disparity as a consequence of conducting experiments on three distinct augmentation techniques. Word embedding is a technique employed to substitute words within a sentence with their corresponding synonyms. In order to achieve a fair distribution of sarcastic and non-sarcastic classes, the technique of repeating sentences is employed to replicate instances of the same content. The accuracy of validation and tests has shown a significant disparity as a consequence of conducting experiments on three distinct augmentation techniques. Among the numerous experiments undertaken, it was seen that the BERT model, when augmented with the repetition of words, achieved a validation accuracy of 0.92. Additionally, when the model was tested using pre-processing techniques that involved converting emojis to strings, a test accuracy of 0.87 was obtained. Nevertheless, there are still symptoms of overfitting present in the model due to the possibility of the augmentation technique being excessively highlighted as a result of the frequent repetition of phrases. In contrast, the external dataset augmentation exhibited notable performance, achieving the highest validation and test accuracy scores of 0.41 and 0.07, respectively. Ultimately, the technique of synonym replacement augmentation demonstrated superior performance, achieving scores of 0.86 and 0.84, respectively. Excessive variance in the external dataset, as well as an overemphasis on certain elements, such as repeated words, might lead to overfitting of the model or a decline in its performance.

The model's capacity to recognize sarcasm may be influenced by the data augmentation of sarcastic sentences [23]. Generative Adversarial Network (GAN) is a potential method for augmenting data. Common applications of GANs in the field of image augmentation include the generation of synthetic data with high levels of similarity to the original data. In this approach, synthetic data can be utilized to expand the range of the original dataset [24]. GAN technique paired with BERT is another method for performing data augmentation for text datasets [20]. Both labeled and unlabeled data are sent to BERT as input for vectorization. Meanwhile, the GAN generator reproduces false data derived from random distribution noise. The discriminator's job is to distinguish between authentic and false data. Training continues until the discriminator is unable to distinguish between genuine and fraudulent data that has been reproduced by the generator. GAN-BERT was tested on two datasets: Stanford Sentiment Treebank with 5 different classes (SST-5) for sentiment analysis and Multi-Genre

Natural Language Inference (MNLI) for natural language inference. GAN-results BERT's improve accuracy by 8.2% on the SST-5 sentiment analysis dataset. There is evidence that using a smaller proportion of labeled data is more beneficial when using GAN-BERT. However, no tests of the fully labeled dataset in cases of sarcasm detection with more complex characteristics have been conducted.

Inverting class labels is another GAN technique [25]. Image data was used in the research. Typically, GAN trains the discriminator to distinguish between real and fake data and requires the generator to produce data that is as close to the original data as possible [26]. Reference [25], however, attempted to reverse the labels so that the discriminator can be viewed as a classifier that learns features from the original data. GANs that perform label inversion can learn more than just the difference between real and fake data. The research was successful in demonstrating another point of view through the use of GANs, but this technique still produces unstable results and has a chance of success only in certain cases.

Another research on data augmentation in sarcasm detection using the synonym replacement and duplication methods with NLPAug was done by [19]. The F1-score was evaluated using BERT, Robustly Optimized BERT Approach (RoBERTa) [27], and DistilBERT [28]. RoBERTa is a BERT-based model that has no next-sentence prediction (NSP) to predict a subsequent sentence prediction. Meanwhile, DistilBERT is a BERT model with smaller parameter values that is faster than BERT but has lower classification performance. The duplication augmentation technique improves performance on datasets iSarcasm [29], Ghosh et al. [30], and SemEval-18 [31]. The results obtained, however, demonstrate that augmentation data enhances model accuracy when it comes to non-sarcastic detection, as shown by an increase in true negatives.

According to previous works, performing augmentation on sarcasm data is challenging due to the unique complexity of the data. Meanwhile, in text data augmentation research, the GAN-based approach appears to be more promising than synonym replacement using NLPAug, repeating words, and external dataset augmentation, and there has been no text data research that used RGAN as a method of data augmentation. The performance of the model in detecting sarcasm is determined by the dataset, appropriate hyper-parameters, and appropriate model architecture. However, the main aim of this research is to propose a novel augmentation strategy for enhancing the sarcastic dataset through the utilization of Reverse Generative Adversarial Networks (RGAN). The characteristics of the data and the results of the sarcasm detection will be investigated thoroughly.

3 Methodology

In this section, the methodological framework for investigating the effects of using Generative Adversarial Networks (GANs) for data augmentation in sarcasm detection is outlined. The methodology serves as the foundation upon which the selection of datasets and subsequent analysis is based.

3.1 Dataset

This research utilizes four unique datasets to support theories and conduct an in-depth analysis of the effects of using GAN as a data augmentation. The dataset is divided into two categories: small (less than 10,000 sentences) and large (more than 30,000 sentences). This research uses iSarcasm [29] and SemEval [31] for small datasets. Each dataset has unique characteristics; for example, iSarcasm is a dataset obtained from Twitter via an online survey. Participants in survey responses provided sarcastic sentences and labels; this allows manual labeling techniques to be avoided because they cannot accurately determine sarcastic sentences from the perspective of the author of sarcasm sentences. In addition, this dataset has an unbalanced number of sarcastic and non-sarcastic. Unlike SemEval-18,

this dataset has a relatively balanced combination of sarcastic and non-sarcastic words. SemEval-18 is a Twitter-sourced dataset. The data is labeled manually using a fine-grained annotation [32] scheme. The annotators are three linguistics students who speak English as a second language.

Ghosh et al. [30] was the dataset from the large category that was used in this research scenario. Ghosh is a Twitter dataset that contains sarcastic and non-sarcastic sentences. Sarcasm classes are collected by searching for the hashtag (#) sarcasm and #not. One example of a sarcastic sentence obtained by removing ‘#not’ is “I #not love when people start rumors about me.” It Becomes “I love it when people start rumors about me.” Meanwhile, when a sentence lacks a positive marker, it is classified as non-sarcastic. The obtained sentences are not in the form of lengthy conversations. Table 2 illustrates the size of the distribution of each class in the dataset used in detail.

Table 2: The quantity of data for each class in the dataset that was used

Dataset	Non-sarcastic	Sarcastic	Sarcasm percentage (%)
iSarcasm	3,584	766	17.62%
SemEval-18	2,379	2,177	49.50%
Ghosh	22,725	18,478	44.84%

3.2 Data Pre-Processing and Augmentation

Fig. 2 shows the data preprocessing scheme up to the input data to perform GAN data augmentation. All collected datasets undertake a cleaning process, such as URL links, hashtags, foreign languages, stop words, non-English ASCII characters, and emojis. After cleaning the dataset, an 80:20 splitting train and validation were performed for each dataset. Only the sarcasm class is used as input to the RGAN model. The main reason for augmentation in the sarcasm class is that sarcasm data is difficult to obtain [30], and augmenting the non-sarcastic class will only increase data inequality between classes. Furthermore, GAN is used on the dataset to perform unbalanced sarcasm data balancing. The augmentation process begins with the embedding process using DistilBERT, which is fed a dataset with only the sarcasm class as input. The generator then generates data in the form of noise from a random distribution. The discriminator uses word embedding and fake data from the generator to distinguish between real and fake data. As a result, the discriminator loss can be fed into the generator to generate data that is as close to the original data as possible. The data generated from the generator is in the form of features whose distribution results are close to the original data.

3.3 Proposed GAN Model

This research proposed a novel pipeline that would make use of GAN for data augmentation based on the Reverse GAN (RGAN) [25], which is commonly used with image data. Using a similar concept, and made several changes to accommodate the sarcasm data. Fig. 3 depicts a more detailed stage of the novel RGAN pipeline proposed. Sarcasm data that has been pre-processed will be used as input for embedding with distilBERT. The class used for embedding only contains fully labeled sarcasm classes. The main reasons for using only one class are: (1) Sarcasm data is difficult to obtain due to its high level of complexity, and (2) Balance sarcasm and non-sarcastic classes in each dataset used. The embedding process then utilizes pre-trained distilBERT, resulting in a high-quality embedding because the pre-trained distilBERT was trained on a large corpus of words. The result of the embedding is feature data, which is labeled as fake data. The original data is marked as fake data in the RGAN concept,

and vice versa. As a result, the noise data generated by the generator is labeled as original data, so the discriminator must be certain that the data generated by the generator is real data, and the original data that is tagged false serves to cause the generator to produce data that is similar to but not an exact duplicate of the original data. The loss discriminator and generator are used as feedback for the generator to produce good data quality, with an indication that the lower the value of the loss generator and discriminator, the better the quality of the resulting data. The hyperparameter used is the learning rate of 0.001 as a result of hyperparameter parameter tuning in the generator and discriminator models and batch size of 16. This RGAN model also employs the Adam optimizer. The activation function used is the Rectified Linear Unit (ReLu). Instead of RGAN, we employ a standard GAN scheme that does not swap real and fake data labels. In the standard GAN, there are various indicators of a good model. For example, if the discriminator loss value is greater, it can indicate that the data generated by the generator is similar to the original data, and the discriminator is unsure whether the data is real or fake.

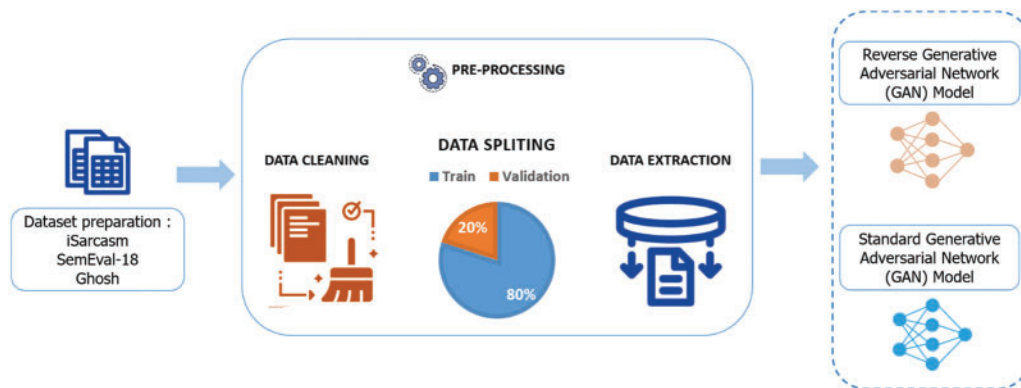


Figure 2: Data pre-processing stage

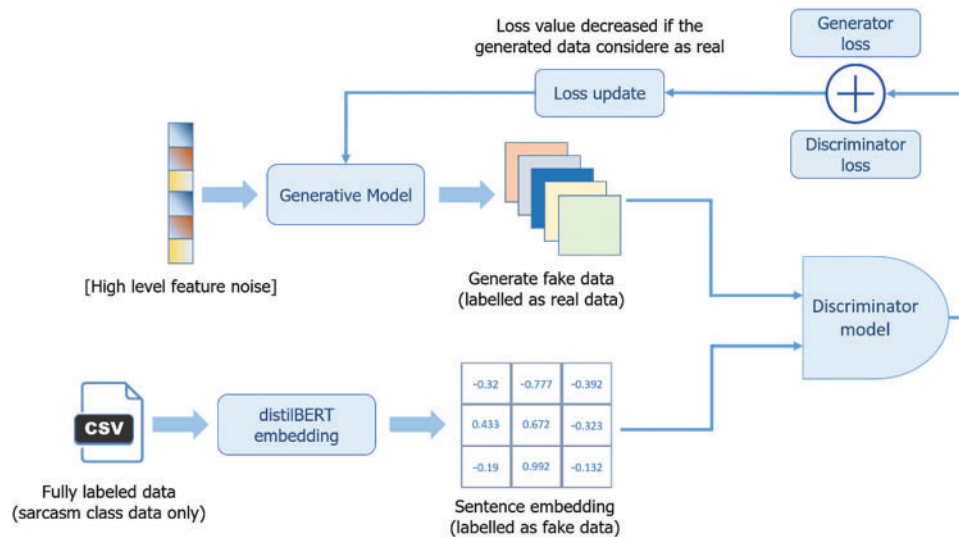


Figure 3: Proposed Reverse GAN (RGAN)

3.4 Experimental Framework

The generator’s new data is divided into three scenarios with data augmentation scales of 15%, 30%, and 45%. The generator’s synthetic data is combined with the original data as input for model training. The outcomes of each scenario will be compared to determine the GAN’s ability to detect sarcasm sentences. More detailed scenarios can be seen in Fig. 4.

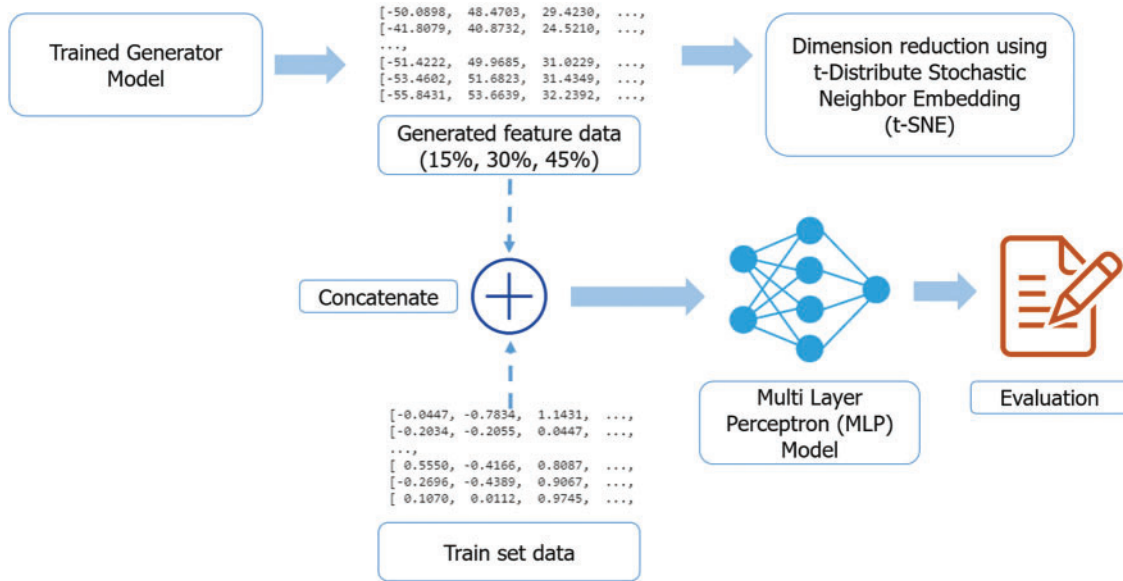


Figure 4: Model training with the data generated by the generator combined with the original data

The final evaluation results will compare the proposed RGAN model to the general GAN model. Aspects of the analysis performed, beginning with the quality of data analysis, the results of the resulting data distribution, and the effects observed when training with original data using the MLP model. Visualization is performed to analyze the data generated by RGAN by reducing the dimensions of the data. Because the characteristics of the RGAN data are quite complex, the t-Distribute Stochastic Neighbor Embedding (t-SNE) [33] algorithm was used to interpret the data visually. The t-SNE algorithm employs the following equation:

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)} \quad (1)$$

$$q_{ij} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2)} \quad (2)$$

The Eq. (1) is used to determine the pairwise similarity of data points in high-dimensional space. It assigns a probability density to each pair of data points based on their Euclidean distance using a Gaussian kernel. Then, Eq. (2) is used to determine the similarity of data points in low-dimensional space. It assigns a probability density to each pair of data points using a Student’s t-distribution. Gradient descent is used iteratively to minimize Kullback–Leibler (KL) divergence by adjusting the position of the data distribution with the following Eq. (3):

$$\frac{\delta C}{\delta y_i} = 2 \sum_j (P_{ij} - q_{ij} + P_{ij} - q_{ij}) (y_i - y_j) \quad (3)$$

Meanwhile, the Multi Layer Perceptron (MLP) model results from the F1-score and the MLP model loss will demonstrate validation of the quality of the data generated by the RGAN. The hyperparameter used in the MLP model is a learning rate of 0.0001 using the Adam optimizer, batch size 16, epoch 100, early stopping with large patience of 10 and a seed value of 200. The augmentation method with RGAN is compared with the augmentation method using the original GAN, where the labels of the original data and synthetic data are not reversed, and NLPAug, which is one of the popular augmentation frameworks on text data [34].

4 Result and Discussion

In this section, the results of the experiments are presented and discussed, focusing on the application of Generative Adversarial Networks (GANs) for data augmentation in sarcasm detection. The primary objective is to evaluate the impact of GAN-based augmentation on model performance, emphasizing the quality of augmented data and its effect on classification accuracy. The analysis begins with an evaluation of data quality following the GAN augmentation process. This evaluation employs loss values of the generator and discriminator as well as dimensionality reduction techniques to visualize the differences between real and synthetic data distributions. Subsequently, the discussion delves into the experimental results, comparing GAN augmentation with alternative techniques, such as NLPAug and unaugmented data. These experiments provide insights into the benefits and limitations of GAN augmentation, particularly in scenarios involving small datasets and class imbalance.

4.1 Augmented Data Quality Evaluation

Multidimensional features are utilized to store the information generated by the GAN generator. The evaluation of data quality takes place subsequent to data generation. Loss values from both the generator and discriminator serve as valuable indicators for assessing data quality. Furthermore, to gain deeper insights into the disparities between the distributions of real and synthetic data, the t-SNE technique is employed for dimensionality reduction. The results of this visualization technique are presented in Figs. 5–7.

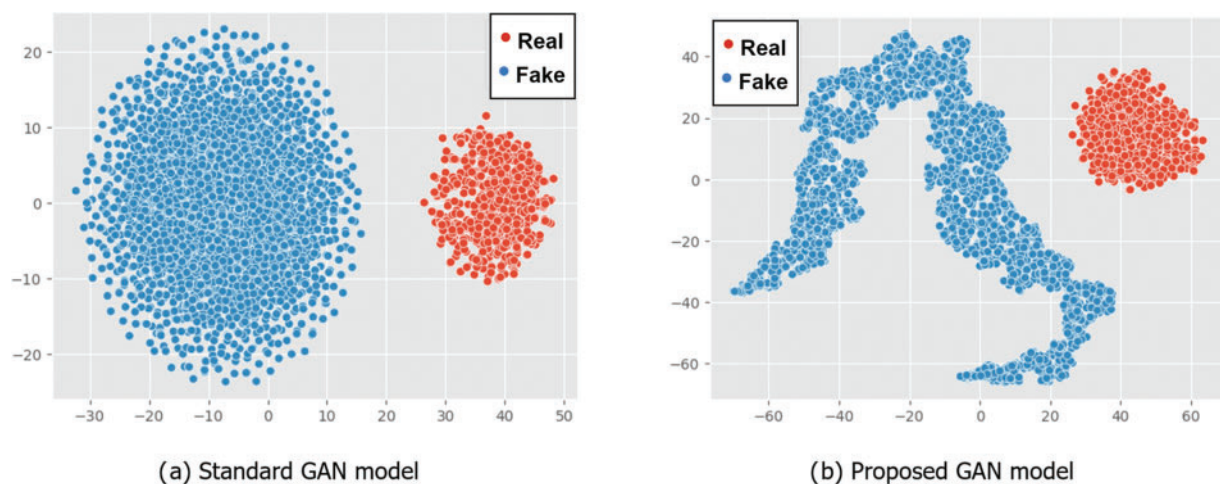


Figure 5: Generated data for iSarcasm dataset

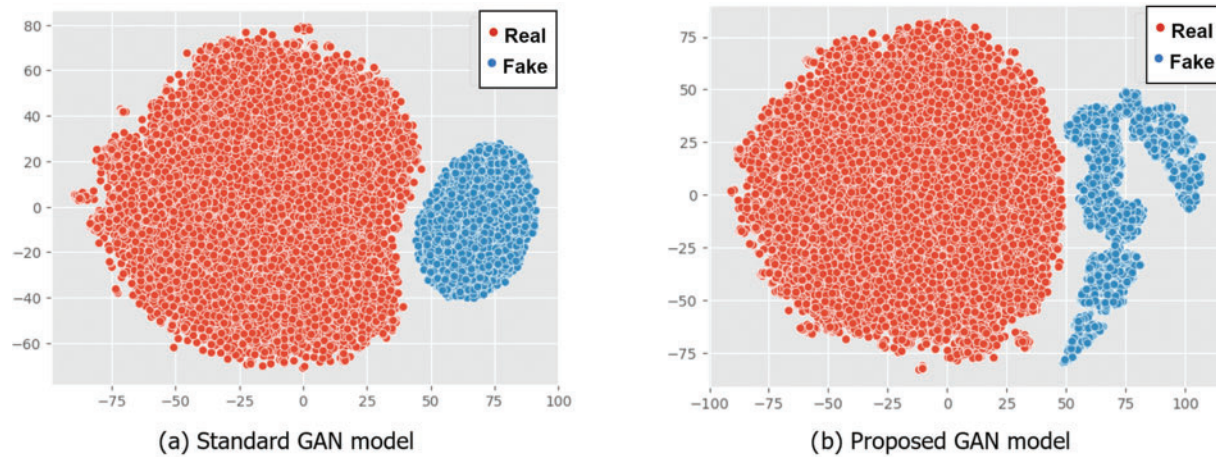


Figure 6: Generated data for SemEval-18 dataset



Figure 7: Generated data for Ghosh dataset

The resulting pattern shows the difference in data distribution between standard GAN and RGAN. The standard GAN tends to follow the original dataset's pattern and has a more defined data center point, whereas the RGAN has a more varied pattern. Both types of GAN produce data that is close enough to the original data to be considered similar. However, if the resulting data exceeds the original data, the RGAN has a high level of outliers. For example, in Fig. 5b, the resulting data far outnumbers the original data. Meanwhile, the data distribution in Fig. 5a is more consistent. The amount of the learning rate has an impact on the data distribution distance produced by the Reverse Generative Adversarial Network (RGAN). A lower learning rate leads to a more realistic representation of the original data, which can introduce noise due to the duplication of several features. On the other hand, in the case of an excessively high learning rate, the resultant feature distance will be significantly greater, leading to an increased data variance that can result in overfitting [35]. Therefore, a learning rate of 0.001 is commonly employed in the Adam optimizer [36] to ensure that the data distribution remains suitably balanced in relation to the original feature.

4.2 Experimental Result

In this section, the focus shifts to the discussion of experimental results and the ramifications of employing GAN augmentation. It is important to note that the dataset augmentation was exclusively applied to the sarcasm class. A comparative analysis is conducted, contrasting the outcomes of GAN augmentation with those of NLPAug and the original dataset. The data generated by the generator in GAN is in the form of a feature that approximates the distribution of sarcasm class features in the DistilBERT-encoded. Meanwhile, NLPAug replaces adjectives in the sarcasm class with synonyms and does not increase the amount of data. Figs. 8–10 are class distributions in each dataset used with changes in the amount of data in the sarcasm class after augmentation with GAN. Table 3 shows the results of experiments on the iSarcasm dataset. The obtained results show that the standard GAN with an augmentation percentage of 45% has the highest F1-score. When compared to RGAN, standard GAN has a more stable data distribution. Experiments on balanced GANs, on the other hand, show that the RGAN has a much higher value than the standard GAN. In the balanced sarcasm class experiment, data augmentation produced nearly 2.1 times fake data based on the original one. The data generated by the standard GAN has a noisy indication, which reduces accuracy. RGAN, on the other hand, has succeeded in producing more varied data, even though there are data points that are further away from the original data points, it does not make the data dirty. However, in order to obtain a low loss value and a high F1-score, the data generated by the reverse GAN must be adjusted to the original dataset. When compared to NLPAug, all augmentation data using GAN has a higher value, implying that GAN augmentation is appropriate for datasets with small amounts of data and unbalanced classes.

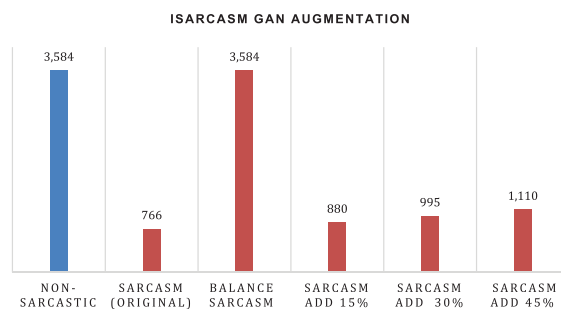


Figure 8: The difference in the amount of data in the sarcasm class in the iSarcasm dataset is based on the percentage of augmentation

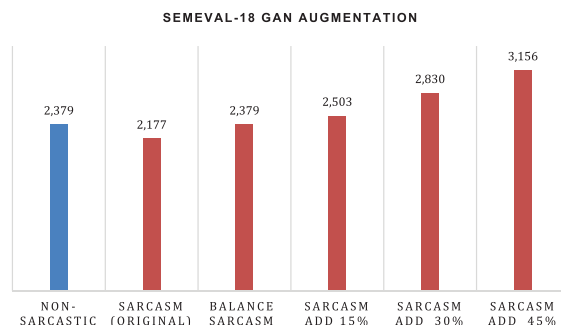


Figure 9: The difference in the amount of data in the sarcasm class in the SemEval-18 dataset is based on the percentage of augmentation

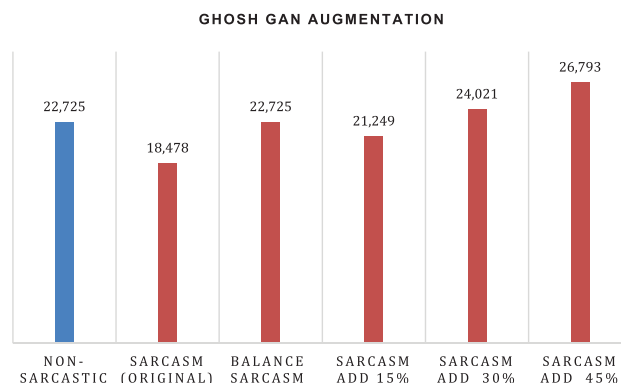


Figure 10: The difference in the amount of data in the sarcasm class in the Ghosh dataset is based on the percentage of augmentation

Table 3: Experimental results on the iSarcasm dataset

Datasets	Experiment	F1-score validation	Val loss
iSarcasm	No augmentation	0.51642	0.43370
	Balance class standard GAN	0.48503	0.43431
	15% standard GAN	0.52672	0.43041
	30% standard GAN	0.49546	0.43826
	45% standard GAN	0.54529	0.43368
	Balance class RGAN	0.52357	0.43766
	15% RGAN	0.52696	0.42709
	30% RGAN	0.50873	0.43531
	45% RGAN	0.52696	0.44045
	15% NLPAug	0.51667	0.43488
30% NLPAug	0.47328	0.43491	
45% NLPAug	0.50937	0.44532	

Experiments with the SemEval-18 dataset yielded the same results in [Table 4](#). Standard GAN has the highest F1-score and the lowest validation loss, implying that GAN augmentation is also suitable for small datasets but has a balanced class. However, if the data generated by GAN causes an imbalance between classes, the accuracy tends to decrease, as in the case of the addition of 45% data, where the sarcastic data far outnumbers the non-sarcastic data. Unlike iSarcasm, the SemEval-18 dataset shows a better GAN standard under balanced sarcasm class conditions because the resulting data is still within reasonable limits below 15%. Meanwhile, the results of data augmentation with NLPAug did not increase the F1-score and tended to decrease the F1-score so that it could be indicated if NLPAug caused noise in the data.

[Table 5](#) shows that augmentation with RGAN has the highest F1-score, at 69.01%, based on experiments on the Ghosh dataset. Balance class is a dataset that contains non-sarcastic numbers with balanced sarcasm and non-sarcastic classes of 22,725 and 22,725 sarcasm. With a relatively balanced

distribution of data for each class in the SemEval-18 and Ghosh datasets, it has the consistency that if the augmentation data of the sarcastic class far exceeds the non-sarcastic class, the accuracy will tend to decrease. In a balanced class condition, the RGAN has a better F1-score, so it shows the same indication as the iSarcasm dataset. Augmentation with a standard GAN yields similar results to augmentation with an RGAN. NLPAug also produces results that are consistent with lower values than GAN augmentation.

Table 4: Experimental results on the SemEval-18 dataset

Datasets	Experiment	F1-score validation	Val loss
SemEval-18	No augmentation	0.63935	0.61887
	Balance class standard GAN	0.64318	0.61675
	15% standard GAN	0.65090	0.61940
	30% standard GAN	0.61275	0.62339
	45% standard GAN	0.64306	0.62498
	Balance class RGAN	0.62422	0.61828
	15% RGAN	0.63116	0.62527
	30% RGAN	0.64805	0.62116
	45% RGAN	0.64295	0.62529
	15% NLPAug	0.62258	0.63037
	30% NLPAug	0.62484	0.63025
	45% NLPAug	0.61624	0.63506

Table 5: Experimental results on the Ghosh dataset

Datasets	Experiment	F1-score validation	Val loss
Ghosh	No augmentation	0.68343	0.56665
	Balance class standard GAN	0.68117	0.57319
	15% standard GAN	0.68850	0.56964
	30% standard GAN	0.67746	0.56686
	45% standard GAN	0.68556	0.57018
	Balance class RGAN	0.69017	0.57399
	15% RGAN	0.68641	0.57073
	30% RGAN	0.67156	0.57993
	45% RGAN	0.68409	0.56540
	15% NLPAug	0.68479	0.57126
	30% NLPAug	0.68708	0.58207
	45% NLPAug	0.67557	0.57799

In general, the proposed data augmentation in this research differs from GAN-BERT. The data generated by the generator is only used as input to the discriminator in GAN-BERT, so it is unclear whether changes in accuracy in classifying sarcasm data are used when using GAN or not. Meanwhile, the generator that is used to generate new data will be filled into the model as input as features with data points that are similar to the original data, with the goal of evaluating the quality of the resulting data. GAN-BERT uses unlabeled labels as supporting data for unbalanced classes on the generator to balance unbalanced data. However, using an excessive amount of unlabeled data may cause the generator to produce data in a balanced class, allowing data to become even more unbalanced. Consequently, the approach here is to train the GAN model with data from unbalanced classes. All experiments showed that GAN-based augmentation improved the model's ability to classify sarcasm compared to NLPAug. Data augmented with NLPAug does not generate new data, so there is still an imbalance between classes. NLPAug is better suited for adding variety to balanced datasets like Ghosh. NLPAug does not improve at all on unbalanced datasets like iSarcasm.

A comparison and discussion between this research and other works in implementing data augmentation can be seen in [Table 6](#).

Table 6: Comparison between this research with other works

Title	Technique	Discussion
Plumeria at SemEval-2022 Task 6: Robust approaches for sarcasm detection for English and Arabic using transformers and data augmentation [18]	BERT	There exist three distinct forms of augmentation employed in this study. The initial method involves the incorporation of an external dataset, which results in a notable decrease in test performance by approximately 30.6%. This decline can be attributed to the introduction of an excessive number of variants. The model has challenges in generalization, resulting in suboptimal performance. The second augmentation technique employs word embedding to facilitate the substitution of synonyms. The efficacy of the second type of augmentation is limited as it does not significantly enhance the quantity of data, hence resulting in little fluctuations and negligible influence. Conversely, the difficulty in the process of synonym substitution resides in the careful selection of an appropriate synonym for every term that is being substituted. Consequently, the incorporation of word embedding led to a decrease in the test outcomes, resulting in an accuracy of 7.78% F1-score. Additionally, the validation test yielded an F1-score of 41.52%. In conclusion, the utilization of word repetition as a means of augmentation yields a validation value of 92.62%, but the corresponding test accuracy experiences a fall to 87%. The ineffectiveness of this augmentation is attributed to its persistent shortfall of 5.62% un comparison to the validation accuracy. In addition to this, the recurrence of words might lead to the occurrence of duplicate feature values, which may introduce noise into the dataset.

(Continued)

Table 6 (continued)

Title	Technique	Discussion
Sarcasm Detection in Twitter—Performance impact while using data augmentation: Word embeddings [19]	RoBERTa	This research primarily employed data augmentation using GloVe word embeddings, which, while effective to some extent, it can introduce certain shortcomings. First, the paper mainly focused on synonym replacement, a relatively simplistic augmentation technique. This approach can potentially alter the context and semantics of the original text, diminishing the effectiveness of sarcasm detection models. Additionally, the research did not thoroughly investigate alternative data augmentation methods, such as word insertion, deletion, or sentence paraphrasing, which have been proven effective in other NLP tasks. While the study demonstrated an increase in the F1-score for sarcasm detection, it is essential to consider that increased accuracy in predicting non-sarcastic texts was observed. To improve the robustness and performance of data augmentation techniques in sarcasm detection, future research should explore a broader spectrum of augmentation strategies, encompassing diverse pre-trained word embeddings and more complex textual transformations.
GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples [20]	BERT	The GAN-BERT model utilizes the Generative Adversarial Network (GAN) framework to produce synthetic data. The dataset utilized consists of both tagged and unlabeled utterances containing sarcasm. The empirical findings indicate a positive correlation between the quantity of labeled data and the efficacy of the model in detecting. The final outcomes demonstrate a disparity in accuracy between BERT and GAN-BERT, with respective values of 48.35% and 62.67%, as observed on the MNIST dataset. This increase is widely regarded as highly successful to implement. Nevertheless, the findings of this study did not demonstrate the efficacy of the proposed approach on either imbalanced or balanced datasets.

In comparison to the studies presented in [Table 6](#), this research offers a distinct and superior approach to sarcasm detection through Generative Adversarial Networks (GANs) for data augmentation. While the mentioned studies have primarily focused on external datasets, synonym replacement, or simple augmentation techniques, our research introduces a novel framework utilizing Reverse GAN (RGAN) in the context of sarcasm detection. The results of our experiments on datasets like iSarcasm, SemEval-18, and Ghosh showcase the efficacy of GAN-based augmentation, particularly when the synthetic data closely aligns with the volume of the original data. Notably, our approach outperforms NLPAug in scenarios with small datasets and class imbalances. Moreover, we demonstrate that RGAN, a less common technique, can achieve performance results comparable to those of standard GAN. The ability to generate synthetic text data that closely matches the original data in features sets our research apart, offering a more balanced and effective approach to augmenting text data for sarcasm detection.

This research emphasizes the versatility and effectiveness of the RGAN technique, providing a robust solution for improving sarcasm detection accuracy in diverse augmentation scenarios.

5 Conclusion

1. The successful application of the suggested novel framework for enhancing text data through the incorporation of additional data features has been demonstrated in its ability to enhance the performance of the model in identifying sarcasm within specific augmentation scenarios.

2. Due to the different characteristics of each dataset, GAN-based augmentation in different datasets could have a different impact on performance. Overall, based on the analysis, it is found that if the synthetic data does not exceed the amount of original data, GAN-based augmentation can improve performance significantly when compared to using NLPAug.

3. The utilization of the Reverse GAN technique, although not commonly practiced, delivered performance outcomes in sarcasm detection that are on par with those achieved using the standard GAN.

In conclusion, this study has introduced a novel framework for enhancing text data through the incorporation of additional data features, demonstrating its success in improving model performance in sarcasm detection within specific augmentation scenarios. This study revealed that the impact of GAN-based augmentation on performance varies across datasets, with a consistent finding that GAN-based augmentation outperforms NLPAug when synthetic data does not significantly exceed the volume of original data. One of the key contributions of this research is the utilization of the Reverse GAN (RGAN) technique, a less common approach, which yielded performance results in sarcasm detection comparable to those achieved using the standard GAN. This suggests the effectiveness and versatility of RGAN in enhancing text data.

Augmentation with GAN in the sarcasm class tends to lose accuracy if the data generated far exceeds that of the non-sarcastic class. Meanwhile, if the augmentation in the sarcasm class produces data that is many times the size of the original data, using an RGAN, as in the iSarcasm dataset, will be more profitable. The SemEval-18 dataset yields the opposite result, demonstrating that using standard GANs is more advantageous when adding data reaches a balance point. However, the Ghosh dataset demonstrates that a relatively balanced dataset does not necessitate a large amount of synthetic data. This is because [Tables 3 and 4](#) show that the best augmentation results were obtained in experiments with generated data less than 45%. Because of the relatively high level of difficulty, the future research potential for augmenting text datasets is still very broad. Producing synthetic text data in the form of text (rather than features) is a challenging task. Currently, the Reverse Generative Adversarial Network (RGAN) lacks the capability to reconstruct feature forms into textual representations. Another challenge is the development of transformer models capable of reading input in the form of features; there are currently only a few state-of-the-art models capable of receiving input data in the form of features.

Acknowledgement: None.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Study conception and design: Derwin Suhartono; data collection: Derwin Suhartono, Alif Tri Handoyo; analysis and interpretation of results: Franz Adeta Junior, Alif Tri Handoyo; draft manuscript preparation: Franz Adeta Junior. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All dataset in this paper is publicly available in GitHub repositories. For iSarcasm dataset (<https://anonymous.4open.science/r/24639225-ac0e-4057-b2d4-16e7e50570d0/README.md>), SemEval-2018 (<https://github.com/Cyvhee/SemEval2018-Task3>), and Ghosh (<https://github.com/MirunaPislar/Sarcasm-Detection>).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] C. I. Eke, A. A. Norman, L. Shuib and H. F. Nweke, “Sarcasm identification in textual data: Systematic review, research challenges and open directions,” *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4215–4258, 2020.
- [2] S. Frenda, A. T. Cignarella, V. Basile, C. Bosco, V. Patti *et al.*, “The unbearable hurtfulness of sarcasm,” *Expert Systems with Applications*, vol. 193, pp. 116398, 2022. <https://doi.org/10.1016/j.eswa.2021.116398>
- [3] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani and B. Wright, “Sarcasm detection using machine learning algorithms in Twitter: A systematic review,” *International Journal of Market Research*, vol. 62, no. 5, pp. 578–598, 2020.
- [4] S. K. Bharti, B. Vachha, R. K. Pradhan, K. S. Babu and S. K. Jena, “Sarcastic sentiment detection in tweets streamed in real time: A big data approach,” *Digital Communications and Networks*, vol. 2, no. 3, pp. 108–121, 2016.
- [5] A. C. Băroiu and Ş. T. Matu, “Automatic sarcasm detection: Systematic literature review,” *Information*, vol. 13, no. 8, pp. 1–17, 2022.
- [6] A. Joshi, V. Tripathi, K. Patel, P. Bhattacharyya and M. Carman, “Are word embedding-based features useful for sarcasm detection?” in *Proc. of 2016 Conf. on Empirical Methods in Natural Language Processing*, Austin, Texas, pp. 1001–1006, 2013.
- [7] R. J. Kreuz and G. M. Caucci, “Lexical influences on the perception of sarcasm,” in *Proc. of Workshop on Computational Approaches to Figurative Language*, Rochester, New York, USA, pp. 1–4, 2007.
- [8] J. Godara, I. Batra, R. Aron and M. Shabaz, “Ensemble classification approach for sarcasm detection,” *Behavioural Neurology*, 2021. <https://doi.org/10.1155/2021/9731519>
- [9] J. Lemmens, B. Burtenshaw, E. Lotfi, I. Markov and W. Daelemans, “Sarcasm detection using an ensemble approach,” in *Proc. of Second Workshop on Figurative Language Processing*, Online, pp. 264–269, 2020.
- [10] R. Misra and P. Arora, “Sarcasm detection using hybrid neural network,” arXiv:1908.07414, 2019.
- [11] N. K. Singh and K. Raza, “Medical image generation using generative adversarial networks: A review,” *Studies in Computational Intelligence*, vol. 932, pp. 77–96, 2021.
- [12] J. Choe, S. Park, K. Kim, J. H. Park, D. Kim *et al.*, “Face generation for low-shot learning using generative adversarial networks,” in *Proc. of IEEE Int. Conf. on Computer Vision Workshops*, Venice, Italy, pp. 1940–1948, 2017.
- [13] Y. Lu, D. Chen, E. Olaniyi and Y. Huang, “Generative Adversarial Networks (GANs) for image augmentation in agriculture: A systematic review,” *Computers and Electronics in Agriculture*, vol. 200, pp. 107208, 2022.
- [14] A. Rahaman, R. Kuri, S. Islam, M. J. Hossain and M. H. Kabir, “Sarcasm detection in tweets: A feature-based approach using supervised machine learning models,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 454–460, 2021.
- [15] K. Sentamilselvan, P. Suresh, G. K. Kamalam, S. Mahendran and D. Aneri, “Detection on sarcasm using machine learning classifiers and rule based approach,” in *IOP Conf. Series: Materials Science and Engineering*, Erode, India, vol. 1055, no. 1, pp. 012105, 2021.
- [16] A. K. Jena, A. Sinha and R. Agarwal, “C-Net: Contextual network for sarcasm detection,” in *Proc. of Second Workshop on Figurative Language Processing*, Kochi, Kerala, India, pp. 61–66, 2020.

- [17] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati and L. B. M. Neti, "Sarcasm detection using multi-head attention based bidirectional LSTM," *IEEE Access*, vol. 8, pp. 6388–6397, 2020.
- [18] M. Shaheen and S. K. Nigam, "Plumeria at SemEval-2022 Task 6: Sarcasm detection for English and Arabic using transformers and data augmentation," in *Proc. of 16th Int. Workshop on Semantic Evaluation (SemEval-2022)*, Seattle, USA, vol. 6, pp. 923–937, 2022.
- [19] A. T. Handoyo, H. Hidayaturrehman, C. J. Setiadi and D. Suhartono, "Sarcasm detection in twitter—Performance impact while using data augmentation: Word embeddings," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 22, no. 4, pp. 401–413, 2022.
- [20] D. Croce, G. Castellucci and R. Basili, "GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples," in *Proc. of 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2114–2119, 2020.
- [21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones *et al.*, "Attention is all you need," in *Proc. of 31st Conf. on Neural Information Processing Systems*, Long Beach, CA, USA, pp. 5999–6009, 2017.
- [22] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. of North American Chapter of the Association for Computational Linguistics*, Minneapolis, Minnesota, USA, vol. 1, pp. 4171–4186, 2019.
- [23] H. Lee, Y. Yu and G. Kim, "Augmenting data for sarcasm detection with unlabeled conversation context," in *Proc. Second Workshop on Figurative Language Processing*, pp. 12–17, 2020.
- [24] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta *et al.*, "Deep learning for visual understanding: Part 2 generative adversarial networks," *IEEE Signal Process Magazine*, vol. 35, no. 1, pp. 53–65, 2018.
- [25] R. M. Weber, "Exploiting the hidden tasks of GANs: Making implicit subproblems explicit," arXiv:2101.11863, 2021.
- [26] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley *et al.*, "Generative adversarial networks," *Communications of the Association for Computing Machinery*, vol. 63, no. 11, pp. 139–144, 2020.
- [27] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi *et al.*, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv:1907.11692, 2019.
- [28] V. Sanh, L. Debut, J. Chaumond and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," arXiv:1910.01108, 2019.
- [29] S. V. Oprea and W. Magdy, "iSarcasm: A dataset of intended sarcasm," in *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 1279–1289, 2020.
- [30] A. Ghosh and T. Veale, "Fracking sarcasm using neural network," in *Proc. of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, San Diego, California, USA, pp. 161–169, 2016.
- [31] C. van Hee, E. Lefever and V. Hoste, "SemEval-2018 Task 3: Irony detection in English tweets," in *Proc. of the 12th Int. Workshop on Semantic Evaluation*, New Orleans, Louisiana, USA, pp. 39–50, 2018.
- [32] C. Van Hee, E. Lefever and V. Hoste, "Guidelines for annotating irony in social media text, version 2.0," *LT3 Technical Report Series*, 2016. https://lt3.ugent.be/media/uploads/publications/2015/annotation_guidelines_irony.pdf (accessed on 18/01/2023)
- [33] C. R. G. Alonso, L. M. P. Naranjo and J. C. F. Caballero, "Multiobjective evolutionary algorithms to identify highly autocorrelated areas: The case of spatial distribution in financially compromised farms," *Annals of Operations Research*, vol. 219, no. 1, pp. 187–202, 2014.
- [34] B. Li, Y. Hou and W. Che, "Data augmentation approaches in natural language processing: A survey," *AI Open*, vol. 3, pp. 71–90, 2022.
- [35] J. Lever, M. Krzywinski and N. Altman, "Points of significance: Model selection and overfitting," *Nature Methods*, vol. 13, no. 9, pp. 703–704, 2016.
- [36] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. of 3rd Int. Conf. on Learning Representations*, San Diego, CA, USA, pp. 1–15, 2015.