International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Sarcasm Identification in Reddit Online Discussion Forum Using Fully Contextual CASCADE

Kevin Hadrian Hadirahardja*[1], Abba Suganda Girsang[2]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract: Sarcasm is a form of figurative language that cannot be easily detected using simple sentiment analysis because of contradictory nature between its literal and true meaning. Sarcasm detection research is conducted using various methods and algorithm, one of those method is ContextuAl SarCasm Detector (CASCADE) which implements content and contextual features to detect sarcasm from comment. The model uses CNN to extract content-based features from the comments, Word2Vec and CNN to extract contextual features for user and discourse embeddings. However, content feature extraction can be further improved by implementing transformer since it can understand connection between words better thus improving contextual knowledge of the comments for better content-based modelling. This study proposes an enhancement for CASCADE as baseline model, replacing its CNN based method for content modelling by using BERT-BiLSTM method to create a better content-based modelling and concatenating it with CASCADE's user and discourse embeddings. The proposed model will then be used to detect sarcasm from REDDIT online discussion forum corpus namely SARC, a dataset for sarcasm research purpose. The proposed method gives a slight increase in accuracy and F1-score compared to the previous research and proven to perform best by training with balanced dataset. This research is still in early stage, and it may get better from hyperparameter tuning and cleaner method, for now it provides a significant increase in Accuracy and F1-Score.

Keywords: *BERT, BiLSTM, CASCADE, CNN, SARC, Word2Vec*

1. Introduction

Sarcasm is one form of figurative language that often used to make a statement to be more meaningful or to state a criticism in a way that contradicts the literal meaning of the word itself. Both irony and sarcasm had this contradictory nature between literal and figurative meaning, but no need to be confused as sarcasm is commonly used as satire or criticism while irony isn't [1]. Based on speech delivery, sarcasm can be divided into two forms which is explicitly and implicitly. Explicit form can be seen by noticing its lexical or pragmatical cues like words that commonly used in sarcasm remarks, punctuation, interjection, contrast change of sentiment, etc [2]. Implicit form differs from explicit sarcasm since it doesn't contain lexical cues, hence contextual knowledge like user comments history data, previous comments, topic, or background are used to determine this kind of sarcasm.

Sentiment analysis can be used for sarcasm detection task, but false results may occur due to contradictory nature of the sarcasm resulting in poor performance. Human can differentiate sarcasm by considering the context, knowledge of norms, and speaker mindset [3], so computer needs further contextual information too[4], one way is to understand the implicit and explicit information of the sarcasm and using it as knowledge to detect the sarcasm, one example that implements this method is ContextuAl SarCasm Detector or CASCADE[5]. This model utilises the users writing style and personality to measure user's tendency in making sarcasm remarks as information in addition to content information from user's comment gathered using Convolutional Neural Network (CNN) based model and currently is the state of art model for detecting sarcasm from Reddit Corpus SARC[6]

Further improvement can be done to this model specifically for content information modelling of CASCADE from using CNN to a recently popular transformer named Bidirectional Encoder Representations from Transformers (BERT)[7] which has proven to be superior in many Natural Language Processing (NLP) task. BERT currently is the most used transformer for NLP task, it works by reading the entire sequence of words bidirectionally (left-to-right or right-to-left) at once, so it increases the knowledge of interconnection of words to each other, making this model has better ability in understanding context within a sentence and suitable to be implemented in sarcasm detection task. The contribution purpose of this research is to enhance the current CASCADE model by changing its content modelling method from using CNN to BERT – BiLSTM resulting in richer model that has greater contextual knowledge for the content-based model.

2. Literature Review

Sarcasm detection task is relatively new in natural language processing field of research, most of the previous research can be divided into two approaches which is content-based modelling that generally utilize the lexical clues of the word or its sentiment, and context-based modelling that improves on content-based method by adding additional information as consideration for text sarcasm classification, like example conversation background or its topic.

In content-based modelling, various work has been done to figure out the pattern or finding the lexical cues for sarcasm detection,

[1]*Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta 11380, Indonesia ORCID ID: 0000-0002-0529-2095*
[2]*Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta 11380, Indonesia ORCID ID: 0000-0003-4574-3679*
** Corresponding Author Email: kevin.hadirahardja@binus.ac.id*

lexical cues are used in form of positive or negative predicate, interjection, and punctuation [8], emojis are also present in comments and can be used as a clue for sarcasm[9], sentence pattern are studied as well to train the model to learn synthetic pattern or form to create sarcasm detection classifier[10], [11]. Sentiment analysis are frequently used together with linguistic theory to detect sarcasm for example defining sarcasm as sentence that has positive sentiment but the sentence contains negative sentiment words [12], and using implicit - explicit context incongruity [13].

Improving on content-based modelling, context-based modelling uses additional information as consideration for sarcasm detection with assumption that sarcasm as figurative language cannot be simply processed because it has contradicting sentiments and usually refers to a specific topic. In social media, historical message can be used to calculate tendency of a person to be sarcastic[14], [15], using the response of the comments referred [16], and taking its discussion topic or user profile as embeddings [5] [17].

The current trend on Natural Language Processing (NLP) study right now is implementing transformer for various kind of task, BERT, and its variant like RoBERTa[18], DistilBERT [19], Multilingual BERT, etc. proven to be superior in these generic tasks or its specific tasks. In sarcasm detection study, BERT can be used as classifier to detect sarcasm [20], [21], as contextual word embeddings for classification[22], combining word embeddings of RoBERTa with BiLSTM [23]or implementing multi-head attention BiLSTM method [24] for sarcasm classification.

3. Proposed Model and Methodology

In this section, the research framework as shown in Figure 1 for the proposed model will be discussed steps by steps from the dataset, the pre-processing steps done, the flow of inputs/outputs until it will be used to detect sarcasm, and how evaluation will be done to the proposed model.

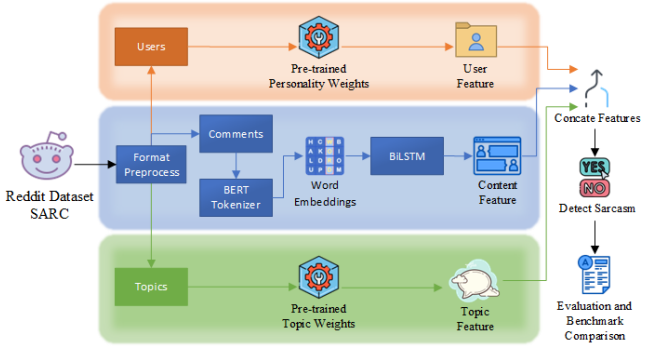

Figure 1. Research Framework

3.1. Dataset

The dataset used is SARC [6], a corpus for sarcasm detection research containing comments from Reddit online discussion forum. There are three types of datasets available which is balanced, unbalanced, and single topic: politic, and all of it will be separated into training and testing data with ratio of 7:3, the amount of data used in this research will be the same exact amount as in previous research [5] to ease the evaluation process and the details for the data ratio will be shown in table 1.

Table 1. Dataset Split Amount

No	Type	Training		Test	
		Sarcasm	Not Sarcasm	Sarcasm	Not Sarcasm
1	Balanced	77.351	77.351	32.333	32.333
2	Unbalanced	77.351	25.784	32.333	10.778
3	Politic	6.834	6.834	1.703	1.703

Only small portion from total SARC data will be used, on balanced dataset the amount of both labels will be the same, unbalanced dataset consist of the same data as balanced dataset but the ratio between two labels will be set as approximately 8:2 for non-sarcasm and sarcasm, and politics dataset is a subset data from the balanced dataset that has politic for its topic. Every training process the dataset will be processed by using cross validation of 10 to make sure the model able to predict the new data, it runs at every end of single epoch run and serve to be a flag to determine whether the model is overfitting, and to give insights on how the model will perform using untrained dataset.

3.2. Format Preprocess

The dataset used consist of two main files, a dictionary in JSON format containing details of all the comments and the dataset file in format of 'postID $commentID_{1\text{-}n}$ | $responseID_{1\text{-}n}$ | $label_{1\text{-}n}$' with each ID referencing a comment in JSON file and the label is '0' for non-sarcasm and '1' for sarcasm. A single row of data consists of one postID, zero or many comments, and one or many responses from the last comment with its label as a pair. Each row may have different amounts of response and label, a format change is essential to easen the data processing, the format will be changed to 'responseID | Label' and it will be focused on each response, if a row consist of two responses, then a single row will be two rows of data in the new format and so on. Later the dataset ID will be replaced with its comment and the details like its author and topic from JSON dictionary and converted into panda dataframe, the comments will be then treated with preprocessing step only to change the case to lowercase.

3.3. Tokenizing

Comments data from panda dataframe format are processed by performing tokenization using three different BERT encoder variants which is DistilBERT, RoBERTa, and multilanguage BERT to get representational vector from each comment. There are no preprocessing steps to clean each word in the comments except converting all case to lowercase, since WordPiece tokenizer of BERT works by reading every single word in a sequence and every word or special characters present in a sarcasm remark may be considered as lexical clue in detecting sarcasm.

The tokenizing process gives three vectors as output, the token embedding, segment embedding, and position embedding. For creating the word embedding only the token embedding are used which contains both the word tokens and its masking tokens except DistilBERT because it didn't do masking process when tokenizing so only word tokens are used.

3.4. Word Embeddings

Vector containing the tokens and masking tokens are then processed as inputs to the pre-trained BERT models to be fitted. The word embeddings are created from the output produced by BERT model's pooling layer consisting vectors of contextual representation for the comments.

3.5. Content Feature

Word embeddings will then be processed as inputs to BiLSTM, a Recurrent Neural Network containing two LSTM units that works both ways making it suitable to understand the contextual information of a comment from both sides, left to right, and right to left. Attention heads will be implemented also to help the BiLSTM understand the most relevant information of the of inputs from the word embeddings. The final output of BiLSTM will be in form of representational vector rich in contextual information of the content from all the comment processed which will be referred as Content Features.

3.6. Topic and User Feature

Both user and topic can be acquired from the panda dataframe data, where each username and discussion topic will act as their own ID. For every user and topic will then be stored into an array each to prevent ordering mistake between the comment-user-topic pair, there will be two arrays containing users and topics in the end. Pretrained weights for both are acquired from the previous research [5] containing representational vectors for each respective user and topic, these pretrained weights will be imported in form of dictionary with the real username and its topic as the ID.
Each user and topic from the arrays will then be processed and referenced to the dictionary by replacing the value of each user and topic ID with its weights vector, the result is two arrays of user weights and the topic weights which will be referred as User Features and Topic Features.

3.7. Sarcasm Prediction and Evaluation

Figure 2 depict the process after collecting three features needed, next step is to concate the three into a concate layer, followed by networks of dense layers with dropout layers in between dense layers, the last layer is a sigmoid layer as the output layer consists of two neurons containing sigmoid probability either a prediction is belonged to either sarcasm or not sarcasm label. For training purpose each sequence's prediction is used to count the loss function using binary cross-entropy.

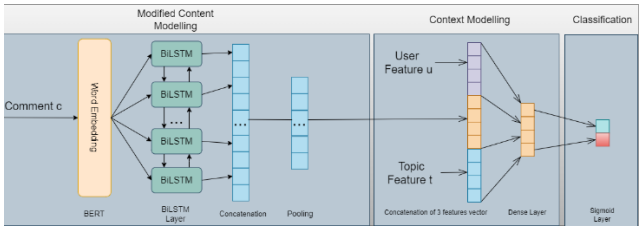

Figure 2. Features Concatenation Scheme

3.8. Proposed Model

The detailed layer by layer architecture of the proposed model shown in Figure 3 alongside its dimension shape for processing a single sequence of comment, first are multiple input layers consisting of word tokens, user weights, and topic weights as showed in striped box, variants of BERT (DistilBERT, RoBERTa, Multilingual BERT) encoder for word embedding, BiLSTM with 64 attention heads, features concatenation by concatenation layer, neural networks consisting of multiple dense and dropout layers, and lastly a sigmoid layer as the output layer with sarcasm prediction as its final output.

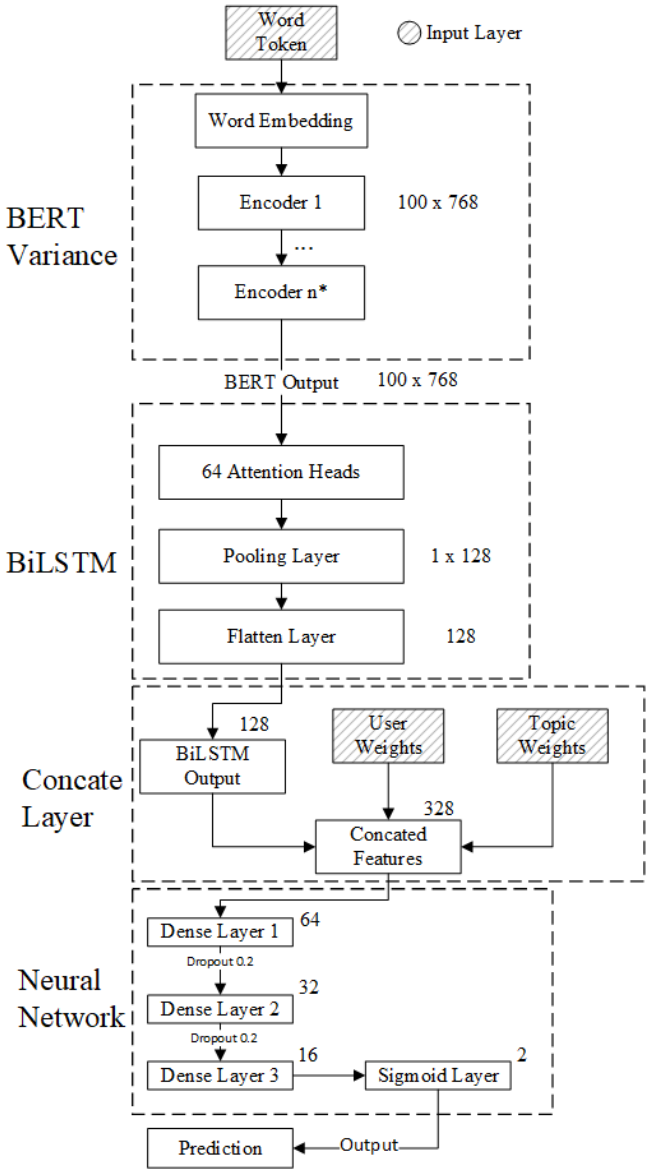

Figure 3. Detailed Structure of Proposed Model

4. Result Analysis

4.1. Enhancement of Previous Research

BiLSTM based method from previous research [23], [24] and the usage of BERT or its variants as word embeddings [20]-[23] has proven to work well in sarcasm detection task using various kind of sarcasm datasets. The idea is to improve the performance of CASCADE [5], the current state of the art model for sarcasm detection using SARC dataset by changing its content modelling method from using Word2Vec word embedding and Convolutional Neural Network to using BERT word embeddings and BiLSTM to get better content-based modelling with stronger contextual knowledge then combining it with user and topic features using CASCADE's pretrained weights to detect sarcasm.

4.2. Model Training Details

For training purpose, each of the models will be using the same learning rate of $5x10^{-4}$ using Adam optimizer, trained for 15 epochs, and 64 attention heads for the BiLSTM layer by using three kinds of BERT variants for word embedding (DistilBERT, RoBERTa, and Multilingual BERT). Training will be done for three types of datasets, each with different variants of word embedding. The initial accuracy result of training and validation

set is shown in Table 2.

Table 2. Initial Accuracy from Training and Validation Data

Model Variants	Balanced		Unbalanced		Politics	
	Train	Valid	Train	Valid	Train	Valid
FC CASCADE DistilBert	81.9	81.2	83.6	82,6	77.9	80.6
FC CASCADE RoBERTa	82.0	81.2	83.5	83.0	77.1	78.1
FC CASCADE MBERT	81.5	81.4	81.9	82.3	74.5	77.0

From the initial accuracy of each model for each dataset, it can be found that the accuracy difference for the train and validation data are slim since the final accuracy are not far off which also proves that the model loss between train and validation data are small.
For the three variants of word embedding, their accuracy is almost similar, the way how BERT and its variants were trained using vast and various dataset may contributing to its ability to understand lexical clues in sarcasm comments better.

4.3. Testing Results

After finishing the training process, each of the models will then be fitted to testing dataset treated with the same steps just like in section 3.2 to 3.6. The label prediction result of the models is gathered and compared to its true labels. These two values will then be used as score in evaluating the model's performance.

4.4. Performance Evaluation

The final model accuracy value will determine the success rate of a model in detecting sarcasm, to be able to get the accuracy of the model for evaluation (1) is used.

$$Accuracy = \frac{Correct\ prediction}{Total\ data\ processed} \quad (1)$$

Since one of the cases of model training is using unbalanced dataset, we cannot rely on accuracy alone to determine its performance. Unbalanced dataset has 8:2 ratio for non-sarcasm-sarcasm data, the high accuracy we get from the training process may biased to one major class only so we need other metrics to solve this issue, other evaluation metrics like Recall, Precision, and F1-Score will be used also. From the prediction result of testing data, a confusion matrix will be created to get four values which is true-positive, true-negative, false-positive, and false-negative. These four values will be used in (2), (3), (4) to find Precision, Recall, and F1-score of the evaluated model. Not all metrics are available in the previous research of baseline model so the unavailable metrics will be filled with 'n/a' as not available.

$$Precision = \frac{True\ Positive}{True\ Positive+False\ Positive} \quad (2)$$

$$Recall = \frac{True\ Positive}{True\ Positive+False\ Negative} \quad (3)$$

$$F\ Score = \frac{2\ x\ (Precision\ x\ Recall)}{Precision+Recall} \quad (4)$$

After evaluation process, the model's evaluation metrics will then be reported, and its performance will be compared between each case and with the baseline models as shown in Table 3.

Table 3. Performance Evaluation of FC CASCADE Models

Balanced Dataset				
Models	Accuracy	Precision	Recall	F1-Score
CASCADE [5]	77.00	n/a	n/a	77.00
FC CASCADE DistilBert*	**81.33**	78.38	86.54	**82.25**
FC CASCADE RoBERTa*	81.15	80.64	81.99	81.31
FC CASCADE MBERT*	80.91	78.79	84.61	81.60
Unbalanced Dataset				
Models	Accuracy	Precision	Recall	F1-Score
CASCADE [5]	79.00	n/a	n/a	**86.00**
FC CASCADE DistilBert*	82.73	68.58	57.08	62.31
FC CASCADE RoBERTa*	**82.98**	66.80	63.46	65.09
FC CASCADE MBERT*	81.97	66.17	57.09	61.29
Politic Dataset				
Models	Accuracy	Precision	Recall	F1-Score
CASCADE [5]	74.00	n/a	n/a	75.00
RoBERTA-RCNN [23]	**79.00**	78.00	78.00	78.00
FC CASCADE DistilBert*	78.77	75.98	84.14	**79.85**
FC CASCADE RoBERTa*	77.51	80.20	73.04	76.46
FC CASCADE MBERT*	76.71	73.45	83.67	78.23

*****Proposed models

The evaluation results in Table 3 shows the proposed model gives a decent increase both in accuracy and F1 Score from the base model. The best performance of FC CASCADE is by using DistilBERT as its word embeddings, it gives a higher Accuracy and F1 Score compared to its base model [5] on balanced and politic dataset, although its Accuracy in politics dataset is slightly lower from previous research [23] but the proposed model still gives a better F1 Score than the other baseline model.
On unbalanced dataset, the best accuracy given by FC CASCADE with RoBERTa word embedding with more or less 4% increase in Accuracy but far lower F1 Score compared to baseline model, we assume it happens because the model only able to predict test data from a major portion of label and got mostly wrong prediction on minor portion of label indicated by the low number of Precision and Recall which impacted the F1 Score overall.

5. Conclusion and Future Works

In Sarcasm Detection task, high accuracy is a sign that the model able to detect sarcasm successfully by considering context from condition or the ongoing topic of the comments given. The usage of BERT as word embeddings combined with BiLSTM proven to increase the performance of the model to detect sarcasm and amongst BERT variants the result is almost similar because the metrics results are competitive to each other. In conclusion, the proposed method capable of delivering better performance with more than 4% increase in Accuracy and F1 Score using balanced dataset and politics dataset, and the model is less capable to detect sarcasm when trained using unbalanced dataset so using balanced dataset on training process is recommended.
Due to restriction in computation resources and scope of work, not all BERT's variant able to be compared in this research. This research can be further improved by implementing larger size of BERT's variant like BERT-Large or XLM-BERT and implementing training configuration method to tune the learning rate by using callback, scheduler, and different optimizer with its configs to improve the quality of training process to make a better trained model. For model training using unbalanced dataset,

further research can be done with the same ratio and using a larger number of dataset or by creating more dataset by using data augmentation to improve the model in learning from unbalanced dataset.

References

R. Filik, A. Ţurcan, C. Ralph-Nearman, and A. Pitiot, "What is the difference between irony and sarcasm? An fMRI study," *Cortex*, vol. 115, pp. 112–122, Jun. 2019, doi: 10.1016/j.cortex.2019.01.025.

P. Carvalho, L. Sarmento, M. J. Silva, and E. de Oliveira, "Clues for detecting irony in user-generated contents," in *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion - TSA '09*, 2009, p. 53. doi: 10.1145/1651461.1651471.

A. Ghosh and T. Veale, "Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual and Very Personal," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 482–491. doi: 10.18653/v1/D17-1050.

B. C. Wallace, D. K. Choe, L. Kertz, and E. Charniak, "Humans Require Context to Infer Ironic Intent (so Computers Probably do, too)," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2014, pp. 512–516. doi: 10.3115/v1/P14-2084.

D. Hazarika, S. Poria, S. Gorantla, E. Cambria, R. Zimmermann, and R. Mihalcea, "CASCADE: Contextual Sarcasm Detection in Online Discussion Forums," May 2018, [Online]. Available: http://arxiv.org/abs/1805.06413

M. Khodak, N. Saunshi, and K. Vodrahalli, "A Large Self-Annotated Corpus for Sarcasm," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.05579

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04805

Maojin. Jiang, ACM Digital Library., Association for Computing Machinery. Special Interest Group on Information Retrieval., and H. and Web. Association for Computing Machinery. Special Interest Group on Hypertext, *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion.* ACM, 2009.

R. González-Ibáñez, S. Muresan, and N. Wacholder, "Identifying Sarcasm in Twitter: A Closer Look," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Jun. 2011, pp. 581–586. [Online]. Available: https://aclanthology.org/P11-2102

O. Tsur, D. Davidov, and A. Rappoport, "ICWSM - A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews," in *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*, 2010. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1495

D. Davidov, O. Tsur, and A. Rappoport, "Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon," Association for Computational Linguistics, 2010. [Online]. Available: http://tinysong.com/cO6i

E. Riloff, A. Qadir, P. Surve, L. de Silva, N. Gilbert, and R. Huang, "Sarcasm as Contrast between a Positive Sentiment and Negative Situation," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, 2013, pp. 704–714. [Online]. Available: https://aclanthology.org/D13-1066/

A. Joshi, V. Sharma, and P. Bhattacharyya, "Harnessing Context Incongruity for Sarcasm Detection," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2015, pp. 757–762. doi: 10.3115/v1/P15-2124.

A. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm detection on twitter:A behavioral modeling approach," in *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, Feb. 2015, pp. 97–106. doi: 10.1145/2684822.2685316.

M. Zhang, Y. Zhang, and G. Fu, "Tweet Sarcasm Detection Using Deep Neural Network," in *COLING*, 2016.

A. Ghosh and T. Veale, "Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual and Very Personal." [Online]. Available: https://liwc.wpengine.com/

S. Amir, B. C. Wallace, H. Lyu, and P. C. M. J. Silva, "Modelling Context with User Embeddings for Sarcasm Detection in Social Media," Jul. 2016, [Online]. Available: http://arxiv.org/abs/1607.00976

Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, [Online]. Available: http://arxiv.org/abs/1907.11692

V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.01108

A. Baruah, K. Das, F. Barbhuiya, and K. Dey, "Context-Aware Sarcasm Detection Using BERT," in *Proceedings of the Second Workshop on Figurative Language Processing*, 2020, pp. 83–87. doi: 10.18653/v1/2020.figlang-1.12.

K. Pant and T. Dadu, "Sarcasm Detection using Context Separators in Online Discourse," Jun. 2020.

A. Khatri, P. P, and Dr. A. K. M, "Sarcasm Detection in Tweets with BERT and GloVe Embeddings," Jun. 2020.

R. A. Potamias, G. Siolas, and A. G. Stafylopatis, "A transformer-based approach to irony and sarcasm detection," *Neural Comput Appl*, vol. 32, no. 23, pp. 17309–17320, Dec. 2020, doi: 10.1007/s00521-020-05102-3.

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, doi: 10.1109/ACCESS.2019.2963630.