Sarcasm Text Detection on News Headlines Using Novel Hybrid Machine Learning Techniques

Neha Singh and Umesh Chandra Jaiswal

Department of ITCA, Madan Mohan Malaviya University of Technology, Gorakhpur, India
nehaps2703@gmail.com, ucjitca@mmmut.ac.inb

KEYWORDS	ABSTRACT
Sentiment analysis; Machine learning; News headlines; Vectorization model	One of the biggest problems with sentiment analysis systems is sarcasm. The use of implicit, indirect language to express opinions is what gives it its complexity. Sarcasm can be represented in a number of ways, such as in headings, conversations, or book titles. Even for a human, recognizing sarcasm can be difficult because it conveys feelings that are diametrically contrary to the literal meaning expressed in the text. There are several different models for sarcasm detection. To identify humorous news headlines, this article assessed vectorization algorithms and several machine learning models. The recommended hybrid technique using the bag-of-words and TF-IDF feature vectorization models is compared experimentally to other machine learning approaches. In comparison to existing strategies, experiments demonstrate that the proposed hybrid technique with the bag-of-word vectorization model offers greater accuracy and F1-score results.

1. Introduction

The current trend in sentiment analysis is geared towards the specialized area of sarcasm detection. For the community of sentiment analysts, it is quickly becoming one of the most difficult and important tasks. Numerous elements, such as context awareness, cultural considerations, and personality qualities, are said to contribute to the difficulty of sardonic content detection (Husain & Uzuner, 2021). Incorrect polarity resolution in sentences is brought on by sarcasm. Thus, sentiment analysis now faces the challenge of sarcasm discernment. Context and accent are key components of sarcasm. For instance, "Wasting my valuable time in traffic is a fantastic experience." Here, the word "fantastic" is a positive word that is used to convey the negative emotion of losing time in a traffic jam. Because there

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

1

are binary target labels, such as sarcastic and non-sarcastic texts, sarcasm detection can be understood as a dual classification task (Verma et al., 2021).

Without considering the likelihood of a reciprocal interaction between the two activities, most researchers nowadays concentrate on either sentiment categorization or sarcasm detection. Sarcasm is commonly used by people to convey a vehemently negative sentiment, so it is obvious that the two behaviours are connected. Through this observation, one of the two activities may be able to easily assist in the development of the other (Majumder et al., 2019). The six-tupled examples of sarcasm are:

$$< S, H, C, U, P, P >$$

Where, S = Speaker, H = Hearer or Listener, C = Context, U = Utterances, P = Literal Proposition, P = Intended Proposition. In essence, Speaker S produces an utterance U in context C that refers to proposition P while the hearer H intends to comprehend the proposition (Katyayan & Joshi, 2019). Figure 1 shows the textonym of sarcasm text detection.

1.1. Sarcasm Types

There are four varieties of sarcasm:

- *Propositional Sarcasm*: Although it seems like a statement that does not involve sentiment, there is a feeling at play.
- *Embedded Sarcasm*: Incorporates words and phrases with disparate sentiments.
- *Like-Prefixed Sarcasm*: Similar expressions imply argument denial. For instance, "your good friend is present".
- *Illocutionary Sarcasm*: It contains quasi cues that suggest an attitude that is contrary to what is actually said (Katyayan & Joshi, 2019).

1.2. Approaches

In general, there are three different types of sarcasm detection techniques:

- *Rule-Based Approaches*: Try to detect sarcasm using credible proof. Rules based on sarcastic indications are used to describe this proof. A benefit of this technique is that they give an evaluation that matches many rules.
- *Statistical Approaches*: The features and classifiers used in statistical methods for sarcasm detection vary. It is divided into two parts: features used and learning algorithms.
  - *Features Used*: It examines the criteria that have been claimed to be useful in detecting statistical sarcasm. Bag-of-words qualities are used in most strategies.
  - *Learning Algorithms*: For sarcasm identification, a range of models have been tested. SVM and logistic regression are the main techniques used in sarcasm detection. They are employed to identify distinctive characteristics.
- *Deep Learning-Based Approaches*: Few of these methods have been described for artificial intelligence sarcasm detection, in structures that use deep learning methods. They utilize word embedding similarities as a criterium for sarcasm detection. (Joshi et al., 2017). Improve word embedding features by comparing them to the most uniform and incompatible word pairings; assess quality gains.

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

2

1.3. Feature Set Analysis

Sarcasm detection uses a feature selection method based on the different kinds of sarcasm that can be found in a text. Additionally, feature extraction is based on the format or structure of the word (Chaudhari & Chandankhede, 2017).

- *Lexical Feature*: Text qualities such as unigrams, bigrams, n-grams, skip-grams, #hashtags, etc. are included in this group of features. They are crucial for sarcasm recognition.

- *Pragmatic Features*: Pragmatic aspects include symbolic and metaphorical messages such as emoticons, smileys, answers, forceful and onomatopoeic phrases, etc. Many authors have used them in their studies.

- *Hyperbole Feature*: This function is essential for sarcasm recognition. Exaggeration is used as a form of speech expression in hyperbole. Hyperbole is the use of text elements such as intensifiers, interjections, quotations, punctuation, etc. Many scholars have produced accurate work by utilizing this capability.

- *Pattern-Based Feature*: High-frequency terms and content words are included in pattern-based features based on how frequently they appear in the text. The pattern used in a text might be quite particular for certain texts or very general for others.

- *Syntactic Feature*: These characteristics include repeating morphosyntactic design patterns. Morphosyntactic traits are those that are important to syntax. They frequently include chronological inequity and PoS-grams.

- *Contextual Feature*: Contextual features are the features designed to identify sarcasm that draws on data outside of the text. In general, any information or background knowledge outside of the word needs to be forecasted. It can be integrated by utilizing auxiliary data or details from the data source system. Consider annotations. Reddit posts with "sarcasm" tags are used by labelled data to request more context.

- *Metaphoric Feature*: Extreme adjectives, proverbs, honorifics, and nouns with a strong positive or negative connotation are all included. The information about the author's feelings in this feature is significant and valuable.

1.4. Issues in Sarcasm Detection

A variety of problems are encountered by the present sarcasm detection techniques, types, and features. This part focuses on a number of significant concerns, including problems with data, problems with features, and problems with classification methods.

- *Issues with Data*: The integrity of the dataset may become vague and questionable even when labelled phrases with hashtags provide facts in a clear and concise manner, such as: "I adore dull food. #not" #not is a sarcastic expression. It might not be taken as sarcastic if they delete the #not and just think of the statement "I adore boring food."

- *Issues with Features*: The sentiment classifier is fooled by sarcastic statements, which could impair the text's classification accuracy. The ground valence of the sentence is a requirement for using sentiment as a predictor feature. Because of this, it is important to investigate new traits and combine them with already-existing features to improve accuracy.

- *Issues with Classification Techniques*: Scholars sometimes use small datasets and other times use vast ones. For classification, however, a dataset need not be evenly distributed, which gives

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

3

rise to balanced and unbalanced datasets. Therefore, the dataset must be subjected to the correct classification method in order to accurately classify utterances into sarcastic and non-sarcastic ones (Chaudhari & Chandankhede, 2017).

1.5. Sarcasm Detection Challenges

The following are the key difficulties in sarcasm detection:

- The quality of the information was another important consideration in sarcasm recognition. Sarcastic statements tend to be unclear and dubious in nature. This ambiguity is eliminated with hashtags that denote sarcasm. However, caustic statements are challenging to grasp without hashtags.

- Another crucial step in sarcasm recognition is feature extraction. As a result, adding new elements and combining them with existing ones can improve the efficacy of sarcasm detection. Choosing a suitable feature should involve a broader investigation of lexical, punctuation-based, hyperbolic features etc.

- Choosing the right classification methods is also important. There are both balanced and unbalanced data sets.

- Sarcasm recognition in noisy text is quite difficult because brief and noisy texts typically contain fewer features and convey less context.

- A sarcastic statement typically conveys a bad feeling while utilizing only pleasant words. As a result, sardonic identification requires more advanced features, such as semantic features and features relating to the author's text (Aboobaker & Ilavarasan, 2020).
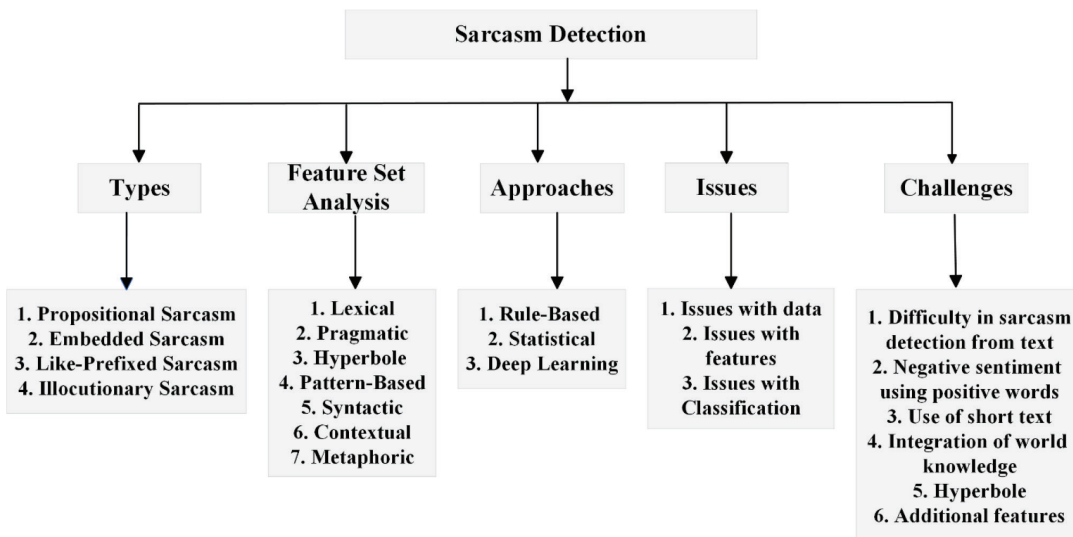

Figure 1. Textonym of sarcasm detection

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

4

The majority of earlier studies on sarcasm recognition used Twitter datasets gathered under hashtag supervision, although these datasets are messy in regard to tags and grammar. Additionally, a lot of tweets are answers to previous tweets, making it difficult to discern sarcasm in them (Barhoom et al., 2022). In comparison to other Twitter datasets, this one has the following advantages:

- Since news headlines are correctly crafted by professionals, there are no spelling errors or colloquial language. As a result, the possibility of discovering previously trained embeddings improves, and sparsity decreases.

- In addition, because The Onion only publishes satirical news, researchers receive high-quality tags with far less distortion compared to Twitter datasets.

- The news headlines were gathered as standalone items, as opposed to tweets that respond to other tweets. This would make it simpler for us to recognize the genuine sarcastic elements.

1.6. Motivation

The purpose of sarcasm text recognition on news headlines is to outline research towards the creation of a method for detecting sarcasm in news headlines. The key aspects of this study were:

- *Sarcasm Text Detection*: The main objective of this research is to create a method that can recognise sarcasm in text, especially in news headlines.

- *News Headlines*: This describes the category of text that is being examined.

- *Novel Hybrid Machine Learning Technique*: This research entails developing a fresh, cutting-edge method for identifying sarcasm in news headlines. The term "hybrid" suggests that many machine learning approaches or techniques are blended to produce the desired result.

The overall goal of the study is to address the problem of sarcasm detection in news headlines using a novel combination of machine learning approaches. This kind of research can be used to improve natural language comprehension models, improve news analysis accuracy, and improve sentiment analysis.

1.7. Contribution

The contribution of the proposed research, "Sarcasm Text Detection on News Headlines Using a Novel Hybrid Machine Learning Technique" is noteworthy in a number of ways:

- *Improved Sentiment Analysis*: Sentiment analysis results can frequently be distorted by sarcasm. A reliable sarcasm detection technique can result in more precise sentiment analysis of news headlines, which is essential for figuring out how the general population feels about particular events.

- *Advancement of Machine Learning Techniques*: An innovative hybrid machine learning method for sarcasm recognition is developed. This contribution has the potential to advance several relevant topics.

- *Addressing Limitations of Existing Methods*: Many current sarcasm detection techniques struggle with linguistic nuance and contextual comprehension. These drawbacks may be overcome, and sarcasm detection accuracy may be increased with an innovative hybrid technique.

- *Innovation in Machine Learning*: It shows creativity in the machine learning field to introduce a novel hybrid technique. Such contributions may encourage additional research towards inventive method fusions for tackling challenging problems.

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

5

The experiment conducted this research demonstrates the efficiency of all seven machine learning techniques, utilizing two vectorization models. The accuracy and f1-score outcomes for the proposed hybrid technique using the bag-of-words vectorization model are 0.85% and 0.77%, respectively, whereas the outcomes for the proposed hybrid model using the TF-IDF vectorization model are lower at 0.53% and 0.60%. Both accuracy and f1-score for random forest models employing the vectorization models bag-of-words and TF-IDF are 0.84% and 0.76%, respectively. The two machine learning techniques likewise produce results with excellent accuracy, ranging between 80% and 81%. The accuracy of the three machine learning techniques is just 0.73% with adaboost, 65% with xgboost, and 0.56% with gradient boosting, respectively. Naive Bayes and logistic regression both have F1 scores of 76%. The F1 scores for xgboost approaches are lower at 0.62%, 0.61% for adaboost, and 0.60% for gradient boosting.

The rest of this article is organised as follows: A review of the earlier research, upon which this study is built, is provided in Section 2. The collection of the data, dataset preprocessing techniques, the vectorization technique, machine learning models, and performance parameter metrics that convert texts into numerical vectors are all covered in Section 3. Section 4 provides a description of the model results. In Section 5, conclusions are listed, and the future lines of research are considered.

2. Literature Review

To categorise material that is sarcastic, researchers employ a wide range of strategies. Tables 1 and 2 contain information about various relevant studies in more depth.

Farha & Magdy (2020) introduced ArSarcasm, a dataset for detecting sarcasm in Arabic that was produced by reannotating existing datasets for sentiment analysis in Arabic. There are a total of 10,547 tweets in the sample, 16% of which are sarcastic. Modern sentiment detectors degrade when presented with ironic content, according to studies. Finally, they use BiLSTM to train a model for sarcasm recognition.

Kumar et al. (2020) presented a multi-attention-based bidirectional long-short memory connectivity method for detecting irony in sentences. The research findings demonstrate that the effectiveness of BiLSTM is improved by a multi-head attention mechanism, and that it outperforms feature rich SVM classifiers.

Razali et al. (2021) focused on sarcasm detection in tweets utilizing handcrafted contextual cues and characteristics retrieved via deep learning. The feature representation of the CNN architecture was retrieved before it was mixed with meticulously constructed feature sets. These individually designed attribute sets were developed on the basis of their corresponding contextual justifications. Each attribute set has been presented expressly with the goal of detecting irony in mind. The goal is to identify the best attributes. Some sets function properly even when used independently. The performance outcomes of all parameters in the study were favorable. For this assignment, it was discovered that logistic regression is the best classification technique.

The study of Pawar & Bhingarkar (2020) proposed a pattern-based method for detecting sarcasm, using data from Twitter. The analysis of tweets as irony or non-irony employed four sets of attributes that contained several instances of precise sarcasm. The proposed attributes were examined, and new cost classes were evaluated.

Abulaish & Kamal (2018) provided a unique strategy for the identification of self-irony. While approaches were utilised for attribute identification and classification, rule-based strategies focus on identifying potential self-irony tweets. The proposed method was assessed using a Twitter dataset and contrasted with other cutting-edge sarcasm detection techniques.

Liu et al. (2019) simulated face-to-face conversation in their research. The researchers used a method that included auxiliary factors such as syntax, numeric, emoji, etc., to increase classification

results. The findings of the research showed that their method performed better in terms of efficiency.

Goel et al. (2022) recognized and comprehended sarcastic behaviour and patterns. This research intended to bridge the gap between the intelligence of humans and machines. The authors used different neural methods to recognize irony on the internet. The dataset was built using various previously trained word-embedding methods, and their prediction accuracy was evaluated to enhance the proposed model's accuracy. The goal was to be able to categories the writer's general attitude as sardonic or non-sarcastic in order to make sure that the intended audience receives the right information. The final analysis demonstrated that the proposed ensemble approach performed better than the other cutting-edge methods.

Nayak & Bolla (2022) examined different computation and machine learning techniques to identify ironic headlines in this study. Experimental tests demonstrated that the recall, accuracy, and F1-score were improved when pre-trained transformer-based embeddings were integrated with LSTM networks.

The study of Yin et al. (2021) proposed a unique network for sentiment classification and irony identification. The network's modules were used by MT_SS to jointly train for both objectives, acquiring local attribute map and global attribute description using convolutional neural networks. The tests demonstrated that the authors' proposed approach was superior to cutting-edge methods.

Kumar et al. (2019) presented an approach which was built on a novel hybrid of sAtt-BLSTM and convNet employing GLoVe for creating semantic word embeddings. Using balanced and unbalanced datasets, the proposed model's robustness was examined.

Misra (2022) researched sarcasm detection. The accessibility of contextual tweets was also necessary because many tweets are in reply to other tweets, making it difficult to discern sarcasm without context. The author compiled the news headline dataset. Unlike HuffPost, which provides accurate news, The Onion seeks to produce satirical representations of current events. The author added links to the news items' original sources to make it more useful and enable further data extraction if required. Along with other use cases, besides sarcasm detection, the author discussed several dataset characteristics in this research.

Bharti, Sathya Babu, et al. (2017) identified sarcasm in Hindi tweets. In this manuscript, the dataset was considered when determining the level of sarcasm. The proposed scheme achieved a 79.4% accuracy rate.

Mandal & Mahto (2019) described a novel CNN and LSTM-based sarcasm detection method and considered the sarcasm in news headlines.

The study of Jariwala (2020) proposed that the best features should be chosen before the data is sent for classification. Therefore, data pre-processing cleans up the information to improve the outcomes. The effectiveness of irony detection utilizing the best feature sets was demonstrated by the findings.

Kumar & Katiyar (2019) described the datasets, feature extraction, and method employed in earlier sarcasm detection techniques. The study demonstrated that SVM is the best method for sarcasm detection, the learning algorithm was a crucial step in building a model, and the News Headline dataset was the best one to use for research because it was created by experts.

In the study of Mohammed et al. (2020), effective short text classification techniques were put to the test for the first time in identifying sarcasm. Additionally, detailed output for each technique was shown, based on various research performance criteria.

Ying et al. (2021) explained the precision of categorization for news texts. This research employed a quantitative methodology established by the author. The accuracy scores acquired using the Naive Bayes approach were compared to those achieved using other techniques in this study to assess the method's efficacy. The results showed that the accuracy value did not reach its optimum and could still be improved upon by restructuring and reevaluating the data.

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

7

Table 1 presents the findings from these investigations and lists the different methods they had employed. A wide range of methods can be applied to evaluate sentiment on various types of feedback, including Naïve Bayes (NB), support vector machines (SVM), k-nearest neighbor (KNN), logistic regression (LR), decision trees (DT), discriminant analysis, random forests (RF), deep neural networks, long-short-term memory (LSTM), gated recurrent units (GRU), A2Text-Net, convolutional neural networks (CNN), bidirectional LSTM (BiLSTM), and CNN-LSTM. When compared with other methods, all strategies provide optimal results, with SVM, Naive Bayes, Random Forest, and hybrid models providing the best findings for sarcasm recognition in text. The objective of this research is to determine the most effective method for sarcasm detection in texts.

Table 1. A comparison of different types of approaches

Ref. No.	Methods	Best Methods	Results
(Farha & Magdy, 2020)	BiLSTM	BiLSTM	The algorithm yields a result of 0.46.
(Kumar et al., 2020)	BiLSTM, SVM	BiLSTM	The outcomes of the research showed that BiLSTM performs better when a multihead attention mechanism is used.
(Razali et al., 2021)	KNN, DT, LR, Discriminant analysis, SVM	Discriminant analysis	All the performance criteria that measure the outcomes of the studies are all excellent.
(Pawar & Bhingarkar, 2020)	RF, SVM, KNN	RF	The best accuracy and F1 score are provided by Random Forest.
(Abulaish & Kamal, 2018)	DT, NB, Bagging	Bagging	For both datasets, bagging had the highest precision values while decision trees had the highest f-score and recall values.
(Liu et al., 2019)	LR, SVM, RF, Deep Neural Networks, LSTM, GRU, A2Text-Net	A2Text-Net	The outcomes of the testing show that the A2Text-Net strategy enhances classification efficiency.
(Goel et al., 2022)	CNN, LSTM, GRU	LSTM	The study found that the LSTM with GloVe embeddings was the classifier that performed the best.
(Nayak & Bolla, 2022)	NB, BiLSTM, LR, Multilayer Perceptron, SVM, LSTM	LSTM	The conducted tests demonstrated how long short-term memory networks improve F1 score, memory, and accuracy.
(Yin et al., 2021)	Single Classifiers, Simple multi-task Bi-GRU, Multi-task Bi-GRU, MTSS, State-of-art	MT_SS	The test results demonstrated that the proposed approach is superior to cutting-edge methods.
(Kumar et al., 2019)	LSTM, BLSTM without Attention, BLSTM with Attention, sAtt-BLSTM convNet	sAtt-BLSTM convNet	The proposed approach outperformed existing models.
(Mandal & Mahto, 2019)	CNN-LSTM, BiLSTM	CNN-LSTM	The CNN-LSTM model yielded the best accuracy results, according to this study's results.

Table 1. A comparison of different types of approaches (continued)

Ref. No.	Methods	Best Methods	Results
(Jariwala, 2020)	SVM without feature selection, SVM with optimal feature selection	SVM with optimal feature selection	A higher efficiency in sarcasm detection utilizing the best feature sets is demonstrated by the outcomes.
(Kumar & Katiyar, 2019)	SVM, NB, RF, Lexical Method, Neural Networks	SVM	The research demonstrates that the best method for detecting sarcasm is a support vector machine.
(Mohammed et al., 2020)	CNN, LSTM, CNN-LSTM	CNN-LSTM	In comparison to other algorithms, the CNN-LSTM model yielded the best results.
(Ying et al., 2021)	NB, Bernoulli Convolution, Naïve Bayes Classification Modeling	NB	The Naive Bayes model has a higher level of accuracy when compared to other approaches.

Explain Table 2's methods and the effectiveness matrix that were utilised to get the predictions. Models such as naive bayes (NB), support vector machine (SVM), logistic regression (LR), random forest (RF), decision tree (DT), k-nearest neighbor (KNN), xgboost, and adaboost are used to predict the results, and efficiency matrices, such as accuracy (A), precision (P), recall (R), and F1-score, are also used (F1).

Table 2. Methodologies and use frequency of performance metrics in studies for sarcasm detextion in text

Ref.No.	Types of Methods							Performance Metrics			
	NB	SVM	RF	LR	DT	KNN	Ada Boost	A	P	R	F
(Razali et al., 2021)		✔		✔	✔	✔		✔	✔	✔	✔
(Abulaish & Kamal, 2018)	✔				✔				✔	✔	✔
(Jariwala, 2020)		✔						✔	✔	✔	✔
(Pawar & Bhingarkar, 2020)		✔	✔			✔		✔	✔	✔	✔
(Chudi-Iwueze & Afli, 2020)	✔	✔		✔				✔			✔
(Novic, 2022)		✔		✔				✔	✔	✔	✔
(Aneja et al., 2022)	✔	✔	✔		✔	✔	✔	✔			✔
(Alexandru & G\^\ifu, 2020)	✔	✔						✔			
(Trystan et al., 2021)	✔	✔	✔	✔		✔		✔			
(Bharti, Pradhan, et al., 2017)	✔	✔			✔			✔	✔	✔	✔
(Parkar & Bhalla, 2023)	✔	✔	✔	✔	✔	✔		✔	✔	✔	✔
(Pal & Prasad, 2023)			✔	✔		✔		✔	✔	✔	✔
(Park & Chai, 2023)		✔	✔	✔				✔	✔	✔	✔
(Mykytiuk et al., 2023)	✔		✔	✔	✔	✔			✔	✔	✔

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

9

3. Methodology

The primary objective of this study is to identify sarcasm in text using a dataset of news headlines, and to contrast the proposed hybrid strategy with existing methods. The Jupyter Notebook platform, Microsoft Visio for creating diagrams, and Originpro for plotting graphs were used to test the proposed method for detecting sarcasm in text in news headlines. Figure 2 displays the proposed approach. The following six subsections make up this section: data source, dataset preparation, feature extraction vectorization model, machine learning techniques, performance parameters, and proposed algorithm. Below is a brief summary of each subsection.
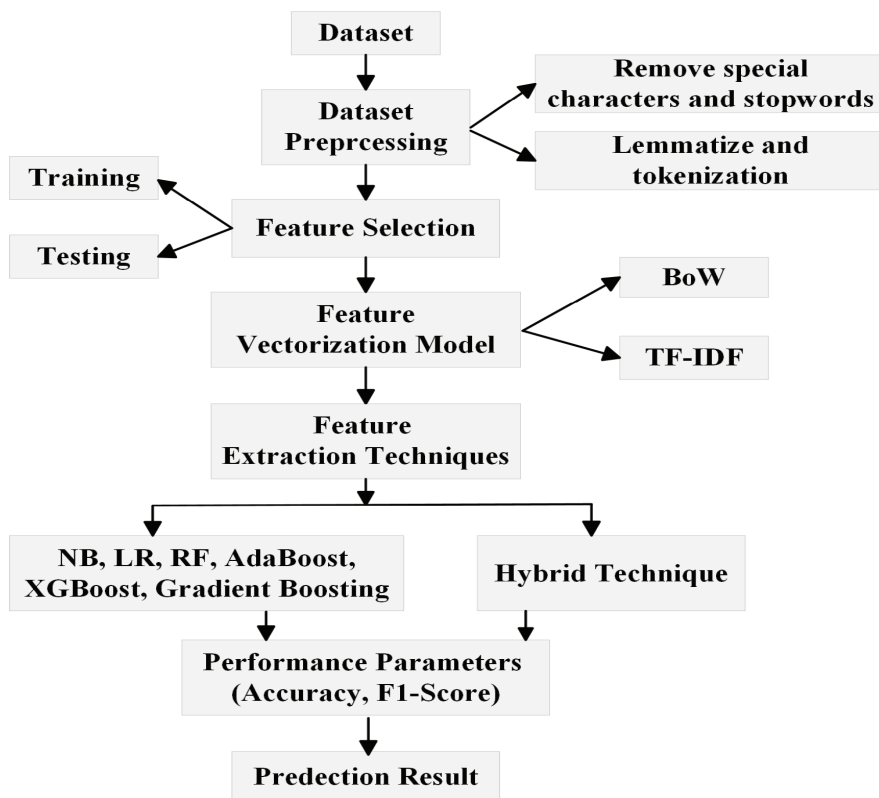

Figure 2. Proposed strategy

3.1. Dataset Sourcing

On 5 September 2022, the news headlines dataset for sarcasm detection from Kaggle (https://www.kaggle.com/datasets/rmisra/news-headlines-dataset-for-sarcasm-detection) was used in this study. A JSON file containing the dataset was provided. Table 3 shows the dataset containing three attribute.

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

10

Table 3. Specifics of the dataset

Sr. No.	Attribute	Details
1.	headline	The headlines of the article.
2.	is_sarcastic	A label parameter in which a sarcastic details is represented by one and a non-sarcastic record by zero.
3.	article_link	Includes a hyperlink to the original news story. This can be used, if necessary, to gather more information.

3.2. Dataset Preprocessing

Preprocessing data sets is crucial to sentiment analysis. The majority of user-generated material is unorganized (Nguyen et al., 2018). The following are the primary actions taken during text normalization:

- Remove special characters and stopwords: Stopwords and many other special characters should be removed in this phase since the accent sign character causes a lot of problems in preprocessing procedures.

- Lemmatize text: The process of getting a word's grammatical root is called lemmatization. Lemmatizing frequently enables computers to distinguish between various tenses of the same word.

- Perform tokenization: Tokenization is a feature extraction method that involves breaking down a sentence's words into smaller pieces. For this task, the NLTK library's tokenize module has been utilised (Nguyen et al., 2018).

A dataset word cloud was created after the dataset preprocessing phase. Figure 3 shows the dataset's word clouds. The two-word clouds depict words that are both sarcastic and non-sarcastic.
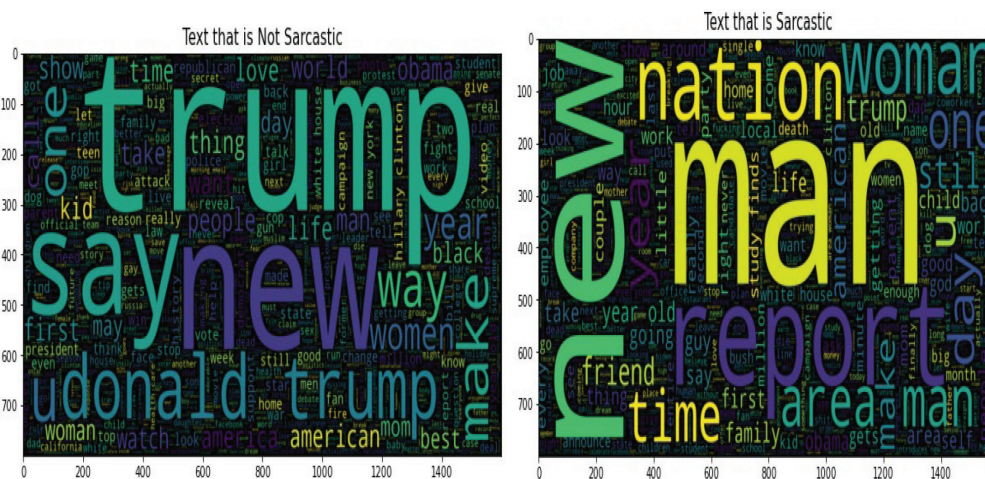

Figure 3. Shows the word clouds of datasets

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

11

3.3. Feature Extraction Vectorization Model

To provide statistical or machine learning algorithms with useful features, raw textual data must first be vectorized. Because the procedure produces a collection of numerical vectors as its outcome, it is also referred to as "vectorization." The bag-of-words (BoW) and term frequency inverse document frequency (TF-IDF) models were used in this study.

- *Bag-of-Words (BoW)*: The number of times a word appears in a text is measured using the BoW metric (Thavareesan & Mahesan, 2019).

$$BoW = occurrences\ of\ the\ word\ "w"$$

- *Term Frequency Inverse Document Frequency (TF-IDF)*: It is determined by multiplying the TF and IDF values, where TF represents the frequency of occurrence in the current document and IDF represents the overall rarity of a term.

$$IDF(w) = log\left[\frac{Quantity\ of\ documents}{Number\ of\ documents\ that\ include\ Word\ w}\right]$$

3.4. Machine Learning Techniques

The methods used in this study are described in detail below.

- *Random Forest*: It is employed in machine learning to address classification and regression problems. However, organizational contexts represent the main application. Similar to this, machine learning's random forest method creates decision trees from the examples provided before predicting the result that fit the situation the best. By combining the outcomes of several decision trees, the overfitting scenario is typically eliminated. Unlike the single decision tree method, the random forest machine learning technique also processes a wide variety of data outputs (Sentamilselvan et al., 2021).

- *AdaBoost*: This approach is regarded as the best schema in the machine learning paradigm. Alternative names for this include adaptive boosting. Adaboost is used by researchers.

- *Gradient Boosting*: For classification and regression problems, it is one of the most widely used machine learning approaches. Using this method, weak estimation techniques can be ensembled (Sentamilselvan et al., 2021).

- *Logistic Regression*: Similar to the majority of keyword-based algorithms, logistic regression comes in two flavors: classification and regression. In a binary classification framework, logistic regression simply shows the probability that the input will result in the desired output. When using a classifier model, a cutoff value is selected, and the class is determined by whether the likelihood of the inputs is higher than the threshold, in which case they will be placed in one class, or lower than the threshold, in which case they will be placed in another class. The standard threshold value of 10 is employed for this investigation (Razali et al., 2021).

- *XGBoost*: Today's most popular boosting method is called XGBoost. It is a method of implementing gradient-boosting engines. Other gradient-boosting methods cannot match XGBoost's speed or accuracy (Bagate & Suguna, 2021). It is one of the ensemble strategies that makes use of the second partial differential equation of the error function, which provides more details about the gradient's direction and how to minimize it.

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

12

- *Naïve Bayes*: It is an improvement of the Bayes theorem and a conditional probabilistic prediction method. The main benefit of a Naive Bayes classifier is that it can categorise a dataset more rapidly and with fewer training examples. The efficiency rate of the classifier is higher than that of other supervised learning methods. In this method, each attribute pair operates independently of the others (Kanakam et al., 2022).

- *Proposed Model*: The proposed approach is a hybrid one that uses two vectorization models to provide a solution to the sarcasm detection problem. The model uses BoW and TF-IDF to determine whether or not the headline text is ironic. The machine learning technique receives the classification result from both models and learns the appended sentence for categorization likelihood. The vectors are combined and then transmitted identify whether the input is sarcastic.

- In a hybrid approach, different classifiers are used to handle a number of problems and boost accuracy. The method for detecting sarcasm in text is based on random forest and naive bayes techniques. The naive bayes-random forest method is a novel optimal method for categorization issues. It fixes the problems with both strategies. A statistical learning approach called "naive bayes" is based on the bayes rule, and the assumption of independence for detecting sarcasm in text is based on the random forest and naive bayes techniques. The naive bayes-random forest method is a novel and optimal method for categorization issues. It fixes the problems with both strategies. A statistical learning approach called "naive bayes" is based on the bayes rule and the assumption of independence. According to the conditional assumption of independence, the qualities are conditionally independent given the class. As a result, the method is effective and scales effectively. The supervised learning method known as "random forest" uses the bagging operation and the random methodology of subspace. As the fundamental classification model, decision trees were used in this method. To offer variability among the base learners, a random selection of attributes was used. In consequence, on datasets with messy or unimportant data, the model can produce promising learning approaches (Onan & Tocoglu, 2020). And lastly, our proposed solution is more accurate than the other strategies we reviewed. The testing of the proposal demonstrated that it performs better than other state-of-the-art approaches in terms of accuracy.

3.5. Performance Parameters

Table 4 shows the metrics used in sarcasm detection. The effectiveness of the proposed model is assessed using two widely used machine learning metrics: Accuracy and the F1-measure (Gul et al., 2022). The metrics are summarized in Table 4.

Table 4. Criteria for performance assessment

Measures of Performance	Formulas
Accuracy	$\dfrac{TP + TN}{TP + TN + FP + FN}$
F-Measure	$\dfrac{2 * (Precision * Recall)}{(Precision + Recall)}$

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

13

Where T.P., TN, F.P., and F.N., accordingly, represent true positive, true negative, false positive, and false negative.

3.6. Proposed Algorithm

Sarcasm Text Detection Algorithm Representation for Proposed Hybrid Machine Learning Approach

Input: Training on processed data using the Hybrid Naïve Bayes - Random Forest Technique,
Output: Sarcasm text detection results,

1) Import the required library
2) **If** the model has not been trained, **then**
3) Load the dataset
4) Read the dataset
5) Check the data for sarcastic or non-sarcastic content
6) Preprocess the dataset

    o   Remove special characters and stopwords

    o   Lemmatization and tokenization

7) Dataset length distribution
8) Generate word clouds that are sarcastic or non-sarcastic
9) Data partition into the train and test set
10) Build the model
11) Sarcasm Feature extraction vectorization model (BoW, TF-IDF)
12) Sarcasm detection using techniques (Proposed Hybrid Approach, Naive Bayes, Logistic Regression, Random Forest, AdaBoost, XGBoost, Gradient Boosting)
13) Set the classification parameter
14) Save trained set
15) **Else**
16) Load trained set
17) **End**
18) Evaluate model performance
19) Obtain the result

4. Result

This section presents the experimental findings from all seven methods for sarcasm detection in news headlines using two vectorization models. Table 5 displays the dataset's categorization findings.

According to Table 5, the results demonstrate that bag-of-words (BoW) models outperform the TF-IDF model. The TF-IDF model outperforms the bag-of-words models when using the random forest technique in the two parameters.

4.1. Statistical Results

Table 5 compares various methods utilizing the BoW and TF-IDF models. This study makes use of two characteristics to gauge the effectiveness of the strategies and select the best one. The effectiveness

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

14

Table 5. An overview of the experimental data for each of the seven classification approaches' assessment parameters

Vectorization Models / Techniques	BoW		TF-IDF	
	Accuracy	F1-Score	Accuracy	F1-Score
Naïve Bayes	0.80907	0.76882	0.51572	0.60430
Logistic Regression	0.81551	0.76531	0.50598	0.60424
Random Forest	0.84696	0.76687	**0.84336**	**0.76755**
AdaBoost	0.73285	0.61005	0.56498	0.60424
Gradient Boosting	0.56483	0.60424	0.56483	0.60424
XGBoost	0.65094	0.62341	0.56648	0.60424
Proposed Hybrid Technique	**0.85369**	**0.77584**	0.53024	0.60439

of detecting sarcasm in text is assessed using several methods. Thee results of their quantitative investigation of sarcastic text identification are reported in Table 5 and the obtained parameters are compared with the parameters of other state-of-the-art techniques. Feature areas where our model outperforms other techniques:

4.1.1. Bag-of-Words Model

- *Accuracy*: Hybrid technique (NB-RF) performance shows a 0.007 % accuracy improvement with random forest, as well as 0.038 % improvement with logistic regression, 0.044 % improvement with naive bayes, 0.121 % improvement with adaboost, 0.203 % improvement with xgboost, and 0.289 % improvement with gradient boosting.
- *F1-Score*: The hybrid technique (NB-RF) outperforms random forest by 0.009 percentage points on the F1-score, and it likewise performs better when naive bayes is used, 0.01 when logistic regression is used, 0.152 when xgboost is used, and 0.165 and 0.171% respectively, when adaboost and gradient boosting are used.

4.1.2. TF-IDF Model

- *Accuracy*: An improvement of 0.313% over the hybrid approach with random forest (NB-RF). Additionally, there was an increase of 0.279 % in the accuracy using adaboost and gradient boosting, 0.277 % using xgboost, 0.328 % using naive bayes, and 0.338 % using logistic regression.
- *F1-Score*: Random forest efficiency outperforms naive bayes, xgboost, logistic regression, gradient boosting, the hybrid model (NB-RF), and adaboost in terms of f1-score by 0.163%.

The proposed hybrid strategy, which uses a bag-of-words vectorization model, outperforms all other techniques, as shown in Figure 4, which compares the accuracy and f1-score of various techniques on the dataset.

The comparison of the techniques is shown in Figure 5. By comparing the accuracy and F1-score of the different techniques' performance on the dataset, the random forest strategy applying the TF-IDF vectorization model surpasses all other strategies.
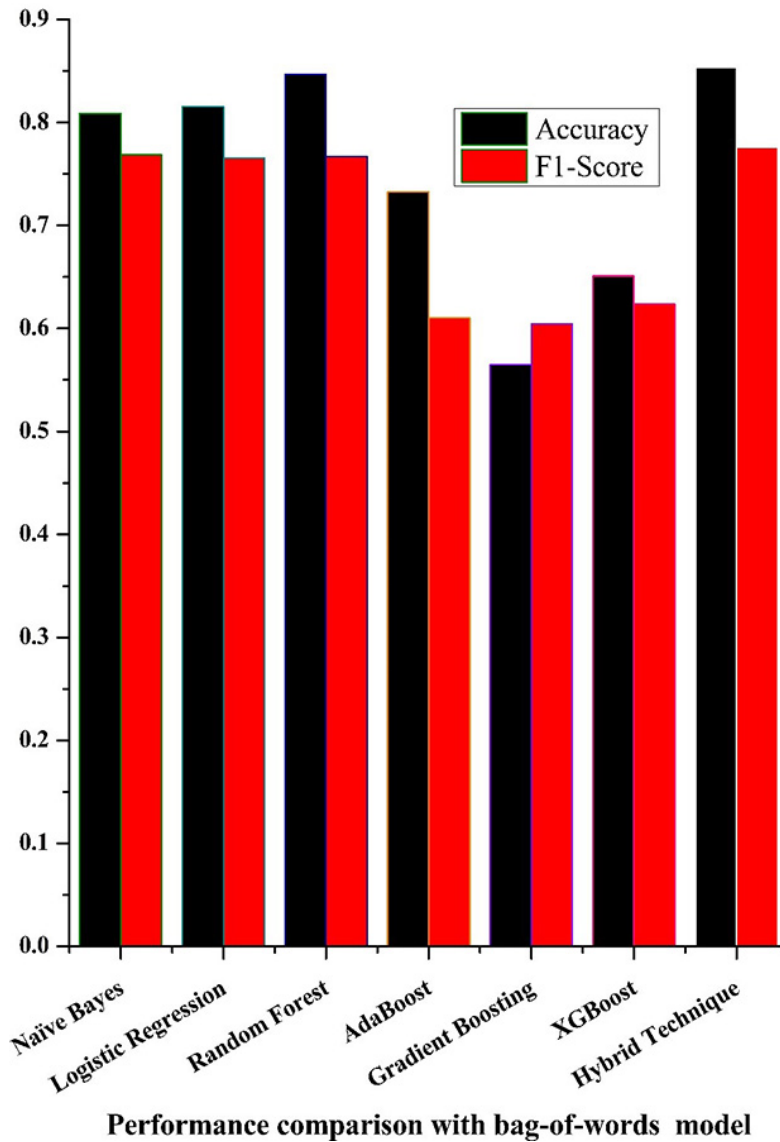
Figure 4. Comparison of various approaches using bag-of-words vectorization model
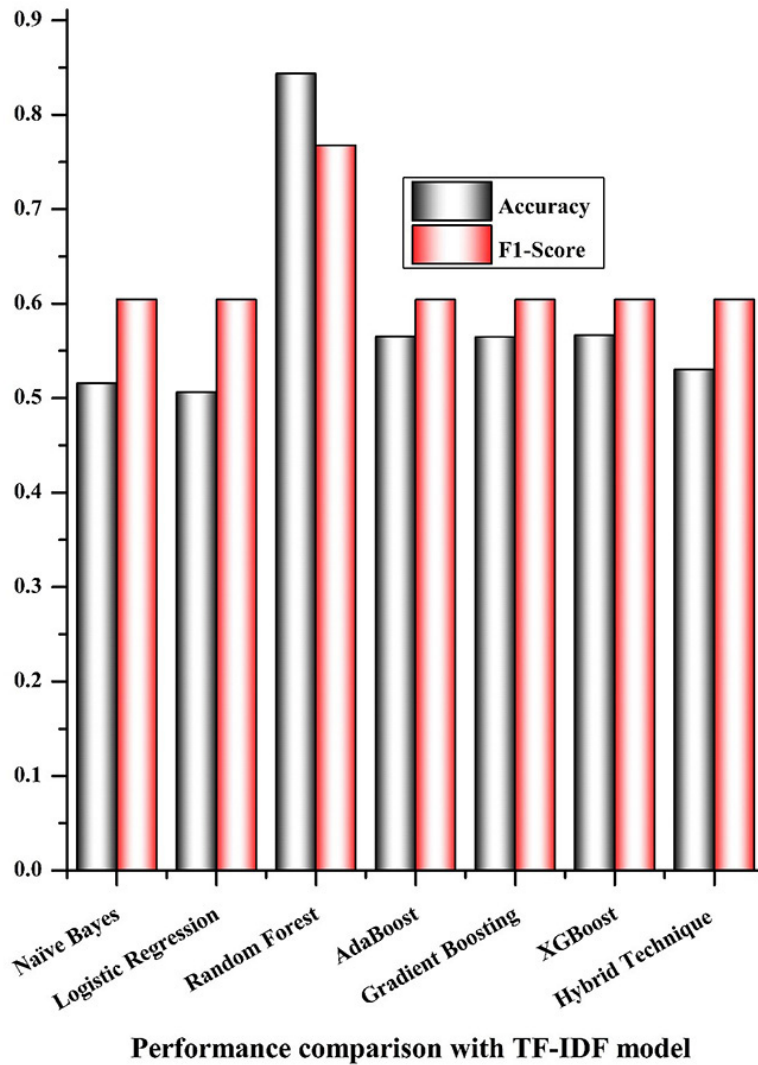
Figure 5. Comparison of the accuracy and F1-score of seven distinct approaches, using the TF-IDF vectorization model

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

17

Table 6 displays a comparison of the results. The proposed hybrid strategy in Table 6 above produces the best outcomes when compared to other methods.

Table 6. A comparative study of the findings

Ref. No.	Accuracy	F1-Score
(Farha & Magdy, 2020)	0.62%	0.46%
(Bharti, Sathya Babu, et al., 2017)	0.79%	0.72%
(Chudi-Iwueze & Afli, 2020)	0.79%	0.76%
(Novic, 2022)	0.79%	0.75%
(Bharti, Pradhan, et al., 2017)	0.48%	0.46%
(Gul et al., 2022)	0.72%	0.77%
(Pelser & Murrell, 2019)	0.69%	0.69%
Proposed Hybrid Technique	**0.85%**	**0.77%**

5. Conclusion and Future Scope

Seven alternative methods for sarcasm identification in the news headline dataset were compared in this study. Additionally, the authors conducted experiments using the BoW and TF-IDF vectorization models on news headlines. The efficacy and F1 score of sarcastic text detection algorithms were investigated through experimental methods. All techniques were also successful at classifying data with good accuracy and favorable F1-score results. Out of the seven methods, the proposed hybrid technique has the maximum accuracy and F1-score, making it the most effective technique overall. The models of both methodologies fared better than one another in text sarcasm detection. In conclusion, the BoW model's results are superior to those of the TF-IDF model. A future study will aim to improve performance in terms of accuracy, precision, recall, and f-score by expanding on the current research through the application of deep learning approaches. The results could also be improved by training with Word2vec, Doc2vec, or Pargraph2vec vectorization and Bert models rather than TF-IDF and BoW.

References

Aboobaker, J., & Ilavarasan, E. (2020). A Survey on Sarcasm detection and challenges. *Proc. of 6th Intl. Conf. on Advanced Computing & Communication Systems*, 1234-1240. https://doi.org/10.1109/ICACCS48705.2020.9074163

Abulaish, M., & Kamal, A. (2018). Self-deprecating sarcasm detection: an amalgamation of rule-based and machine learning approach. *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 574-579. https://doi.org/10.1109/WI.2018.00-35

Alexandru, D., & Gîfu, D. (2020). Tracing humor in edited news headlines. *Ludic, Co-Design and Tools Supporting Smart Learning Ecosystems and Smart Education: Proceedings of the 5th International*

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

18

Conference on Smart Learning Ecosystems and Regional Development, 187-196. https://doi.org/10.1007/978-981-15-7383-5_16

Aneja, S., Aneja, N., & Kumaraguru, P. (2022). Predictive linguistic cues for fake news: a societal artificial intelligence problem. *ArXiv*. Preprint ArXiv:2211.14505. https://doi.org/10.11591/ijai.v11.i4.pp1252-1260

Bagate, R. A., & Suguna, R. (2021). Sarcasm detection of tweets without\# sarcasm: data science approach. *Indonesian Journal of Electrical Engineering and Computer Science*, *23*(2), 993-1001. https://doi.org/10.11591/ijeecs.v23.i2.pp993-1001

Barhoom, A., Abu-Nasser, B. S., & Abu-Naser, S. S. (2022). Sarcasm Detection in Headline News using Machine and Deep Learning Algorithms. *International Journal of Engineering and Information Systems (IJEAIS), 6*(4), 66-73.

Bharti, S. K., Pradhan, R., Babu, K. S., & Jena, S. K. (2017). Sarcasm analysis on twitter data using machine learning approaches. *Trends in Social Network Analysis: Information Propagation, User Behavior Modeling, Forecasting, and Vulnerability Assessment*, 51-76. https://doi.org/10.1007/978-3-319-53420-6_3

Bharti, S. K., Sathya Babu, K., & Jena, S. K. (2017). Harnessing online news for sarcasm detection in hindi tweets. *International Conference on Pattern Recognition and Machine Intelligence*, 679-686. https://doi.org/10.1007/978-3-319-69900-4_86

Chaudhari, P., & Chandankhede, C. (2017). Literature survey of sarcasm detection. *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2041-2046. https://doi.org/10.1109/WiSPNET.2017.8300120

Chudi-Iwueze, O., & Afli, H. (2020). Detecting Sarcasm in News Headlines. *CERC*, 100-111.

Farha, I. A., & Magdy, W. (2020). From arabic sentiment analysis to sarcasm detection: The arsarcasm dataset. *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, 32-39.

Goel, P., Jain, R., Nayyar, A., Singhal, S., & Srivastava, M. (2022). Sarcasm detection using deep learning and ensemble learning. *Multimedia Tools and Applications*, *81*(30), 43229-43252. https://doi.org/10.1007/s11042-022-12930-z

Gul, S., Khan, R. U., Ullah, M., Aftab, R., Waheed, A., & Wu, T. Y. (2022). Tanz-Indicator: A Novel Framework for Detection of Perso-Arabic-Scripted Urdu Sarcastic Opinions. *Wireless Communications and Mobile Computing*, *2022*. https://doi.org/10.1155/2022/9151890

Husain, F., & Uzuner, O. (2021). Leveraging offensive language for sarcasm and sentiment detection in Arabic. *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, 364-369.

Jariwala, V. P. (2020). Optimal feature extraction based machine learning approach for sarcasm type detection in news headlines. *International Journal of Computer Applications*, *975*, 8887.

Joshi, A., Bhattacharyya, P., & Carman, M. J. (2017). Automatic sarcasm detection: A survey. *ACM Computing Surveys (CSUR)*, *50*(5), 1-22. https://doi.org/10.1145/3124420

Kanakam, R., Mohmmad, S., Sudarshan, E., Shabana, S., & Gopal, M. (2022). A survey on approaches and issues for detecting sarcasm on social media tweets. *AIP Conference Proceedings*, *2418*(1). https://doi.org/10.1063/5.0082034

Katyayan, P., & Joshi, N. (2019). Sarcasm Detection approaches for English language. *Smart Techniques for a Smarter Planet: Towards Smarter Algorithms*, 167-183. https://doi.org/10.1007/978-3-030-03131-2_9

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

19

Kumar, A., & Katiyar, V. (2019). A comparative analysis of sarcasm detection. *Int J Recent Eng Res Dev (IJRERD)*, *4*(08), 104-108.

Kumar, A., Narapareddy, V. T., Srikanth, V. A., Malapati, A., & Neti, L. B. M. (2020). Sarcasm detection using multi-head attention based bidirectional LSTM. *Ieee Access*, *8*, 6388-6397. https://doi.org/10.1109/ACCESS.2019.2963630

Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., Abdel-Basset, M., & others. (2019). Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE Access*, *7*, 23319-23328. https://doi.org/10.1109/ACCESS.2019.2899260

Liu, L., Priestley, J. L., Zhou, Y., Ray, H. E., & Han, M. (2019). A2text-net: A novel deep neural network for sarcasm detection. *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*, 118-126. https://doi.org/10.1109/CogMI48466.2019.00025

Majumder, N., Poria, S., Peng, H., Chhaya, N., Cambria, E., & Gelbukh, A. (2019). Sentiment and sarcasm classification with multitask learning. *IEEE Intelligent Systems*, *34*(3), 38-43. https://doi.org/10.1109/MIS.2019.2904691

Mandal, P. K., & Mahto, R. (2019). Deep CNN-LSTM with word embeddings for news headline sarcasm detection. *16th International Conference on Information Technology-New Generations (ITNG 2019)*, 495-498. https://doi.org/10.1007/978-3-030-14070-0_69

Misra, R. (2022). News headlines dataset for sarcasm detection. *ArXiv Preprint ArXiv:2212.06035*.

Mohammed, P., Eid, Y., Badawy, M., & Hassan, A. (2020). Evaluation of different sarcasm detection models for arabic news headlines. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2019*, 418-426. https://doi.org/10.1007/978-3-030-31129-2_38

Mykytiuk, A., Vysotska, V., Markiv, O., Chyrun, L., & Pelekh, Y. (2023). *Technology of Fake News Recognition Based on Machine Learning Methods*.

Nayak, D. K., & Bolla, B. K. (2022). Efficient deep learning methods for sarcasm detection of news headlines. In *Machine Learning and Autonomous Systems: Proceedings of ICMLAS 2021*, 371-382. Springer. https://doi.org/10.1007/978-981-16-7996-4_26

Nguyen, H., Veluchamy, A., Diop, M., & Iqbal, R. (2018). Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. *SMU Data Science Review*, *1*(4), 7.

Novic, L. I. (2022). *A machine learning approach to text-based sarcasm detection* [Master theses, City University of New York]. CUNY Academic Works.

Onan, A., & Tocoglu, M. A. (2020). Satire identification in Turkish news articles based on ensemble of classifiers. *Turkish Journal of Electrical Engineering and Computer Sciences*, *28*(2), 1086-1106. https://doi.org/10.3906/elk-1907-11

Pal, M., & Prasad, R. (2023). Sarcasm Detection followed by Sentiment Analysis for Bengali Language: Neural Network \& Supervised Approach. *2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS)*, 1-7. https://doi.org/10.1109/AICAPS57044.2023.10074510

Park, M., & Chai, S. (2023). Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques (Jan 2023). *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3294613

Parkar, A., & Bhalla, R. (2023). Analytical comparison on detection of Sarcasm using machine learning and deep learning techniques. *International Journal of Computing and Digital Systems*, *14*(1), 1615-1625. https://doi.org/10.12785/ijcds/1501114

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using
Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

20

Pawar, N., & Bhingarkar, S. (2020). Machine learning based sarcasm detection on Twitter data. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 957-961. https://doi.org/10.1109/ICCES48766.2020.9137924

Pelser, D., & Murrell, H. (2019). *Deep and Dense Sarcasm Detection*. http://arxiv.org/abs/1911.07474

Razali, M. S., Halin, A. A., Ye, L., Doraisamy, S., & Norowi, N. M. (2021). Sarcasm detection using deep learning with contextual features. *IEEE Access*, *9*, 68609-68618. https://doi.org/10.1109/ACCESS.2021.3076789

Sentamilselvan, K., Suresh, P., Kamalam, G. K., Mahendran, S., & Aneri, D. (2021). Detection on sarcasm using machine learning classifiers and rule based approach. *IOP Conference Series: Materials Science and Engineering*, *1055*(1), 12105. https://doi.org/10.1088/1757-899X/1055/1/012105

Thavareesan, S., & Mahesan, S. (2019). Sentiment analysis in Tamil texts: A study on machine learning techniques and feature representation. *2019 14th Conference on Industrial and Information Systems (ICIIS)*, 320-325. https://doi.org/10.1109/ICIIS47346.2019.9063341

Trystan, S., Matiushchenko, O., & Naumenko, M. (2021). Method Of Recognition Sarcasm In English Communication With The Application Of Information Technologies. *CEUR, 3200*.

Verma, P., Shukla, N., & Shukla, A. P. (2021). Techniques of sarcasm detection: A review. *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 968-972. https://doi.org/10.1109/ICACITE51222.2021.9404585

Yin, C., Chen, Y., & Zuo, W. (2021). Multi-task deep neural networks for joint sarcasm detection and sentiment analysis. *Pattern Recognition and Image Analysis*, *31*, 103-108. https://doi.org/10.1134/S105466182101017X

Ying, Y., Mursitama, T. N., & others. (2021). Effectiveness of the News Text Classification Test Using the Naïve Bayes' Classification Text Mining Method. *Journal of Physics: Conference Series*, *1764*(1), 12105. https://doi.org/10.1088/1742-6596/1764/1/012105

Neha Singh and Umesh Chandra Jaiswal

Sarcasm Text Detection on News Headlines Using Novel Hybrid Machine Learning Techniques

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 13 (2024), e31601
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

21