Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 14, Issue. 6, June 2025, pg.120 – 126

Sarcasm Detection in News Headline using Stack Ensemble Technique

Mohammed Ali Kawo[1]; Dr. Garba Muhammad[2]; Dr. Danlami Gabi[3]; Dr. Musa Sule Argungu[4]

[1]Department of Computer Science, Federal University Gusau, Zamfara State, Nigeria
[1,2,3,4]Department of Computer Science, Kebbi State University of Science and Technology, Aliero, Nigeria
[1] ali_mak2001@yahoo.com; [2] garbamga@gmail.com; [3] gabsonley@gmail.com; [4] sm279arg@gmail.com

DOI: https://doi.org/10.47760/ijcsmc.2025.v14i06.012

Abstract: The challenges and benefits of sarcasm detection in opinion mining has led to curiosities in researches; to distinguish sarcasm in a sentence has become strenuous task for humans and it causes a lot of conflicts when analysing sentiments, text analysis and in computational linguistics. The individual base classifiers tend to predict in a biased form with a decline in accuracy scores. Sarcasm obstructs a lot of effective models in machine learning techniques. This research stresses in addressing the challenges of sarcasm and to effectively generate a Stack ensemble technique to detect sarcasm in news headline dataset. Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Tree (DT) and a Logistic Regression (LR) Meta-Stack ensemble classifier were classified with the following metrices such as accuracy, precision, recall and F1-scores on a news headline dataset from kaggle.com to ascertain the best performance classifier in dealing with sarcasm detection. The stack ensemble model outperforms all other classifiers by achieving a significant accuracy score of 87%, precision for no-sarcasm is 85%, while for sarcasm is 81%, recall for no-sarcasm is 86%, while for sarcasm is 79% and the F1-score for no-sarcasm is 85% and for sarcasm is 80%. These revealed results have depicted how effective stack ensemble model performed in detecting sarcasm in news headline dataset. However, it is essential to exploit other advanced techniques such as deep learning in respect of sarcasm, so as to improve in accuracy scores.

Keywords: Stack Ensemble, Confusion Matrix, Sarcasm, Sentiment Analysis

1. INTRODUCTION

In using statements that generally designates the real words in order to ridicule someone or show despise for individuals is known as sarcasm. The essential intention is to recognize sarcasm in every lingual statement that conveys any form of intimidation to an affair (Rajani, 2023). Sentiment analysis is a domain of natural language processing that desire to deduce the opinion immersed in a textual data. Regardless of its benefits, sentiment analysis is tremendously difficult in nature due to various problems and challenges (Bordoloi & Biswas, 2023). Sarcasm is a special form of sentiment of opinion mining but it's not exactly sentiment analysis. Even though

sarcasm is a problem in sentiment analysis because it leads to fake information with an intent to overturn sentiment from the expression that they seek to report.

Sarcasm is speech that are in the form of positive statement that represent negative or undesirable character. Sarcasm detection has many implications by rendering sentiment analysis; natural language processing and text classification to be bias in most cases (Sarsam et al., 2020).

According to Sarsam et al. (2020), in other to characterize sarcasm in a social platform data, conflict between positive situation and negative situation will be interpolated. In this regards sarcasm detection plays an exciting role in natural language processing, sentiment analysis and mutual communication between individuals and machines.

This research is to detect sarcasm using stack ensemble technique on news headline dataset to ascertain which headline is sarcastic and which is not. Four different classifiers such as Decision Trees (DT), K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Naïve Bayes (NB) were compared with the intended proposed model which is Stack Ensemble Classifier to decide which classifier outperformed effectively.

In text classification, features such as regularly words occurrence are most likely to be considered (Rajani, 2023). These features and many others are tackled in the preprocessed stage of the model. The significant importance of sarcasm detection in social platforms, text mining and diverse datasets is severely advanced compared to just sentiment analysis. Some of the chaos that sarcastic data inherits are as follows:

i) Words misinterpretation, which most at times leads to ambiguity by given different meanings to a single word.

ii) Most sarcasm carries a detailed emotional quality that cannot be easily apprehended in writing. Except if it is accompanied by a face-to-face interaction.

iii) Context sensitivity, every culture has a way of perceiving meanings to any given word in communication. What some may think its sarcastic, may be earnest in others which can lead to confusion.

iv) Most precisely when sarcastic words are frequently used in expression, it tends to become more difficult to interpret the real meaning of some words or statements which may lead to bias translations.

2. RESEARCH METHOD

This research presents a model depicted in figure 1 below, which consist of four stages for detecting sarcasm using stack ensemble techniques. The uppermost stage is the data collection layer, where dataset is collected from kaggle.com; second stage encompasses data preprocessing layer where feature engineering, noise reduction and tokenization of dataset is done, while the third stage is data validation stage where 5-fold cross validation is used on the dataset and the final stage is where all classifier are pipelined to form a meta classifier to be applied on the dataset for final sarcasm prediction. The illustration is also shown in Figure 1 below.

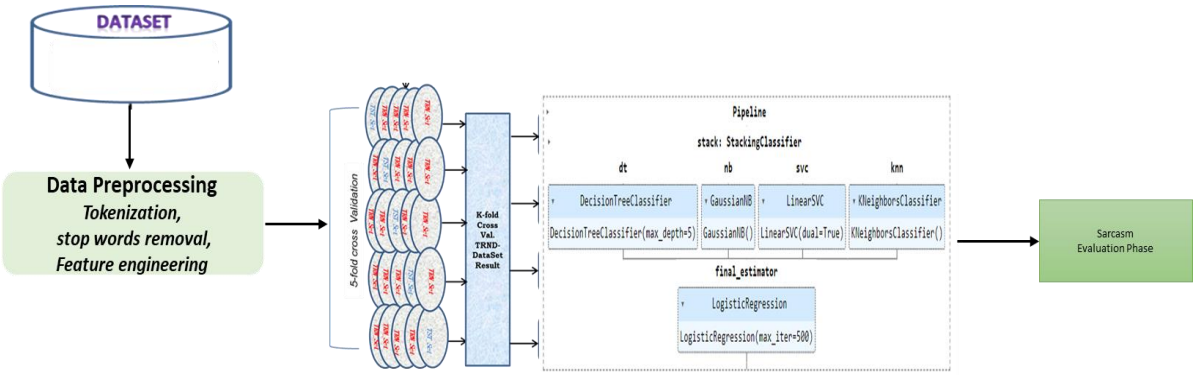

Fig 1: Proposed Model for Sarcasm Detection

2.1. Dataset

The research utilizes a news headlines dataset that was retrieved from kaggle.com which contains 28619 sentences with non-null values, the dataset is of JSON source types. The proposed model which is stack ensemble technique will be used to detect sarcasm on the dataset. Table 1 below depicts few information of the news headlines sarcasm dataset from Jupyter source code. while figure 2 below display the number of new headlines by is sarcastic or not in a bar chart.

TABLE 1:
News headlines sarcasm data

```
: data
t[7]:
```

	article_link	headline	is_sarcastic
0	https://www.huffingtonpost.com/entry/versace-b...	former versace store clerk sues over secret 'b...	0
1	https://www.huffingtonpost.com/entry/roseanne-...	the 'roseanne' revival catches up to our thorn...	0
2	https://local.theonion.com/mom-starting-to-fea...	mom starting to fear son's web series closest ...	1
3	https://politics.theonion.com/boehner-just-wan...	boehner just wants wife to listen, not come up...	1
4	https://www.huffingtonpost.com/entry/jk-rowlin...	j.k. rowling wishes snape happy birthday in th...	0
...	...	...	...
26704	https://www.huffingtonpost.com/entry/american-...	american politics in moral free-fall	0
26705	https://www.huffingtonpost.com/entry/americas-...	america's best 20 hikes	0
26706	https://www.huffingtonpost.com/entry/reparatio...	reparations and obama	0
26707	https://www.huffingtonpost.com/entry/israeli-b...	israeli ban targeting boycott supporters raise...	0
26708	https://www.huffingtonpost.com/entry/gourmet-g...	gourmet gifts for the foodie 2014	0

26709 rows × 3 columns

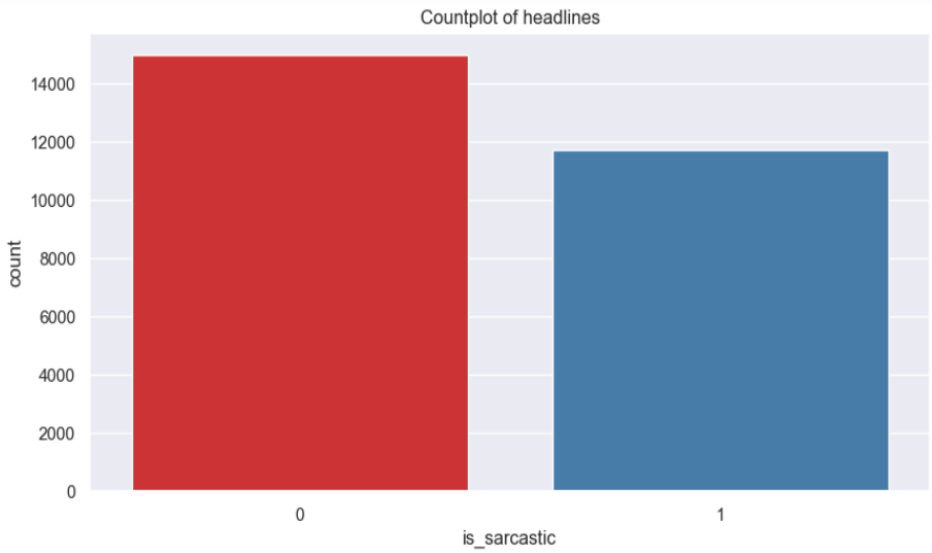

Fig 2: Bar chart checks for number of headlines by is sarcastic

The above Bar chart provides information on the news headline data showing the figures with sarcasm and without sarcasm, where the zero label with red colour bar is without sarcastic and label one with blue colour bar is sarcastic.

2.2. Data Transformation and Feature Selection

In other to train any model, data transformation and feature selection are crucial steps in data preprocessing. Tokenization, stemming and vectorization of dataset are considered in most researches. In this research, Tokenization plays an important role by breaking down text sequence into smaller units called tokens, these tokens range in size from little characters to full words (Allam et al., 2021). The extraction of stop words to eliminate words without polarity so as to have accurate results was also carried out in the preprocessing stage. In sarcasm detection and sentiment analysis, feature extraction is an essential phase in the preprocessing of the dataset, it entails selecting or adjusting the most relevant and instructive aspects (variables or attributes) from the raw dataset in order to create an enhanced model (Bahrawi, 2019). The aim is to reduction the dimensionality of the dataset while keeping all of its significant information intact.

After cleaning and preprocessing of the dataset, data labelling is also checked to verify if both classes are of the same sample size. Figure 3 bar chart displays a balanced dataset for news headlines sarcasm where both the minority and majority classes are homogeneously equal.

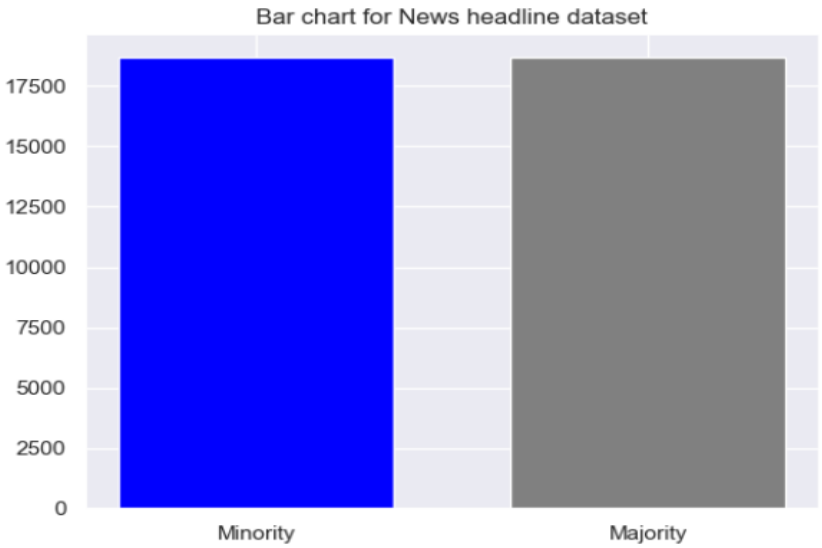

Fig 3: Bar chart for a balanced news headlines sarcasm

2.3. Ensemble technique

The ensemble approach is a machine learning methodology that combines the predictions of several models to improve forecasting accuracy and robustness (F. Zhou et al., 2023). It uses the collective intelligence of the ensemble to reduce errors or biases that can be present in individual models. Ensemble methods train multiple learners to solve the same problem. Pintelas & Livieris (2020), unlike individual base learning approaches, which aim to create a single learner from training data, ensemble methods attempt to create a group of learners and merge them. Using the diversity of classifiers, ensemble learning increase's reliability and generalization capacity. The ensemble model (classifier) is less prone to overfitting because of the diversity of the base model (Feng et al., 2023). The ensemble technique is generally categorized into four major models as shown in figure 4 below.

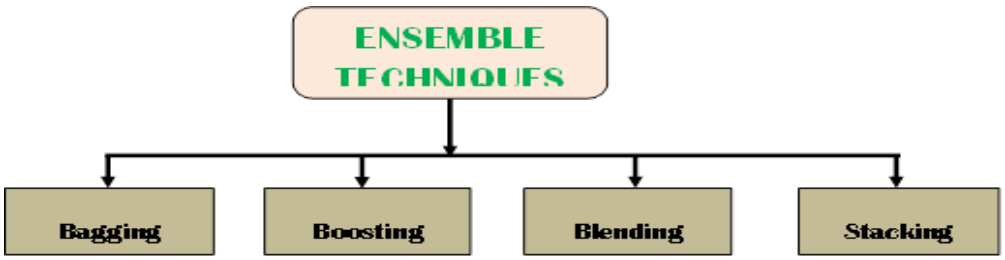

Fig 4: Ensemble Techniques

2.3.1 Bagging Ensemble Technique

The name Bagging was originated from the abbreviation of Bootstrap AGGregatING. As the term suggests, aggregation and bootstrap are the two essential components of bagging (Pintelas & Livieris, 2020). Bagging is a technique that generates a generalized outcome by combining the output of multiple models. However, the dataset is not given to each model individually. As an alternative, subsets of the original dataset are produced using the bootstrapping technique (Jawale, 2019).

2.3.2 Boosting Ensemble Technique

A family of algorithms known as "Boosting" has the capacity to turn weak learners into strong learners. It makes intuitive sense that a weak learner performs about as well as a random guess, whereas a great learner performs almost exactly as intended (Dietterich, 2000). A boosting strategy is used to tackle problems where every new model in the boosting process attempts to fix the flaws in the prior model. As a result, the subsequent models rely on the earlier model (Obaidat et al., 2022).

2.3.3 Blending Ensemble Technique

The blending is a similar strategy to stacking in terms of prediction accuracy. Predictions are solely made on the validation set, in contrast to stacking. These forecasts are applied to the test set and are used to train and construct the meta model. Prior to blending, the training dataset is divided into two sets: the validation set and

the base-training set (Kumar et al., 2023). Using this base-training dataset, the base models are trained. Predictions are performed on the validation set using these trained models. The meta-classifier uses these predictions as features, and the test set is utilized to generate final predictions (Jawale, 2019).

2.3.4 Stack Ensemble Technique

A sophisticated technique for ensemble learning called stacking involves stacking the predictions made by each individual model and using that stack as input to train the meta-model. Predictions are made on the test set using this meta-model (Dietterich, 2000). According to Feng et al. (2023) stated that, stacking basic concept involves training a number of base classifiers with a single training set, and then having each base classifier create a new training set in order to train the meta-classifier.

2.4 Experimental Settings

The entire experiment was performed on HP Laptop, AMD Ryzen 3, 3250U with Radeon 2.60Ghz graphics and a minimum of 8 GB RAM, on Windows 10 Pro, X64 bits-based processor version 22H2. The language used in carrying out the experiments in this research is an open-source Python distributor popularly known as Jupyter notebook (Version 6.5.4 IDE) from Anaconda, where Panda and NumPy are used for handling the datasets and dataframe, while libraries like scikit-learn, matplotlib and seaborn for training, testing, mathematical calculation and visualization of the entire results.

3. RESULTS AND DISCUSSION

Stack ensemble is an advanced ensemble learning technique where the predictions from individual model are stacked and used as input to train the meta-model. Due to the fact that sarcasm is largely used in most social platforms to ridicule or denounce in a difficult manner that even human barely detect what it really meant. The metaphorical nature of sarcasm makes it very challenging for sentiment analysis (Ladoja & Afape, 2024). Table 2 below displays a performance metrics results on sarcasm detection using news headline dataset for stack ensemble model and four individual base classifiers.

TABLE 2:
Performance evaluation on Stack ensemble with base classifiers on news headline dataset for sarcasm.

Classifiers	Accuracy	Precision	Recall	F1-score
Naïve Bayes	84%	84% 83%	89% 79%	86% 82%
S-V-Machine	82%	83% 80%	85% 78%	84% 79%
Decision Tree	75%	77% 71%	77% 71%	77% 71%
K-N-Neighbour	63%	61% 71%	93% 25%	74% 37%
Stack Ensemble	87%	85% 81%	86% 79%	85% 80%

In table 2 above, the stack ensemble technique outperformed all other classifiers with the accuracy score of 87 percent followed by naïve bayes with 84 percent, while support vector machine has an accuracy score of 75 percent then k-nearest neighbour with 63 percent respectively. The F1-score, recall and precision result in the table portray excellent scores for both sarcastic and non-sarcastic news headlines which indicates the maximum usage of the news podium.

The figure 5 below, depict the bar chart result from table 2 respectively, for clearer visualization and effective comparison of the classifiers at a glance.

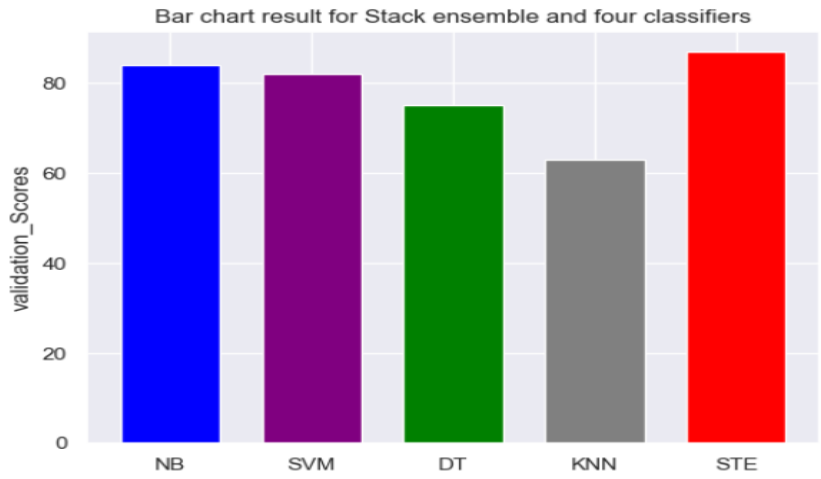

Fig 5: Bar Stack ensemble and four classifiers on news headline for sarcasm

The confusion matrix for stack ensemble model in figure 6 below, provides additional insight in determining the strength and weakness of the stack ensemble model. And from all indication the model performance is effectively remarkable.

According to (Sarsam et al., 2020), in other to characterize sarcasm in a social platform data, conflict between positive situation and negative situation will be interpolated. In this regards sarcasm detection plays an exciting role in sentiment analysis.

This document is a template. An electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the conference publications committee as indicated on the conference website. Information about final paper submission is available from the conference website.

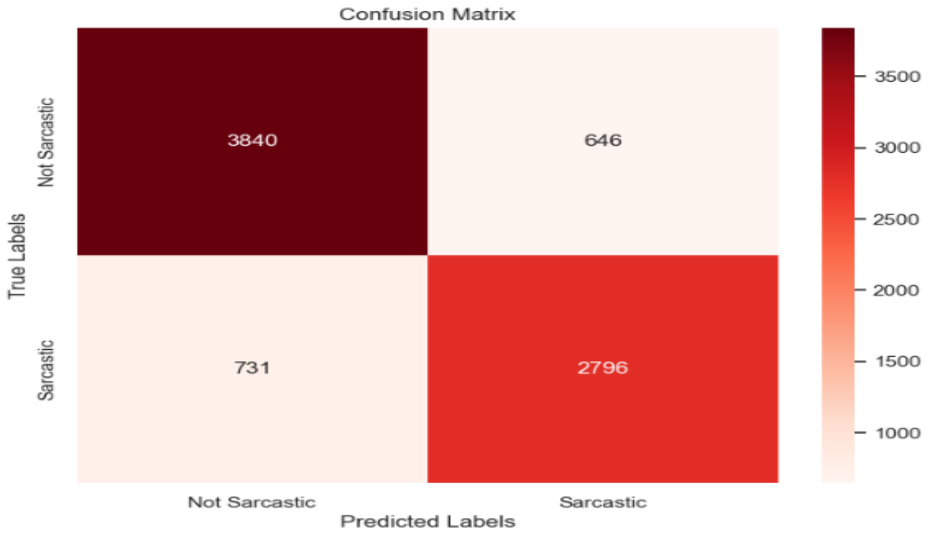

Fig 6: Confusion Matrix for Stack ensemble technique on news headline for sarcasm

4. CONCLUSION

In this research, stack ensemble technique was used in detecting sarcasm in a news headline dataset through the utilization of machine learning algorithms such as Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Tree (DT) and logistic Regression (LR) as Meta-Stack Ensemble classifier. The outcome of the result in this research contribute to the conclusion that stack ensemble model is a benchmark for sarcasm detection. This research is significantly important for several advancements which includes:

i) Enhancement in the overall analysis accuracy by combining the characteristics of numerous base models, frequently employing alternative algorithms or feature representations.

ii) In employing various based models which have been trained on various data subsets or using various methodologies. Due to its diversity, the research will be more flexible and more capable in managing different sarcasm and sentiment analysis tasks.

iii) By lessening the influence of individual model flaws or biases, the research will help in increasing the overall robustness of the poor performance of the base model by leveraging the capabilities of other models.

iv) It will also play a crucial role in handling diverse and noisy datasets. By aggregating predictions from multiple models, they can capture different aspects of the underlying data distribution, making it suitable for different types of problems and providing a more stable and reliable output.

ACKNOWLEDGEMENT

With the outstanding performance obtained by the stack ensemble technique in news headline dataset for sarcasm detection, it could be a useful tool for real world applications by providing aid in sentiment analysis suggestions obtained from predicted results to make clearer decisions for an appropriate action. The fact still remains that stack ensemble technique has the strength of handling complex regression or classification problems regardless of the dataset type. In the future, ensemble techniques and deep learning methods should be compared in predicting sarcasm, sentiment analysis and so on in other to standardize a model for detecting sarcasm and sentiment analysis.

REFERENCES

[1]. Allam, A. H., Abdallah, H. M., Amer, E., & Nayel, H. A. (2021). Machine Learning-Based Model for Sentiment and Sarcasm Detection. *WANLP 2021 - 6th Arabic Natural Language Processing Workshop, Proceedings of the Workshop*, *Ml*, 386–389.

[2]. Bahrawi, B. (2020). *SENTIMENT ANALYSIS USING RANDOM FOREST ALGORITHM ONLINE SOCIAL SENTIMENT ANALYSIS USING RANDOM FOREST ALGORITHM-*. *January*, 5–10.

[3]. Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. In *Artificial Intelligence Review* (Vol. 56, Issue 11). Springer Netherlands. https://doi.org/10.1007/s10462-023-10442-2

[4]. Dietterich, T. G. (2000). Ensemble methods in machine learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1857 LNCS*, 1–15. https://doi.org/10.1007/3-540-45014-9_1

[5]. Feng, W., Gou, J., Fan, Z., & Chen, X. (2023). An ensemble machine learning approach for classification tasks using feature generation. *Connection Science*, *35*(1). https://doi.org/10.1080/09540091.2023.2231168

[6]. Jawale, S. (2019). *Sentiment Analysis using Ensemble Learning*. *May*.

[7]. Kumar, S., Kaur, N., Kavita, & Joshi, A. (2023). *Tweet sentiment analysis using logistic regression*. *July*, 332–336. https://doi.org/10.1049/icp.2023.1801

[8]. Ladoja, K. T., & Afape, R. T. (2024). *Sarcasm Detection in Pidgin Tweets Using Machine Learning Techniques*. *17*(5), 212–221. https://doi.org/10.9734/AJRCOS/2024/v17i5450

[9]. Obaidat, M. A., Alexandrou, A., & Sanacore, S. (2022). *Machine Learning Stacking Ensemble Model for Predicting Heart Attacks Machine Learning Stacking Ensemble Model for Predicting Heart Attacks*. *August*.

[10]. Pintelas, P., & Livieris, I. E. (2020). Ensemble Algorithms and Their Applications. In *Ensemble Algorithms and Their Applications*. https://doi.org/10.3390/books978-3-03936-959-1

[11]. Rajani, B. (2023). *INTELLIGENT SYSTEMS AND APPLICATIONS IN Design and Analyze the Machine Learning Based Sarcasm Prediction Model for Social Media Context*.

[12]. Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., & Wright, B. (2020). Sarcasm detection using machine learning algorithms in Twitter: A systematic review. *International Journal of Market Research*, *62*(5), 578–598. https://doi.org/10.1177/1470785320921779

[13]. Zhou, F., Fan, H., Liu, Y., Zhang, H., & Ji, R. (2023). Hybrid Model of Machine Learning Method and Empirical Method for Rate of Penetration Prediction Based on Data Similarity. *Applied Sciences (Switzerland)*, *13*(10). https://doi.org/10.3390/app13105870

[14]. Zhou, Z.-H. (2012). Ensemble Method Foundation and Algorithms. In *Taylor & Francis Group LLC, CRC Press. A chapman & Hil Book* (Vol. 13, Issue 2).

