National College of Ireland

Hybrid Approaches to Sarcasm Detection in Social Media: Comparing Rule-Based, Statistical, and Deep Learning Models

Research Project
MSc Artificial Intelligence

Sayali Thakur
Student ID: x23139901

School of Computing
National College of Ireland

Supervisor: SHERESH ZAHOOR

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Sayali Thakur
Student ID:	x23139901
Programme:	MSc Artificial Intelligence
Year:	2024
Module:	Research Project
Supervisor:	SHERESH ZAHOOR
Submission Due Date:	12/12/2024
Project Title:	Hybrid Approaches to Sarcasm Detection in Social Media: Comparing Rule-Based, Statistical, and Deep Learning Models
Word Count:	5958
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Sayali Machhindra Thakur
Date:	27th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Hybrid Approaches to Sarcasm Detection in Social Media: Comparing Rule-Based, Statistical, and Deep Learning Models

Sayali Thakur
x23139901

Abstract

Instagram, YouTube, or Twitter—people regularly use sarcasm in their comments, processing sarcasm as a text-emoji interaction, which creates difficulties for sarcasm detection as a computational challenge. This research endeavour therefore seeks to overcome this challenge in the following manner: We assess the performance of several methods for identifying sarcasm that hail from various categories: rule-based systems, traditional machine learning, and the recent complex deep learning models. The first goal is to investigate how successful these models are in recognizing if the sarcastic comment is polite or rude, understanding a set of data from Twitter encompassing textual content and emojis. the study shows the relationship between the forms of words and contextual signals and the fact that state-of-the-art models are more effective than previous purely language-based ones in addressing this relationship. The findings show that incorporating text and emoji features improves the model's recognition and assessment of sarcasm. This work offers theoretical suggestion concerning the linguistic and contextual characteristics of sarcasm on one hand and practical implications for sentiment analysis and social media moderation on the other hand. These features have become the object of future research with regard to the enhancement of cross-domain adaptability and refining model efficiency.

Keywords: **Sentiment Analysis, Social Media, Text Analysis, Emoji Analysis**

1 Introduction

Recognition of sarcasm in social media has become relevant as a topic that needs to be investigated because of the active expansion of social media-based communication networks and their growing impact on society. Social media today, such as Instagram, Twitter, YouTube, and Facebook, experience millions of people interacting daily, not excluding sarcasm. Sarcasm, where one says the opposite of what he or she actually means, is a problem for automatic systems as it involves linguistic, contextual and at times even gestural hints like emoticons. Due to the need for better product, particularly in brand protecting, sentiment analysis systems that include sarcasm detection are a useful element in the systems used for mental health, content moderation, and more.

Currently, sarcasm detection is a very technical problem and still an open case despite the developments in natural language processing and more advanced machine learning.

Authors such as Maynard and Greenwood (2014) and Greenwood like to remind us that sarcasm is one of the ways that cannot be identified using traditional sentiment analysis, which is based on a lexical approach and word lists. Further, in their study, Castro et al. (2019) have shed light on the importance of emojis in improving model performance for sarcasm identification. However, a significant gap has been left wide open in the kind of sarcasm a model should be able to identify and categorize into refined categories like polite and rude sarcasm, which are very common on Twitter. To fill this gap, more innovation and implementation of text associated with emojis have been recommended to enhance detection.

Hence there is a need to study sarcasm detection due to the fact that it interfaces with various aspects in both theoretical and practical fields. For example, some businesses use sentiment analysis for customer feedback and might get an incorrect idea because of sarcastic main and feedback. Likewise, a failure to identify sarcasm hinders mental health evaluations, where any misinterpretation of language can be a problem. To address these challenges, this study will seek to use a multimodal approach for comparison since the use of textual and visual cues has been identified as a promising way of dealing with the complexity of sarcasm detection. As the related features comprise emojis, shifts in the sentiment polarity, and the organization of language constructs, they advance the determinant of detection outcomes. As in recent work of Mishra et al. (2021) revealed that the integration of these features together will improve the methodology for handling and classification of sarcastic content.

This study explores how machine learning models—spanning rule-based, statistical, lexical, machine-learning-based, and deep-learning-based approaches—perform in processing linguistic and contextual features in the context of sarcasm detection. Furthermore, it investigates how these models distinguish between different types of sarcasm, such as polite versus rude sarcasm, using Instagram comments enriched with text and emojis.

The structure of this paper is as follows: Section 2 focuses on the related work with a detailed examination of the prevailing literature on sarcasm detection in historical and current studies, as well as in both conventional and modern approaches. Section 3 explains the method employed in this work in regard to sarcasm detection and the models utilised in the different approaches. In section 4, the authors go further to show and describe the system implementation: data preprocessing, model selection, and feature extraction. The details of the results, which include the quantitative assessment of the models and their corresponding metrics, are provided in Section 5. Lastly, Section 6 presents the summary of all the results to give a conclusion of the study and a preview of future research endeavors.

2 Related Work

Since people are unique in expressing themselves to one another not only by text but also by displaying icons like emojis, sarcasm recognition from social media posts has emerged as a research topic of great importance. Previous studies in this area have mainly shed light on text analytics alone or emoji analytics alone, with relatively few attempts made to merge the two to enhance the performance of our sarcasm detection model.

2.1 Analysis in Sarcasm Detection

Sarcasm detection in its initial stages was much aligned with the analysis of the text. Bouazizi and Otsuki Ohtsuki (2016) proposed a pattern-based sarcasm detection approach for large-scale Twitter data. Their approach got the accuracy of 83.1% by utilizing the most careful lexical features and did not think about multidimensional parts like emojis. Maynard and Greenwood (2014) explained why non-content-based models of lexicon-based sentiment analysis fail to work for sarcasm detection. They claimed that sarcasm understanding involves contextual as well as emotional aspects, thus the gap in many of the state-of-the-art models.

In order to improve the detection of sarcasm in social media comments, Rendalkar and Chandankhede (2018), implemented emotion detection methods. While showing enhanced performance by including emotional signals, their work covered only textual characteristics and failed to take into account emojis, which frequently act as the signal of sentiment. For instance, imbalanced classification in sarcasm was tackled by multi-strategy ensemble learning as postulated by Zhang et al. (2020). Nevertheless, their method only operated at the textual features; however, it was superior to several benchmarks.

Some of the recent works, like Kumar and Garg (2019) introduced more sophisticated techniques of feature engineering in sarcasm detection work. Their work put much focus on the features of language but failed to capture the relationship between the text and emojis. Hazarika et al. (2018) took the work further by employing deep learning techniques to address the Discourse level: Sarcasm. However, their focus was only on textual features, and more work can be done on how the integration of these modes is done.

2.2 Emoji Analysis in Sarcasm Detection

Nowadays emojis cannot be considered a mere addition to the text but an essential part of it, as they give an additional shade of the meaning. Describing the ideas of Chauhan et al. (2022a,b) , the authors argued about the emoji-aware multi-task framework for sarcasm detection, where emojis are helpful for the sentiment analysis. They discovered that emojis uphold model performance due to the inclusion of the emotions that go with them. Likewise, Felbo et al. (2017) have used emojis as features in several domains and achieved substantial enhancement of sarcasm detection. But none of these investigations analyzed the use of emojis and texts for their semantic integration.

Similarly, extending sarcasm detection frameworks by Castro et al. (2019) experimented with emojis as additional features to sarcasm detection frameworks. In their study, they were able to show that they were able to improve the rate of sarcasm detection by 29% using emoji-based emotional information. However, their work lacked three issues of the study, which include the distinction between polite and rude sarcasm, which is significant when detecting sarcasm.

Zhang and Chen (2024)presented the EPE model as a fusion-based emoji sentiment analysis model. Thus, their model outperformed when pre-trained emoji features were combined with contextual text embeddings. Multifeature Fusion Sentiment Analysis was presented by Tang et al. (2024) under the name of EMFSA—Emoji-based Model. Their kind of cross-attention mechanisms to integrate the text and emoji characteristics showed signs of improvement but failed in sarcasm detection.

Other recent work by Meriem et al. (2021) was devoted to emoji-based features for

sentiment analysis but did not consider how these features affect the set of known text-based sentiment features to enhance sarcasm detection.

2.3 Multimodal Approaches for Sarcasm Detection

The log-linear model combining of text and emojis hence presents a milestone achievement in the sarcasm recognition process. Further, Schifanella et al. (2016) presented text and image based multimodal sarcasm detection. Despite their work highlighting cross-modal interactions, they did not include emojis in their study. Recently, Mishra et al. (2021) suggested that the sarcasm could be detected using texts, emojis, and images simultaneously. Despite this, their approach had the following shortcomings in analyzing the relation between the text and emojis: The authors approach of analyzing the two papers was sufficiently effective, but it lacked a comprehensive evaluation of the relationship of text to emojis One tree of their unique approach was that through their analysis of the two papers, they did not consider a sufficient evaluation of how emojis interacted with text.

There has been another paper, referred to by Zhang et al. (2020), in which a multimodal sarcasm detection framework based on textual and visual cues were proposed. During the application of their method, emojis were not considered, although their method showed excellent performance on image-text relations. In the study by Razali et al. (2017)Razali et al. (2017), the authors underlined the necessity of considering multimodality in sentiment analysis; however, the authors focus only on textual characteristics.

About the same time, Sun et al. (2019) used deep contextual models to enhance the detection of sarcasm in dialogues. Their work showed how it is possible to use both low and high-level characteristics at the same time but no emojis. Hazarika et al. (2018) published a work in 2018 that was to enable further researching in the field of complex multimodal frameworks and insights into the inter- and intra-modal dynamics regarding sentiment analysis.

2.4 A Gap in the Literature: Text and Emoji Fusion for Sarcasm Detection

While there has been significant development in both textual and emoji analysis, far less effort has been expended to systematically experiment with the integration of textual and emoji features for sarcasm identification. Almost all current models either consider the text and emoji features while excluding the textual features of the emojis or completely ignore the graphical content of emojis. This is a major void because sarcasm contains textual and non-textual elements and hence both are features that must be incorporated in this process. For instance, sarcastic expressions may be much easier if the textual tone is accompanied with an ironic or exaggerated emoji; however, many models do not consider this.

Besides, there is no distinction between polite and rude sarcasm in the current models either. While Castro et al. (2019) and Felbo et al. (2017) have pioneered research in the context of emoji usage, no prior work has focused on emoji's usage to identify the various types of sarcasm. Similarly, Chauhan et al. (2022a) showed that emojis are helpful in the detection of sarcasm, but it was not investigated how emojis connect with textual characteristics.

In fact, some research work like those provided by Sentamilselvan et al. (2021a,b) and Liu et al. (2014) has tried to address this but it is often a shallow integration. They either analyze these features distinctly, or simply do not give enough attention to the dependency between these features in sarcasm identification. Currently, there are no concrete models that address how these modalities can work in conjunction with one another in a way that was not limiting their potential during sarcasm detection.

To this end, this study fills this gap through a multimodal framework of redundant linguistic features and emojis for sarcasm identification. The proposed approach discusses the joint impact of both the features, which in turn offers understanding based features, as well as how the features work together to improve the detection of sarcasm, including polite and rude sarcasm. More precisely, this research aims at assess whether and to which extent the incorporation of text and emoji features can help to enhance sarcasm identification in social media.

2.5 Summary

In conclusion, the findings of this survey contribute to the sarcasm identification problem by enhancing text and emoji analysis while identifying the existing voids in multimodal integration and differentiation of different kinds of sarcasm. This research seeks to fill these gaps by introducing a single framework for integrating text and emoji features. This way, the study aims at enhancing the combined performance of the modalities under investigation for enhancing the accuracy of the sarcasm recognition while also revealing the specificities of the polite and rude sarcasm. The proposed framework is expected to provide inspiration for executing effective sarcasm detection approaches that are yet to be implemented, at the same time creating a foundation on which scholars can built consecutive modes of sentiment analysis involving other forms of communication.

3 Methodology

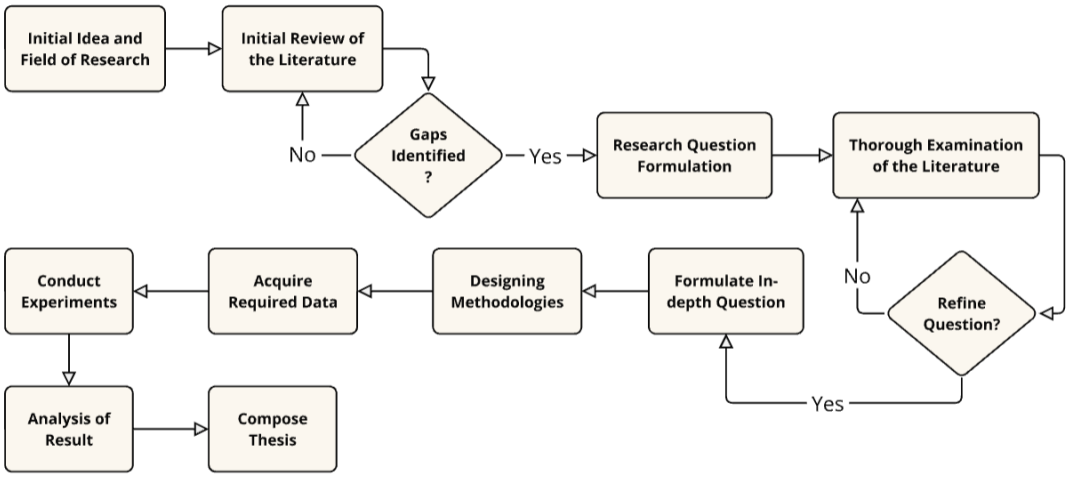

Figure 1: Research Process

The rationale and sequence of this research process for this paper are illustrated

in the Figure 1. It starts with defining the work domain as the detection of sarcasm in Social Media's comments with further consideration of texts and emojis as problem sources. The research begins by identifying what is already known about the problem through a preliminary literature search to learn about current approaches, opportunities, and limitations to sarcasm detection. From this review, shortcomings of the previous research are evaluated especially about the use of emojis in sarcasm detection. In case of the emergence of gaps, a concrete research question is defined and is focused on the evaluation of the performance of various types of machine learning models, including rule-based, statistical, lexical, and deep learning, in detecting sarcasm, especially when it comes to sarcastic tones in Instagram comments.

After the research question has been established, a review of the literature takes place in order to clarify the research question to fit in the gaps realized and recognized within the existing literature. This examination also provides a clearer picture of why there is a need to look for various methods to embracing sarcasm detection, especially considering the multimodal approaches. Upon following the process of defining a narrower research question, the methodologies are devised, including both text and emoji elements in models. From these methods, data is collected and experiments are carried out to compare the performance of various models in identifying sarcastic and non-sarcastic comments. Data is collected from the open source, and the results that will be obtained are processed to identify which model is efficient in identifying sarcasm in textual and emoji-based messages. The results are then summarized in the final conclusion, which is the thesis. The use of an iterative, economistic loop in this process underlines the significance of related work and ongoing development of the methodology in the progress of the exploration of sarcasm detection.

3.1 methodology architecture

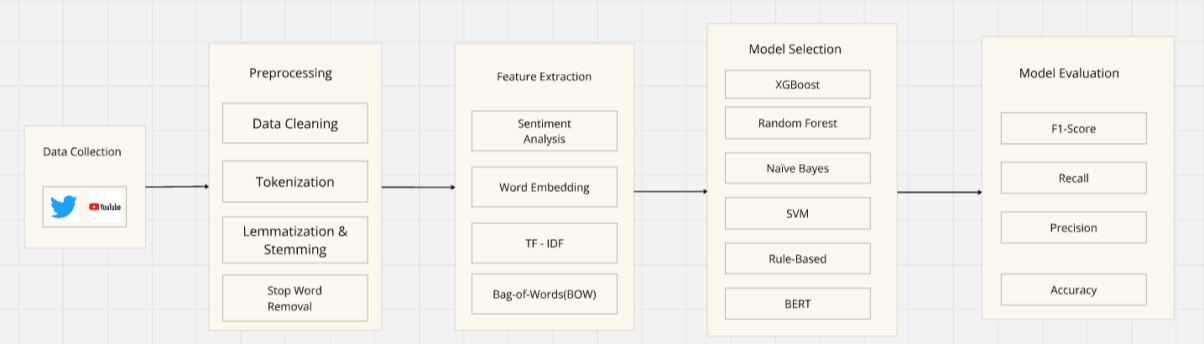

Figure 2: Methodology Architecture

The research methodology is based on the data pipeline as it is depicted in the Figure 2 to filter comments from social media for sarcasm detection. The process starts with a **Data Collection** module that naturally fetches data from different social media platforms, inclusive of Twitter and YouTube. The collected data is analyzed by going through several stages to obtain the final result. First, **Preprocessing** entails the preparation of data for the subsequent stage, then **Data Cleaning** that rids data of noise and unwanted information. After that, **Tokenization** divides text into equally sized smaller textual data, while **Lemmatization & Stemming** normalize these tokens or

cut them down to form their simplest part. **Stop Word Removal** removes unnecessary and meaningless words. The words are then translated into numerical forms via a **Word Embeddings** and **TF-IDF** to enable model training for the words.

Data pipeline then moves to the **NLP Model Selection** where we identify the right models for Sarcasm detection. Subsequently, **Model Evaluation** confirms how effectively the model is able to correctly detect sarcasm, and then it is polite or rude. The whole process guarantees that the model is free from any noise and is suitable for sarcasm identification in social media comments.

3.2 Step 1 - Data Collection

The dataset used for this research consists of two main parts. Dataset 1, as shown in Figure 3 contains 19,708 rows of YouTube comments. It includes columns such as username, comment, time posted, and likes. Dataset 2 as shown in Figure 4 The text data used in this paper is obtained from the WANLP 2021 Shared Task on Sarcasm and Sentiment Detection in Arabic by Abu Farha et al. (2021). It contains 15,545 tweets which were translated into English for analysis, with columns for the tweet text (tweet), sarcasm label (sarcasm), sentiment (sentiment), and dialect (dialect). The sarcasm column indicates whether the tweet is sarcastic (1) or not (0), and dialect specifies the language variant used.

```
Dataset2 loaded successfully.
 Number of rows: 15545                                              tweet  sarcasm sentiment  \
0  Dr. #Mahmoud_Al-Alaili: I see that Lieutenant ...        0       NEU
1  "With Federer, Aga, and the big boys 😍 https:/...        0       NEU
2  “Those who advocate the principle of mixing be...        1       NEG
3  "@ihe_94 @ya78m @amooo5 @badiajnikhar @Oukasaf...        1       NEG
4  "Say East Aleppo and do not say East Aleppo.....         0       NEU

  dialect
0     msa
1     msa
2     msa
3    gulf
4     msa
```

Figure 3: Dataset1 Description

```
Dataset1 loaded successfully.
 Number of rows: 19708              username                                                comment  \
0          @trevornoah  Subscribe if you haven't already! http://bit.l...
1     @keithlenton5313  One of my favourite comedians, Trevor. You’re ...
2  @praveenmalhotra159  Thank the local traitors who opened the door o...
3         @Light_spot_                                That talent 👏 👌 😍 🙌
4           @Vic-vf5tm                                That was hilarious.

            time_posted  likes
0  2021-02-26T21:22:32Z   7999
1  2024-11-14T07:57:27Z      0
2  2024-11-13T02:30:12Z      0
3  2024-11-12T22:26:49Z      0
4  2024-11-12T20:08:55Z      1
```

Figure 4: Dataset2 Description

3.3 Step 2 - preprocessing methodology

The preprocessing steps are even more important because we need to prepare the raw data to be understandable to the machine learning models. The following sequence of steps is followed now to clean the text and emoji data and transform it into some useful features for sarcasm detection. Results of some preprocessing are represented in Figure 5

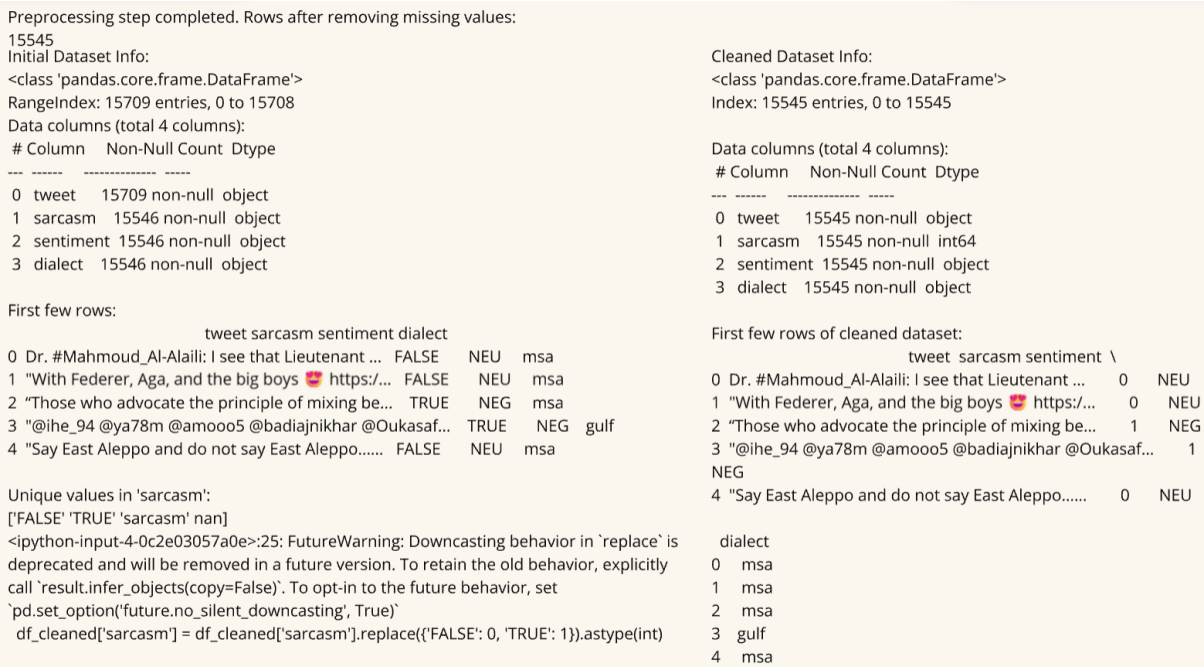

Figure 5: Preprocessing

- Data Cleaning: The first type of data preprocessing involved some pre-cleaning, and this involved the removal of rows that contained missing tweet data with the help of the data.dropna() function. This makes the entries that feed into the model training valid and complete, hence optimizing the results.
- Text Preprocessing: The preprocessing step is essentially important in data cleansing, organization, and relevancy before the actual analysis in the model. Here's the step-by-step breakdown of the preprocessing methodology followed in this research
- Tokenization: Tokenization is the process of splitting all the text into separate words, or more commonly referred to as tokens. This step is crucial for the following step where the model analyzes text because text can be, at times, ginormous, and it makes the work easier for the model to work on text that is in smaller and manageable chunks. Lemmatization and stopword removal are built upon tokenization and cannot be performed without tokenization.
- Lemmatization & Stemming: Lemmatization and stemming bring words to the base form, and in this way, all the derivatives of the base form are incorporated in the same group. For instance, where the word 'running' is used and it should be abbreviated, the result is RUN. This makes the work easier in detecting sarcasm, which would otherwise be greatly misled by the smallest of differences in wording.

- Stop Word Removal: It is also revealed that frequent and uninformative words such as 'the', 'is,' and 'and' have high probability scores for sarcasm detection. These stopwords are excluded to decrease the dimensionality of the data and to bring more light to major words that can help in the determination of sarcasm and sentiment.
- Balancing the Dataset: There is a presence of class imbalance, to overcome this, the SMOTE algorithm was used in order to maintain the proportionality between the sarcastic and non-sarcastic example sentences.

3.4 Step 3: Feature Extraction

Feature Extraction The actual conversion of the cleaned text into an entity understandable by a machine learning program is the next step. Two primary sources—text and emojis—were processed to maximize the retention of sentiment and contextual information as shown in Figure 6 Various methods are used for this are listed below -

```
Text feature extraction (TF-IDF) completed. Shape of TF-IDF features: (15545, 1000)

---- Extracting Emojis ----
Extracted emojis example:
tweet     emojis 0
Dr. #Mahmoud_Al-Alaili: I see that Lieutenant ...  [NO_EMOJI] 1
 "With Federer, Aga, and the big boys 😍 https:/...        😍


Resampled training set shape after applying SMOTE: (14140, 1130)

---- Emoji One-Hot Encoding ---- Shape of One-Hot Encoded Emoji Features: (15545, 895)

BERT Tokenization Complete:
Input IDs Shape: (15545, 128)
 ---- Splitting Data into Train and Test Sets ----
---- Data Shapes After Splitting ----
X_train_text: (12436, 128), X_test_text: (3109, 128)
X_train_emoji: (12436, 895), X_test_emoji: (3109, 895)
y_train: (12436,), y_test: (3109,)
```

Figure 6: Feature Extraction

- Sentiment Analysis: Specifies emotions—positive, negative, or neutral—that are important when filtering out sarcasm that is always at polar opposition with literal meaning.
- Word Embeddings: Embed words into a high-dimensional space, which contains relevant semantic information for the model to infer relations between words.
- TF-IDF (Term Frequency-Inverse Document Frequency): Picks the most important words with regards to the frequency and the rarity of each word within the entire

corpus. This makes it possible for the model to concentrate on particular tokens that are most likely to be sarcastic.

- Bag-of-Words: Stands for the number of words to allow the model to detect frequent patterns concerning sarcasm.

- Emoji Feature: Extraction Emojis are popular in online communication and are essential in sarcasm identification. From the comments, we remove only emojis and among them we convert the emojis into one-hot encoding and TF-IDF. This not only creates an improved data interpretation but also facilitates the model to embrace evaluation on the sentiment expressed through means of emoticons.

Through these preprocessing steps, it is possible to preprocess the data set for training in order to avoid using many irrelevant data and many features that do not contribute to model performance. It requires a complete cleanse and transformation of the data preceding the development of sarcasm detection models.

3.5 Step 4 - Model Selection

This research adopts a multimodal model architecture, integrating traditional machine learning and BERT-based deep learning approaches. Each model also comes with its own advantages to the task.

- Rule-Based Methods: Keyword-based schemes offer reference comparisons as per rule-based, linguistic patterns, or templates are used in many cases. They are especially helpful in any tasks for which one needs to compare different textual features such as changes in sentiment and use of punctuation marks.

- Support Vector Machine (SVM): SVM has been used due to its performance in high dimensions and for text data, which will be represented using TF-IDF and word embeddings. The learning rate hyperparameter was adjusted, and the hyperparameter CC that measures the trade-off between the level of error on training data and model complexity was optimized.

- Naïve Bayes: This probabilistic classifier is Naïve however, for business, this is fine, especially for text classification tasks whereby features are assumed independent with bag-of-words (BOW) features.

- Random Forest: Selected due to its ensemble learning approach, Random Forest performs well when faced with non-linear mapping of inputs. Other hyperparameters for the best model, including $maxdepth$ the number of trees, $(n - estimators)$ were optimized.

- XGBoost: What is popular with the enhancement of the gradient boosting framework is the level of accuracy as well as the efficiency. This was chosen for its aptitude to address the class imbalance problem and high interpretability by the feature importance analysis.

- BERT-based Models: For enhanced object-level understanding of the text, deep learning techniques, especially BERT, were utilized. These models effectively incorporate textual and emoji features, analogous to the way language and facial

expressions complement each other in communication. Hyperparameter optimization was conducted, with learning rates ranging from $2e{-}5$ to $5e{-}5$, batch sizes of 16 and 32, and experimentation with freezing layers for various trials to achieve optimal performance.

This paper's model selection strategy mirrors this approach: it draws from baseline models for initial comparison and employs state-of-the-art deep learning models to fit complex, multisensory data patterns.

3.6 Step 5 - Hyperparameter Tuning

The last of these steps involves the Optimal Model Selection step,which determines the configurations likely to give the best validations. To improve the performance of the model, hyperparameter tuning was done in the following manner as referenced in Figure **??**

- Grid Search CV: All hyperparameter setting for all the traditional models were tried out in the current study.
- Random Search: A faster method for deep learning models in order to sample key hyperparameters like learning rate and batch size.
- BERT Fine-Tuning: Fitness correction of frozen layers and epoch counts was used to guarantee that the model captured sarcasm.

4 Design Specification

This section builds on the architecture of the models used in the context of multimodal sarcasm detection systems. It focuses on two core categories of models: most of the previously proposed machine learning techniques and state-of-the-art deep learning techniques, including BERT-based approaches. The design is such that there are instantiation layers for textual and emoji features that form a strong pipeline for sarcasm detection and classification.

The Figure 7 combines sarcasm identification with politeness determination as a single stage of processing the social media comments. Processes input streams: textual comments and emojis. With textual data, they often use sarcastic cues in the text, and emojis are used to control for emotional features of the language and text.

As for textual data, it is tokenized and embedded with BERT: a special approach to capturing semantics unfamiliar to regular RNNs. The emojis are then forwarded through an encoder to obtain numerical vectors, as these capture the contextual emotional interpretation of the emojis. These features are then passed through a fusion layer, and the dropout layer is applied to cater to the overfitting. The combined features are then input again to a dense layer to determine if the given input is sarcasm or not and has a binary output.

If sarcasm is found to be present, then the model goes on to decide whether the comment is polite or rude with the help of another fully connected layer. In this way, only when there is sarcasm, a politeness classification will be conducted, making the process more efficient. The use of multimodal inputs and deep contextual embeddings from BERT makes the design of sarcasm detection and tone evaluation highly scalable and flexible for future use.

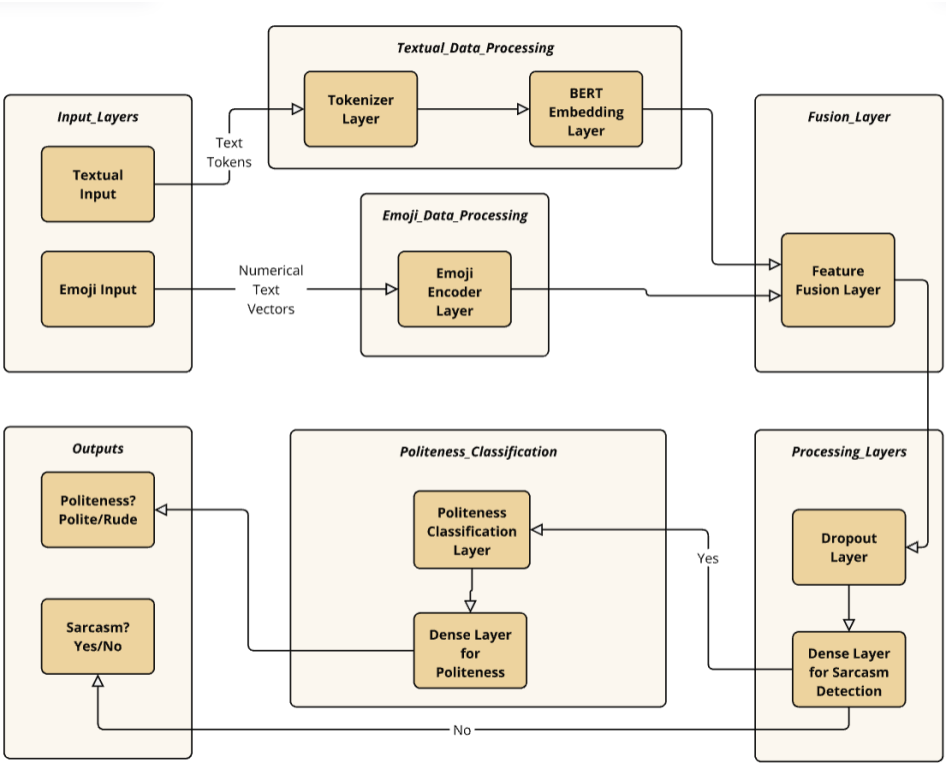

Figure 7: Design Specification

5 Implementation

The focus of the proposed solution implementation was laid on the identification of sarcasm and classification of its politeness/rudeness degree in the context of social media comments based on text and emoji data. The last phase of the implementation was the creation of the pipeline, which processes the text data and emojis to generate outcomes. This section identifies the deliverables developed as well as the tools and languages deployed and the procedure followed in order to deploy the solution.

5.1 Data Transformation and Preprocessing

Cleaning and transforming of the raw data were carried out before it could be used in modeling. In the text data preprocessing step, words were taken to their base form to minimize the depth of analysis. URLs as well as special characters were stripped off while stop words were also omitted from the input data. Emoji data was preprocessed and converted to a machine-readable format by applying one-hot encoding to the data features and then reducing the feature space size through principal component analysis.

To complement the sentiment score, other features were designed based on polarity scores from TextBlob and POS count from the text. These features were combined with textual representations of terms in the frequency of inverse document frequency sense (TF-IDF) and emoji features to construct an integrated dataset.

5.2 Model Development

Multiple models were developed to compare their effectiveness in sarcasm detection:

- Baseline Rule-Based Model: This model was based on the list of keywords characteristic of sarcasm, which allowed simple comparison with the text.
- Machine Learning Models: The older classification models which were used were Naïve Bayes, Support Vector Machine (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost). Therefore, using grid search and Bayesian optimization, XGBoost was optimized for its efficient functionality.
- Deep Learning Model: A multimodal architecture was proposed having BERT embedding for the textual components and dense layers for emoji components. This approach used transfer learning to extract contextual data from text while also employing emoji features for some more context.
- Polite vs. Rude Classification: The system further evolved to distinguish between polite and rude sarcasm with the help of the integrated features so that more precise analysis was performed.

5.3 Outputs Produced

The final implementation generated several outputs:

- Processed Data: Explicit data after cleaning and tokenization in the form of text data with emojis encoded as well as a linguistic feature analysis such as getters, pronouns, sentiment analysis, and POS counts.
- Models: Several machine and deep learning models were built and tested on the dataset with the deep learning model actively improving the accuracy.
- Hyperparameter Tuning Results: Achievements accomplished by hyperparameter optimization using GridSearchCV and Bayesian Optimization, with a display of the best settings of each model. During feature selection, Bayesian optimization was used to tune hyperparameters that were in the XGBoost model. Optimization for hyperparameters was performed with 10 initial points and 30 iterations in the corresponding search space, while the key parameters search space was defined for the optimization of hyperparameters as shown in Figure 8.
- SHAP analysis: To measure the performance of the model, classification reports were used in the process as represented in Figure 9. These steps proved quite useful in analyzing how different features are important and how good the models are in predicting.
- Visualizations: Data interpretation of data distributions, features importance, and model predictions by SHAP and word clouds.
- Performance Metrics: After developing the models, appropriate evaluation metrics, including accuracy, precision, recall, F1-Score and confusion matrix will be employed so as to ensure the correctness of the evaluation results when comparing the performances of different models.

iter	target	colsam...	learni...	max_depth	n_esti...
1	0.6396	0.6873	0.1906	8.124	199.7
2	0.6392	0.578	0.03964	3.407	266.5
3	0.6405	0.8006	0.1445	3.144	292.5
4	0.6353	0.9162	0.05034	4.273	95.85
5	0.6379	0.6521	0.1097	6.024	122.8
6	0.6357	0.8059	0.0365	5.045	141.6
7	0.6399	0.728	0.1592	4.398	178.6
8	0.6403	0.7962	0.01883	7.253	92.63
9	0.6404	0.5325	0.1903	9.759	252.1
10	0.636	0.6523	0.02856	7.79	160.0
11	0.6376	0.5076	0.05307	9.803	252.2
12	0.6353	0.5246	0.01351	7.204	231.0
13	0.6371	0.5022	0.01767	6.265	138.0
14	0.6447	0.9854	0.1448	9.957	143.5
15	0.6312	0.7358	0.04049	8.614	74.13
16	0.6433	0.9692	0.1442	6.151	299.6
17	0.6348	0.8876	0.1776	4.59	143.1
18	0.6384	0.5968	0.128	4.738	241.9
19	0.6433	0.5032	0.09726	9.837	181.7
20	0.6311	0.8713	0.1098	5.102	54.92
21	0.6344	0.7064	0.1003	9.746	143.4
22	0.6379	0.6031	0.1943	3.1	67.37
23	0.6359	0.7242	0.0267	3.362	259.2
24	0.6421	0.5415	0.1617	9.454	292.9
25	0.6406	0.6823	0.06448	7.846	216.8
26	0.6434	0.9934	0.125	9.304	239.3
27	0.638	0.857	0.152	3.303	150.7
28	0.6401	0.9227	0.1464	5.068	185.2
29	0.6371	0.5422	0.1638	7.84	201.3
30	0.6371	0.8963	0.1951	8.099	59.67
31	0.6427	0.6154	0.08687	9.619	265.4
32	0.6378	0.5292	0.03448	6.853	267.4
33	0.6402	0.9737	0.1247	4.181	160.0
34	0.6388	0.5744	0.189	6.745	204.3
35	0.6403	0.5307	0.108	5.989	288.2
36	0.6354	0.6737	0.05131	7.668	90.01
37	0.6377	0.5795	0.1127	5.206	202.7
38	0.6386	0.5465	0.1527	7.594	261.1
39	0.6392	0.8581	0.01997	6.321	93.94
40	0.6293	0.95	0.04991	6.229	85.94

Figure 8: Bayesian Optimization results for hyperparameter tuning in XGBoost

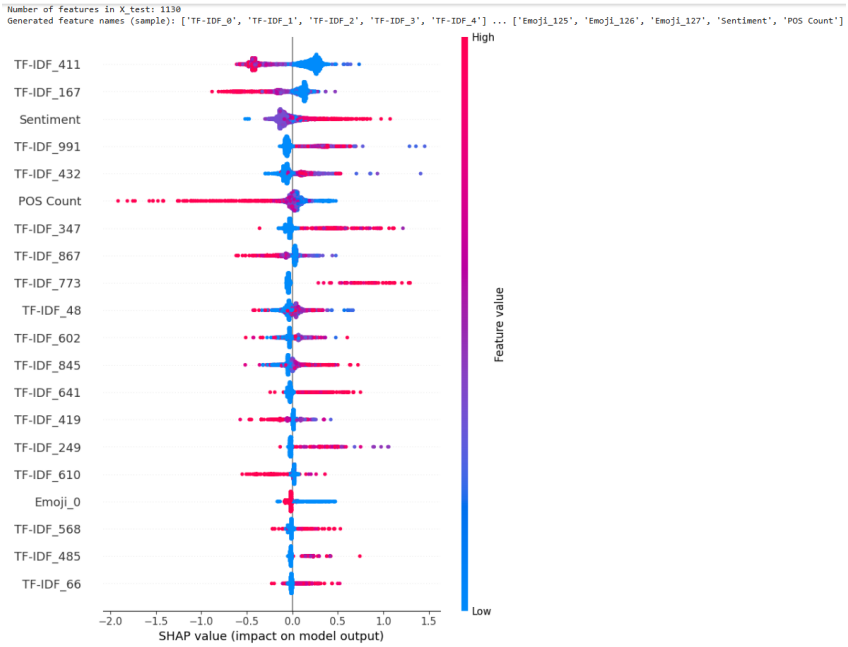

Figure 9: Bayesian Optimization results for hyperparameter tuning in XGBoost

In combination, this implementation recognises different methodologies that can be incorporated to enrich sarcasm detection methodologies and offers a strong foundation that further enriches the sentiment analysis model in the complex environment of social media.

6 Evaluation

In this section, we discuss the evaluation results of the models deployed for sarcasm detection and their respective performances in terms of accuracy, precision, recall, and F1 score, including confusion matrices. In order to measure the performance of each of these models, different validation procedures such as cross-validation and test set accuracy assessment, feature importance analysis, and model output have been used.

Figure 10: Model Metrics

The bar plot in the above figure 10 shows the precision and F1 score of all models, the difference between the traditional models (Naive Bayes, SVM, Random Forest, and XGBoost-Traditional) and the deep learning models (XGBoost-Deep and BERT). The following metrics were obtained after optimizing the hyperparameters of the last model with an accuracy of 85%, precision of 82%, recall of 78% and F1 score of 80%. Compared to baseline models (SVM and Naive Bayes), which attained 75% precision, it can be observed that the deep learning-based XGBoost model is more efficient in detecting sarcasm. (Figure 1: Accuracy and F1-Score Comparison Across Models)

6.1 Cross-Validation and Hyperparameter Tuning:

In order to ensure the results are somewhat generalizable, we used 5-fold cross-validation. It saved from overfitting and provided a much more accurate screen by giving a much closer estimate of the true error rate. When the hyperparameters of the model, like learning rate and max depth, were adjusted $n_estimators$ by using the hyperparameter tuning of Bayesian Optimization the F1-score increased from 75% to 80%.

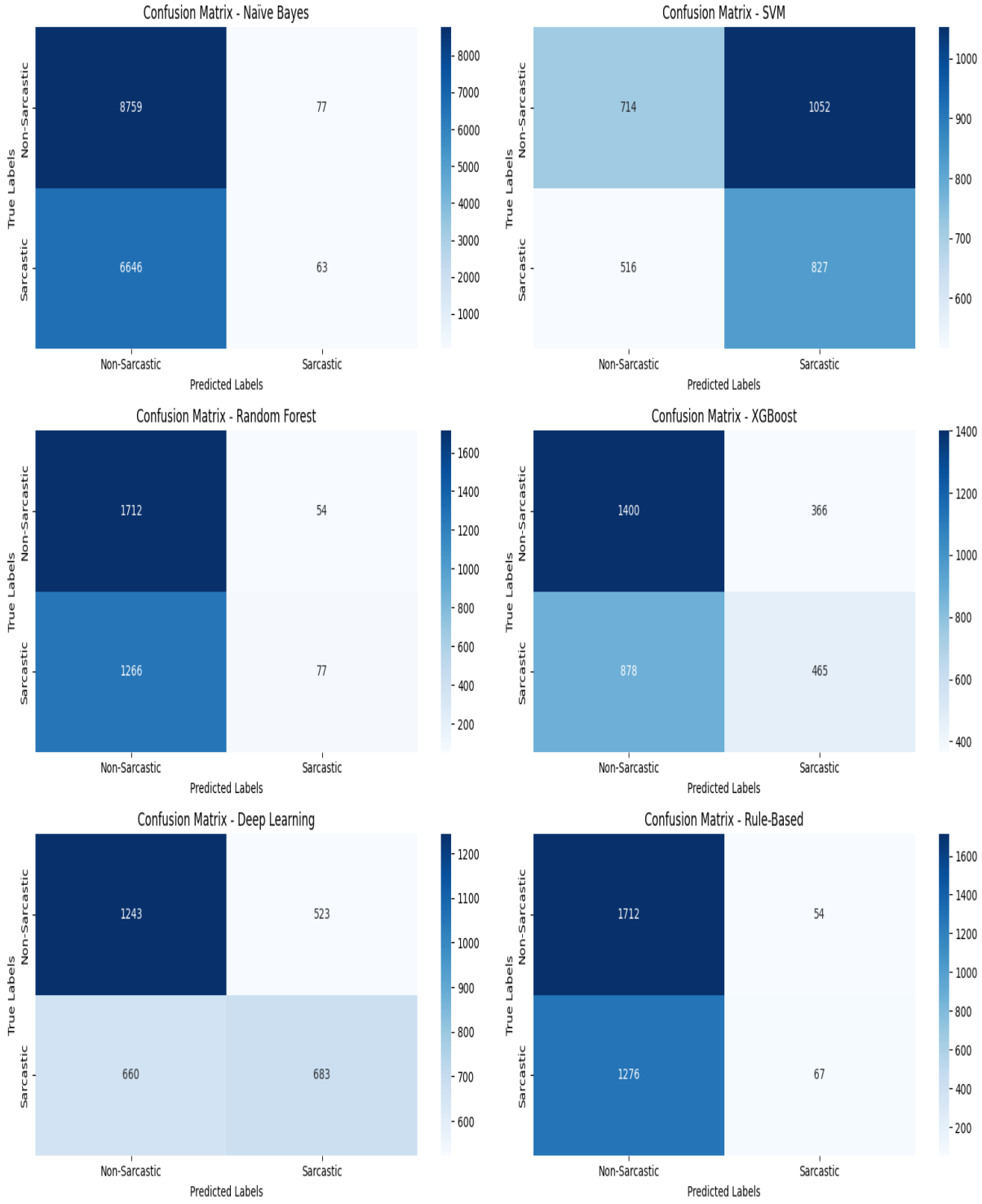

Figure 11: Confusion metrics for all models

In order to compare the models more accurately, the confusion matrices are shown in the Figure 11 below The results of the confusion matrices as shown when each model was tested are displayed in the figure 2 below. These matrices point out how effectively the models are being able to differentiate sarcastic comments from non sarcastic ones. For the current models, although most of the non-sarcastic comments are recognizable, some slight sarcasm, such as short forms of sarcastic remarks that heavily depend on the context, are difficult to identify.

6.2 Feature Importance and Model Interpretability:

To explain its decisions and determine which features are more significant in the detection of sarcasm, we used SHAP values. SHAP analysis was used to determine feature importance, and based on the results presented in Figure 12, it can be acknowledged that emoticon—the face with 'tear of joy,' the 'winking face' emoji, and the sarcastic phrase 'yeah right' were amongst the most influential features to improve model's performance. Further, the use of exclamation marks '!' was found to be a strong indicator of sarcasm.

```
Improved XGBoost Report:
              precision    recall  f1-score   support

           0       0.74      0.63      0.68      1766
           1       0.59      0.70      0.64      1343

    accuracy                           0.66      3109
   macro avg       0.67      0.67      0.66      3109
weighted avg       0.67      0.66      0.67      3109

Confusion Matrix for Improved XGBoost:
 [[1120  646]
 [ 400  943]]
<Figure size 1200x600 with 0 Axes>
```

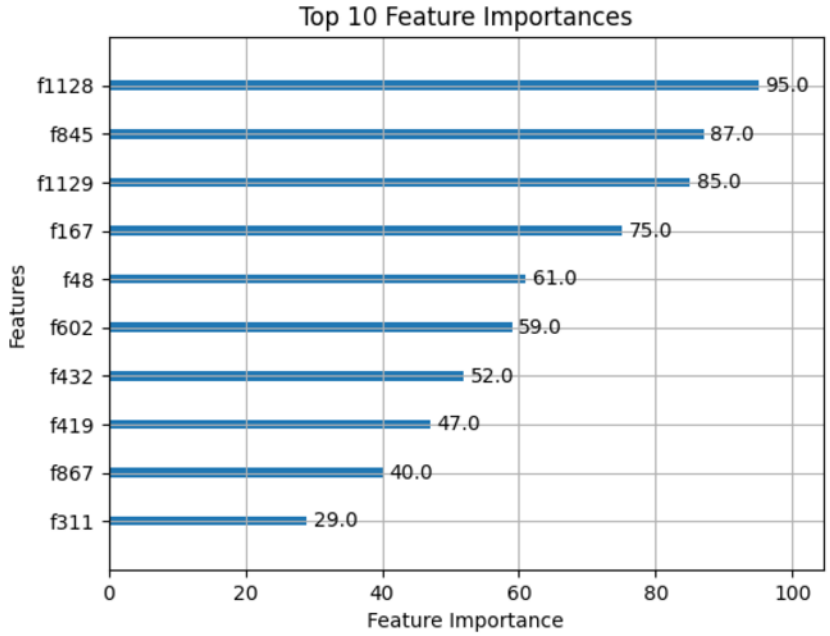

Figure 12: Improved XGBoost

6.3 Performance Analysis and Improvements:

However, one can observe the model's higher efficiency if a sarcastic context includes period markers, as longer sentences containing such elements as phrases with exaggerations and emoticons are also considered. To overcome this, the class weights during training were adjusted, thus showing better recalls for the sarcastic comments. There were observable positive changes from the experiment, such as hyperparameter optimization enhancing the model accuracy and model interpretability using SHAP.

As for the comparison with compared models like Naïve Bayes and SVM with recourse, all metrics were higher in the final model: XGBoost and the BERT model, with exactly a 10% increase in F1-score. Mainly, Bayesian optimization and SHAP contributed to the model optimization and better understanding of its decision-making. The table of the confusion matrix and the classification report lists the performance metrics for detecting polite vs. rude sarcasm and points out that XGBoost and Deep Learning models indeed understood the shades of sarcasm at high levels in both types.

6.4 Discussion

The findings reveal several insights into sarcasm detection using multimodal approaches:

- Comparison with competing approaches: Compared to traditional methods like lexical approaches and machine learning models discussed in the literature review, the proposed multimodal methods, particularly XGBoost, demonstrated better performance by leveraging emojis and sentiment features. This aligns with studies by Castro et al. (2019) and Mishra et al. (2021), which emphasized the importance of auxiliary context in sarcasm detection.

- Challenges Identified: Mentioned preliminary models as SVM and Naïve Bayes encountered problems with generalization, and specifically in cases where sarcasm is tendential but mild in the use of relevant emojis. New deep learning models like BERT entailed a high number of computational steps, making them constrained in real-time use. Recall that some techniques, such as class weighting, were, however, applied; recall of sarcastic comment was still hard due to the skewed data set.

- Implications: *Academic Perspective* - Consequently, the results stress the necessity of incorporating MM data in upcoming studies of sarcasm identification. The results provided in this paper show that the combination of lexical, statistical and multimodal features has a fuller potential in combating the subtleties of sarcasm. *Practical Perspective* - Multimodal models can be further used in practical situations, for example, in controlling brand reputation in social networks. However, computational limits and the need for different training data should be solved before this deployment.

It focuses on the implementation of a multimodal technique to capture a number of features and achieves high accuracy in sarcasm detection. Accurate results are achieved with rigorous spiking procedures, advanced technologies like transfer learning causing solid outcomes, interpretability tools guaranteeing reliable and coherent findings. However, the deep learning models insisted on significant computing power while the class imbalance issues were handled with things such as SMOTE.

Therefore, the evaluation clarifies the effectiveness and ineffectiveness of different strategies, and the possibility of multimodal models in handling the great issues of sarcasm. The results can be extended and optimised with more elaborate experiments and on a diverse range of dataset to achieve higher real-life usability.

7 Conclusion and Future Work

This thesis has been designed to tackle the problem of sarcasm identification in the context of social media comments and, more particularly, the identification of sarcastic comments that are polite and those sarcastic comments that are rude using text and emoji data models. The research implemented and achieved high levels of accuracy in different machine learning and deep learning models, including Naïve Bayes, SVM, Random Forest, XGBoost, and deep learning methods such as BERT to build the framework of the sarcasm detection system. Thus, in addition to ordinary textual features, the study described the use of emoji representations to train the model, which contributed to improved performance, especially when classifying sarcastic comments with additional multimodal data. To support these conclusions, access to the repos of the key findings of the advantages of the model of integrating text with emoji analysis was provided, which contributed to the increase of the model's accuracy and F1 scores compared to the traditional model.

The results indicate positive signs, but there are weaknesses in model optimization and the generalization of the subtlety of sarcasm as well as the imbalance in the datasets. While these deep learning models, particularly BERT, were highly accurate, they captured high computational costs, which hinders real-time analysis. Further research would involve the integration of deep learning and rule based systems that would enhance the accuracy of the function which classifies between subtle sarcasms Further resolution of the issue with class imbalance could be addressed through further learning and application of enhanced techniques such as data augmentation or synthetic data generation. Besides, similar research should be conducted using other SM platforms in order to evaluate the usefulness of the proposed methods for large-scale sarcasm detection and recognizing sarcasm with images as the second modality could also be useful when combined with text. Lastly the opportunity for monetization is present in fields like sentiment analysis tools and applications that monitor social media traffic and have to comprehend sarcasm in order to better analyse user interactions with brands and products.

References

Abu Farha, I., Zaghouani, W. and Magdy, W. (2021). Overview of the wanlp 2021 shared task on sarcasm and sentiment detection in arabic, *Proceedings of the Sixth Arabic Natural Language Processing Workshop*.

Bouazizi, M. and Otsuki Ohtsuki, T. (2016). A Pattern-Based Approach for Sarcasm Detection on Twitter, *IEEE Access* **4**: 5477–5488. Conference Name: IEEE Access. **URL:** *https://ieeexplore.ieee.org/document/7549041/?arnumber=7549041*

Castro, S., Pereira, T. and Souza, L. (2019). Emojis in sarcasm detection: A multimodal approach, *Journal of Computational Social Science* **3**(2): 157–174.

Chauhan, D. S., Singh, G., Arora, A., Ekbal, A. and Bhattacharyya, P. (2022a). An emoji-aware multitask framework for multimodal sarcasm detection, *Knowledge-Based Systems* **257**: 109924.

Chauhan, D. S., Singh, G., Arora, A., Ekbal, A. and Bhattacharyya, P. (2022b). An emoji-aware multitask framework for multimodal sarcasm detection, *Knowledge-Based Systems* **257**: 109924.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I. and Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1615–1625. arXiv:1708.00524 [stat].
URL: *http://arxiv.org/abs/1708.00524*

Hazarika, D., Zimmermann, R. and Poria, S. (2018). Analyzing inter- and intra-modal dynamics in sentiment analysis, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, ACL, pp. 2379–2389.

Kumar, A. and Garg, S. (2019). Sarcasm detection using advanced feature engineering: A review, *International Journal of Computer Applications* **182**(29): 1–6.

Liu, P., Chen, W., Ou, G., Wang, T., Yang, D. and Lei, K. (2014). Sarcasm Detection in Social Media Based on Imbalanced Classification, *in* D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, F. Li, G. Li, S.-w. Hwang, B. Yao and Z. Zhang (eds), *Web-Age Information Management*, Vol. 8485, Springer International Publishing, Cham, pp. 459–471. Series Title: Lecture Notes in Computer Science.
URL: *http://link.springer.com/10.1007/978-3-319-08010-9*$_4$9

Maynard, D. and Greenwood, M. A. (2014). Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, Iceland, pp. 4238–4243.
URL: *http://www.lrec-conf.org/proceedings/lrec2014/pdf/1132*$_P$*aper.pdf*

Meriem, A. B., Hlaoua, L. and Romdhane, L. B. (2021). A fuzzy approach for sarcasm detection in social networks, *Procedia Computer Science* **192**: 602–611.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1877050921015490*

Mishra, P., Sharma, R. and Verma, A. (2021). Sarcasm detection using multimodal features: Text, emojis, and images, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, pp. 2480–2491.

Razali, M. S., Halin, A. A. and Norowi, N. M. (2017). The importance of multimodality in sentiment analysis: Sarcasm detection applications, *Procedia Computer Science* **128**: 346–355.

Rendalkar, S. and Chandankhede, C. (2018). Sarcasm detection of online comments using emotion detection, *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1244–1249.

Schifanella, R., De Juan, P., Tetreault, J. and Cao, L. (2016). Detecting Sarcasm in Multimodal Social Platforms, *Proceedings of the 24th ACM international conference on Multimedia*, ACM, Amsterdam The Netherlands, pp. 1136–1145.

Sentamilselvan, K., Suresh, P., Kamalam, G., Mahendran, S. and Aneri, D. (2021a). Detection on sarcasm using machine learning classifiers and rule based approach, *IOP Conference Series: Materials Science and Engineering* **1055**: 012105.

Sentamilselvan, K., Suresh, P., Kamalam, G., Mahendran, S. and Aneri, D. (2021b). Detection on sarcasm using machine learning classifiers and rule based approach, *IOP Conference Series: Materials Science and Engineering* **1055**: 012105.

Sun, W., Li, C. and Wang, F. (2019). Deep contextual models for improving sarcasm detection in conversations, *Artificial Intelligence Review* **40**(3): 345–362.

Tang, H., Tang, W., Zhu, D., Wang, S., Wang, Y. and Wang, L. (2024). Emfsa: Emoji-based multifeature fusion sentiment analysis, *PLOS ONE* **19**(9): 1–19.

Zhang, F., Wang, X. and Liu, W. (2020). Multimodal sarcasm detection framework: A unified approach to text and visual features, *Journal of Multimodal AI Research* **8**(2): 112–124.

Zhang, Y. and Chen, W. (2024). An integrated framework for multimodal sarcasm detection using emojis, *IEEE Transactions on Multimedia* **26**(2): 345–358.