International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

Expressivity of Tweets on Social Issues Using Aspect Based Text Classification

Elizabeth Leah Thomas[1], Subashini Parthasarathy[2]
[1]Research Scholar, Department of Computer Science. Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India
[2]Professor Department of Computer Science. Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India
18phcsp003@avinuity.ac.in[1], subashini_cs@avinuty.ac.in[2]

Abstract— Social discussions about crime on Twitter and open forums aims to understand the barriers that hinder people from expressing their concerns or aligning with popular opinions. A curated dataset spanning three months in 2023 was collected, covering categories like crimes and Gender Equality and Violence Against Women.

The study employs aspect-based sentiment analysis to classify sentiment polarity in tweets, utilising a comprehensive framework involving three text feature classification stages. The initial stage analyses individual words, phrases, and tweet patterns to classify text features based on specific linguistic elements. In the subsequent step, semantic relations explore a better understanding of the core sentiment and infer relationships between different text keywords. This stage enhances the analysis by considering the meaning and contextual nuances of the language used in the tweets. The final stage incorporates transformer-based models for effective multilabel classification to view the diversity present in the dataset. The study's quantitative analysis reveals that the Ensemble learning approach demonstrates an impressive precision measure of 93%. By integrating the three stages of text feature classification, the study enhances the accuracy and comprehensiveness of sentiment analysis in social discussions about crime on Twitter.

Index Terms— Crime and Gender Equality tweets, Ensemble learning classification algorithms, Multilabel text feature classification, Semantic Text features, Precision score.

I. INTRODUCTION

Social media is a ubiquitous platform for individuals to communicate and share their thoughts and experiences. Developing countries face the challenge of having large populations but limited resources to address the multitude of crime-related tweets and similar content, including reports of accidents and fraud. Additionally, there is a particular emphasis on addressing issues such as Murder, raids, and robbery and the broader objective of promoting gender equality and combating violence against women addressing concerns related to femicide, gender violence, and rape.

The tweets retrieved from the Twitter API through web scraping were analysed using several classification methods in the proposed work. By utilising machine-learning methods, it is possible to identify the most trending subjects on Twitter. Moreover, it is essential to note that Twitter and other popular social media platforms have effectively raised awareness of societal issues and advocated change.

Crime-related studies[1]-[5] proposed different frameworks for analysing and classifying crime-related tweets using Naive Bayes, SVM, and Decision Trees based on performance metrics such as Precision, Recall, F1 Score, and Accuracy. The best accuracy achieved was 83.9% using the SVM algorithm, with several features representing the tweets, including Bag-of-Words, TF-IDF, and Word Embeddings. [6] -[8] analysed geolocated tweets and their correlation with crime rates and the presence of tourists and commuters in different locations, with insights into crime patterns and trends. [9], [10] have conducted research to help law enforcement agencies and policymakers better understand crime patterns and respond to emerging threats in real-time.

Early studies on violence against women[11]and[12] adopted sustainable development goals to transform the world.[13] Considering the target goal, this study focuses on gender violence, Rape and Femicide as primary keywords in social media studies. Recent studies on femicide emphasise the UN's focus on measuring femicide as a quantifiable indicator of progress in addressing gender-based violence, obscures other forms of violence, and fails to address the root causes of gender inequality.[14] In previous studies, these tweets under investigation were more commonly observed in areas where intense mating rivalry was expected among men due to male-biased sex ratios, a lack of single women, substantial wealth disparity, and minor gender income differences.

Aspect-based sentiment analysis (ABSA) is a natural language processing technique used to investigate opinions and sentiments related to specific aspects or features of a given topic. Aspect-based tweets can help identify and analyse thoughts and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

ideas on factors such as the type, location, impact on victims, public perceptions, and attitudes toward the system. [15], [16] The Aspect of the tweet expressed determines the tweet polarity, that is, whether the user's opinion is Positive or Negative. The biggest challenge in identifying relevant tweets corresponds to the semantics of textual content such as adjectives, adverbs, verbs, or phrasal verbs. The tasks for analysing sentiments previously relied on means such as Naive Bayes or Support Vector Machines. Nonetheless, there has recently been a growing trend towards using deep learning approaches that combine basic nonlinear methods. This research aims to identify keywords highlighting social issues, focus on sustainable development goal 5(SDG-5) against gender violence, and identify public perceptions and attitudes of Tweet users. A curated dataset of over 30,000 tweets was collected daily for three months in 2023 for robust evaluation and feature extraction. Aspect-based sentiment analysis is performed for the final results after pre-processing tweets collected from two separate datasets divided into two classes: #accidents and #frauds. #Murder, #raid, #robbery under the category Crime and from Gender violence group with keywords such as #femicide #gender violence #rape

The three opinion classes of tweets from the cleaned dataset with semantic relations, where term frequency representations are in the form of word clouds, word frequencies in each tweet, and highly used tweets, can be used for tweet categorisation and selection. Four classifier-based approaches using machine learning and deep learning models for text feature identification were based on aspects and semantics.

The rest of the study delves into Section 2, Related Works, which highlights the recent research. Section 3 explains the dataset, Section 4 outlines the proposed methodology and experimental results, and Section 5 discusses the analysis and suggestions. Section 6 concludes the paper and outlines avenues for future work.

II. RELATED WORK

Enormous studies and work have been previously carried out in this domain of Aspect level sentiment analysis. The current paper focus on the ablation studies that can be highlighted under the trending Deep Learning models with Transformer models. [17] proves that by employing a combination of lexicon-based and rule-based methods alongside Deep Learning models like Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) and in terms of accuracy for sentiment analysis tasks on the dataset, the mixed method outperformed both the standalone RNN method (which yielded a score of only (79.4%) the lexicon-based technique. Recent studies [17] The comparison of various deep learning algorithms is made by the authors, which included Convolutional Neural Networks (CNN), Recurrent neural networks (RNN)and transformer-based models like Bidirectional Encoder Representations (BERT) and Generative Pre-trained Transformer (GPT3), and multiple content analysis assignments such as sentiment analysis, text and topic are to evaluate the efficacy of these algorithms. [19], [20] The researchers examine how transformer models like BERT and GPT3 outperform concerning accuracy when used on various text classification tasks. Suggests a study which generates coherent topics and remains competitive across various benchmarks involving classical models following the more recent clustering approach to the issue. Experiments by [21]-[25] mention that the current topic-guided language models generate problems that are conceivably coherent compared to those of the regular Latent Dirichlet Allocation (LDA) topic model. The authors of the studies [26] -[29] propose a text feature extraction algorithm based on a deep neural network that improves the efficiency of modern social media. Generative Adversarial Networks (GAN) have received much attention and produced impressive results. One such model is the (VAE) [30], [31]which is a type of generative model that learns a compressed representation of data, called latent variables, that can be used to generate new data samples

III. METHODOLOGY

A. Dataset

For the experimentation purpose, the Crime and gender violence-related tweets were collected and queried with the Twitter Application Programming Interfaces (API). The Tweets posted from the first week of March 2023 till 13 weeks were gathered as relevant tweets for the analysis. The dataset had the following challenges: the tweets were fewer on certain days owing to the trend, and some days had high trending tweets for a particular keyword. The retrieved also showed specific imbalances or biases.

Data collection	March 2023- May 2023 Total number of tweets collected till May 2023
Crime- Keywords	#accidents and #frauds. #Murder, #raid, #robbery
SDG5- Keywords	#Gender violence # Femicide # Rape
Dataset description fields	'Date', 'Likes Count', 'Source', 'User', 'Retweets Count' 'TweetURL' 'Tweet Content' 'Language'
Parameters considered	'Tweet_ID',' Entity',' Sentiment',' Tweet content

TABLE I. SUMMARY OF ANNOTATED DATASETS

B. Empirical Analysis

Aspect-level sentiment analysis is performed with VADER (Valence Aware Dictionary and sEntiment Reasoner), which is a lexicon- and rule-based. VADER quantifies the degree Of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

positive or negative sentiment of the post, in addition to identifying the binary positive or negative opinion of a tweet. On a tweet that detects polarities. The analysis found tweets with several distinct sentences conveying contrasting sentiments. If multiple emotions are seen within a tweet, they will impact the single final score through an averaging process of all included feelings. One drawback to sentiment classification is that it decreases specific information at the aspect level. The focus of the study is on aspect based sentiment analysis where the keywords are linked for Tweet popularity based on tweets, retweets, and User popularity based on the type of sensation of the incident occurred and the mixed user feelings displayed in the form of topics frequency or keywords that are maximum among the collected tweets.

a) Aspect Feature Extraction on tweets

The study aims to build a multiclass classification model that extracts strong sentiments and the top words related to that keyword. The aim is, therefore, to analyse the dataset from a social perspective to help understand people's perceptions on the topic. The mentioned machine learning model is used to classify tweets and opinions into different classes by assigning a polarity value.

Data Tweet Pre-processing: The primary approach is to remove numbers, punctuation marks, URLs and symbols in which all mentions (e.g.,@, users), hashtags, unknown signs, and emojis. Tokenisation, Lemmatisation, and stop word removal are performed using the NLTK package.

b) Feature Extraction

The feature extraction is done using two methods of NLTK python and SpaCy methods as an evaluation step in validating the extracted features and then validating through some known machine learning and Deep Learning classifiers to find the accuracy. The evaluation parameters provide a practical approach to validating and understanding the features of the associated terms. The features of the obtained tweets are grouped into social audience types and Tweets with Retweet or social count maximum. We characterise various features like term frequency-inverse document frequency (TF-IDF) and n-gram features for word embeddings to be accurate. Data visualised through Word Cloud answers questions about the most common words in the complete dataset, the most common words in the dataset for negative and positive tweets, the number of hashtags present in a tweet, trends associated with the dataset, trends associated relating to the feelings of one or the other and their compatibility with emotions.

C. Empirical Methodology

The proposed Methodologies, as shown in Figure 1, are developed with a dataset of tweets collected daily for the study from March 2023, which is then pre-processed and cleaned to add sentiment analysis of the tweets.

Phase 1: Text Feature Classification: The proposed methodology categorises tweets based on their textual characteristics or features. It involves analysing individual words, phrases, or patterns within the text to determine the category or label that best represents the content of each tweet. Machine learning algorithms, including Naive Bayes, SVM classifier, and Random Forest Classifier, to categorise the tweets into positive and negative.

Phase 2: Semantic Text Feature Classification: It focuses on understanding the meaning and context of the words or phrases used in the tweets. This involves language processing techniques to capture the semantic relationships between words and infer the underlying intent or sentiment expressed in the text. Semantic analysis considers the overall meaning and implications of the language used to determine each tweet's appropriate category or label. Deep learning algorithms such as CNN, RNN, and LSTM provide powerful semantic analysis capabilities to extract meaning from text.

Phase 3: Multi label Text Feature Classification: Assigning multiple labels or categories to each tweet. Unlike single-label classification, where a tweet is given to only one type, multilabel classification recognises that a tweet can simultaneously encompass multiple topics or themes. This allows for a more nuanced representation of the content and accommodates the diversity of the dataset. For example, a tweet may be classified as belonging to both the Crimes and SDG5 categories, reflecting its relevance to both subjects. The classification model is designed to predict the presence or absence of multiple labels for each tweet, providing a more comprehensive understanding of its content. The study employed ensemble learning classification algorithms, including Multinomial Naïve Bayes Classifier, Random Forest Classifier, Extra Trees classifier, Gradient Boosting Classifier, Ridge Classifier, and Stochastic Gradient Classifier. These algorithms were evaluated based on precision, recall, and f1-score metrics to classify tweets into multiple categories.

D. Performance Classification Measure

The associated confusion matrix, which was obtained by testing sets with the four elements true positive (TP), true negative (TN), false positive (FP), and false negative (FN), was used to assess the performance of the classifiers. Metrics for classification performance included

Precision (PPV = TP TP+FP) (1)

Recall (TPR = TP TP+FN), (2)

F1 score = (2(precision × recall))/(precision + recall) (3)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

Accuracy (ACC = TP+TN TP+TN+FP+FN). (4)

High scores for ACC, F1, PPV, and TPR revealed robust model performance, proving the validity of the categorisation models. These measurements also allowed us to compare several text embedding and classification models to determine the most precise and dependable.

E. Hyperparameter Tuning

Hyperparameters, a crucial part of every ML technique, are

frequently in charge of significant performance improvements. Mainly, the K-fold cross-validation method—where K is the number of folds—is commonly utilised; depending on the size of the dataset, 5-fold or 10-fold cross-validation is frequently used. These folds show how datasets have been divided into several portions; for example, the 5-fold suggests separating into five components. The training and testing sets for the K-fold cross-validation are all the same for all features. Collecting hyperparameter values yielding the maximum accuracy is obtained using accuracy as the performance metric in the cross-validation. Following that, we use these hyperparameter values to implement the relevant ML.

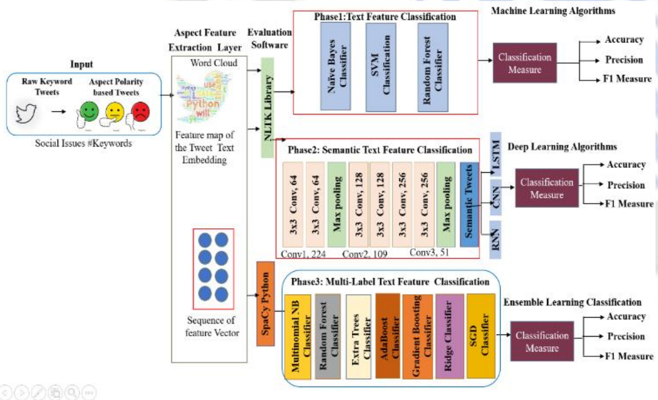

Figure 1. Proposed Framework

IV. EXPERIMENTAL RESULTS

A. ***Data source and sampling*** The findings from the research study in Table II indicate a notable distribution of sentiment in tweets collected, particularly concerning crime and Sustainable Development Goal 5 (gender violence, women, and rape). The sentiment analysis using VADER revealed that a significant portion of the crime-related tweets (33%) expressed a positive sentiment toward the derived keyword. In comparison, 25% said a negative opinion of feelings for the incident, and 42% remained neutral about keywords on general news and topics of world news.

TABLE II: DATASETS SUMMARY OF ASPECT SENTIMENT POLARITY TWEETS

S.No	#Keyword	Processed Tweets	Positive Tweets	Negative Tweets	Neutral Tweets
1	Accident	29999	9256	8558	12185
2	Fraud	34000	11282	8029	14689
3	Murder	32000	10746	10605	10649
4	Raid	35032	12968	7992	14072
5	Robbery	36458	10785	10908	14765
6	Gender Violence	16877	6788	7213	2876
7	Femicide	9743	1496	1296	6951
8	Rape	18000	5825	4728	7447

This distribution suggests a general reluctance among Twitter users to openly discuss their opinions on crime, leading to a potential bias in the expressed sentiment. Regarding tweets related to Sustainable Development Goals, the sentiment analysis demonstrated that negative views accounted for 43% of the dataset, positive feelings for 40%, and a marginal 20% as neutral. This skewed sentiment distribution reveals a similar bias in the perception of gender violence, women, and rape. It indicates hesitancy among Twitter users to address and share their emotions on these sensitive subjects openly.

In Table III, the tweet counts vary among the categories, with femicide having the lowest count of 9,743 and robbery having the highest count of 36,458. Secondly, memory usage differs across categories, ranging from approximately 1.0+ MB for Femicide to 3.9+ MB for Robbery. Thirdly, the vocabulary sizes vary, with fraud having the most extensive vocabulary of 29,727 unique words and femicide having the smallest vocabulary of 4,420 uncommon words.

Moreover, the maximum sequence length for the tweet lists also varies across categories. The accident has the most extended sequence length of 91, followed by Murder at 88, Fraud at 82, and Raid, Robbery, and Gender Violence at 69. Femicide has the shortest maximum sequence length of 41, while rape has a full sequence length of 69. However, considering data within two standard deviations from the average sequence length reduces the maximum sequence lengths, ranging from 20 to 24. This reduced range covers a significant portion of the data, providing a more manageable size for analysis. Lastly, the calculated numbers reveal data coverage by the maximum sequence length within two standard deviations. The scope ranges from approximately 92.57% for Gender Violence to 94.28% for robbery. This information is valuable for planning effective data processing and modelling strategies tailored to each tweet category.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

TABLE III. STATISTICS OF THE DATASETS

Characteristics	Accident	Fraud	Murder	Raid	Robbery	Gender Violence	Femicide	Rape
Tweet Count	29999	34000	32000	35032	36458	16877	9743	18000
Memory usage:	3.2+ MB	3.6+ MB	3.4+ MB	3.7+ MB	3.9+ MB	1.8+ MB	1.0+ MB	1.9+ MB
The vocabulary of the dataset is	29197	29727	31496	28611	26237	10778	4420	15692
The maximum length of the sequence in the list (Tweet)	91	82	88	69	88	66	41	69
The maximum length of the sequence data(SD-Avg)	23	24	23	23	23	22	20	22
Dataset Coverage in Empirical Analysis	93.23 % of the data	93.63 % of the data	92.63 % of the data	93.85 % of the data	94.28 % of the data	92.57 % of the data	93.38% of the data	93.9 % of the data

B. Phase1: Text Feature Classification

In this study, we have collected vital phrases of word vectors. The classification using the selected features is performed using machine learning algorithms such as Naive Bayes (NB), Support Vector Machine (SVM) Random Forest (RF). The evaluation measures used are accuracy (Acc.), precision (Prec), recall, and f1-score (F1). [32]

Algorithm 1. Pseudocode for Text Feature Classification

Input: Training Dataset: (X_train), (y_train), Testing Dataset: (X_test), (y_test)

#tweets.csv: Dataset containing tweets related to crime and gender violence

Output: accuracy: Accuracy (total accuracy) #the sentiment classification model

Procedure:

1. Load the dataset tweets.csv containing tweets related to crime and gender violence.
2. Extract the tweet text as features (X) and the dataset's sentiment labels (y).
3. Split the dataset into training and testing sets with a specified test size (e.g., 20%) and a random state (for reproducibility).
4. Create a count vectoriser to convert the tweet text into numerical features.
5. Fit the CountVectorizer on the training set (X_train) and transform the training set and the testing set (X_test) into numerical features.
6. Train a model on the training set using the transformed training features (X_train_features) and the corresponding sentiment labels (y_train).
7. Predict the test set's sentiment labels by applying the trained model to the transformed testing features (X_test_features).

#Model Evaluation

8. Calculate the accuracy of the sentiment classification by comparing the predicted labels with the true labels (y_test).
   Total accuracy.append(validationAcc)
9. End for
10. return total accuracy
11. End Procedure

This algorithm uses a CountVectorizer to convert the tweet text into numerical features and then trains a model to classify the sentiments and calculate the classification accuracy, as shown in Table IV and the graph in Figure 2.

TABLE IV: TEXT FEATURE IDENTIFICATION MODELS AND ACCURACY

Algorithms (Accuracy in %)	**Naïve Bayes Classifier**	**SVM Classifier**	**Random Forest Classifier**
Category: Crime			
Accident	0.77	0.81	0.79
Fraud	0.75	0.79	0.77
Murder	0.78	0.81	0.79
Raid	0.83	0.86	0.85
Robbery	0.83	0.86	0.84
Category: SDG-5			
Gender Violence	0.88	0.91	0.91
Femicide	0.85	0.89	0.87
Rape	0.82	0.84	0.82

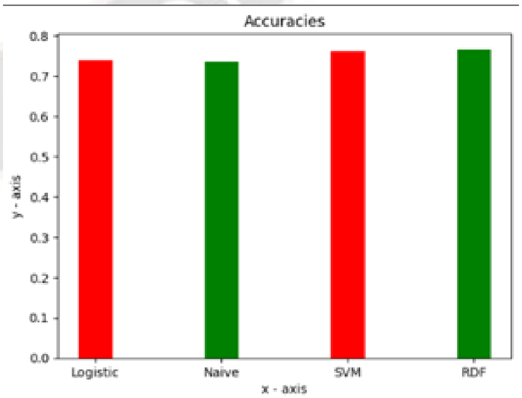

Figure 2: Result Analysis of Phase 2: Semantic Text Feature Classification

Based on the provided accuracy scores, the SVM Classifier and Random Forest Classifier algorithms consistently perform better in classifying tweets related to both Crime and SDG-5

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

categories than the Naïve Bayes Classifier. The accuracy scores for all crime categories are relatively close for the SVM Classifier and Random Forest Classifier, indicating their effectiveness in classifying crime-related tweets. Notably, the Raid category achieves the highest accuracy scores across all algorithms, suggesting that tweets about raids are more easily organised. Similarly, for SDG-5 classes, the SVM Classifier and Random Forest Classifier outperform the Naïve Bayes Classifier, and all three algorithms exhibit consistently high accuracy scores. The Gender Violence category has the highest accuracy scores among all SDG-5 types, implying its relatively more straightforward classification than femicide and rape.

C. Phase 2: Semantic Text Feature Classification

Extracting linguistic features from tweets involves applying deep learning techniques to social media datasets, and these features play a vital role in categorising tweets into different informative groups. It provides powerful semantic analysis capabilities to extract meaning from text. The Key idea is to identify semantic text features in the relevant tweets that can outline the impact of the crowd's feelings, emotions, reactions, and pieces of evidence among the global and local society, such as systematic discrimination against Women, the problem of violence with health promotion perspective and programme indicators for monitoring and surveillance.

Algorithm2: Pseudocode for Semantic Text Feature Classification with spaCy

Input:

X_train: List of training texts

X_test: List of testing texts

y_train: List of corresponding sentiment labels for the training texts

y_test: List of related sentiment labels for the testing texts

Output:

Accuracy: Accuracy of the semantic text feature classification model

1. Load the spaCy model.
2. Define a function extract_semantic_features that takes a text as input and returns a list of extracted semantic features using spaCy.
    a. Process the text using spaCy (NLP).
    b. Initialise an empty list of semantic_features.
    c. Iterate over each token in the processed document.
    d. If the token is not a stop word and its part-of-speech tag is either 'ADJ', 'NOUN', or 'VERB':
    e. Append the lemma of the token to semantic_features.
    f. Return the semantic_features list.

#Extract Semantic Features

3. Apply the extract_semantic_features function to each text in X_train and X_test to extract semantic features.
    a. Store the results in X_train_semantic and X_test_semantic, respectively.
    b. Convert the extracted semantic features (X_train_semantic and X_test_semantic) into numerical vectors using a vectoriser (e.g., CountVectorizer).
    c. Store the transformed features in X_train_semantic_features and X_test_semantic_features, respectively.

#Train the Model

4. Train a model on the training set using the transformed semantic features (X_train_semantic_features) and the corresponding sentiment labels (y_train).
5. Predict the test set's sentiment labels by applying the trained model to the transformed semantic features (X_test_semantic_features).
6. Store the predicted labels in y_pred.

#Model Evaluation

7. Calculate the accuracy of the semantic text feature classification by comparing the predicted labels (y_pred) with the true labels (y_test).
8. Store the accuracy value in accuracy.
9. Output the accuracy of the semantic text feature classification model.
10. End Procedure

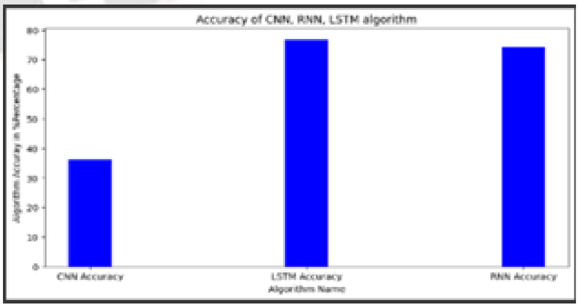

Figure 3: Result Analysis of Phase 2: Semantic Text Feature Classification

Based on the accuracy scores of different deep learning algorithms in classifying Crime and SDG-5-related tweets, the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

LSTM consistently performs better than the CNN and RNN for both categories. The LSTM achieves the highest accuracy scores across all crime-related and SDG-5 categories, followed by RNN and CNN. These results suggest that LSTM and RNN are more effective in accurately classifying tweets related to Crime and SDG-5 than CNN. Furthermore, the Raid category exhibits the highest accuracy scores for all deep learning algorithms in the crime category, indicating that tweets associated with raids are relatively easier to classify accurately. The Gender Violence category, on the other hand, garners the highest accuracy scores among all SDG-5 types, emphasising that tweets related to gender violence are comparatively easier to classify correctly.

D. Phase 3: Multi label Text Feature Classification

The proposed study employs the NLTK library and utilises data-driven techniques with transformer-based models like BERT to effectively implement multilabel classification, facilitating a nuanced representation of the tweet content. By considering the diversity inherent in the dataset, this stage aims to accommodate various perspectives and viewpoints. By integrating these three stages of text feature classification, this study endeavours to enhance the accuracy and comprehensiveness of sentiment analysis for tweets.

Algorithm: Multilabel Text Feature Classification with NLTK and BERT

Input:

X_train: List of training texts

X_test: List of testing texts

y_train: DataFrame of corresponding category labels for the training texts

y_test: DataFrame of related category labels for the testing texts

Output:

accuracy: Accuracy of the multilabel text feature classification model

PROCEDURE BEGIN:

1. Import the necessary libraries: NLTK, transformers, torch, sklearn.metrics.accuracy_score.
2. Load the NLTK library and BERT tokeniser and model from the transformers library.
3. Load the dataset containing the text features (X) and labels (y) from your source.
   a. X_train = [...] # List of training texts
   b. X_test = [...] # List of testing texts
   c. y_train = [...] # DataFrame of corresponding category labels for training texts
   d. y_test = [...] # DataFrame of corresponding category labels for testing texts
4. Split the dataset into training and testing sets using the train_test_split function, specifying the desired test size and random state.
   a. X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
5. Tokenise the training and testing texts using the BERT tokeniser.
   a. tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
   b. X_train_tokens = tokenizer(X_train, padding=True, truncation=True, return_tensors='pt')
   c. X_test_tokens = tokenizer(X_test, padding=True, truncation=True, return_tensors='pt')
6. Convert the tokenised texts into numerical features using BERT's input encoding scheme.
7. Train a multilabel classifier using the BERT model on the training set features and corresponding labels.
   a. classifier = MultiOutputClassifier(Classifier())
   b. classifier.fit(X_train_features, y_train)
8. Predict the labels for the test set using the trained model on the testing selected features.
   a. y_pred = classifier.predict(X_test_features)
9. Calculate the accuracy of the multilabel text feature classification by comparing the predicted labels with the true labels.
   a. accuracy = accuracy_score(y_test, y_pred)
   b. print("Multi-label Text Feature Classification Accuracy:", accuracy)
10. Output the accuracy of the multilabel text feature classification model.

Ensemble learning enhances performance by combining predictions from multiple models. The two popular methods are bagging and boosting. Bagging trains independent models in parallel using random subsets of the data. Boosting Algorithm trains models sequentially, with each model learning from the mistakes of the previous one. Ensemble learning reduces bias and variance, resulting in more flexible and robust models. The classifiers compared are the AdaBoost classifier, which increases the weight of misclassified data points, while Gradient Boosting learns from residual errors. The ridge classifier trees take the text classification as a regression task of assigning two classes. Stochastic Gradient Descent (SGD) is a training

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

algorithm for linear classifiers such as Support Vector Machines (SVM) and logistic regression. It has been designed to handle the training data's dense or sparse arrays of floating-point feature values. Figure 3, 4,5 relates to the above.

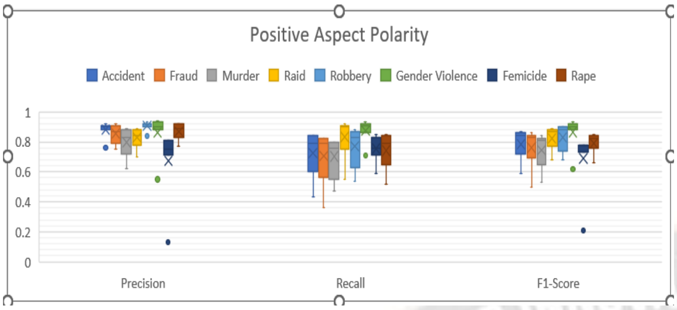

Figure:4 Performance Evaluation of Tweets on Positive Aspect Polarity

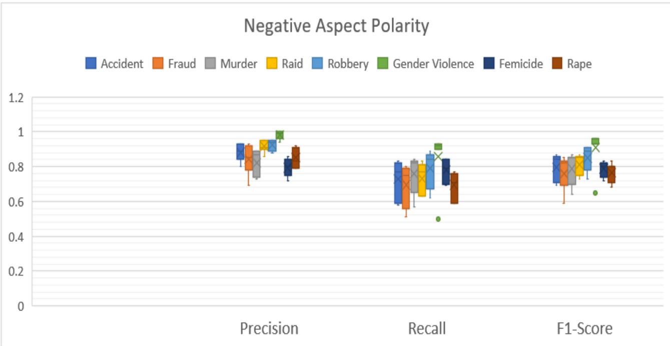

Figure:5 Performance Evaluation of Tweets on Negative Aspect Polarity

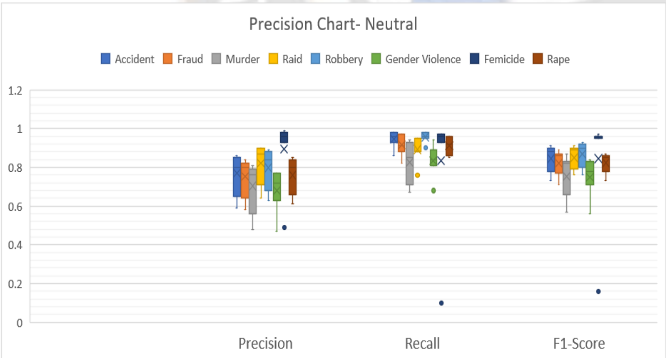

Figure:6 Performance Evaluation of Tweets on Neutral Aspect Polarity

E. Multilabel Text Feature Classification

Multilabel classification suits problems where texts have multiple aspects or belong to various categories. The user opinion and dimension to replying to tweets, retweets, mentions and likes indicate gender based discrimination to create a more significant political focus on violence against women and gender equality. The random forest classifier and Extra Trees Classifier achieves high accuracy scores in most categories, indicating that they can correctly classify positive tweets. The AdaBoost classifier has high accuracy in specific types, such as fraud and robbery, but lower accuracy and recall scores for Murder and rape. Gradient Boosting classifiers indicate a high probability of false negatives, especially in categories such as Murder and robbery. By assigning multiple labels, it captures the complexity and diversity of textual content. The common approaches to this multilabel classification include Feature selection of identifying the most informative features or words in a text that contribute significantly to the prediction of multiple labels. Utilising word embeddings or contextualised embeddings (e.g., BERT) to capture the semantic relationships between words. Inter-label relationships: Analysing the correlations or dependencies between labels and their associated features. Importance ranking: Ranking the importance or contribution of features for each tag. This analysis helps identify the most significant elements for each label and prioritise them accordingly in the classification process.

Aspect Sentiment: Positive

Topic Aspect: Accident, Raid, Femicide

Influencer Aspect: Celebrities, TV professionals, Public Personnel and victims of the crime

Keyword	**Unigrams and Top 10 Words**
Accident	'accident', 'money', 'person', 'insurance': 'daily', immediately', 'Super', 'boying' ', Tucker', 'good', 'incredulous.'
Raid	'Raid': 'FBI', 'police', 'all', 'play', 'people', 'stream', 'special', 'Thank', 'game', 'debut.'
Femicide	'femicide',' killed', 'women', 'man',' legal', 'identify', 'sentence', 'girls', 'gay', 'hires', 'surrogate.'

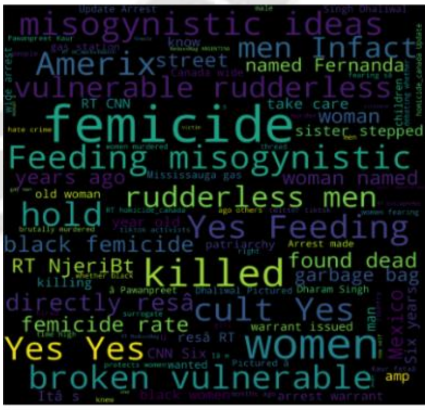

Figure:7 Word cloud of Positive " #Femicide" tweet

Femicide and other forms of violence have five gender dimensions: the gender of the victim, the gender of the abuser, is the abuser a family member, partner or other family member) of the victim, if there is a sexual aspect (for example, rape) and sexual motives. Tweets indicate that choosing one gender aspect

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

or another is the consequence of a careful balance between universalism and particularism, between feminist strategies found in different situations. The prevalence of crime across places and periods, an association between commuting populations and corruption, and famous personalities and fan clubs received the highest number of tweets.

Aspect Sentiment: Negative

Topic Aspect: Accident, Fraud, Robberry, Femicide, Gender violence

Influencer Aspect: Political, Govt Agencies, Crime, Gender discrimination

Keyword **Unigrams and Top 10 Words**

Accident 'WallStreetSilv', 'Turkmenistan', 'Soviet', Westbound',' Ignites', 'Alhambra', 'Explosion', 'Bridge', 'Crashes',

Fraud 'fraud', 'election', 'investigation', 'politicians', 'Nigeria', 'MuellerSheWrote', 'RitchieTorres', 'Trump',

Robbery 'NYPD', 'prison', 'conviction', 'lied', 'framed', 'Victim', 'money', 'Jordan', 'car', 'Armed', 'chase', 'ROBBERY.'

Femicide femicide', 'misogynistic',' broken', 'vulnerable', 'Canada', 'family', 'gender', 'men', 'lose', 'millions', 'dollars.'

Gender violence 'violence', 'DarrigoMelanie', 'gender', 'sexual', 'exploitation',' refugee', 'camps', 'Anika', 'Krstic', 'director',

Figure:8 Word cloud of Negative " #Femicide" tweet

In the Neutral aspect polarity, The Multinomial Naive Bayes algorithm recall values are also relatively high, suggesting an excellent ability to capture true positives. However, the algorithm struggles with precision for categories like femicide and rape. A Random forest classifier has the ability to capture true positives. The AdaBoost classifier has relatively low precision and recall values for most classes, which is unsuitable for this study. Overall, the results suggest that algorithms such as Random Forest Classifier, Extra Trees Classifier, and Ridge Classifier perform consistently well across multiple categories for classifying neutral tweets.

Aspect Sentiment: Neutral

Topic Aspect: Accident, Fraud, Robberry, Femicide, Gender violence

Influencer Aspect: Biased crime words, Politics, Person Names, places and vehicles used in crime

Keyword **Unigrams and Top 10 Words**

Accident 'motorcycle', 'car', 'bumper', 'Fuel', 'Tanker', 'Truck'

Fraud 'Britain', 'CORRUPT', 'RACIST', 'CHILD', 'MOLESTER', 'election"Giuliani', Sidney', 'FRAUD',

Rape 'Rape shelter', 'sexual assault', 'rape victim', 'baby beginner', 'sexual assault', 'murder', 'impeached,

Femicide 'woman', 'JulieSLalonde', 'pandemic', 'Canada', 'legislation', 'deadlier', 'fault', 'divorce', 'abortion

Figure:9 Word cloud of Neutral " #rape" tweet

F. Evaluation Performance Scores

Overall, based on the provided accuracy scores, the SVM Classifier and Random Forest Classifier algorithms perform better in classifying tweets related to both Crime and SDG-5 categories than the Naïve Bayes Classifier. However, further research and evaluation are necessary to consider other factors such as precision, recall, and F1 scores and the specific dataset and features used in the classification process.

TABLE V: COMPARISON OF ACCURACY MEASURES WITH ENSEMBLE LEARNING

Accuracy Measure	Category: Crime				
Classifier Algorithm	Accident	Fraud	Murder	Raid	Robbery
Multinomial NB	0.79	0.77	0.74	0.8	0.84

Random Forest Classifier	0.86	0.85	0.83	0.88	0.89
Extra Trees Classifier	0.89	0.87	0.86	0.89	0.91
Ada Boost Classifier	0.69	0.63	0.58	0.73	0.74
Gradient Boosting Classifier	0.75	0.73	0.67	0.77	0.78
Ridge Classifier	0.87	0.84	0.83	0.88	0.91
SGD Classifier	0.86	0.84	0.83	0.87	0.89

Accuracy Measure	Category: SDG-5		
Classifier Algorithm	Gender Violence	Femicide	Rape
Multinomial NB	0.87	0.9	0.79
Random Forest Classifier	0.9	0.91	0.83
Extra Trees Classifier	0.92	0.91	0.85
Ada Boost Classifier	0.62	0.26	0.7
Gradient Boosting Classifier	0.88	0.89	0.76
Ridge Classifier	0.92	0.92	0.83
SGD Classifier	0.91	0.91	0.83

V. ANALYSIS AND SUGGESTIONS

Multilabel text feature classification helps and its prominent feature analysis:

Handling multiple categories: Multilabel text feature classification allows tweets to be assigned numerous labels or categories simultaneously. The study enables the variety of tweets into various types, such as Accident, Fraud, Murder, Raid, Robbery, Gender Violence, Femicide, and Rape. This approach captures the diverse nature of incidents and crimes discussed on Twitter.

Nuanced analysis: Multilabel classification enables a more nuanced analysis of tweets by considering multiple categories. It captures the complexity and diversity of the content, allowing for a comprehensive understanding of the expressibility of tweets across various incidents and crimes. The study can gain insights into the co-occurrence of different categories within tweets and their interrelationships.

Improved accuracy: By considering multiple labels simultaneously, multilabel text feature classification enhances the accuracy of tweet classification. The study's findings indicate that algorithms such as the Random Forest Classifier, Extra Trees Classifier, Ridge Classifier, SGD Classifier, and LSTM model exhibited high expressibility and consistently achieved accuracy scores above 0.83. These algorithms effectively capture the features and patterns of different categories, leading to the accurate classification of tweets.

Prominent feature analysis: In multilabel text feature classification, analysing major features involves identifying each category's most informative and relevant segments. This analysis helps understand which words, phrases, or patterns contribute significantly to classifying tweets into specific categories. By determining the prominent features, the study can gain insights into the language and content characteristics of different incidents and crimes discussed on Twitter.

Overall, multilabel text feature classification allows for a comprehensive analysis of tweets across multiple categories, improves classification accuracy, and provides insights into the prominent features associated with each class. It enables a deeper understanding of the expressibility of tweets in the context of various incidents and crimes, as highlighted in the study's findings.

VI. CONCLUSION AND FUTURE SCOPE

In this study on the expressibility of tweets on social issues, the topic prevalence highlighted the public opinions based on sentiments and discovering hidden patterns among the social events with a sparsity of short text as tweets. The dataset collection reflected the trend and hidden ways of social events. The research took a lot of time in data cleaning with the NLTK library, but SpaCy parsing tweets with the BERT transformer model improvised the feature extraction and topic modelling of tweets. The aspect-based sentiment analysis methods are categorised into three Phases machine learning, deep learning and Ensemble multi label text feature classification. The multilabel text feature classification aims to automate the categorisation process, enabling efficient analysis and organisation of large volumes of text data related to these specific incidents of SDG5 on Gender issues or crimes.

REFERENCES

[1] Tucker, R. O. (2021). Who ‘Tweets’ Where and When, and How Does it Help Understand Crime Rates at Places? Measuring the Presence of Tourists and Commuters in Ambient Populations. Journal of Quantitative Criminology, 37, 333 - 359. https://doi.org/10.1007/s10940-020-09487-1

[2] Hannach, H. &. (2018). WordNet based Implicit Aspect Sentiment Analysis for Crime Identification from Twitter. International Journal of Advanced Computer Science and Applications., 9. https://doi.org/10.14569/IJACSA.2018.091222

[3] Prathap, M. V. (2023). Crime Analysis and Forecasting Using Twitter Data in the Indian Context. 2023 International Conference on Artificial Intelligence and Smart Communication (AISC) (pp. 919-924). Greater Noida, India. https://doi.org/10.1109/AISC56616.2023.10085282

[4] Benkhalifa, H. E. (2018). WordNet based Implicit Aspect Sentiment Analysis for Crime Identification from Twitter. International Journal of Advanced Computer Science and Applications(IJACSA), 9(12). http://dx.doi.org/10.14569/IJACSA.2018.091222

[5] Sangeeta Lal, L. T. (2020). Analysis and Classification of Crime Tweets. Procedia Computer Science,, 167, 1911-1919. https://doi.org/1016/j.procs.2020.03.211

[6] Rao, L. (2022). Sentiment Analysis of English Text with Multilevel Features. Scientific Programming. https://doi.org/10.1155/2022/7605125

[7] Kulkarni, V. L. (2022, May 03). CTM - A Model for Large-Scale Multi-View Tweet Topic Classification. North American Chapter of the Association for Computational Linguistics. https://aclanthology.org/2022.naacl-industry.28.pdf

[8] P. Vyas, M. R. (2022, June). Automated Classification of Societal Sentiments on Twitter With Machine Learning. IEEE Transactions on Technology and Society., 3(2), 100-110. https://doi.org/10.1109/TTS.2021.3108963

[9] Fitri Andri Astuti, B. P. (2022). Antisocial Behavior Monitoring Services of Indonesian Public Twitter Using Machine Learning. Proceedings of 2021 International Conference on Data Science and Official Statistics (ICDSOS), Vol. 2021. https://doi.org/10.34123/icdsos.v2021i1.181

[10] Bouadjenek MR, S. S. (2022). A longitudinal study of topic classification on Twitter. PeerJ Computer Science., 8. https://doi.org/10.7717/peerj-cs.991

[11] Ellsberg, M. (2006). Violence against women and the Millennium. International Federation of Gynecology and Obstetrics, 94, 325—332. https://doi.org/10.1016/j.ijgo.2006.04.021

[12] Babu BV, K. Y. (2017). Violence against Women and Girls in the Sustainable Development Goals. Health Promotion Perspectives, 7(1), 1-3. https://doi.org/10.15171/hpp.2017.01

[13] Gu, X. &. (2022). Research on gender prediction based on short texts on Chinese social platforms. 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE).

[14] Brooks, R. C.-B. (2022). Incel Activity on Social Media Linked to Local Mating Ecology. Psychological Science, 33(2), 249–258. https://doi.org/10.1177/09567976211036065

[15] Ismet, H. M. (2022). Aspect Based Sentiment Analysis of Product Review Using Memory Network. Scientific Journal of Informatics., 9(1). https://doi.org/10.15294/sji.v9i1.34094

[16] Trang Uyen Tran, H. T. (2022). Toward a multitask aspect-based sentiment analysis model using deep learning. IAES International Journal of Artificial Intelligence (IJ-AI), 11 ((2)), 9. https://doi.org/10.11591/ijai.v11.i2.pp516-524

[17] Ray, P. A. (2022). A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis. Applied Computing and Informatics, 18, 163-178. https://doi.org/10.1016/j.aci.2019.02.002

[18] Rezaeenour, J. A. (2023). Systematic review of content analysis algorithms based on deep neural networks. Multimed Tools Appl, 82, 17879–17903. https://doi.org/10.1007/s11042-022-14043-z

[19] Rahman M, I. D. (2020). A deep learning approach based on convolutional LSTM for detecting diabetes. Comput Biol Chem. https://doi.org/10.1016/j.compbiolchem.2020.107329

[20] Grootendorst, M. (2022). BERTopic: Neural topic modelling with a class-based TF-IDF procedure. arXiv:2203.05794.

[21] Martoglia, R. &. (2022). A Predictive Method to Improve the Effectiveness of Twitter Communication in a Cultural Heritage Scenario. ACM Journal on Computing and Cultural Heritage (JOCCH), 15. https://doi.org/10.1145/3470786

[22] Hourrane, O. &. (2022). Topic-Transformer for Document-Level Language Understanding. Journal of Computer Science., 18(1). https://doi.org/10.3844/jcssp.2022.18.25

[23] Bozarth, L. &. (2022). Keyword expansion techniques for mining social movement data on social media. EPJ Data Science., 11. https://doi.org/10.1140/epjds/s13688-022-00343-9

[24] Kaur, G. S. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J Big Data, 10(5). https://doi.org/10.1186/s40537-022-00680-6

[25] Kunapuli, G. (2023). Ensemble Methods for Machine Learning. (9781617297137, Ed.)

[26] Zhang, R. (2022). Research on Social Media Feature Learning Algorithm Based on Deep Neural Network. IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), (pp. 590-594). Shenyang, China. IEEE. https://doi.org/10.1109/ICPECA53709.2022.9719080

[27] Li Xiaoyan, R. C. (2022). GloVe-CNN-BiLSTM Model for Sentiment Analysis on Text Reviews. Journal of Sensors, 12. https://doi.org/10.1155/2022/7212366

[28] Liang, Y. L. (2021). Fusion of heterogeneous attention mechanisms in multi-view convolutional neural network for text classification. Information Sciences, 548, 295-312. https://doi.org/10.1016/j.ins.2020.10.021

[29] Dimosthenis Antypas, A. P.-C. (2023). Negativity spreads faster: A large-scale multilingual Twitter analysis on the role of sentiment in political communication. Online Social Networks and Media, 33. https://doi.org/10.1016/j.osnem.2023.100242.

[30] Xu, S. &. (2023). CNN-VAE: An intelligent text representation algorithm. The Journal of Supercomputing, 1-26. https://doi.org/10.1007/s11227-023-05139-w.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7528
***Article Received**: 26 April 2023 **Revised:** 12 June 2023 **Accepted:** 29 June 2023*

[31] Ghazi Felhi, J. L. (2021). Challenging the Semi-Supervised VAE Framework for Text Classification.https://doi.org/10.48550/arXiv.2109.12969
[32] Ahmad S, A. M. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Hum Cent Comput Inf Sci, 9(24). https://doi.org/10.1186/s13673-019-0185-6