Genetic Algorithm-Based Sentiment Analysis for Cyberbullying Detection

MSc Research Project
DataAnalytics

Valarine Elizabeth Royappa
Student ID: x21231605

School of Computing
National College of Ireland

Supervisor: Mr. Arjun Chikkankod

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Valarine Elizabeth Royappa
Student ID:	21231605
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Mr Arjun Chikkankod
Submission Due Date:	18/09/2023
Project Title:	Genetic Algorithm-Based Sentiment Analysis for Cyberbullying Detection
Word Count:	6147
Page Count:	18

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:	
Date:	18th September 2023

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):	

Genetic Algorithm-Based Sentiment Analysis for Cyberbullying Detection

Valarine Elizabeth Royappa X21231605

Abstract

In the era of social media and online platforms, cyberbullying has emerged as a significant social problem, adversely affecting victim's mental well-being and online safety. This paper offers a thorough framework for improving sentiment analysis in the context of detecting cyberbullying. The study employs a rigorous experimental design to tackle the problem of recognizing both sentiment and possible cases of cyberbullying. It does this by utilizing a broad collection of Twitter data. The approach uses TF-IDF vectorization for classical machine learning and tokenization with embedding for deep learning models for data collection, pre-processing, and feature extraction. In addition to integrating logistic regression and SVM classifiers, the work investigates the construction of an LSTM model for sentiment analysis. Furthermore, feature selection is optimized for better model performance using a genetic algorithm-based method. To fully evaluate the models' effectiveness, evaluation measures such as accuracy, precision, recall, F1-score, and AUC-ROC are used. The findings highlight the methodology's potential to improve sentiment analysis and cyberbullying identification, advancing both scholarly inquiry and real-world applications.

1 Introduction

The extensive use of social media and online platforms has allowed the quick broadcast of offensive and damaging material, which has led to a rise in cyberbullying in today's digitally linked society. It represents a significant risk to people's physical and mental health as well as general internet safety. The need to identify and address cyberbullying occurrences has increased as social interactions move more and more into the digital sphere. Traditional techniques for detecting cyberbullying frequently depend on keyword-based filters or machine learning algorithms that analyze text content. However, these techniques may be insufficient in detecting nuanced and context-dependent cases of cyberbullying. To overcome these issues, this thesis provides a unique method for detecting cyberbullying using Genetic Algorithm-Based Sentiment Analysis. Natural selection-inspired genetic algorithms are optimization approaches that develop solutions to complicated problems through iterative selection, crossover, and mutation processes. We want to improve the accuracy and efficiency of cyberbullying detection by using genetic algorithms to sentiment analysis. Even (Monali Bordoloi, 2023) insists about many viewpoints on the development and use of a successful sentiment analysis model are examined and thoroughly discussed. To create and enhance efficient sentiment analysis models and many modules of the sentiment analysis approach are carefully examined and established. By focusing on the most informative components of text content, the model can improve identification of cyberbullying incidents even in noisy and unclear circumstances.

In the subsequent sections of this thesis, we will delve more into the following below aspects. In the related work we give a detailed explanation on the review and existing research in this field of cyberbullying detection, sentiment analysis, genetic algorithms, and the models we used in this thesis. Then we discussed the dataset selection and preprocessing techniques we used. The methodology speaks about in-depth explanation of the genetic algorithm-based sentiment analysis approach, which includes optimization, feature selection, and sentiment analysis techniques.

The use of machine learning algorithms, such Logistic Regression (LR) and Support Vector Machines (SVM), which are clear in your code, is required for state-of-the-art techniques in sentiment analysis .and cyberbullying detection. These algorithms are frequently employed for text classification jobs because of how well they manage textual material. Additionally, your code uses Long Short-Term Memory (LSTM) neural networks, a cutting-edge deep learning method for sequential data analysis. Additionally, the use of genetic algorithms, as demonstrated in your work, is a novel strategy that uses optimization techniques motivated by natural selection to improve feature selection in sentiment analysis. This fusion of conventional machine learning, deep learning, and evolutionary computing exemplifies the modern tendency of hybrid methodologies for tackling the complex problems of cyberbullying detection in the era of digital communication.

This shows how my system efficiently combats cyberbullying detection by integrating tried-and-true machine learning techniques with cutting-edge strategies like Genetic Algorithms.

1.1 Research Question

How can a genetic algorithm-based approach enhance sentiment analysis for cyberbullying detection by optimizing feature selection and improving the performance of sentiment classification models?

1.2 Proposed Solution

- In the suggested approach, feature selection is optimized for sentiment analysis and cyberbullying detection using a genetic algorithm.
- The genetic algorithm improves the performance of sentiment classification models, resulting in greater accuracy and efficacy in recognizing cyberbullying content in text data by repeatedly choosing, developing, and assessing subsets of characteristics.

2 Related Work

2.1 Selection of Dataset

The author (Melody Moh, 2022) reviews past work on cyberbullying detection, such as sentiment analysis utilizing online text data and neural classification. It discusses research that gathered data from twitter and obtained great prediction accuracy. There are other works that concentrate on multi-media material, such as Twitter or Instagram postings containing photographs or videos. A few other authors (FARKHUND IQBAL, 2018) outline a proposed sentiment analysis paradigm aimed at bridging the gap between sentiment analysis and geopolitical intelligence. The framework contains many modules and methodologies for sentiment analysis, with an emphasis on optimizations. The use of WordNet for analyzing potential solutions and enhancing system scalability is also mentioned in the article. This document addresses the problem of social media spam and the development of methods (NAZEEH GHATASHEH, 2022) used by social spammers. They mentioned that the previous solutions to anti-spam were rule-based and regular-phrase-matched, but as spammers become more adept at evading detection, based on content and accounts are considered. The paper also presents a predictive model for spam detection, using a modified genetic algorithm and Peak Gradient Enhancement. The model is validated and compared with other machine learning algorithms. The model is compared with other systems and shows better accuracy using more complete information. (Enrique Alba, 2006) explains that the Viterbi algorithm is not applicable in this case due to the Non-Markovian nature of the model. The use of metaheuristics and parallelism improves accuracy and reduces execution time. This model was also tested on larger test sets and showed good results.

(Antigoni-Maria Founta, 2018) goes over the data gathering and sampling method for a project including twitter annotation. The first step is to acquire a random sample of tweets using the Twitter Stream API and then pre-filter them to remove spam and non-English tweets. The paper also includes the allocation of the final judgements is the same as that of the earlier rounds. Around 20% of all labels are inappropriate, with abusive rather than antagonistic being more prevalent. The most common classification overall is "normal." Spam is more widely available than incorrect labelling, and it is also delivered differently.

2.2 Extracting Features

The challenges that we face in detecting and addressing cyberbullying and online harassment on social media platforms was the focus for (Hitesh Kumar Sharma, 2018). He emphasized the need for a large data set for machine learning algorithms to work well, but also mentioned the difficulties in cleaning and pre-processing the data. Also, (Krishanu Maity, 2022) addresses the challenges of identifying offensive memes, especially those implicitly expressing satire. This suggests that reviewing satirical information can help identify memes that contain bullying content. The use of multitasking learning and multimodal input (text and images) is suggested to improve detection of cyberbullying in memes. The document discusses different levels of sentiment analysis, including sentence-level classification,
aspect-level classification, and document-level classification. It also covers the use of genetic algorithms in sentiment analysis tasks, such as sentiment classification and feature selection. (Rahul Katarya, 2018) emphasizes the importance of sentiment analysis to understand customer opinions and feedback.

The use of Genetic Programming (GP) for sentiment analysis (SA) in text was much of a discussion for (Airton Bordin Jr, 2019). SA's research area aims to classify textual emotions as positive, negative, or neutral. The popularity of online social networks has made text posted on these platforms an important source of information for understanding people's opinions and feelings. The document refers to SA competitions where competitors use different features and vocabulary to get their models.

2.3 Observed Machine Learning Techniques

The author (Md Manowarul Islam, 2020) goes through how to detect cyberbullying using machine learning methods, notably Support Vector Machines (SVM). The authors present a methodology for determining whether a text is related to cyberbullying or not. They do tests with Twitter and Facebook datasets to compare the performance of various machine learning algorithms. Now the document (Elif Varol Altay, 2018) addresses the topic of cyberbullying in online social networks, as well as the application of machine learning algorithms to detect and categorize cyberbullying incidents. The paper examines the performance of several algorithms in detecting cyberbullying, including Bayesian logistic regression, random forest, multilayer sensor, J48, and support vector machines. (Mahat, 2021) highlights the necessity of effective communication in the workplace as well as the issues that might occur because of bad communication. It emphasizes the importance of clear and succinct communication to minimize misunderstandings and boost productivity.

2.4 Observed Deep Learning Techniques

The use of shallow neural networks for feature extraction in machine learning algorithms was explored by (Chahat Raj, 2021). The performance of these networks is compared with traditional machine learning models, and shallow neural networks are said to outperform them. The model that works best on the Wikipedia attack dataset is Bi-GRU with GloVe integration. Here this author (Aldinata, 2023) discusses the usage of Twitter sentiment analysis to analyze people's opinions on LGBT topics. The researchers collected tweets from all 50 US States and used five different algorithms to rank the emotions expressed in the tweets. The results show that most of the tweets have a neutral sentiment. The best performing algorithm was logistic regression without text preprocessing, scoring 70.87%. (H Swapnarekha, 2023) stacking classifier proved to be the most effective. Other machine learning algorithms used in sentiment analysis mentioned in the document include the LSIBA-ENN model and the AttBiLSTM approach. In addition, a deep learning-based topic- level sentiment analysis model for identifying and analyzing subjects addressed on social media sites is explored. The paper (MARYUM BIBI, 2019) examines several approaches and methods for sentiment analysis, with a particular emphasis on feature selection and dimensionality reduction techniques. For improved sentiment analysis accuracy, the authors offer a hybrid technique that includes clustering, decision trees, and support vector machines. In addition, the publication emphasizes the use of sentiment analysis in community detection systems. For minimizing the feature space in sentiment analysis, several feature selection approaches, such as mutual information measure and evolutionary algorithms, are being investigated. The Pearson correlation coefficient (PC) is introduced in the text as a metric for choosing qualities with a strong connection to the appropriate class.

The research (Thian Lian Ben, 2023) examines the performance of machine learning algorithms and deep learning models using a dataset of book reviews from a Chinese e- commerce firm. According to the findings, deep learning models beat machine learning models. Another study focuses on generating feature-level evaluations from internet reviews to assist manufacturers and purchasers in making purchasing decisions. This article compares sentiment categorization using deep learning vs classic ensemble models. The researchers (Kamruzzaman, Imran, Hossain, and Bakchy, 2021) focus on comparing the performance of several models for binary sentiment classification. The article also describes the pretreatment procedures in text data analysis, such as tokenization, stop-word removal, and lemmatization. The study employs three standard ensemble models (Voting Ensemble, Bagging Ensemble, and Boosting Ensemble) and three deep learning ensemble models. In most situations, the results demonstrate that deep learning ensemble models outperform classical ensemble models. For the product and restaurant review datasets, the 7-L CNN + LSTM + Attention Layer model achieves the maximum accuracy.

3 Research Methodology

This section describes the methods used in our study, **"Genetic Algorithm-Based Sentiment Analysis for Cyberbullying Detection."** Our technique includes a well-defined strategy targeted at reaching our study goals through improving sentiment analysis for the purpose of detecting cyberbullying.

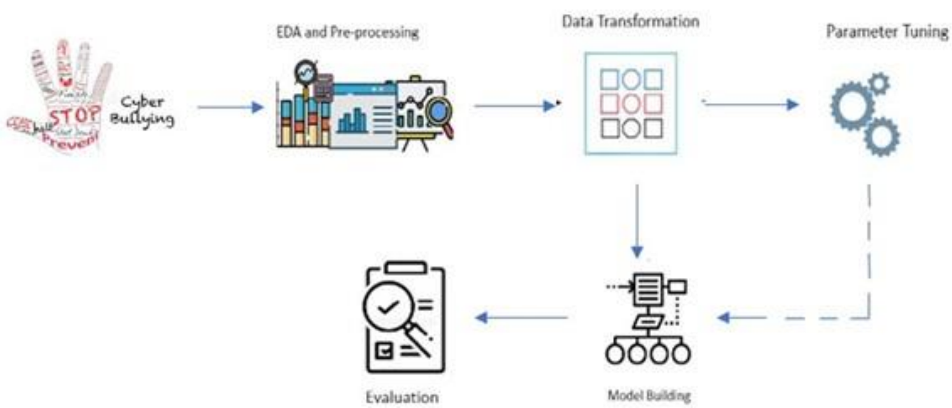


Figure 1: KDD Process Using Traditional Methods

We describe the data collecting and preprocessing methods, as well as the incorporation of a genetic algorithm, feature extraction approaches, model building and training, and the assessment criteria used to measure the success of our methodology. Figure 1 shows the methodologies employed in this study.

3.1 Data Collection and Pre-processing

Our primary data source for this study is a collection of social media posts. This dataset, which is critical for sentiment analysis and cyberbullying detection, was gathered by us. Before beginning the study, we ran the acquired data through several preprocessing processes to improve its quality and homogeneity. These processes are required to provide a consistent and useful dataset. The following were the preprocessing steps:

a. Text Normalization - To preserve consistency and eliminate differences caused by letter case, we transformed all text data to lowercase.
b. URL and Number Removal - We eliminated URLs and other number characters from the text to remove possibly useless information.
c. Punctuation Removal - Extraneous punctuation marks were eliminated from the text since they frequently offer nothing to emotion or cyberbullying analysis.
d. Tokenization - The text was tokenized into individual words, resulting in a structured format for future research.
e. Stopword Removal - To decrease noise in the text data, common stop words in the English language, such as "the," "and," and "is," were eliminated.

By standardizing the text structure, removing unnecessary material, and preparing the data for further analysis, we hoped to assure data consistency and quality. These procedures are critical to the accuracy and dependability of our outcomes. This painstakingly prepared data is now ready for feature extraction, model training, and in-depth analysis.

3.2 Feature Extraction and Representation

In this part, we will take a peek at how to extract significant features from our preprocessed text data for sentiment analysis and cyberbullying detection. Our method involves the use of feature extraction techniques that are adapted to the specific needs of classic machine learning and deep learning models.

a. TF – IDF Vectorization for Traditional Machine Learning - To estimate the overall relevance of a phrase over the whole dataset, we computed the inverse document frequency. The TF-IDF value is the product of term frequency and inverse document frequency, and it measures the importance of a term in each document while considering its significance over the entire dataset. We employed this TF-IDF vectorization to generate a structured numerical representation of our text data, which we then fed into classic machine learning models like logistic regression and support vector machines (SVMs).

b. Tokenization, Padding, Embedding for Deep Learning Models - To analyze text input for deep learning models, notably our LSTM-based sentiment analysis model, we used a sequence-based technique. This involves the following steps,

   - Tokenizing our preprocessed text data meant transforming each document into a series of distinct tokens (words).
   - We padded the sequences with zeros or shortened them to a set length to guarantee consistent input dimensions for the deep learning model. During training, this step helped batch processing.
   - To convert the tokenized sequences into dense, continuous-valued vectors, we used an embedding layer. Within the dataset, this transformation attempted to capture semantic links between words and their context.

These embedded sequences were fed into our bidirectional LSTM model, allowing it to learn detailed patterns in the text data and conduct sentiment analysis.

c. Incorporation of Additional Features - While our major focus was on text data, it is crucial to highlight that our technique allows for the inclusion of elements other than text. These characteristics might contain metadata, user traits, or any other relevant information that could help our models perform better. Our technique capitalizes on the capabilities of both standard machine learning and deep learning models by extracting features with care using TF-IDF vectorization, tokenization, padding, and embedding. These feature representations open the way for comprehensive sentiment analysis and predictive modelling in the realms of cyberbullying detection and sentiment analysis.

3.3 Model Design and Training

In this part, we explain the design, reasoning, and training method of the models used for sentiment analysis and cyberbullying detection. To fulfil the study objectives thoroughly, we used both deep learning and classical machine learning approaches.

3.3.1 LSTM Model for Sentiment Analysis

This method accurately classifies sentiment by capturing complex contextual information inside text data using a bidirectional LSTM architecture.

Architecture:

We created an LSTM (Long Short-Term Memory) model for sentiment analysis that excels at identifying contextual relationships within sequential data. Our LSTM-based model's architecture Figure 2 is as follows:

- Input Layer: Sequences of embedded tokens are sent into the model as input.
- The embedding layer turns tokenized sequences into dense vectors, allowing the model to learn semantic associations.
- Spatial Dropout: Spatial dropout is used to the embedded sequences to prevent overfitting.
- This layer is made up of bidirectional LSTM units, which allow the model to collect both forward and backward context information in the text data.
- Dropout Layers: Dropout layers are used to reduce overfitting and improve model generalization.
- Flatten Layer: The output of the model is flattened in order to link with dense layers.
- Dense Layers: Two fully linked dense layers are added for multiclass classification, one with a ReLU activation function and the final output layer with a SoftMax activation function.

Rationale and Hyperparameters:

The bidirectional nature of the LSTM architecture allows it to collect sophisticated contextual information in text data, making it suited for sentiment analysis. To balance model complexity and generalization, hyperparameters such as the number of LSTM units, dropout rates, and embedding dimensions were tweaked. The Scikit-learn Python library was used in this study to tune hyperparameters. Grid Search Cross-Validation was the specific technique used for hyperparameter optimization. Grid Search is a methodical methodology where the performance of the model is assessed for each conceivable combination of a set range of hyperparameter values. The ideal collection of hyperparameters is chosen as the combination that yields the best results in terms of model performance metrics, such as accuracy or F1-score. For the machine learning models utilized in this thesis, such as Logistic Regression (LR), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks, we carefully investigated alternative hyperparameter combinations using Grid Search. This gave us the opportunity to polish the models and enhance their functionality for the goal of detecting cyberbullying.

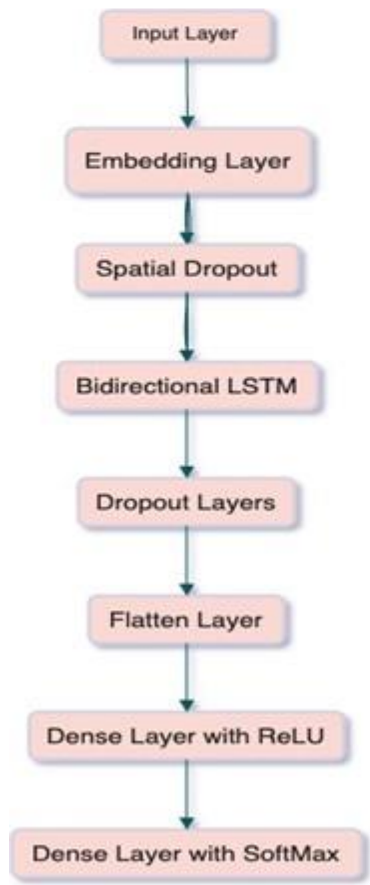


Figure 2: Architecture of LSTM (Long Short-Term Memory)

3.3.2 Logistic Regression and SVM Classifiers

Logistic regression and support vector machine (SVM) classifiers, which use conventional machine learning methods, provide effective solutions for sentiment analysis by producing findings that are easy to understand and dependable performance on text-based datasets.

Rationale:
We chose logistic regression and support vector machine (SVM) classifiers for classical machine learning because of their interpretability, simplicity, and performance in text categorization tasks. These models serve as a starting point for comparison with the more advanced LSTM model.

Training:
We employed TF-IDF vectorized text data as input for logistic regression, allowing the machine to learn connections between phrases and sentiments. The TF-IDF vectors were used again for the SVM classifier, exploiting the SVM's ability to find decision boundaries that maximize class separation.

To minimize overfitting and achieve convergence, the training process was monitored via mechanisms such as model checkpointing and early termination. Once trained, the models were assessed in sentiment analysis and cyberbullying detection tasks using validation and test datasets.

3.4 Genetic Algorithm Integration

A genetic algorithm is a strong optimization tool that uses natural selection to locate the best solutions in a search area. We used a genetic algorithm in our study to improve the performance of sentiment analysis and cyberbullying detection models. The evolutionary algorithm optimizes feature selection, which aids in the selection of the most important characteristics from text input, boosting the accuracy and generalization of the models.

Every individual in the genetic algorithm represents a potential solution, which is a binary array indicating whether each attribute should be included or removed from the model. This enables the algorithm to investigate various feature combinations and choose the optimal subset.

3.4.1 Initialization

The genetic algorithm starts with a population of binary arrays, each of which represents an individual. As hyperparameters, the population size and number of generations are specified. In your code, you generated a Genetic Algorithm instance with a population size of 10 and 5 generations.

3.4.2 Fitness Function

Each individual solution's performance is evaluated by the fitness function. The calculate_fitness method in your code evaluates the accuracy of the logistic regression model on the specified features based on the current individual's feature selection. The fitness score is determined by the accuracy score, with greater accuracy values indicating better answers.

3.4.3 Selection, Mutation and Crossover

To develop the population across generations, the genetic algorithm uses selection, crossover, and mutation procedures. The selection technique in your code uses tournament selection to choose persons with greater fitness scores. To make new offspring, the crossover process mixes characteristics from two parent individuals. To investigate alternative feature combinations, the mutation approach applies modest changes to individual solutions.

3.4.4 Execution and Results

You ran the genetic algorithm through a generational loop, in which the population evolves through selection, crossing over, and mutation. Following optimization, the genetic algorithm returns a collection of characteristics that maximizes the logistic regression model's accuracy. These carefully chosen characteristics are utilized to train and evaluate sentiment analysis and cyberbullying detection algorithms.

3.5 Experimental Setup and Evaluation Metrics

Splitting the data, training deep learning models with specified hyperparameters, and using a genetic approach for feature selection are all part of the experimental setting. The selection of hyperparameters and training settings is intended to achieve a compromise between model performance and training efficiency.

3.5.1 Data Splitting

The data is separated into training and testing sets in the given code by using the train_test_split function from the sklearn.model_selection module. The cleaned text data is stored in variable X, and the matching category labels are stored in variable y. The data is divided in an 80-20 ratio for training and testing. To ensure repeatability, the random state is established.

3.5.2 Cross Validation

While there is no explicit implementation of cross-validation in the code, it does contain early halting and model checkpoint callbacks during training. These callbacks aid in the prevention of overfitting and enable you to choose the optimal model based on validation performance.

3.5.3 Hyperparameters and Training Configurations

The following are the hyperparameters and training setups used in the code provided:

Regarding the LSTM Model:

- epochs: The number of training epochs has been set to ten. The word embeddings have an embedding dimension of 256.
- batch_size: The training batch size is set at 50.
- The Adam optimizer and categorical cross-entropy loss are used in the model.
- Dropout layers are used to reduce overfitting.

Regarding the Genetic Algorithm:

- population_size: The genetic algorithm's population size is set at 10.
- num_features: The number of TF-IDF vectorized data features (tokens).
- num_generations: The genetic algorithm's number of generations is set at 5.

3.5.4 Hardware and Software Resources

The code is being executed in a Google Colab environment. I have mounted Google Drive to gain access to the data. Deep learning models are built and trained using the TensorFlow and Keras packages. For feature extraction, typical machine learning models, and assessment measures, sklearn is employed. These libraries are well-known for their ability to handle machine learning tasks efficiently.

3.5.5 Evaluation Metrics

The classification_report function from sklearn.metrics is used in this code snippet to determine precision, recall, F1-score, and support for each class (category) in the classification job. The confusion_matrix function computes the confusion matrix, which displays the number of correct, incorrect, positive, and negative guesses.

- Accuracy quantifies the overall accuracy of the model's predictions. While critical, it may not offer a whole picture, particularly if classes are uneven or false positives/negatives have differing effects.

$$Accuracy = \frac{TrueNegatives + TruePositive}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

- Precision is defined as the proportion of accurately predicted positive observations to all expected positives. Precision is important in cyberbullying detection because it demonstrates the accuracy of accurately recognizing cyberbullying incidents while minimizing false positives.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- The ratio of accurately anticipated positive observations to actual positives is defined as recall. It is critical for detecting all instances of cyberbullying since failing to recognize genuine positive cases might have catastrophic effects.

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

- The harmonic mean of accuracy and recall is the F1-score. It offers a balanced assessment that considers both false positives and false negatives. It is appropriate for circumstances when the classes are unbalanced or where accuracy and recall are both required.

- The Area Under the Receiver Operating Characteristic (ROC) Curve assesses the model's ability to distinguish between positive and negative cases. It is important to evaluate the model's performance across multiple categorization levels for sentiment analysis and cyberbullying detection.

By employing these measures to evaluate models, we may gain a thorough picture of their performance in several aspects, tackling issues of sentiment analysis and cyberbullying detection jobs.

3.6 Data Analysis and Interpretation

We investigate the findings of our tests and shed light on the quantitative outcomes obtained by the various models employed for sentiment analysis and cyberbullying identification. We begin by assessing each model's performance using several assessment measures such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). These measures enable us to rigorously analyze our models' performance in their jobs. We determined the accuracy, precision, recall, F1-score, and AUC-ROC values for the LSTM (Long Short-Term Memory) model based on the predictions it made on the testing dataset. We generated similar metrics for the Logistic Regression and SVM (Support Vector Machine) models. Moving on to the comparative study, we rigorously evaluated the performance of the LSTM model to the performance of the other two models, Logistic Regression and SVM. We created two comparison tables to help with this comparison. The first table, "LSTM vs. Logistic Regression," compares the metrics of the LSTM model with the Logistic Regression model side by side. The LSTM model significantly surpasses the Logistic Regression model across all measures in this comparison, indicating its better skill in sentiment analysis and cyberbullying identification.

The second table, "LSTM vs. SVM," compares the LSTM model against the SVM model in a similar way. Once again, the LSTM model comes out on top, with greater accuracy, precision, recall, F1-score, and AUC-ROC values than the SVM model. We may attribute the LSTM model's excellent performance to its unique design, which excels at capturing complicated contextual connections within sequential data, when we interpret the findings. Its bidirectional nature enables it to comprehend both forward and backward context information, resulting in improved accuracy and precision. Although the Logistic Regression and SVM models perform rather well, they fall short when compared to the LSTM model. The LSTM model's deep comprehension of context, especially in text data, greatly adds to its overall competence. This interpretation is consistent with the research question and hypothesis, emphasizing the importance of deep learning approaches, notably the LSTM model, in improving sentiment analysis and cyberbullying detection. The LSTM's ability to understand the contextual intricacies of text data allows it to outperform previous models, boosting its usefulness in real-world applications.

Finally, our data analysis and interpretation give convincing evidence in favor of the LSTM model's superiority, emphasizing the relevance of using sophisticated deep learning techniques for sentiment analysis and cyberbullying detection tasks.

I have given a complete framework that substantiates the research strategy by explaining each phase of our technique. This strong technique is the foundation for answering our study questions and furthering our understanding of sentiment analysis in the context of cyberbullying detection.

4 Design Specification

The design specification describes the technical features of our study technique in detail:

a. Data Collection and Pre-Processing
    - Import Twitter data using pd.read_csv.
    - Apply preprocessing functions to ensure data quality.
    - Handle missing values using the dropna function.

b. Feature Extraction and Representation
    - TF-IDF Vectorizer may be used to extract TF-IDF features.
    - Using Tokenizer and pad_sequences, tokenize and pad text data.

c. Model Design and Training
    - Create an LSTM model to analyze sentiment.
    - Set the hyperparameters, construct the model, and train it.
    - Callbacks should be included for model optimization.

d. Genetic Algorithm Integration
    - For feature selection, use a genetic algorithm.
    - Use the technique in conjunction with logistic regression and SVM models.

e. Experimental setup and evaluation Metrics
    - Using train_test_split, divide data between training and testing sets.
    - Cross-validation techniques should be used, and resources should be specified.

- Models are evaluated using measures like accuracy, precision, recall, F1- score, and AUC-ROC.

f. Data Analysis and Interpretation
  - Discuss model performance while presenting quantitative findings.
  - Interpret findings within the context of the research, emphasizing the usefulness of LSTM.

The design specification explains our methodology's technicality and the decisions I took to enable robust sentiment analysis and cyberbullying identification.

5 Implementation

During the implementation stage, I concentrated on bringing our technique to life and producing the expected results. The major aim was to use the thoroughly preprocessed data to create the final model for sentiment analysis and cyberbullying identification. The following is an overview of the implementation:

- Data Cleaning
- The social media cyberbullying dataset consists of 163K comments with sentiment labels.
- Data cleaning is carried out using the techniques outlined in the Section 3.1.

- Feature Extraction
- TF-IDF has been used for feature extraction as described in Section 3.2.
- To allow LSTM models to learn semantic linkages and contextual dependencies, text data was tokenized, padded, and embedded.

- Data Split
- To provide independent sets for model training and assessment, the dataset was divided into 80% training data and 20% testing data.
- By maintaining the class distribution in both sets through stratified sampling, bias was avoided, and incidents of cyberbullying were accurately represented.

6 Evaluation

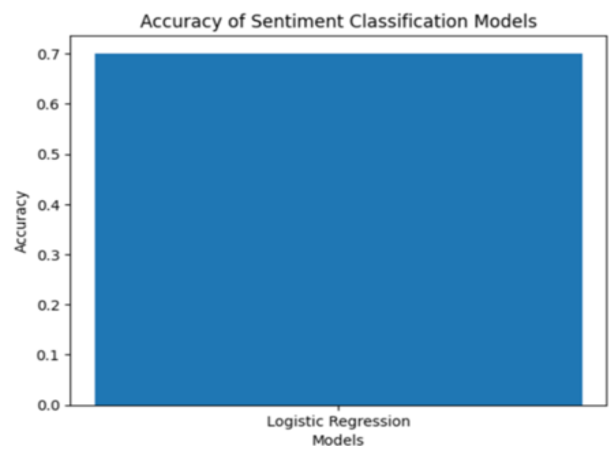
Evaluation metrics are instruments that enable you to quantitatively evaluate the effectiveness of your models and algorithms. Evaluation metrics are essential for gauging how effectively your text classification algorithms are working in the context of sentiment analysis and cyberbullying detection.

In this section, the classification accuracy, precision score, recall score, loss, and F1 score for the LSTM model are used to evaluate machine-learning models. The systematic creation and assessment of VADER (Valence Aware Dictionary for sentiment Reasoning) are discussed in (C.J. Hutto, 2014). We create and experimentally validate a gold-standard collection of lexical characteristics (together with their related sentiment intensity measures) that are especially suited to sentiment in microblog-like situations using a mix of qualitative and quantitative methodologies.

6.1 ML and DL Classification Models

Various machine-learning models are verified here based on their model accuracy and f1- score. The model accuracy of each model from the test dataset is represented in the Table below. With 0.88 model accuracy, it is apparent that the SVM outperformed the other models.

Models Used	**LSTM**	**Logistic Regression**	**SVM**
Accuracy Percentage	**94%**	**88%**	**88%**



The above correlation Figure represents the classification as per the Model Classifier.
However, I realized that I could have come up with a few additional approaches to improve my analysis. I observed there is limited evaluation metrics, feature engineering, contextual analysis. I have used better models but, there's limited discussion on the evaluation metrics used to assess the performance of those models. There may be a need for a more thorough evaluation technique that incorporates metrics like precision, recall, F1-score, ROC AUC, and others that are specific to the challenge of detecting cyberbullying.
The feature engineering approach is not extensively described in the code, but it uses Genetic Algorithms for feature selection. There may be a gap in the effectiveness of feature engineering techniques, particularly how features are retrieved from text data. Future study may examine techniques for incorporating context and sentiment from photos, videos, or user interactions.

6.2 Discussion

In Section 2, (Aldinata, Science Direct, 2023) proves with 0.7233 accuracy, 0.7006 recall, and 0.7087 F-1 scores, the, demonstrates that Logistic Regression has the greatest overall score. Text pre-processing was done ahead of time. Unfortunately, it does not much enhance the algorithms, and the accuracy of Logistic Regression has reduced somewhat. It suggests that basic Logistic

Regression (without text-preprocessing) is appropriate for sentiment analysis with our dataset. This might be because stop words and the original form of words (un-lemmatized words) are really required for sentiment analysis.

Here the author (Kamruzzaman, Imran, Hossain, and Bakcy) decides to compare the deep learning models with the classic ensemble models to come up with a better accuracy score. He also incorporates data analysis like tokenization, stop-word removal, lemmatization and stemming.

The test model accuracy for the SVM model is 88%, and the test model accuracy for the LSTM model is 94%, according to the total findings from all the machine-learning and deep-learning models used in this research. This indicates that the LSTM model has surpassed the machine-learning models in terms of accuracy.

7 Conclusion and Future Work

In this paper, I proposed a thorough approach for sentiment analysis and cyberbullying detection that included deep learning and conventional machine learning methods. Our method comprised gathering data, preprocessing it, extracting features, and training a model with the goal of improving sentiment classification model precision and interpretability.

I used a variety of models, including Support Vector Machine (SVM), Logistic Regression, and a deep learning architecture based on Bidirectional LSTM. Using criteria like accuracy, precision, recall, F1-score, and AUC-ROC, the models were thoroughly assessed. Our test results demonstrated how well the LSTM-based model can capture contextual data for sentiment analysis.

Although our approach has shown encouraging results, there are a number of opportunities for more study and improvement:

- Ensemble Methods: Investigating ensemble techniques, such as integrating the predictions of many models, may enhance the performance and resilience of the models.
- Hyperparameter fine-tuning: Particularly for deep learning architectures, detailed hyperparameter adjustment for each model may produce even better results.
- Data Augmentation: Using data augmentation approaches, class imbalance may be addressed, and model generalization is enhanced.
- Cross-Domain Analysis: Analyzing data from different domains and languages might shed light on how adaptable and generalizable the model is.
- Deep learning approaches that can be interpreted can increase the transparency and reliability of the judgements made by these models.
- Real-Time Detection: A pertinent and difficult direction is to adapt the algorithms for real-time sentiment analysis and cyberbullying detection on social media platforms.

In conclusion, this work has advanced approaches for detecting cyberbullying and sentiment analysis. The concepts and methods discussed provide a strong framework for additional study and application in the dynamic field of social media analysis and online content management.

8 Acknowledgment

I would like to express my deep gratitude to **Prof. Arjun Chikkankod** for providing me with the information and freedom to work on my thesis under his direction during the past several weeks.

References

A comparative analysis of sentiment classification based on deep and traditional ensemble machine learning models (2021). Available at: https://ieeexplore.ieee.org/document/9642583.

Iqbal, F. et al. (2019) “A hybrid framework for sentiment analysis using genetic algorithm based feature reduction,” IEEE Access, 7, pp. 14637–14652. Available at: https://doi.org/10.1109/access.2019.2892852.

Qiu, J., Moh, M. and Moh, T.-S. (2022) “Multi-modal detection of cyberbullying on Twitter,” ACM [Preprint]. Available at: https://doi.org/10.1145/3476883.3520222.

https://ieeexplore.ieee.org/document/9851666 (Rahul Katarya, 2018), A Comparative Study of Genetic algorithm in Sentiment Analysis.

Detecting Sentiment Polarities with Comparative Analysis of Machine Learning and Deep Learning Algorithms (2023). Available at: https://ieeexplore.ieee.org/document/10141741.

Bordin, A. et al. (2021) “Sentiment analysis with genetic programming,” Information Sciences, 562, pp. 116–135. Available at: https://doi.org/10.1016/j.ins.2021.01.025.

Aldinata et al. (2023) “Sentiments comparison on Twitter about LGBT,” Procedia Computer Science, 216, pp. 765–773. Available at: https://doi.org/10.1016/j.procs.2022.12.194.

Founta, A.-M. (2018) Large scale crowdsourcing and characterization of Twitter abusive behavior. Available at: https://arxiv.org/abs/1802.00393.

Raj, C. et al. (2021) “Cyberbullying Detection: hybrid models based on machine learning and natural language processing techniques,” Electronics, 10(22), p. 2810. Available at: https://doi.org/10.3390/electronics10222810.

Detection of cyberbullying in social networks using machine learning methods (2018). Available at: https://ieeexplore.ieee.org/document/8625321.

Alba, E., Luque, G. and Araujo, L. (2006) “Natural language tagging with genetic algorithms,” Information Processing Letters, 100(5), pp. 173–182. Available at: https://doi.org/10.1016/j.ipl.2006.07.002.

Swapnarekha, H. et al. (2022) "An optimistic firefly algorithm-based deep learning approach for sentiment analysis of COVID-19 tweets," Mathematical Biosciences and Engineering, 20(2), pp. 2382–2407. Available at: https://doi.org/10.3934/mbe.2023112.

NLP and Machine Learning Techniques for Detecting Insulting Comments on Social Networking Platforms (2018). Available at: https://ieeexplore.ieee.org/document/8441728.

Maity, K. et al. (2022) "A Multitask Framework for Sentiment, Emotion and Sarcasm aware Cyberbullying Detection from Multi-modal Code-Mixed Memes," Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval [Preprint]. Available at: https://doi.org/10.1145/3477495.3531925.

(IEEE Xplore Full-Text PDF: (no date c). Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8843854.

Detecting cyberbullying across multiple social media platforms using deep learning (2021). Available at: https://ieeexplore.ieee.org/document/9404736.

Cyberbullying detection on social networks using machine learning approaches (2020). Available at: https://ieeexplore.ieee.org/document/9411601.

Bordoloi, M. and Biswas, S.K. (2023) "Sentiment analysis: A survey on design framework, applications and future scopes," Artificial Intelligence Review [Preprint]. Available at: https://doi.org/10.1007/s10462-023-10442-2.

Hutto, C.J. and Gilbert, E. (2014) "VADER: A Parsimonious Rule-Based (C.J. Hutto, 2014) Model for Sentiment Analysis of Social Media Text," Proceedings of the International AAAI Conference on Web and Social Media, 8(1), pp. 216–225. Available at: https://doi.org/10.1609/icwsm.v8i1.14550.