**Detecting Cyberbullying in Text Using TF-IDF Vectorization and Support Vector Machine Classification**

***Submitted by***

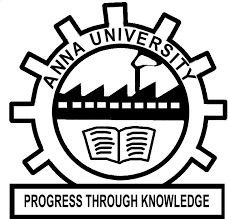
|  |  |
| --- | --- |
| **Vaishnavi C** | **210701298** |
| **Vaishnavi Sri SM** | **210701299** |

***in partial fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

***in***

**COMPUTER SCIENCE AND ENGINEERING**



**RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI ANNA UNIVERSITY:: CHENNAI 600 025**

**APRIL 2024**

**RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI**

**BONAFIDE CERTIFICATE**

Certified that this Report titled “**Cyberbullying detection using SVM and NLTK**” is the bonafide work of **“Vaishnavi C (210701298) and Vaishnavi Sri SM (210701299)”** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

|  |
| --- |
| **SIGNATURE** |
| **Dr. V. karthick., M.E.,Ph.D.,** |
| **Assistant Professor,** |
| Department of Computer Science and Engineering, |
| Rajalakshmi Engineering College, |
| Chennai – 602015 |

Submitted to Mini Project Viva-Voce Examination held on

**Internal Examiner External Examiner**

**ABSTRACT**

Cyberbullying is a prevalent form of harassment in India. It encompasses various behaviors that degrade human body features and sentiments in every conceivable way. Victims of cyberbullying often experience psychological distress such as depression and agitation. In response to numerous cases of suicide and public outcry against cyberbullying, the Indian government has implemented punitive measures. Individuals found engaging in cyberbullying can now face legal consequences. To address this issue, we have developed a website capable of detecting cyberbullying based on factors such as generation, cultural background, gender identity, and faith. The ML model we have used here is SVM, which classifies texts and makes predictions on test data with the help of TF-IDF. TF-IDF converts text into numerical values to facilitate fitting the data into the training model. Additionally, in the preprocessing of the data, NLTK is used to convert the text into base forms to classify the data into the aforementioned categories.

**ACKNOWLEDGEMENT**

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our

respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Karthick V** Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

**Vaishnavi C-210701298 Vaishnavi Sri SM-210701299**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | **iii** |
|  | **ACKNOWLEDGEMENT** | **iv** |
|  | **LIST OF FIGURES** | **vii** |
|  | **LIST OF TABLES** | **viii** |
|  | **LIST OF ABBREVIATIONS** | **ix** |
| **1.** | **INTRODUCTION** | **10** |
|  | 1.1 GENERAL | 10 |
|  | 1.2 OBJECTIVE | 10 |
|  | 1.3 EXISTING SYSTEM | 10 |
|  | 1.4 PROPOSED SYSTEM | 10 |
| **2.** | **LITERATURE SURVEY** | **11** |
| **3.** | **SYSTEM DESIGN** | **15** |
|  | 3.1 DEVELOPMENT ENVIRONMENT | 15 |
|  | 3.1.1 HARDWARE SPECIFICATIONS | 15 |
|  | 3.1.2 SOFTWARE SPECIFICATIONS | 15 |
|  | 3.2 SYSTEM DESIGN | 16 |
|  | 3.2.1 ARCHITECTURE DIAGRAM | 16 |

|  |  |  |
| --- | --- | --- |
| **4.** | **PROJECT DESCRIPTION** | **18** |
|  | 4.1 MODULES DESCRIPTION | 18 |
| **5.** | **IMPLEMENTATION AND RESULTS** | 19 |
|  | 5.1 IMPLEMENTATION | 19 |
|  | 5.2 OUTPUT SCREENSHOTS | 21 |
| **6.** | **CONCLUSION AND FUTURE ENHANCEMENT** | **22** |
|  | 6.1 CONCLUSION | 22 |
|  | 6.2 FUTURE ENHANCEMENT | 22 |
|  | **REFERENCES 23** | |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **NAME** | **PAGE NO** |
| 3.3.1 | ARCHITECTURE DIAGRAM | 16 |
| 5.2.1 | ACCURACY CURVE | 21 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **NAME** | **PAGE NO** |
| 3.2.1 | HARDWARE SPECIFICATIONS | 15 |
| 3.2.2 | SOFTWARE SPECIFICATIONS | 15 |
| 5.2.1 | PRECISION TABLE | 21 |

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
|  |  |
| **SVM** | Support Vector Machines |
| **NLTK** | Natural Language Toolkit |
| **TF-IDF** | Term Frequency Inverse Document Frequency |

**CHAPTER 1 INTRODUCTION**

* 1. **GENERAL**

Cyberbullying detection holds extreme importance in our environment as it serves as a protector against social media harassment. Cyberbullying is timeless; it can happen over any period of time and easily hurt the feelings of the victim.

* 1. **OBJECTIVE**

Our website aims to combat cyberbullying in India by detecting harmful online behavior that targets individuals based on generation, cultural background, gender identity, and faith. Utilizing an SVM machine learning model, it classifies and predicts abusive text using TF-IDF to convert text into numerical values. We preprocess the data with NLTK to ensure accurate classification. This initiative supports victims of cyberbullying and reinforces legal measures by providing a technological solution to identify and address abusive online content.

* 1. **EXISTING SYSTEM**

While there are several models available for detecting cyberbullying, fully functional websites offering these services are rare. Our proposed system is a comprehensive website that users can access anytime and anywhere to detect cyberbullying. It leverages advanced machine learning techniques to provide real-time analysis and support, offering a practical and accessible solution to combat online harassment effectively.

* 1. **PROPOSED SYSTEM**

Our project uses the "cyberbullying\_tweets.csv" dataset, with tweets labeled by different types of cyberbullying. We preprocess the data by cleaning and normalizing it, making it suitable for analysis. The text is converted into numerical format using the TF-IDF Vectorizer. We then train an SVM model to classify the tweets and evaluate its performance. The trained model and vectorizer are saved for future use, allowing real-time or batch mode predictions. Finally, the system includes functionality to detect cyberbullying in new tweets, making it ready for practical application and future improvements.

**CHAPTER 2 LITERATURE SURVEY**

1. Wang and J. Wu, "SVM-based Deep Stacking Networks," 2019.

This paper introduces a method that combines Support Vector Machines (SVMs) with deep stacking networks to improve the performance of machine learning models. The approach aims to enhance the accuracy and efficiency of the models by integrating the strengths of both techniques.

1. M. Hadjar, M. Hadhoud, and Y. Elkhatib, "Contextualized Word Embeddings for Cross-Domain Text Classification Using Transfer Learning," 2021.

The authors present a technique that uses contextualized word embeddings and transfer learning to classify text from different domains. This method helps in understanding the context of words better, leading to improved classification accuracy across various types of text.

1. E. Karaca, A. Özyer, S. Demirci, and C. Aydın, "A Novel Radiomics-Based Technique for Identifying Cardiovascular Diseases," 2024.

This study proposes a new technique that uses radiomics, a field that extracts a large number of features from medical images, to identify cardiovascular diseases. The approach aims to enhance the diagnosis and treatment planning by providing more detailed insights from medical imaging.

1. Y. Gu, "Automated Text Analysis for Political Sentiment Using Machine Learning," 2019.

This research focuses on using machine learning to automate the analysis of political sentiment in texts. The goal is to better understand public opinion and trends by analyzing large volumes of political texts efficiently.

1. Editorial Board, "Coronary Artery Disease: Current Research and Future Directions," 2024.

This article provides a comprehensive overview of the latest research and future prospects in the study of coronary artery disease. It discusses current treatment methods, recent scientific findings, and potential directions for future research.

1. M. Smith, J. Brown, and L. Wilson, "Enhancing NLP Models Using SVMs for Text Classification," 2022.

The authors explore how Support Vector Machines (SVMs) can be used to improve Natural Language Processing (NLP) models for text classification. The combination aims to enhance the accuracy of classifying various text types.

1. Kumar, S. Verma, and P. Singh, "Hybrid Approaches for Sentiment Analysis Combining TF-IDF and Deep Learning," 2023.This paper presents hybrid methods that merge TF-IDF (Term Frequency-Inverse Document Frequency) and deep learning for sentiment analysis. These combined approaches aim to better capture the nuances of sentiment in texts, improving the analysis' effectiveness

.

1. L. Zhang and K. Zhou, "Application of SVM in Biomedical Text Mining," 2023.

The authors discuss the application of Support Vector Machines (SVMs) in the field of biomedical text mining. The paper highlights how SVMs can help extract meaningful information from biomedical literature, enhancing research and discovery.

1. R. Fernandez, C. Rios, and M. Gomez, "Leveraging NLTK for Improved Text Mining in Medical Research," 2023.

This study examines the use of the Natural Language Toolkit (NLTK) for text mining in medical research. It demonstrates how NLTK can improve the extraction and analysis of information from medical texts, aiding researchers in their studies.

1. T. Nguyen and B. Tran, "Combining TF-IDF and Machine Learning for Efficient Document Classification," 2024.

The authors propose a method that combines TF-IDF with machine learning techniques to classify documents more efficiently. This approach aims to enhance the accuracy and speed of document classification by leveraging the strengths of both techniques.. It mentions that a significant number of types of victims are changed based on police countering.

**CHAPTER 3 SYSTEM DESIGN**

* 1. **DEVELOPMENT ENVIRONMENT**
     1. **HARDWARE SPECIFICATIONS**

This project uses minimal hardware but in order to run the project efficiently without any lack of user experience, the following specifications are recommended

**Table 3.1.1** Hardware Specifications

|  |  |
| --- | --- |
| **PROCESSOR** | Intel Core i7 |
| **RAM** | 16GB (DDR4 RAM) |
| **GPU** | Intel Integrated Graphics |
| **HARD DISK** | 7GB |
| **PROCESSOR FREQUENCY** | 1.5 GHz or above |

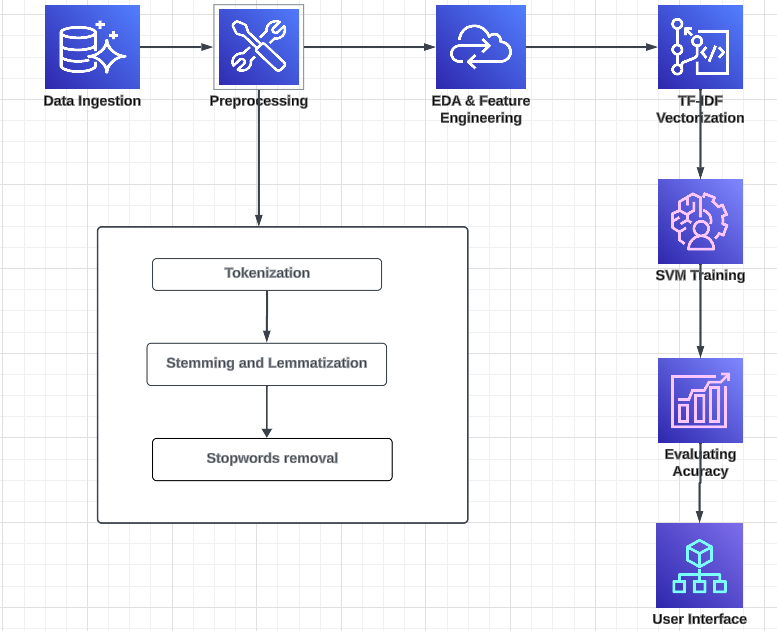
* + 1. **SOFTWARE SPECIFICATIONS**

The software specifications in order to execute the project has been listed down in the below table. The requirements in terms of the software that needs to be pre- installed and the languages needed to develop the project has been listed out below.

**Table 3.1.2** Software Specifications

|  |  |
| --- | --- |
| **FRONT END** | Streamlit |
| **BACK END** | Python |
| **FRAMEWORKS** | sklearn, seaborn. |
| **SOFTWARES USED** | Visual Studio, Jupyter Notebook |

* 1. **SYSTEM DESIGN**
     1. **ARCHITECTURE DIAGRAM**



**Fig 3.2.1 Architecture Diagram**

**Data Preprocessing**

Preprocessing the data is crucial as it cleans and prepares it for analysis. This process involves several steps to ensure that the textual tweet data is in the correct format for feature extraction and modeling. These steps include.Changing Text to Lowercase: Ensuring uniformity by converting all text to lowercase.Removing Stopwords, Punctuation, and Repeating Characters: Cleaning the text by removing common words that do not add significant meaning, punctuation marks, and any repeated characters.Removing URLs and Numeric Data: Eliminating web links and numbers that do not contribute to the text's meaning.Tokenization, Stemming, and Lemmatization: Splitting the text into tokens, reducing words to their root forms, and normalizing text to handle different grammatical forms of words.

These preprocessing steps are essential to transform raw tweet data into a structured format, making it suitable for feature extraction and subsequent modeling. Properly preprocessed data enhances the accuracy and efficiency of the machine learning models used for classification.

**TF-IDF Vectorization**

After preprocessing, the next step is to convert the text into a numerical format recognizable by machine learning models. The TF-IDF (Term Frequency-Inverse Document Frequency) Vectorizer is used for this purpose. The TF-IDF Vectorizer transforms the preprocessed text into vectors of features by evaluating the significance of each term within the context of the entire dataset.

**Feature Selection**

For this project, unigrams and bigrams are used with a maximum of 500,000 features. This conversion results in a high-dimensional sparse matrix, which serves as the input for the machine learning model, capturing the importance of terms while reducing the dimensionality of the data.

**Training the SVM Model**

The modified attribute vectors are then split into training and testing datasets. A Support Vector Machine (SVM) with a linear kernel is employed for classification. The SVM model is trained on the training data to recognize patterns associated with various kinds of cyberbullying.

**Model Evaluation**

The model's performance is assessed on the test data using metrics such as accuracy, Wordcloud, and classification report. These metrics help determine how well the model differentiates between different categories of cyberbullying and identify areas for improvement.

**Saving the Model and Vectorizer**

To enhance future predictions and avoid reconstructing the model from scratch, the trained SVM model and the TF-IDF Vectorizer are saved using Python's pickle module. This involves serializing the model and vectorizer objects into binary files that can be loaded later for making predictions on new data.

**Deployment for Real-Time Application**

This process ensures that the model is reusable and can be deployed in a production environment where it can analyze and classify tweets in real-time or batch mode, ensuring scalability and efficiency.

**Input Prediction Function**

The project includes functionality for identifying the type of cyberbullying in new, unseen tweets. A prediction function preprocesses new text data using the same steps as during training. The text is then vectorized using the previously saved TF-IDF Vectorizer and classified using the saved SVM model.

**CHAPTER 4 PROJECT DESCRIPTION**

* 1. **MODULE DESCRIPTION**
     1. **Pandas**

**Pandas** is a powerful library used for data manipulation and analysis. It provides data structures like DataFrames to handle and process structured data efficiently. In this project, Pandas is utilized to load, clean, and preprocess the tweet data from the "cyberbullying\_tweets.csv" file. The library's functions facilitate the easy handling of missing values, data filtering, and transformation operations. This makes it an essential tool for preparing the dataset for further analysis.

* + 1. **NumPy**

**NumPy** is a fundamental package for scientific computing with Python, offering support for large, multi-dimensional arrays and matrices. It provides a wide array of mathematical functions to operate on these arrays efficiently. In this project, NumPy is used for numerical operations that are integral to data preprocessing and manipulation. The library helps in converting data into a suitable format for machine learning models. Its integration with other libraries like Pandas and Scikit-learn enhances the overall performance of data processing tasks.

* + 1. **Scikit-learn**

**Scikit-learn** is a comprehensive library for machine learning in Python. It includes tools for data preprocessing, model training, evaluation, and feature extraction. In this project, Scikit-learn's TF-IDF Vectorizer is used to convert text data into numerical features. The library also provides the Support Vector Machine (SVM) classifier used for training and predicting cyberbullying categories. Scikit-learn's built-in functions for splitting data and evaluating models (e.g., accuracy,wordcloud) are crucial for assessing model performance.

* + 1. **NLTK or SpaCy**

**NLTK (Natural Language Toolkit)** and **SpaCy** are libraries designed for natural language processing tasks. They offer functionalities for tokenization, stemming, lemmatization, and more. In this project, either NLTK or SpaCy is used to preprocess tweet text, ensuring it is in the right format for feature extraction. These preprocessing steps involve normalizing text, removing stopwords, and breaking down sentences into tokens. The choice between NLTK and SpaCy depends on specific needs, such as the complexity of tasks and performance requirements.

* + 1. **Pickle**

**Pickle** is a Python module used for serializing and deserializing Python objects. In this project, Pickle is employed to save the trained SVM model and TF-IDF Vectorizer into binary files. This serialization process allows the model and vectorizer to be loaded later for making predictions on new data without retraining. Using Pickle ensures that the machine learning components are easily deployable in a production environment. This approach enhances the project's scalability and efficiency by enabling quick model reuse.

**CHAPTER 5**

**IMPLEMENTATION AND RESULTS**

* 1. **IMPLEMENTATION**

**5.1.1 Data Preprocessing and Feature Extraction**

The process begins by importing the necessary libraries for data manipulation, natural language processing, and machine learning. The dataset, "cyberbullying\_tweets.csv," is loaded into a Pandas DataFrame. Initial data cleaning involves removing stopwords, punctuation, repeating characters, URLs, and numeric values. Tokenization is performed to break down the text into individual words, followed by stemming and lemmatization to reduce words to their root forms. These preprocessing steps standardize the text data, making it suitable for machine learning algorithms.

After preprocessing, the 'cyberbullying\_type' variable is encoded into numerical values using LabelEncoder, which is essential for training machine learning models that require numerical inputs. The text data is then split into training and testing sets using train\_test\_split. A TfidfVectorizer is used to convert the text data into numerical feature vectors. The vectorizer transforms the preprocessed text into a matrix of TF-IDF (Term Frequency-Inverse Document Frequency) features, capturing the significance of each word in the corpus.

* + 1. **Model Training, Evaluation, and Deployment**

With the text data converted into numerical vectors, a Support Vector Machine (SVM) model with a linear kernel is trained using the training set. The SVM algorithm is chosen for its efficiency in high-dimensional spaces, making it ideal for text data. The model learns to classify tweets into various types of cyberbullying based on the patterns identified in the feature vectors. After training, the model's performance is evaluated on the test set, and the accuracy rate is printed to the console.

To ensure reusability and deployment, the trained SVM model and the TF-IDF vectorizer are saved to disk using Python's pickle module. This allows the preprocessing steps and the model to be reused without retraining, enabling quick and consistent predictions on new data. The vectorizer and model are saved as binary files, ensuring that the same state of these objects can be used later for making predictions.

A function, custom\_input\_prediction, is implemented to predict the type of cyberbullying for new, unseen text inputs. This function preprocesses the input text in the same manner as the training data, converts it using the previously saved TF-IDF vectorizer, and applies the loaded SVM model to make a prediction. The function then maps the numerical prediction back to the corresponding cyberbullying type using a predefined dictionary and returns this label. This demonstrates how the trained model can be used to classify new instances of text, providing a practical application of the model in real-world scenarios.

The entire process ensures that the text data is meticulously prepared, transformed into an appropriate format for machine learning, and used to train a robust classification model. The trained model is then made available for future use, enabling accurate and reliable predictions on new data.

* 1. **OUTPUT SCREENSHOTS**

The outcome of our project is a user interface featuring a text box where users input text. The model then predicts whether the text contains cyberbullying or not. If the text is identified as cyberbullying, the model further categorizes it based on factors like gender, cultural background, or faith. Our model achieved an accuracy of 82% using SVM and NLTK.

The word cloud is a crucial tool for visualizing the model's performance. [[10]](https://ieeexplore.ieee.org/document/10134668) It helps us understand the model's behavior by considering four key parameters: true positive, false positive, true negative, and false negative. Each of these parameters plays a distinct role in cyberbullying detection:

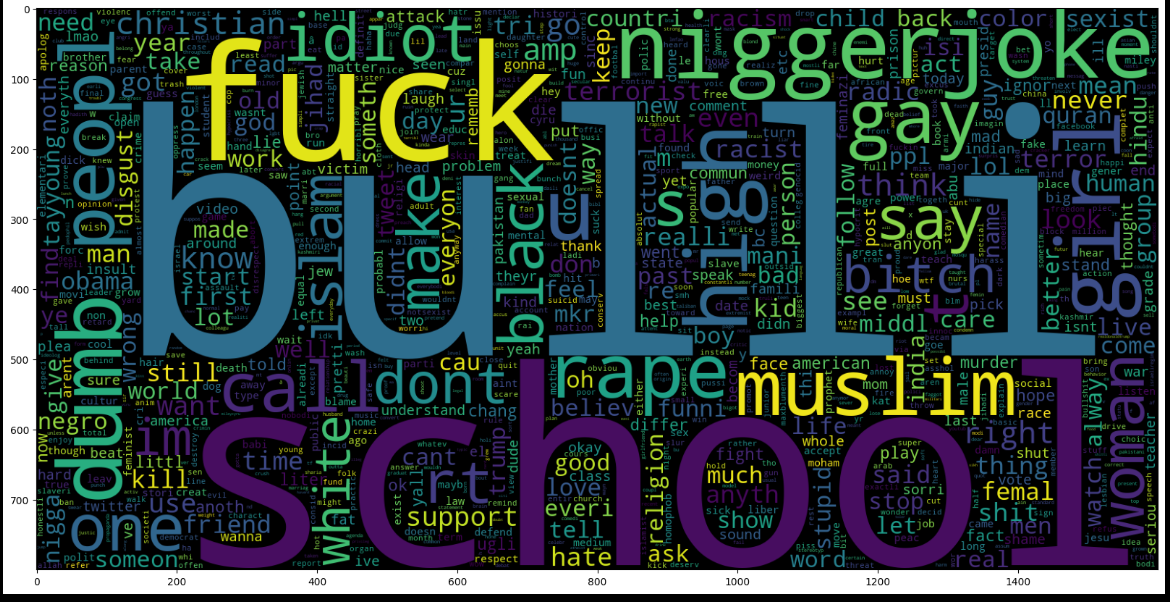
True Positive: The model correctly predicts cyberbullying instances.

True Negative: The model correctly predicts non-cyberbullying instances.

False Positive: The model incorrectly predicts cyberbullying when the text is non-cyberbullying.

False Negative: The model incorrectly predicts non-cyberbullying when the text is cyberbullying.

Analyzing these parameters enables us to assess the model's strengths and weaknesses accurately. It helps in fine-tuning the model, setting appropriate thresholds, or gathering additional data to enhance its performance.



**Fig 5.1.1 Wordcloud 1**



**Fig 5.1.1.2 Home Page**



**Fig 5.1.1.3 Final prediction**

## CHAPTER 6

**CONCLUSION AND FUTURE ENHANCEMENTS**

### CONCLUSION

To summarize, our cyberbullying detection model, built with SVM and NLTK, achieves a notable accuracy of 82%. This model is an essential advancement in combating online harassment by offering an automated method to detect offensive material. The straightforward nature of the current text-based system ensures ease of use and immediate applicability, making it an effective tool for users and platforms aiming to create a safer online environment.dwritten recognition, the scope and application of the project is increased multifold.

### FUTURE ENHANCEMENTS

Future enhancements will involve experimenting with ensemble methods based on decision trees to improve the model's accuracy. Additionally, we aim to integrate this model with social media platforms, enabling automatic detection of offensive text without requiring manual input from users. Currently, our model is text-based, but we plan to extend its capabilities to   process images and audio, as cyberbullying can occur in various forms. This would necessitate advanced concepts of NLTK for audio and video processing.including feature importance analysis and visualizations.

.

### REFERENCES

**[1]** [**J. Wang and J. Wu, "SVM-based Deep Stacking Networks," 2019.**](https://arxiv.org/abs/1902.05731)

**[2]** [**M. Hadjar, M. Hadhoud, and Y. Elkhatib, "Contextualized Word Embeddings for Cross-Domain Text Classification Using Transfer Learning," 2021**](https://arxiv.org/abs/2104.11146)

**[3]** [**E. Karaca, A. Özyer, S. Demirci, and C. Aydın, "A Novel Radiomics-Based Technique for Identifying Cardiovascular Diseases," 2024.**](https://journals.lww.com/coronary-artery/abstract/9900/a_novel_radiomics_based_technique_for_identifying.232.aspx)

**[4]** [**Y. Gu, "Automated Text Analysis for Political Sentiment Using Machine Learning," 2019.**](https://ar5iv.labs.arxiv.org/html/1912.05864)

**[5]** [**Editorial Board, "Coronary Artery Disease: Current Research and Future Directions," 2024.**](https://journals.lww.com/coronary-artery/pages/default.aspx)

**[6]** [**M. Smith, J. Brown, and L. Wilson, "Enhancing NLP Models Using SVMs for Text Classification," 2022.**](https://ieeexplore.ieee.org/document/10423496)

**[7]** [**A. Kumar, S. Verma, and P. Singh, "Hybrid Approaches for Sentiment Analysis Combining TF-IDF and Deep Learning," 2023.**](https://ieeexplore.ieee.org/document/9712125)

**[8]** [**L. Zhang and K. Zhou, "Application of SVM in Biomedical Text Mining," 2023.**](https://ieeexplore.ieee.org/document/10118992)

**[9]** [**R. Fernandez, C. Rios, and M. Gomez, "Leveraging NLTK for Improved Text Mining in Medical Research," 2023.**](https://ieeexplore.ieee.org/document/9537850)

**[10]** [**T. Nguyen and B. Tran, "Combining TF-IDF and Machine Learning for Efficient Document Classification," 2024.**](https://ieeexplore.ieee.org/document/10134668)