JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering

Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

Implementation of Natural Language Processing for Bullying Complaints with Voice to Text Conversion

Miftahul Ilmi[1], Dasril Aldo[2], Sapta Eka Putra[3], Adanti Wido Paramadini[4], Yohani Setiya Rafika Nur[5]

[1]Institut Teknologi dan Bisnis Indobaru Nasional, Batam, 29444, Indonesia
[2,4,5]Institut Teknologi Telkom Purwokerto, Jl. Panjaitan, Purwokerto, 53147, Indonesia
[3]Universitas Tamansiswa, Jl. Tamansiswa, Padang, 25171, Indonesia

ARTICLE INFO

Article historys:

Received : 26/07/2024
Revised : 18/08/2024
Accepted : 12/09/2024

Keywords:

Bullying; Natural Language Processing (NLP); Support Vector Machine (SVM); Mobile Application; Anonymous Reporting

ABSTRACT

Bullying in high school is frequent and negatively affects students psychological well-being. The lack of effective reporting mechanisms makes students hesitant to report bullying cases for fear of their identity being exposed, which can lead to stigma or retaliation. The lack of data on bullying incidents also hampers prevention and intervention measures. The study designed and implemented a mobile application that uses natural language processing (NLP) for speech-to-text conversion, enabling anonymous and convenient reporting of bullying cases. The app ensures the anonymity of whistleblowers and facilitates the collection of accurate data on bullying incidents, helping schools respond with appropriate preventive measures. Software engineering methodologies are used with a focus on requirements analysis, system design, implementation, and application testing. NLP technology is used to interpret verbal instructions into text, with the Support Vector Machine (SVM) method for text classification, ensuring high accuracy in detecting bullying incidents. The trial application in several high schools showed the relevance and effectiveness of the application. Ethical and security considerations are top priorities, with an emphasis on whistleblower identity protection and data security. The test results showed that the application achieved 92% accuracy, 90% precision, and 88% recall, demonstrating its effectiveness in collecting bullying reports anonymously and accurately.

Corresponding Author:

Dasril Aldo
Institut Teknologi Telkom Purwokerto, Jl. Panjaitan, Purwokerto, 53147, Indonesia
Email: dasril@ittelkom-pwt.ac.id

1. INTRODUCTION

One form of adolescent violence that often appears is bullying behavior [1]. Many in high school, bullying has become a persistent problem that negatively impacts students' psychological well-being. Deviant treatment that can have a bad effect on others. The high incidence of bullying in schools shows that this educational institution has become a common location for this behavior. Of the number of students recorded, 18.5% had been victims of bullying, 29.4% had faced acts of physical violence, and 3.1% had experienced sexual harassment. This phenomenon interferes with the learning process and can lead to long-term consequences for the victim. Although its existence has long been recognized, efforts to address bullying are often hampered by the lack of effective and reliable reporting mechanisms. Concerns about identity disclosure have discouraged many students from reporting bullying cases, which in turn perpetuates a cycle of violence and insecurity [2, 3].

Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

Advances in information system-based technology [4] and integrated with AI, can be combined in the field of Natural Language Processing (NLP) [5] and a voice-to-text conversion system [6, 7], the ability of NLP to convert speech to text and text to speech [8] offers the opportunity to develop innovative solutions that can facilitate anonymous reporting of bullying. Mobile applications that utilize this technology can be an important tool to collect accurate data about bullying incidents, schools can respond with appropriate prevention and intervention measures. By ensuring the anonymity of the reporter, the app can reduce students' fear of the negative impact of reporting a bullying case, such as stigma or retaliation.

This research is important to overcome bullying that has an impact on student welfare. The app, developed using NLP and speech-to-text conversion, aims to facilitate anonymous reporting of bullying, increase the number of reports and enable more appropriate school actions. The study will also examine the effectiveness of technology in changing reporting mechanisms, contributing to a safer learning environment.

To address the problem of bullying in high school, this study will adopt several strategies. The approach to solving problem formulation involves the integration of advanced technology by applying Natural Language Processing (NLP) to support verbal reports into text. It provides an intuitive method and is similar to speaking directly, but with maintained privacy. Data analytics will be used for proactive response, where data from the app will be analyzed for bullying patterns, assisting schools in planning prevention and data-driven interventions. In addition, the development of private and secure applications is a priority, with a focus on anonymity and data security, as well as reducing psychological barriers through user-friendly design.

A problem-solving strategy includes several steps. First, the development of an application with advanced NLP algorithms will be carried out, simplifying the reporting process through speech-to-text conversion. Second, data optimization for schools will be carried out by utilizing data to facilitate prevention and intervention measures, as well as improving the accuracy and availability of data through efficient reporting features. Finally, the design of the app will focus on users with the implementation of features that support the anonymity of the complainant, as well as training and education for the safe use of the app. This strategy is expected to facilitate bullying reporting, increase the number of reports received, and allow for more appropriate preventive measures and interventions from the school, thereby creating a safer school environment.

This study will improve previous research by focusing on bullying complaint applications using NLP technology and speech-to-text conversion. Previous research, such as voice sentiment analysis [9], a voice-to-text app for police [10], increasing the resilience of NLP systems to ASR errors [11], the development of virtual bots with human-like conversations [12], and a voice-to-text conversion system in Ukrainian [13], despite making a significant contribution, did not specifically handle bullying complaints. The study offers a more precise solution by integrating the latest technology for accurate and efficient speech-to-text conversion, specifically designed to help individuals report bullying cases in real-time, thereby improving the response and handling of bullying cases.

2. RESEARCH METHOD

In facing the challenges of bullying in high school, this research is directed to develop innovative technology-based solutions. This flowchart presents a comprehensive view of the planned research methodology, from pre-research to final output. The focus is on creating mobile applications that leverage advances in Natural Language Processing (NLP) to simplify and improve the bullying reporting process.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

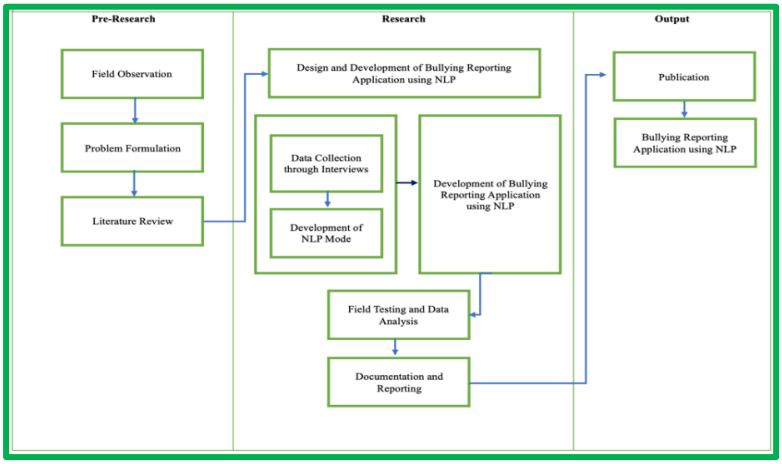


Figure 1. Research methodology

Based on the flow chart shown in Figure 1, the research methodology can be explained by the following steps:

2.1 Pre-Research

At this stage the researcher conducts a series of initial activities before application development.

1. Field Observation: The researcher will conduct observations in the high school environment to understand the context and dynamics of bullying that occurs. This helps in identifying the specifications needed for the application.
2. Problem Formulation: Based on field observations, researchers formulate specific problems that the application will address, such as the lack of efficient and safe reporting mechanisms for students.
3. Literature Review: Before starting development, researchers will conduct a literature study to understand existing research on bullying, reporting applications, and NLP technology. This study will provide a theoretical and practical foundation for application development.

2.2 The core phase of the research includes the development of NLP-based applications.

1. Bullying and NLP Reporting Application Design and Design: Designing the application architecture and interface. The researcher will determine the main features based on the results of observations and literature reviews.
2. Interview Data Collection: Conduct interviews with stakeholders (students, teachers, and school administration) to collect data to be used in the creation of NLP models and to ensure the application meets the needs of users.
3. NLP Model Creation: Develop an NLP model that will be used to analyze and interpret spoken bullying reports. This model will be tested and adjusted for accuracy in recognizing bullying terminology and context. Shown food Figure 2.

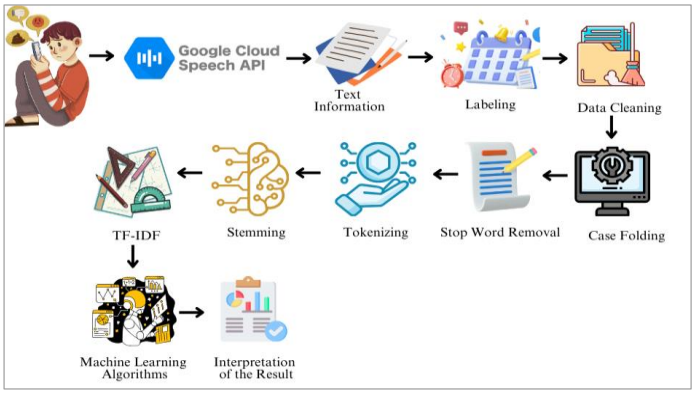


Figure 2. NLP Process

Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

From Figure 2. The following is an explanation of the stages of the NLP process in this study based on figure 2 given:

a. Voice Input from Users: Users, in this case victims or witnesses of bullying, use mobile devices to record their complaints in the form of voice.
b. Voice to Text Conversion: Voices recorded by users are sent to the Google Cloud Speech API to be converted into text. The Google Cloud Speech API uses speech recognition technology to convert speech to text with high accuracy.
c. Text Information Collection: is the stage where the text of the voice conversion is collected and stored in a structured format for further analysis [14]. At this stage, the text is given additional labels such as bullying categories, locations, and other relevant information to organize the data properly. This stage ensures that the data is ready for the cleaning process and subsequent analysis.
d. Labeling: Each generated text is labeled accordingly [15], such as the bullying category (0 = not bullying, 1 = verbal, 2 = physical, 3 = social, 4 = emotional, 5 = online).
e. Data Cleanup:
   - Case Folding: Lowercase all letters in text for consistency [16].
   - Stop Word Removal: Removes common words that do not have significant meaning in the analysis (such as "and", "or", "which") [17].
   - Tokenization: Breaking down text into small units called tokens (usually individual words) [18].
   - Stemming: Reducing words to their basic form (e.g., "hit" to "hit") to unify variants of the same word [19].
f. TF-IDF: After preprocessing the text, including case folding, stopword removal, tokenization, and stemming, we implement TF-IDF (Term Frequency-Inverse Document Frequency) to convert the text into numerical features [20].
g. Machine Learning Algorithm Application: The cleaned and processed data is used to train a machine learning algorithm that will help in detecting bullying patterns from the generated text. These algorithms can be either classification or clustering models depending on the needs of the analysis.
h. Interpretation of Results: The results of machine learning algorithms are interpreted to gain insights and information regarding the incidence of bullying. This includes identifying the types of bullying that often occur, locations prone to bullying, and profiles of perpetrators and victims.

4. Field Testing and Data Analysis: Once the app is developed, field testing will be conducted to see how well the app performs in a real environment. Data from these tests will be analyzed to see how effective the app is.
5. Documentation and Reporting: Create complete documentation of the application development process and test results. This will include reporting on the effectiveness of the app, usage in the field, and feedback from users.

2.3 Output

The final results of this study include two main aspects.

1. Publications: Researchers will write and publish research results in accredited journals or conferences to share knowledge and findings with the wider community.
2. NLP-based Bullying Reporting Application: The end product is a fully functional application that allows anonymous and secure reporting of bullying, ready for adoption by schools participating in the study.

This methodology emphasizes a systematic and iterative approach, which ensures that the developed application is tailored to the needs of the user and has a strong foundation in best practices from the existing literature as well as input from the user.

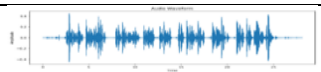
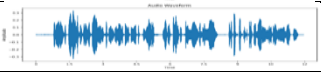
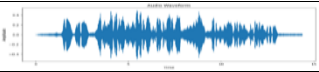
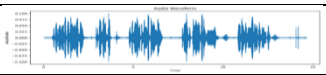
JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

3. RESULTS AND DISCUSSION

3.1 Input Voice Data from User

In this study, the recorded data of bullying complaints received will be presented in Table 1. The data contains reports in Indonesian made by victims or witnesses of bullying. Once the data is recorded, the next step is to translate it into english. The results of the voice data input are shown in Table 1.

Table 1. Voice Input

No	Name	Sound Visualization
1	File14.wav	
2	file13.mp3	
3	file10.mp3	
…	…	…
1000	File1000.mp3	

3.2 Google Cloud Speech API

Voice recordings are converted to text using the Google Cloud Speech API. This technology converts voice into text information that can be further processed. The number of data displayed is only 5 data out of 1000 data obtained, shown in Table 2.

Table 2. Conversion Results

No	Teks
1	They spread lies about me, saying I did things I didn't do. Now everyone believes them and avoids me. It's destroying my social life.
2	Every time there is a music lesson, I am always laughed at by my classmates. They call me names and say I'm useless. It's really affecting my self-esteem. I feel so worthless.
3	Every time I make a mistake, they point it out and laugh. They say I'm a failure and that I'll never succeed. It's crushing my confidence. I feel so hopeless.
4	I used to have a group of friends, but now they ignore me. They act like I'm not there and talk behind my back. It's so painful.
5	I get shoved into lockers almost every day. They say it's just a joke, but it hurts. I have bruises from it. I want it to stop but I'm afraid to say anything.

3.3 Text Information

From the data in Table 2 is the stage where the text of the voice conversion is collected and stored in a structured format for further analysis. At this stage, the text is given additional labels such as bullying categories and other relevant information to organize the data well. This stage ensures that the data is ready for the cleaning process and subsequent analysis.

```
df = pd.read_excel('latif.xlsx', sheet_name='DATA')
df.head()
```

	No	Filename	Bullying text	Category
0	959	file959.mp3	They spread lies about me, saying I did things...	Social Bullying
1	176	file176.mp3	Every time there is a music lesson, I am alway...	Verbal Bullying
2	686	file686.mp3	Every time I make a mistake, they point it out...	Emotional Bullying
3	927	file927.mp3	I used to have a group of friends, but now the...	Social Bullying
4	618	file618.mp3	I get shoved into lockers almost every day. Th...	Physical Bullying

Figure 3. Results of Ketogori Administration

Figure 3 shows how bullying data is loaded using Python from an Excel file named 'latif.xlsx'. The pd.read_excel command is used to read data from a sheet named 'DATA', and df.head() displays the first five rows of the loaded DataFrame. The data consists of the 'No' column which contains the serial

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

number of each entry, 'Filename' which contains the name of the audio file of the bullying complaint recording, 'Bullying text' which contains the text of the transcription from the voice recording, and 'Category' which contains bullying categories such as 'Social Bullying', 'Verbal Bullying', 'Emotional Bullying', and 'Physical Bullying'. By loading and displaying this data, we can understand the structure and content of bullying reports, which is an important first step before conducting further text analysis or applying machine learning techniques.

3.4 Labeling

Labeling bullying data is an important process for classifying the types of bullying detected in the text of the voice conversion. Each bullying text that has been collected is numerically labeled based on its category. This label aids in further analysis, especially in the application of machine learning algorithms to detect bullying patterns.

	No	Filename	Bullying text	Category	Label
0	959	file959.mp3	They spread lies about me, saying I did things...	Social Bullying	3
1	176	file176.mp3	Every time there is a music lesson, I am alway...	Verbal Bullying	1
2	686	file686.mp3	Every time I make a mistake, they point it out...	Emotional Bullying	4
3	927	file927.mp3	I used to have a group of friends, but now the...	Social Bullying	3
4	618	file618.mp3	I get shoved into lockers almost every day. Th...	Physical Bullying	2

Figure 4. Labeling Results

Figure 5 shows bullying data that has been given numerical labels for each bullying category, consisting of 'No', 'Filename', 'Bullying text', 'Category', and 'Label' columns. The folder function of Pandas is used to change the bullying category to numeric labels, such as 'Social Bullying' to 3, 'Verbal Bullying' to 1, and so on. This process is essential for converting qualitative data into quantitative, facilitating further analysis and application of machine learning algorithms. With numerical labeling, bullying data becomes more structured and ready to be used in machine learning models to detect text-based bullying patterns.

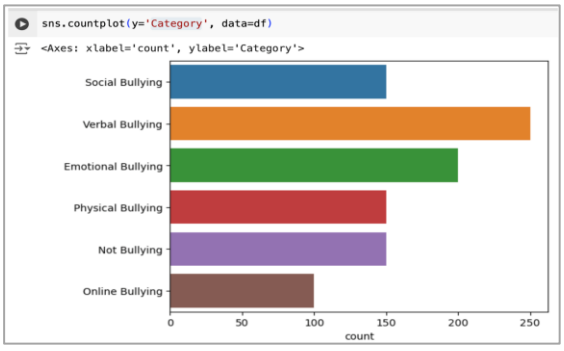


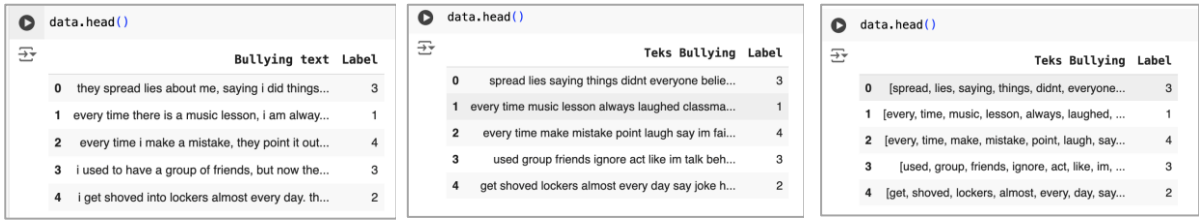
Figure 5. Results of Count Against Category

Figure 5 shows a bar chart created using Seaborn to visualize the distribution of bullying categories in a dataset. The 'sns.countplot' function is used to generate this graph, with the parameter 'y='Category'' indicating that the bullying category is displayed on the y-axis, and the parameter 'data=df' indicating that the data is retrieved from a DataFrame named 'df'. Analysis of this graph shows that 'Verbal Bullying' has the highest number of incidences among other bullying categories, followed by 'Social Bullying' and 'Emotional Bullying'. The 'Online Bullying' category has the fewest number of incidences. This distribution provides an overview of the prevalence of each type of bullying in the dataset, which can help in determining the focus of interventions and more effective prevention strategies.

3.5 Data Cleaning

Data cleansing involves several steps to ensure that the data used is clean and ready for further analysis. The process is shown in Figure 6.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506



data.head()

	Bullying text	Label
0	they spread lies about me, saying i did things...	3
1	every time there is a music lesson, i am alway...	1
2	every time i make a mistake, they point it out...	4
3	i used to have a group of friends, but now the...	3
4	i get shoved into lockers almost every day. th...	2

data.head()

	Teks Bullying	Label
0	spread lies saying things didnt everyone belie...	3
1	every time music lesson always laughed classma...	1
2	every time make mistake point laugh say im fai...	4
3	used group friends ignore act like im talk beh...	3
4	get shoved lockers almost every day say joke h...	2

data.head()

	Teks Bullying	Label
0	[spread, lies, saying, things, didnt, everyone...	3
1	[every, time, music, lesson, always, laughed, ...	1
2	[every, time, make, mistake, point, laugh, say...	4
3	[used, group, friends, ignore, act, like, im, ...	3
4	[get, shoved, lockers, almost, every, day, say...	2

(a) (b) (c)

Figure 6. Case Folding Results (a), Stop Word Results (b), Tokenization Results (c)

Figure 6 (a) shows bullying data that has undergone a case folding process, which is displayed using the data.head() function. The casefolding function is used to convert text to lowercase, which ensures consistency in the text and helps to reduce variations caused by differences in case and lowercase. The implementation of this function uses the .lower() method on the text identified as a string. In the DataFrame data, the 'Text Bullying' column is casefolded for each text entry. This case folding process is important in text preprocessing because it allows machine learning algorithms to perceive words like "Bullying" and "bullying" as the same entity, improving the accuracy and efficiency of subsequent text analysis. These results ensure that all bullying texts in the dataset are lowercase, ready for the next step of data cleaning and analysis.

Figure 6 (b) shows the results of bullying text after going through the stop word cleaning stage using the clean function. This function cleans up the text by removing various irrelevant elements such as @mentions, hashtags, URLs, numbers, single characters, punctuation, emojis, HTML entities, and non-ASCII characters. In addition, this function also lowercase all letters, remove certain phrases such as "translated by Google", and combine some spaces into one. This cleanup process results in cleaner and more consistent text, making it easier for further analysis or use in machine learning models. The end result is a more structured bullying text that is free of bullying elements, thus improving the accuracy of text analysis.

Figure 6 (c) shows the results of bullying text after going through the tokenization stage using the tokenize function. This function converts the text into a list of tokens (individual words) using word_tokenize from the NLTK library. Each bullying text in the 'Bullying Text' column is tokenized function, which generates a list of words from each text. Tokenization breaks down each sentence or phrase into individual words, making it easier to further analyze such as word frequency, n-grams, and the application of machine learning algorithms. For example, the text "They spread lies about me saying I did things I didn't do. Now everyone believes them and avoids me." is changed to [spread, lies, saying, things, didn't, everyone]. This process ensures that the text is more structured and ready for subsequent preprocessing steps or more in-depth analysis.

data.head()

	Bullying text	Label
0	[spread, lies, saying, things, didnt, everyone...	3
1	[every, time, music, lesson, always, laughed, ...	1
2	[every, time, make, mistake, point, laugh, say...	4
3	[used, group, friends, ignore, act, like, im, ...	3
4	[get, shoved, lockers, almost, every, day, say...	2

Figure 8. Stemming Results

Figure 8 shows the results of the bullying text after going through the stemming stage using the stemming function, which uses the Porter Stemmer of the NLTK library to convert each word into its basic form. This function checks if the input is a string, then breaks the text into tokens and applies stemming to each token. The result is more consistent and simple text, with words like "running" becoming "run" and "saying" becoming "say". This word normalization helps to reduce word variation, improve the accuracy of text analysis, and reduce the number of unique features in the text, making it easier to further analyze and apply machine learning models for bullying detection and classification.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

3.6 TF-IDF

The TF-IDF (Term Frequency-Inverse Document Frequency) process will be applied to convert bullying text into a numerical feature that reflects the importance of the words in the document relative to all other documents. The steps to be taken include: initialization of TF-IDF Vectorizer, transformation of bullying text into TF-IDF features, and conversion of TF-IDF matrices into DataFrames for further analysis shown in Figure 10.

tfidf_df

	account	act	affecting	afraid	almost	always	annoying	anymore	anything	appearance	...	waits	want	wants	wasnt	weird	well	whenever	worthless	yesterday	youre
0	0.0	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
1	0.0	0.0000	0.294004	0.000000	0.000000	0.161894	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.294004	0.0	0.0
2	0.0	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
3	0.0	0.2861	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	0.0	0.0000	0.000000	0.282265	0.324855	0.000000	0.000000	0.0	0.228719	0.0	...	0.0	0.170542	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	0.0	0.2861	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
996	0.0	0.0000	0.294004	0.000000	0.000000	0.161894	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.294004	0.0	0.0
997	0.0	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
998	0.0	0.0000	0.000000	0.000000	0.000000	0.000000	0.275118	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
999	0.0	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

1000 rows × 225 columns

Figure 10. TF-IDF Results

Figure 10 shows the TF-IDF matrix of the processed bullying text. Each row in this matrix represents a single document or text entry of bullying, while each column represents a unique word contained in the entire document set. The values in the matrix show the TF-IDF score for each word in each document. This score reflects the importance of the word in a particular document, taking into account how often it appears and how common the word is throughout the document. For example, the word "affecting" in the second line has a high TF-IDF score, indicating that the word is significant in the document. Next, the results of WordCloud are displayed in Figure 11.

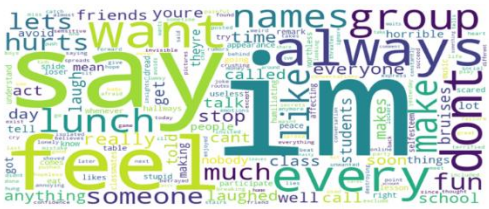


Figure 11. Word Cloud

Figure 11 shows a visualization of the words that most often appear in the processed bullying text, with the word size reflecting their relative TF-IDF score. Words such as "say", "i'm", "feel", "names", "group", "always", "want", "every", and "make" appear larger, indicating that they appear frequently and have a high TF-IDF score, so they are considered important in bullying texts. These words reflect different aspects and contexts of bullying, such as ridicule or name-calling ("names", "called"), feelings of the victim ("feel", "hurt", "isolated"), and the social context in which bullying occurs ("group", "always", "everyone"). This Word Cloud helps in identifying common keywords and themes in bullying texts, provides a quick and easy-to-understand visual overview of the most significant words, and facilitates initial data exploration and further analysis.

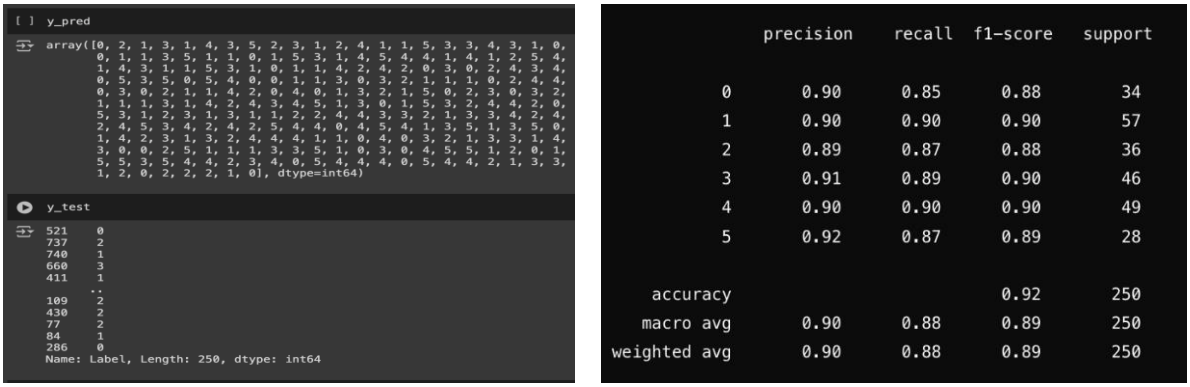
3.7 Machine Learning Algorithm Applications:

Support Vector Machine (SVM) is a highly effective machine learning algorithm for classification tasks, including in natural language processing (NLP). SVM works by looking for hyperplanes that separate data from different classes by maximum margin, and is particularly effective for high-dimensional data that is often found in text. In the context of NLP, SVM can be used for tasks such as text classification, sentiment analysis, and spam detection. These algorithms can handle non-linear data through the use of kernel functions, which transform the data into higher dimensions where it can be separated linearly.

For SVM implementations in NLP, text is usually converted into numerical features using methods such as TF-IDF before being applied to SVMs. After that, the data was divided into training data and testing data with a ratio of 75% for training and 25% for testing. The SVM model is initialized with

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

SVC(), trained using data training (X_train and y_train) with svm.fit(X_train, y_train), and used to predict the results on the test data with svm.predict(X_test). The model performance evaluation is carried out using metrics such as accuracy, precision, recall, and f1-score generated from classification_report(y_test, y_pred), to ensure the model performs well on the given NLP task. Furthermore, an evaluation was carried out on the model with the results as shown in Figure 12.



```
[ ] y_pred
array([0, 2, 1, 3, 1, 4, 3, 5, 2, 3, 1, 2, 4, 1, 1, 5, 3, 3, 4, 3, 1, 0,
       0, 1, 1, 3, 5, 1, 1, 0, 1, 5, 3, 1, 4, 5, 4, 4, 1, 4, 1, 2, 5, 4,
       1, 4, 3, 1, 1, 5, 3, 1, 0, 1, 1, 4, 2, 4, 2, 0, 3, 0, 2, 4, 3, 4,
       0, 5, 3, 5, 0, 5, 4, 0, 0, 1, 1, 3, 0, 3, 2, 1, 1, 1, 0, 2, 4, 4,
       0, 3, 0, 2, 1, 1, 4, 2, 0, 4, 0, 1, 3, 2, 1, 5, 0, 2, 3, 0, 3, 2,
       1, 1, 1, 3, 1, 4, 2, 4, 3, 4, 5, 1, 3, 0, 1, 5, 3, 2, 4, 4, 2, 0,
       5, 3, 1, 2, 3, 1, 3, 1, 1, 2, 2, 4, 4, 3, 3, 2, 1, 3, 3, 4, 2, 4,
       2, 4, 5, 3, 4, 2, 4, 2, 5, 4, 4, 0, 4, 5, 4, 1, 3, 5, 1, 3, 5, 0,
       1, 4, 2, 3, 1, 3, 2, 4, 4, 4, 1, 1, 0, 4, 0, 3, 2, 1, 3, 3, 1, 4,
       3, 0, 0, 2, 5, 1, 1, 1, 3, 3, 5, 1, 0, 3, 0, 4, 5, 5, 1, 2, 0, 1,
       5, 5, 3, 5, 4, 4, 2, 3, 4, 0, 5, 4, 4, 4, 0, 5, 4, 4, 2, 1, 3, 3,
       1, 2, 0, 2, 2, 2, 1, 0], dtype=int64)

y_test
521    0
737    2
740    1
660    3
411    1
      ..
109    2
430    2
77     2
84     1
286    0
Name: Label, Length: 250, dtype: int64
```

(a)

	precision	recall	f1-score	support
0	0.90	0.85	0.88	34
1	0.90	0.90	0.90	57
2	0.89	0.87	0.88	36
3	0.91	0.89	0.90	46
4	0.90	0.90	0.90	49
5	0.92	0.87	0.89	28
accuracy			0.92	250
macro avg	0.90	0.88	0.89	250
weighted avg	0.90	0.88	0.89	250

(b)

Figure 12. Prediction Results (a) Evaluation Results (b)

Figure 12 (a) shows the prediction results (y_pred) and actual values (y_test) of the SVM model for bullying data. y_pred contains the model's predicted labels, while y_test contains the actual labels. Correct predictions occur when both values match; differences indicate errors. Classes include 0, 1, 2, 3, 4, and 5. Analyzing these differences helps calculate accuracy and identify error patterns. Figure 12 (b) presents the SVM model evaluation using a classification report with precision, recall, f1-score, and support metrics from scikit-learn's classification_report function. Precision ranges from 0.89 to 0.92, recall from 0.85 to 0.90, and F1-score from 0.88 to 0.90. Support shows instance counts, such as 34 for class 0 and 57 for class 1. The overall accuracy is 0.92. Macro and weighted average precision, recall, and f1-score are 0.90, 0.88, and 0.89. This report highlights the model's high accuracy and balanced precision and recall across classes.

3.8 Interpretation of Results

The final stage in our process is the "Interpretation of Results". This involves analyzing the output generated by the machine learning algorithms to derive meaningful insights and conclusions regarding the instances of bullying reported through the application.

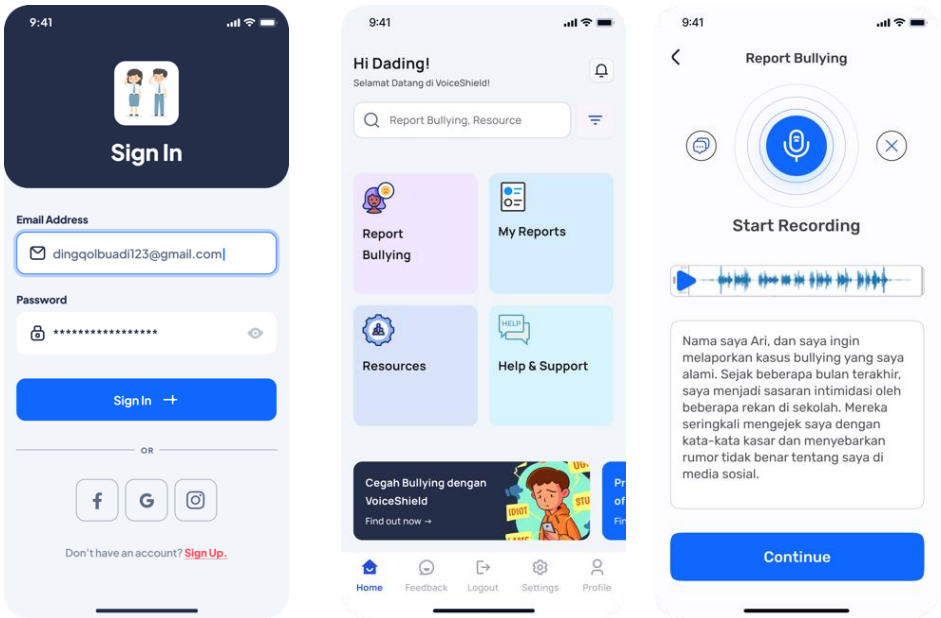


Figure 13. VoiceShield Application Implementation Results

Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

The mobile app "VoiceShield: Bullying Protector App with NLP" is designed to help high school students report bullying incidents and get support. The login page of this application displays a high school student icon above the login form, giving a special impression for high school students. After logging in, users are redirected to the home page which contains menus such as Home, Report Bullying, My Reports, Resources, Help & Support, Profile, Settings, Feedback, and Logout. On the complaints page, users can record their complaints using the voice recorder button, and the results of the recording and transcription of the complaint are displayed for review. For example, the transcription shown in the image depicts the bullying incidents experienced by users in the field. In addition, there is an admin page that contains a list of incoming complaints with a "play" action button to process the complaint. With this application, users are expected to be able to report bullying incidents more easily and get the necessary support.

4. CONCLUSION

The "VoiceShield: Bullying Protector App with NLP" app is designed to help high school students report bullying incidents effectively and get the support they need. The app has several important features, including a login page, a main page with important menus, a complaints page that allows users to record and transcribe their complaints, and an admin page to process complaints. The main process in this application is the conversion of speech to text using Natural Language Processing (NLP), which is then further analyzed. The results of the performance evaluation of the SVM model applied in the application show very satisfactory results, with an accuracy rate of 92%, precision of 90%, recall of 88%, and F1-score of 89%. This evaluation shows that the SVM model is able to classify bullying-related texts with a high degree of accuracy, as well as a balance between precision and recall. As such, the app not only provides users with useful tools for reporting bullying cases, but also ensures that the generated reports can be processed and analyzed with high accuracy, providing more effective support for bullying victims.

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to the Directorate of Research and Community Service (DRTPM) for providing research funds that make this research possible. Financial support from DRTPM was instrumental in the development and implementation of "VoiceShield: Bullying Protection App with NLP". We would also like to thank all the speakers who have shared their experiences and insights, being a valuable source of information and inspiration for this research.

REFERENCES

[1] A. Y. Azhari, D. L. N. Janah, F. E. Meyliana, and B. Setiawan, "The Influence of the Development of Character Education in Overcoming the Problem of Bullying in Indonesia," *Sinar Dunia J. Ris. Sos. Hum. and educator knowledge.*, vol. 2, no. 4, hlm. 257–271, Nov 2023, doi: 10.58192/sidu.v2i4.1588.

[2] N. Aristiani, M. Kanzunnudin, and N. Fajrie, "Bullying Behavior in Elementary School Children in Gribig Village, Kudus," *J. type pedagog.*, vol. 4, no. 2, Des 2021, doi: 10.24176/jpp.v4i2.5989.

[3] Y. Siswati and M. Saputra, "The Role of the School Anti-Bullying Task Force in Overcoming the Phenomenon of Bullying in Senior High Schools," *Cive J. Researcher. Educators. Pancasila and Citizenship*, vol. 3, no. 7, hlm. 216–225, Jul 2023, doi: 10.56393/decive.v3i7.1656.

[4] M. Ilmi, D. R. Habibie, and Y. Arifin, "Analysis and Design of the Monitoring Information System for Street Vendor Students at SMK Permata Harapan," *JOINS J. Inf. Syst.*, vol. 8, no. 2, hlm. 177–187, Nov 2023, doi: 10.33633/joins.v8i2.9233.

[5] M. Furqan, S. Sriani, and M. N. Shidqi, "Telegram Chatbot Using Natural Language Processing," *Walisongo J. Inf. Technol.*, vol. 5, no. 1, HLM. 15–26, Jun 2023, doi: 10.21580/wjit.2023.5.1.14793.

Volume 11, Issue 2, October 2024, pp. 159-169
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v11i2.4506

[6] R. M. Suryadi, "Design and Build Glasses to Convert Voice to Text," *J. Tek. Machine Learning*, Vol. 4, No. 1, HLM. 53, Jul 2021, doi: 10.17977/um054v4i1p53-61.

[7] M. P. R, M. Anu, dan D. S, "Building A Voice Based Image Caption Generator with Deep Learning," dalam *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India: IEEE, Mei 2021, hlm. 943–948. doi: 10.1109/ICICCS51141.2021.9432091.

[8] A. Anand, A. A. Rastogi, R. A. Chadichal, A. Surana, Dr. S. G, dan Dr. L. N. R, "Handwritten Text Recognition and Conversion to Speech," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 6, hlm. 3904–3914, Jun 2023, doi: 10.22214/ijraset.2023.54317.

[9] A. Apturkar, A. I. Iliev, A. Anand, A. Oli, S. Reddy Siddenki, dan V. Reddy Meka, "Sentiment Analysis of Speech with Application to Various Languages," *Digit. Present. Preserv. Cult. Sci. Herit.*, vol. 10, pp. 103–118, Sep 2020, doi: 10.55630/dipp.2020.10.6.

[10] R. A. K, T. Triveni, V. N R, V. K, dan R. B M, "Speech to Text App Customized for Police Functioning in Different Languages," dalam *2023 4th International Conference for Emerging Technology (INCET)*, Belgaum, India: IEEE, May 2023, hlm. 1–4. doi: 10.1109/INCET57972.2023.10170687.

[11] T. Cui, J. Xiao, L. Li, X. Jiang, dan Q. Liu, "An Approach to Improve Robustness of NLP Systems against ASR Errors." arXiv, 25 Maret 2021. Diakses: 12 Juli 2024. [Daring]. Tersedia pada: http://arxiv.org/abs/2103.13610

[12] Lviv Polytechnic National University, Y. Tyshchuk, V. Vysotska, Lviv Polytechnic National University, O. Vlasenko, dan Zhytomyr Ivan Franko State University, "Information system for converting audio in Ukrainian language into its textual representation using nlp methods and machine learning," *Vìsn. The National University of L'vìvì Polìhnìka serìâ ìformation sist. Ta Merezì*, vol. 12, HLM. 23–51, As of 2022, doi: 10.23939/sisn2022.12.023.

[13] G. P. Ashok, "Virtual Bot Powered by Machine Learning and NLP Technologies: Emulating Human-Like Conversations through Speech-to-Text Conversions," *INTERANTIONAL J. Sci. Res. Eng. Manag.*, vol. 07, no. 07, Jul 2023, doi: 10.55041/IJSREM24835.

[14] A. Mishra, A. Sahay, M. A. Pandey, dan S. S. Routaray, "News text Analysis using Text Summarization and Sentiment Analysis based on NLP," dalam *2023 3rd International Conference on Smart Data Intelligence (ICSMDI)*, Trichy, India: IEEE, Mar 2023, hlm. 28–31. doi: 10.1109/ICSMDI57622.2023.00014.

[15] L. Cui, Y. Li, dan Y. Zhang, "Label Attention Network for Structured Prediction," *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 30, HLM. 1235–1248, 2022, doi: 10.1109/TASLP.2022.3145311.

[16] R. P. Dias, C. S. L. Vidanapathirana, R. Weerasinghe, A. Manupiya, R. M. S. J. Bandara, dan Y. P. H. W. Ranasinghe, "Automated use case diagram generator using NLP and ML." arXiv, 2023. doi: 10.48550/ARXIV.2306.06962.

[17] F. Liu, H. Huang, Z. Yang, Z. Hao, dan J. Wang, "Search-Based Algorithm With Scatter Search Strategy for Automated Test Case Generation of NLP Toolkit," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 5, no. 3, HLM. 491–503, Jun 2021, doi: 10.1109/TETCI.2019.2914280.

[18] A. M. Maatuk dan E. A. Abdelnabi, "Generating UML Use Case and Activity Diagrams Using NLP Techniques and Heuristics Rules," dalam *International Conference on Data Science, E-learning and Information Systems 2021*, Ma'an Jordan: ACM, Apr 2021, hlm. 271–277. doi: 10.1145/3460620.3460768.