Assessment 2: Advanced Data Mining Techniques in Cyberbullying Detection

Sachin: 23235298 — Yash Mehta: 23145127 — Mohammed Mufid Shaikh: 23228726 [* 1]

Abstract

The emergence of cyberbullying is a serious problem in the globalised world of today, impacting individuals across different social media platforms. This study explores how data mining methods may be used to detect and stop cyberbullying. Our goal is to find patterns and links in the data by applying association rule analysis, clustering, and classification techniques. Datasets from Kaggle, which classify occurrences of cyberbullying by age, gender, ethnicity, and religion, form the basis of the study. To categorise various forms of cyberbullying, we used machine learning models such as Support Vector Machine (SVM), Random Forest, and Logistic Regression. Clusters of cyberbullying incidents were formed using techniques like K-Means, with text converted into numerical representations through TF-IDF. To streamline the data, PCA was applied for dimensionality reduction. The optimal number of clusters was identified using the Elbow Method, revealing distinct themes within the data. Furthermore, underlying patterns were uncovered through association rule mining, with the reliability of these patterns assessed using metrics such as lift, confidence, and support. This integrated approach provides a deep understanding of cyberbullying, offering valuable insights that can help in developing more effective detection systems and fostering safer online spaces.

1. Introduction

Cyberbullying involves using platforms like Twitter, Facebook, Instagram, or online forums to harass or humiliate others, often leading to anxiety, depression, and social isolation. Unlike traditional bullying, it can happen anytime, often anonymously, and affects victims emotionally, academically, and in their personal lives. Addressing cyberbullying requires education, timely support, and collaboration across various platforms.

Involvement in cyberbullying can happen in three main ways: as a victim, who receives hurtful online messages; as a perpetrator, who sends those harmful messages; or as a bystander, who witnesses the bullying. People who experience cyberbullying are often also involved in traditional bullying, with studies showing a strong connection between the two (Giumetti & Kowalski, 2022).

This study focuses on creating automated systems that use data mining techniques to detect and prevent cyberbullying, helping to make online spaces safer for users. To do this, we will perform various Exploratory Data Analysis and apply datamining techniqueslike clustering, classification, and association. Despite there being various countermeasures for cyberbullying it is still difficult to handle with the increase in cases with the advancement of the digital world.

2. Domain Description

A type of evil law known as ”cybercrime” was created due to advancements in the digital world. Cyberbullying is one of the modern forms of crime on the world wide web. Cyberbullying is when an internet user uses information technology like social media platforms like Facebook, twitter, Instagram, TikTok, Online games etc. or a mobile device to purposefully threaten, intimidate, or embarrass a person or group of users. Cyberbullying is the intentional and persistent use of digital technologies such as email, mobile phones, chat rooms, social networking, and personal messaging, with the goal of harming the other party (Riadi et al., 2017).

Bullying and other undesirable behaviors are becoming common because of people of all ages and genders using social media and other technology more frequently. One of the setbacks that can happen to a person is bullying, especially if it occurs when they are young. Bullying typically affects women, children, and teenagers. Bullying can damage a person’s mental and emotional health as well as their personality (Al-Khater et al., 2020).

3. Problem Definition

Cyberbullying is defined as harmful online conduct that is capable of hurting the emotions and minds of people. Changing abuse is not easily captured by traditional moderation. The purpose of this study is to develop automated detection systems based on data mining and NLP techniques, which can efficiently recognize and avoid cyberbullying to provide secure online spaces for users (Al-Garadi et al., 2016).

4. Dataset Description

The cyberbullying project uses information from two main Kaggle sources. First, a multiclass detection model is developed by authors to address the heightened threat of cyberbullying during COVID-19 in the *Fine-Grained Balanced Cyberbullying* Dataset. The dataset is divided into six categories in which the comments are classified into these cyberbullying categories. They used a semi-supervised machine learning process to obtain more than 47,000 balanced tweets for the dataset (Wang et al., 2020). The *Fine-Grained Balanced Cyberbullying Dataset* includes the following types of cyberbullying, each of them having values 0 and 1, where 1 means presence of cyberbullying and 0 means absence: Age, Ethnicity, Gender, Religion, Other types of cyberbullying, Not cyberbullying.

Table 1. Dataset Summary

Features	Data Size	Minimum	Mean	Maximum
Age	7992	0	0.1655	1
Ethnicity	7961	0	0.1649	1
Gender	7973	0	0.1651	1
Religion	7998	0	0.1656	1
Other_cyberbullying	7823	0	0.1620	1
Not_cyberbullying	7945	0	0.1645	1
Sexism	592	0	0.0123	1

The second dataset, *Twitter_Sexism_Parsed_Dataset*, is a subset of the larger dataset Cyberbullying datasets that was gathered by Fatma Elsafoury in 2020. This compilation contains data from several social media sites, such as YouTube, Wikipedia Talk pages, and Twitter. Hate speech, hostility, insults, and toxicity are all included in the scope of cyberbullying that has been documented. A subset that was specifically focused on Twitter was taken out and used for this research.

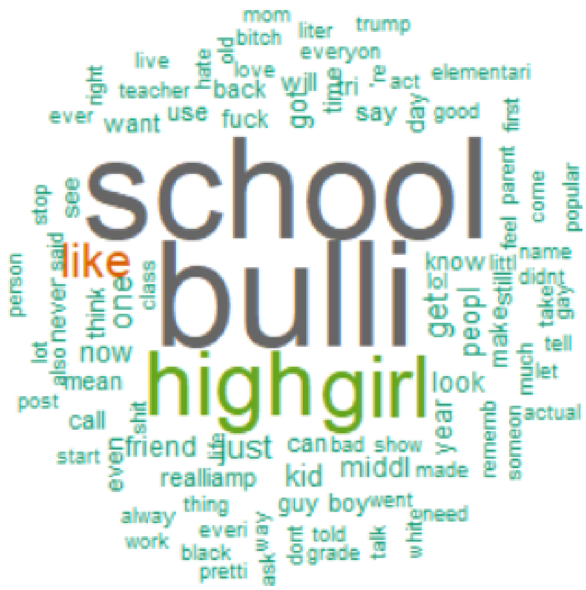
5. Dataset Pre-Processing and EDA

5.1. Pre-Processing

After loading in the dataset from a CSV file, the Cyberbully classification types (*"not_cyberbullying", "gender", "religion", "age", "ethnicity", "sexism", or "other_cyberbullying"*) are converted into numbers using a mapping dictionary (*encoding_dict*) that assigns a unique integer value to each category of cyberbullying.

For text preprocessing, first the text is converted uppercase to lowercase to ensure consistency. Next, punctuation and stopwords are removed . Stopwords are common words that do not contribute much meaning and are often removed to reduce noise in text analysis. A set of custom stopwords specific to the context of tweets is also included such as ‘rt’, ‘don’, ‘im’, etc. Finally, stemming is applied, which reduces words to their base or root forms, thereby standardizing the text and reducing its dimensionality.

5.2. Exploratory Data Analysis (EDA)



(a). Age

(b). Ethnicity

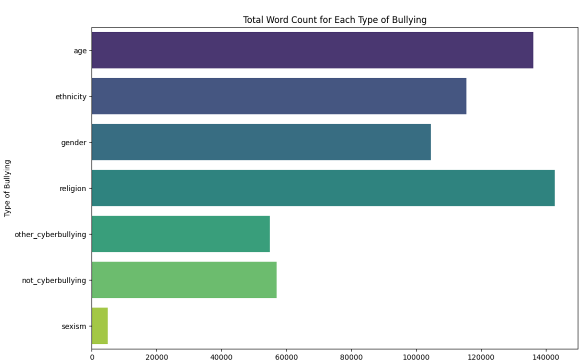
(c). Gender

(d). Not Cyberbullying

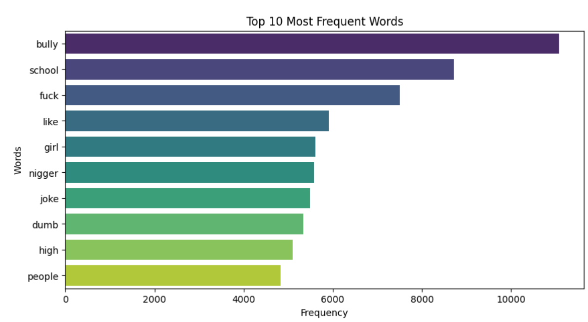
(e). Sexism

(f). Religion

Cyberbullying themes are often depicted by the word clouds, which display offensive language pertaining to age, ethnicity or gender all of which go further to emphasize the all-encompassing and dangerous nature of online harassment.



(g). Total Word Count for Each Type of Bullying



(h). Top 10 Most Frequent Words

Figure (g) The distribution of bullying discussions (by years-old) is shown in this chart and reveals that the major interest focuses on Age, Religion. He finds the least sexism there, but that could also mean it is less visible or under reported.

Figure (h) The top 10 most frequently words represent significant themes such as bullying, school and swearing revealing the core language used during cyberbullying events that can guide focused intervention efforts.

6. Experiments

6.1. Classification

Three prediction models were created to categorize different forms of cyberbullying, such as those that fall under the age, racial, religious, and other categories. Several machine learning approaches were used by the model, such as Logistic Regression, Random Forest, and Support Vector Machine (SVM). To manage the categorization of more than two unique categories for the Logistic Regression component, a multinomial parameter was used. With this method, the distinct types of cyberbullying in the dataset were intended to be precisely identified and distinguished from one another.

6.1.1. Random Forest

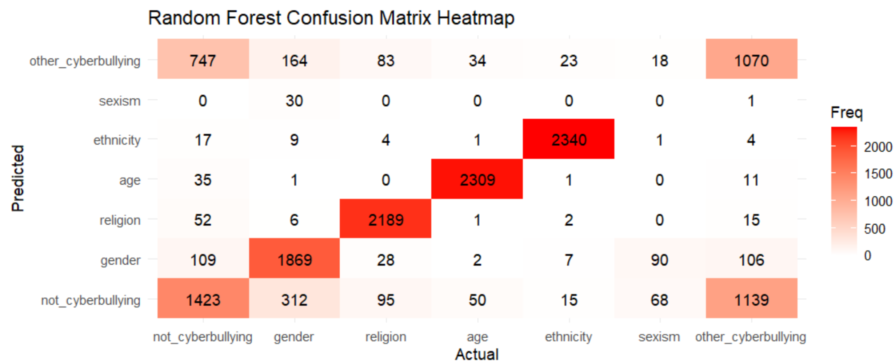


Figure 1.

When compared to conventional classifiers, the random forest classifier provides a higher degree of classification accuracy. Since the number of trees that can be formed is infinite and the generalization error converges consistently without the requirement for tree pruning, it successfully nullifies the risk of over-fitting (Thangarasu & Alla, 2023).

The random forest model performed well in detecting cyberbullying related to age and ethnicity of recall 97.97% and 98.85%, while performing poorly in predicting category "Not_cyberbullying" poorly as compared to other models of about 60% and almost completely unable to detect and predict "sexism" category.

Metric	Not_cyberbullying	Gender	Religion	Age	Ethnicity	Sexism	Other_cyberbullying	Avg
Recall	0.59715	0.7817	0.9125	0.963	0.9799	0	0.4561	0.670093
Specificity	0.86122	0.9717	0.9937	0.996	0.997	0.99783	0.91191	0.961338
Pos Pred Value	0.45874	0.8453	0.9664	0.98	0.9848	0	0.50023	0.676439
Neg Pred Value	0.91563	0.9575	0.9828	0.993	0.996	0.98775	0.89661	0.961284
Prevalence	0.16456	0.1651	0.1657	0.166	0.1649	0.01222	0.16201	0.142856
Detection Rate	0.09827	0.1291	0.1512	0.16	0.1616	0	0.07389	0.110509
Detection Prevalence	0.21421	0.1527	0.1564	0.163	0.1641	0.00214	0.14771	0.142866
Balanced Accuracy	0.72918	0.8767	0.9531	0.98	0.9885	0.49892	0.684	0.815728

Table 2. Performance Metrics of Random Forest

6.1.2. Logistic Regression

This model was applied using "multinomial" parameter in order to classify more than two categories for our dataset. In the project "Non-linguistic Features for Cyberbullying Detection on a Social Media Platform Using Machine Learning" Logistic Regression was applied and received the best results in terms of F1-Measure, Precision, Accuracy, and AUC (Liu et al., 2019).

The logistic regression model shows promising results in detecting cyberbullying related to age and ethnicity, with balanced accuracies of 97.92% and 98.12%, respectively, and high Recall or the score at which the model predicted correctly was 96% and 97%.

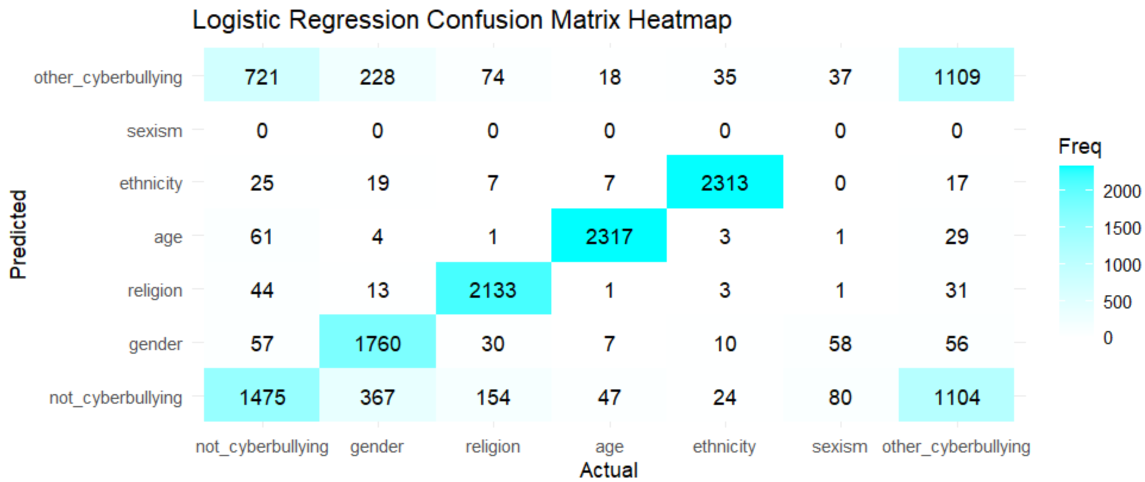


Figure 2.

Metric	Not_cyberbullying	Gender	Religion	Age	Ethnicity	Sexism	Other_cyberbullying	Avg
Recall	0.619	0.7361	0.8891	0.967	0.9686	0	0.47272	0.664589
Specificity	0.8532	0.982	0.9923	0.992	0.9938	1	0.90828	0.960197
Pos Pred Value	0.4537	0.8898	0.9582	0.959	0.9686	NaN	0.4991	0.788067
Neg Pred Value	0.9191	0.9495	0.9783	0.993	0.9938	0.98778	0.89909	0.960139
Prevalence	0.1646	0.1651	0.1657	0.166	0.1649	0.01222	0.16201	0.142861
Detection Rate	0.1019	0.1215	0.1473	0.16	0.1597	0	0.07658	0.109569
Detection Prevalence	0.2245	0.1366	0.1537	0.167	0.1649	0	0.15344	0.142849
Balanced Accuracy	0.7361	0.859	0.9407	0.979	0.9812	0.5	0.6905	0.812386

Table 3. Performance Metrics of Logistic Regression

6.1.3. Support Vector Machine

An initiative to tackle detection of cyberbullying in the Bangla language. They proposed machine learning methods that combined user-specific data with linguistic subtleties and socio-emotional behavior analysis. SVM outperformed other algorithms due to its flexibility in handling text data (Purnachandra Rao et al., 2024). With balanced accuracy close to 98% and extremely high recall of 96.66% and 96.86%, respectively.

According to the confusion Matrix heatmaps all the models had similar issues regarding "sexism" where the model wasn't able to predict it correctly, similarly in the case of "Other_cyberbullying" the models incorrectly predicting as "Not_cyberbullying" category recall ranging between 36% to 47% score.

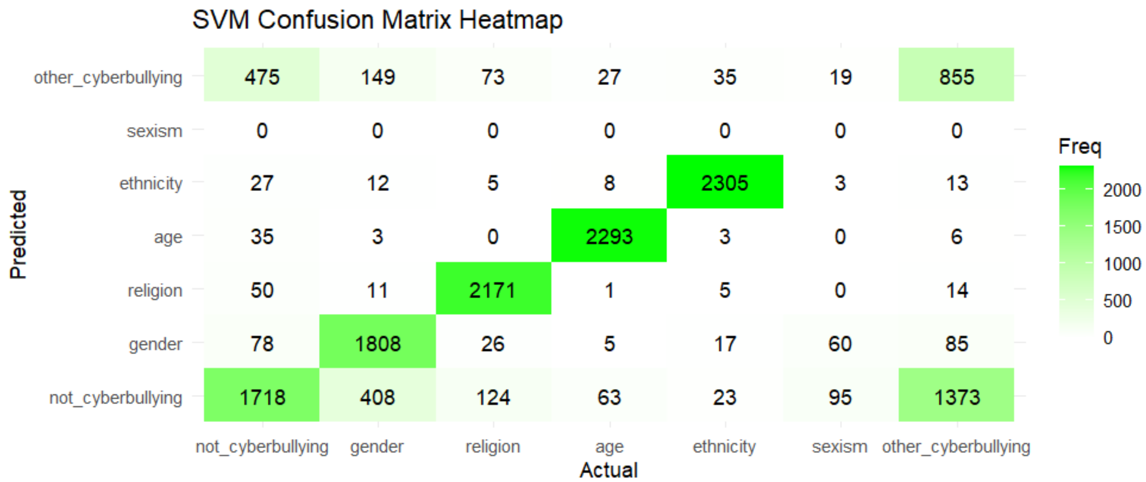


Figure 3.

Metric	Not_cyberbullying	Gender	Religion	Age	Ethnicity	Sexism	Other_cyberbullying	Avg
Recall	0.7209	0.7562	0.905	0.957	0.9652	0	0.36445	0.666907
Specificity	0.8276	0.9776	0.9933	0.996	0.9944	1	0.93589	0.960699
Pos Pred Value	0.4516	0.8696	0.964	0.98	0.9713	NaN	0.52358	0.79333
Neg Pred Value	0.9377	0.953	0.9814	0.991	0.9931	0.98778	0.88395	0.96119
Prevalence	0.1646	0.1651	0.1657	0.166	0.1649	0.01222	0.16201	0.142861
Detection Rate	0.1186	0.1249	0.1499	0.158	0.1592	0	0.05904	0.109907
Detection Prevalence	0.2627	0.1436	0.1555	0.162	0.1639	0	0.11277	0.142867
Balanced Accuracy	0.7743	0.8669	0.9491	0.976	0.9798	0.5	0.65017	0.81381

Table 4. Performance Metrics of SVM

6.2. Clustering

Combined outputs to learn from—which allows clustering to work its insights on data points whose relationships and structure clustering is the method of organizing data points into groups based on their interrelationships. This is done without supervision—meaning it doesn't rely on predeter es are not well known. Clustering can produce a ”family resemblance” among groups of varied data points. To group its points, a clustering method first assesses the degree of similarity between all the pairs of points. Then it uses the computed similarity scores to put the points into the appropriate clusters(Romsaiyud et al., 2017).

6.2.1. TFIDF

A widely used algorithm converts text into a significant numerical representation that can be ”understood” by a machine-learning algorithm, enabling it to make predictions(Murato et al., 2024).

The processed tweets underwent the vectorization transformation of the TF-IDF process. This means that the essence of each tweet was captured in a vector that emerged from the following steps. First, some tweet terms were eliminated that were not ”significant,” which reduced the dataset to the 1,000 most significant and helpful terms. Then, those terms were weighted for importance in a calculation that rendered a ”term frequency-inverse document frequency,” or TF-IDF for short.

6.2.2. Choosing K Value

To find the optimal number of clusters (k) for K-Means clustering, we utilized the commonly accepted Elbow Method. This method takes k as the x-variable and the inertia, or within-cluster sum of squares (WSS), as the y-variable. The inertia measures how well the data is clustered; lower values mean better clustering(Syakur et al., 2018).

When plotting the inertia against k, you see the first decrease in the rate of inertia from k=1 to k=2, the second decrease from k=2 to k=3, and then a much less steep decrease from k=3 to k=4. These decreases give the appearance of an ”elbow.” Although a human is responsible for seeing the ”elbow,” the k=4 choice clearly gives a compact, well-separated set of four clusters.

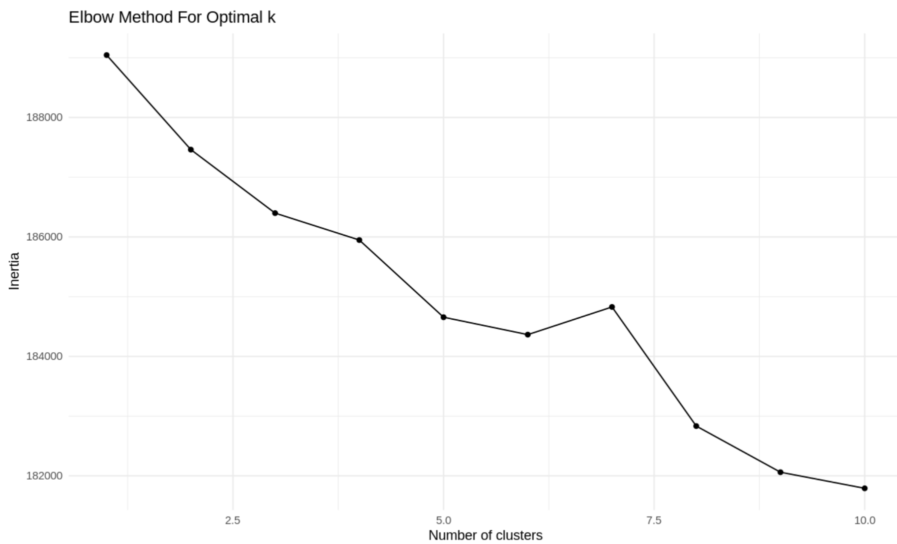


Figure 4. Elbow Method to find optimal value of K

6.2.3. PCA

Reducing the number of features in an ML model is one way to improve it. A model with fewer features can attain a comparable level of explainability to that of a model using the full dataset. And, of course, a model with fewer features uses less memory and requires less computation. We can reduce the number of features by employing PCA. PCA helps us to select the n features that preserve most of the variance(Ding & He, 2004).

6.2.4. K Means Clustering

The K-Means algorithm is perhaps the most basic but also the most used of the unsupervised learning methods. In contrast to supervised learning, where one can provide labeled and pre-categorized training data, K-Means does its work without any foreknowledge of what categories might exist.

The first letter of the name, ”K,” stands for the quantity of categories to be formed. If the user has $K = 2$, for example, the system will turn up two grouped categories. There is a method for determining the best or most appropriate K for a given data set, but in practice one often must simply take a best guess and then refine that guess if necessary (Ledesma et al., 2024).

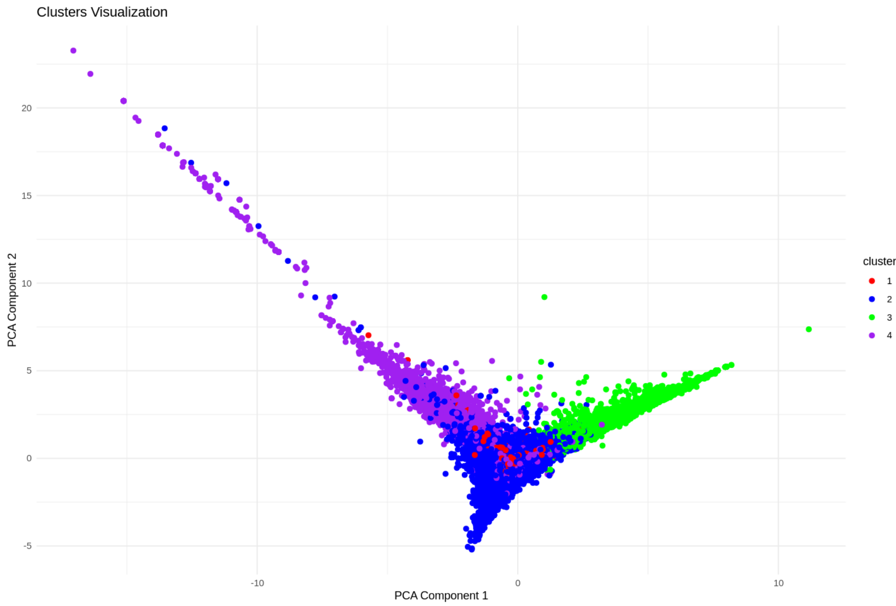


Figure 5. 2D Clustering

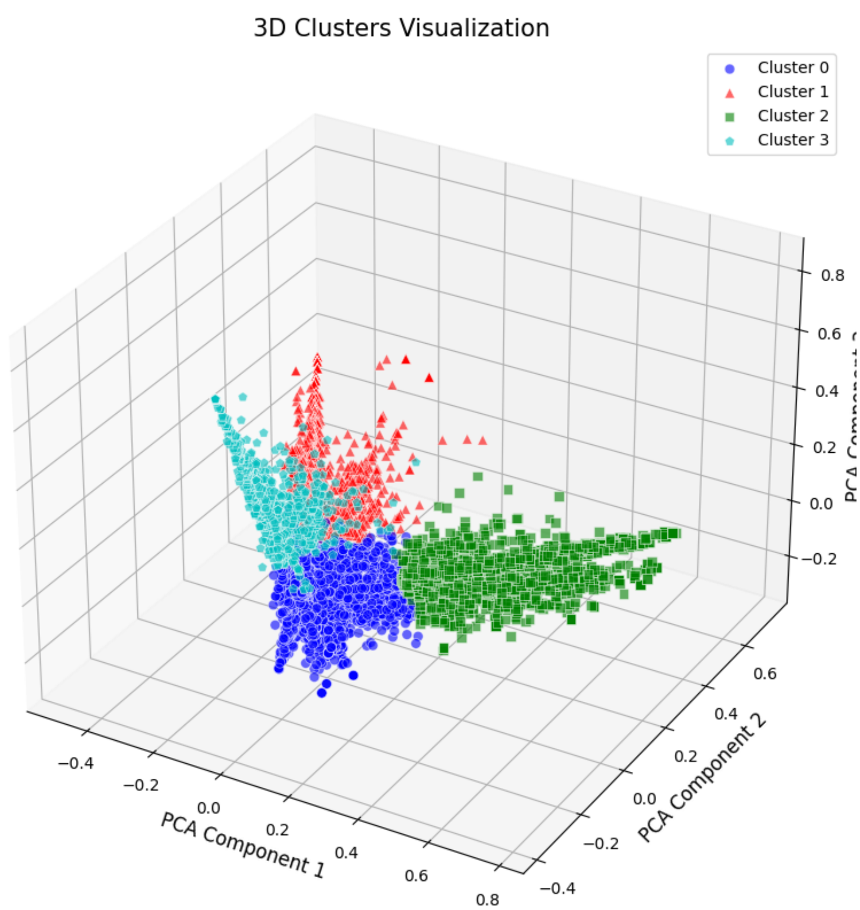
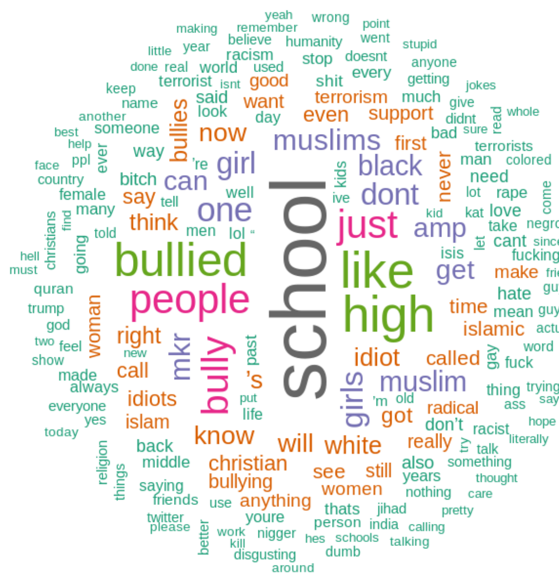


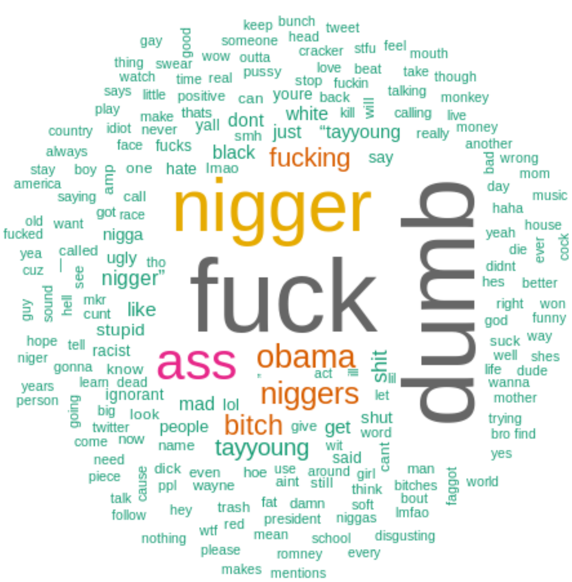
Figure 6. 3D Clustering

The images we provided from our two- and three-dimensional PCA (principal component analysis) offered an enticing view of our data—that is, they showed our data in reduced dimensions. The impression given by the 2D images, however, that our dataset could be cleanly divided into 2 or 3 distinct groups was illusory. The truth, as the 3D PCA showed, is that the data don't fall into obvious clusters. There were evident overlaps among some of these clusters (which a 2D image cannot begin to hint at), and several combinations of different principal components also suggested the existence of more than just a handful of distinct groupings. This was especially important in terms of a cyberbullying detector, in which slight differences in language among neighboring clusters could theoretically result in big differences in how our system works.

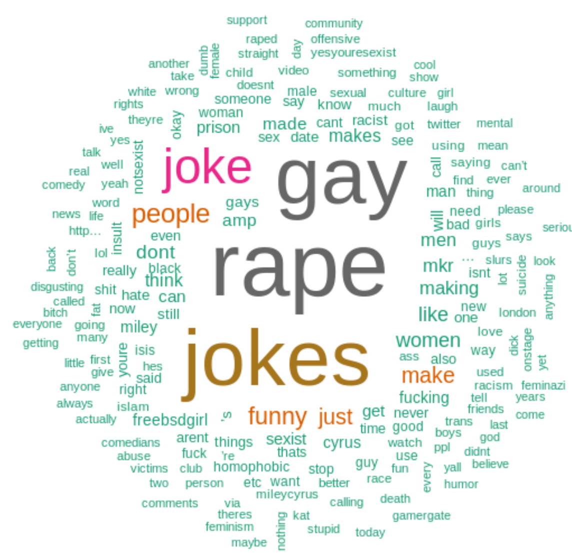
(a). Cluster 1



(b). Cluster 2



(c). Cluster 3



(d). Cluster 4

The visual representations of the data provide significant information about the particular topics and kinds of cyberbullying that exist in each cluster. They allow for an even clearer comprehension of the content contained in each group.

6.2.5. Cluster Centroid Analysis

The most representative words for each of the four clusters were pulled out in the cluster centroid analysis. This highlights the distinct themes within the data set.

Cluster 1 concerns itself with either religious or political discourse. The most prominent term is ”Muslim” which signals that this cluster involves a prominent discourse for a specific religious group that may also have political elements to it.

Cluster 2 gets into school-related bullying and features the terms ”bully” and ”school,” which clearly indicate this is a discussion of cyberbullying incidents or promotions that happen during the school year or related to age.

Cluster 3 involves severe offensive language related to ethnicity.

Cluster 4 kind of revolves around jokes and things related to sexual orientation.

Table 5. Top features for each cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	retweet	bully	fuck	joke
2	muslim	school	dumb	rape
3	idiot	high	nigger	gay
4	im	girl	ass	funny
5	like	like	obama	people
6	dont	middle	retweet	make
7	people	got	bitch	making
8	one	im	shit	makes
9	get	one	mad	prison
10	know	would	get	made

6.3. Association

Association in the context of data mining and machine learning refers to a technique that identifies relationships between variables in large datasets. The most common example of this technique is **association rule mining**, which is used to discover interesting correlations, patterns, associations, or causal structures among a set of items in transactional data. This method is widely used in market basket analysis to identify items that frequently co-occur in transactions(Srikant & Agrawal, 1997)

Key Concepts

- **Support**: Measures how often a particular itemset appears in the dataset.
- **Confidence**: Indicates the likelihood that a particular rule (e.g., ”If A, then B”) is true given the dataset.
- **Lift**: Evaluates the strength of an association rule, measuring how much more likely the consequent is, given the antecedent, compared to its baseline probability.

6.3.1. Association Rules:

Association rules are a fundamental concept in data mining, specifically in the context of association rule mining. They are used to uncover relationships between items in large datasets. An association rule is an implication of the form $A \rightarrow B$, where A (the antecedent) and B (the consequent) are disjoint itemsets. The rule suggests that if a transaction contains A, it is likely to also contain B(Srikant & Agrawal, 1997; Giudici & Passerone, 2002).

antecedents	consequents	support	confidence	lift	leverage	conviction
frozenset({'ethnicity'})	frozenset({'age'})	1	1	1	0	0
frozenset({'ethnicity', 'sexism'})	frozenset({'not_cyberbullying', 'religion', 'other_cyberbullying'})	1	1	1	0	0
frozenset({'sexism', 'not_cyberbullying', 'gender'})	frozenset({'ethnicity', 'other_cyberbullying'})	1	1	1	0	0
frozenset({'gender', 'not_cyberbullying', 'other_cyberbullying'})	frozenset({'ethnicity', 'sexism'})	1	1	1	0	0
frozenset({'sexism', 'other_cyberbullying', 'gender'})	frozenset({'ethnicity', 'not_cyberbullying'})	1	1	1	0	0
frozenset({'sexism', 'not_cyberbullying', 'other_cyberbullying'})	frozenset({'ethnicity', 'gender'})	1	1	1	0	0
frozenset({'ethnicity', 'not_cyberbullying', 'gender'})	frozenset({'sexism', 'other_cyberbullying'})	1	1	1	0	0
frozenset({'ethnicity', 'sexism', 'gender'})	frozenset({'not_cyberbullying', 'other_cyberbullying'})	1	1	1	0	0
frozenset({'ethnicity', 'not_cyberbullying', 'sexism'})	frozenset({'gender', 'other_cyberbullying'})	1	1	1	0	0

Figure 7. Association Rules

6.3.2. Key Metrics for Evaluating Association Rules

There are several key metrics used to evaluate the strength and utility of association rules:

- **Support**: This measures how frequently the items in a rule appear together in the dataset. It is calculated as the proportion of transactions that contain both the antecedent and the consequent. Mathematically, support is defined as:

$$\text{Support}(A \to B) = \frac{\text{Number of transactions containing } A \cup B}{\text{Total number of transactions}}$$

- **Confidence**: This measures the reliability of the inference made by the rule. It is the ratio of the number of transactions that contain both A and B to the number of transactions that contain A. Mathematically, confidence is defined as:

$$\text{Confidence}(A \to B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

  Confidence tells us how likely B is to be purchased when A is purchased.

- **Lift**: This measures the strength of a rule over the random co-occurrence of A and B, given their individual supports. A lift value greater than 1 indicates a positive correlation between A and B, while a lift value less than 1 indicates a negative correlation. Lift is defined as:

$$\text{Lift}(A \to B) = \frac{\text{Confidence}(A \to B)}{\text{Support}(B)} = \frac{\text{Support}(A \cup B)}{\text{Support}(A) \times \text{Support}(B)}$$

- **Leverage**: This measures the difference between the observed frequency of A and B appearing together and the frequency that would be expected if A and B were independent. Leverage is defined as:

$$\text{Leverage}(A \to B) = \text{Support}(A \cup B) - (\text{Support}(A) \times \text{Support}(B))$$

- **Conviction**: This measures the degree of implication of the rule. It is the ratio of the expected frequency that A occurs without B (i.e., A and not B) to the observed frequency of A without B. Conviction is greater than 1 when A is positively correlated with B. It is defined as:

$$\text{Conviction}(A \to B) = \frac{1 - \text{Support}(B)}{1 - \text{Confidence}(A \to B)}$$

- **Antecedents:**The item(s) on the left-hand side (LHS) of an association rule, representing the ”if” part of the rule. In the rule $A \to B$, A is the antecedent.

- **Consequents:**The item(s) on the right-hand side (RHS) of an association rule, representing the ”then” part of the rule. In the rule $A \to B$, B is the consequent.

antecedents	consequents	support	confidence	lift	leverage	convictior
frozenset({'word_1', 'word_3', 'word_10', 'word_5', 'w	frozenset({'word_6', 'word_0', 'word_11', 'word_4', 'word_8', 'word_9'})	0.010208	0.675373	36.51133	0.009928	3.023479
frozenset({'word_1', 'word_3', 'word_10', 'word_4', 'w	frozenset({'word_6', 'word_0', 'word_11', 'word_8', 'word_7', 'word_9'})	0.010208	0.655797	36.45327	0.009928	2.852997
frozenset({'word_1', 'word_3', 'word_10', 'word_5', 'w	frozenset({'word_6', 'word_0', 'word_11', 'word_4', 'word_8', 'word_7', 'word_9'})	0.010208	0.635088	36.44458	0.009927	2.69263
frozenset({'word_6', 'word_1', 'word_3', 'word_10', 'w	frozenset({'word_0', 'word_11', 'word_5', 'word_8', 'word_7', 'word_9'})	0.010208	0.667897	36.44044	0.009927	2.955922
frozenset({'word_6', 'word_1', 'word_3', 'word_10', 'w	frozenset({'word_0', 'word_11', 'word_4', 'word_5', 'word_8', 'word_7', 'word_9'})	0.010208	0.646429	36.3888	0.009927	2.77804
frozenset({'word_6', 'word_1', 'word_3', 'word_10', 'w	frozenset({'word_0', 'word_11', 'word_4', 'word_5', 'word_8', 'word_9'})	0.010208	0.683019	36.37024	0.009927	3.095517
frozenset({'word_1', 'word_3', 'word_10', 'word_2', 'w	frozenset({'word_6', 'word_0', 'word_4', 'word_5', 'word_8', 'word_7', 'word_11'})	0.010208	0.712598	36.30976	0.009926	3.411166
frozenset({'word_6', 'word_1', 'word_3', 'word_10', 'w	frozenset({'word_0', 'word_11', 'word_4', 'word_8', 'word_7', 'word_9'})	0.010208	0.672862	36.26504	0.009926	3.000102
frozenset({'word_1', 'word_10', 'word_4', 'word_5', 'w	frozenset({'word_6', 'word_0', 'word_3', 'word_11', 'word_8', 'word_7', 'word_9'})	0.010208	0.639576	36.2331	0.009926	2.725535
frozenset({'word_1', 'word_3', 'word_10', 'word_2', 'w	frozenset({'word_6', 'word_0', 'word_11', 'word_4', 'word_5', 'word_8', 'word_9'})	0.010208	0.641844	36.13072	0.009925	2.742479
frozenset({'word_0', 'word_1', 'word_3', 'word_5', 'wo	frozenset({'word_6', 'word_4', 'word_10', 'word_8', 'word_7', 'word_11'})	0.010208	0.637324	36.10552	0.009925	2.708611
frozenset({'word_6', 'word_3', 'word_10', 'word_4', 'w	frozenset({'word_0', 'word_1', 'word_11', 'word_5', 'word_8', 'word_9'})	0.010208	0.688213	36.10471	0.009925	3.146181
frozenset({'word_6', 'word_10', 'word_4', 'word_2', 'w	frozenset({'word_0', 'word_1', 'word_3', 'word_11', 'word_5', 'word_8', 'word_9'})	0.010208	0.672862	36.04591	0.009924	2.999757
frozenset({'word_1', 'word_10', 'word_5', 'word_2', 'w	frozenset({'word_6', 'word_0', 'word_3', 'word_11', 'word_4', 'word_8', 'word_9'})	0.010208	0.658182	36.02123	0.009924	2.872076
frozenset({'word_3', 'word_10', 'word_5', 'word_2', 'w	frozenset({'word_6', 'word_0', 'word_1', 'word_11', 'word_4', 'word_8', 'word_9'})	0.010208	0.655797	36.00184	0.009924	2.852342

Figure 8. Output: Mathematical Values for Formulated Rules

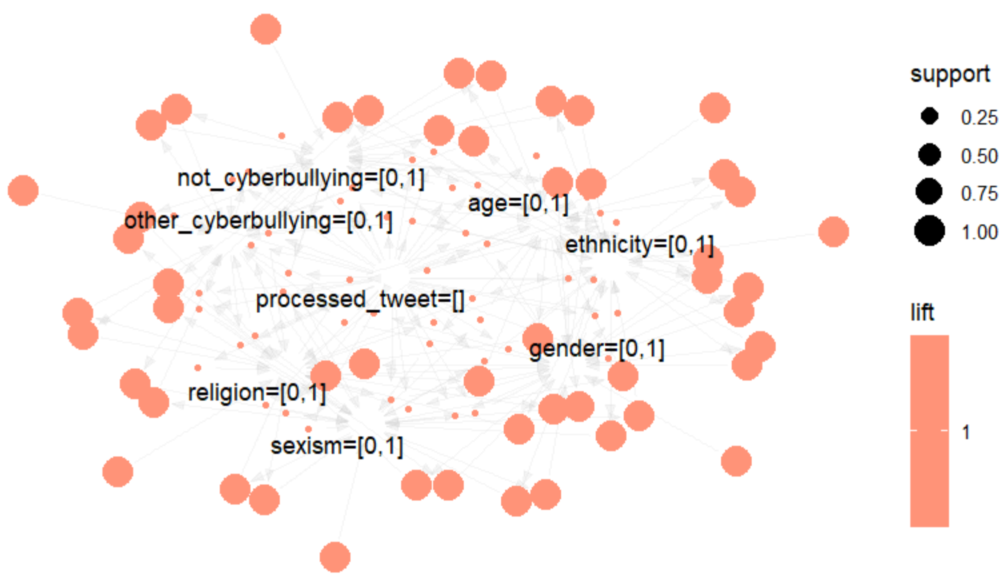


Figure 9. Support and Lift

6.3.3. Importance of Association Rules in Data Mining

Association rules play a crucial role in data mining by helping to uncover hidden patterns, correlations, and relationships in large datasets. These rules are particularly useful in market basket analysis, where businesses can use them to identify products that are frequently bought together. For example, discovering a rule such as $\{\text{Racist}\} \rightarrow \{\text{Black}\}$ might suggest that these words are related to commonly related to cyber bullying(Zhang & Wu, 2011).

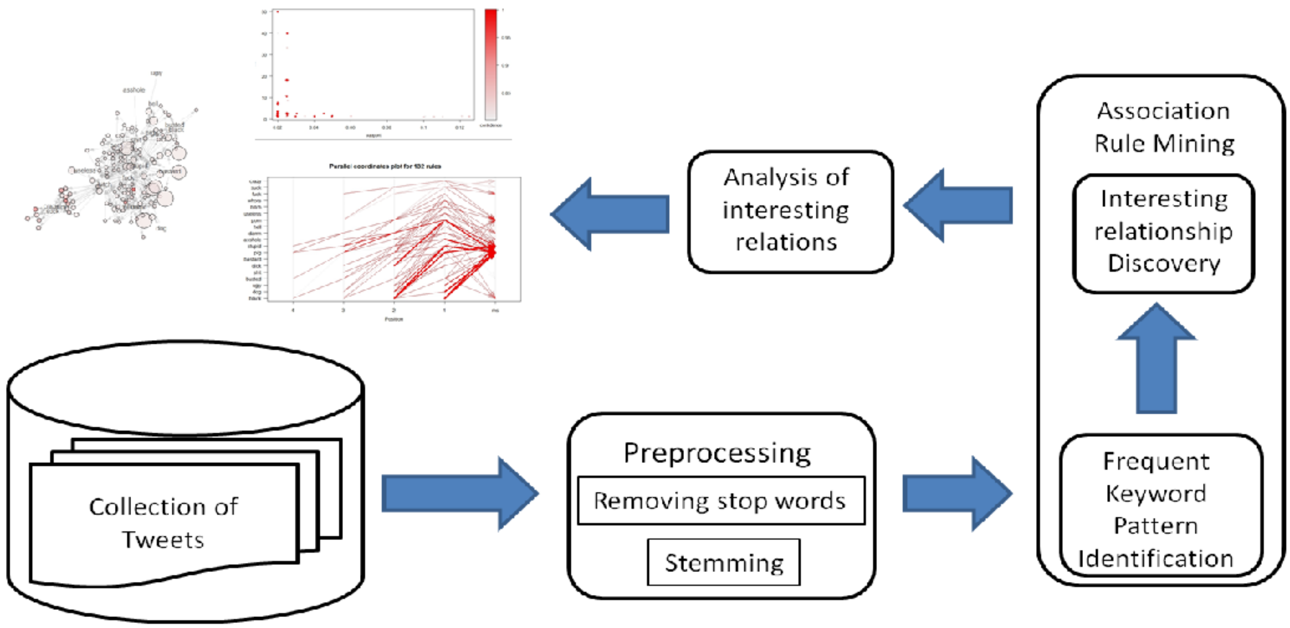


Figure 10. Flow Diagram of Association

6.3.4. Algorithm for Mining Association Rules

There are several algorithms designed to efficiently mine association rules from large datasets:

- **Apriori Algorithm**:.

  Apriori is a data mining technique used in market basket analysis to identify frequent itemsets and generate association rules. It operates iteratively, starting with individual items and combining them into larger sets, while pruning non-frequent itemsets to enhance efficiency. This process continues until no further frequent itemsets are found.(Al-Maolegi & Arkok, 2014; Al-Khater et al., 2020)

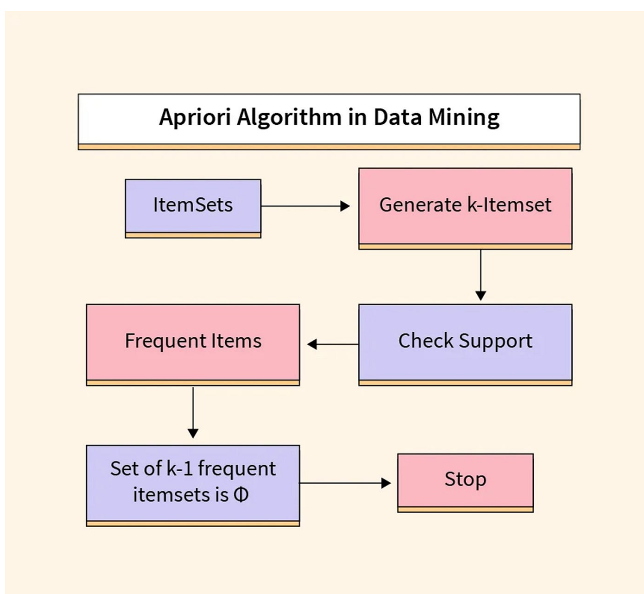


Figure 11. Apriori Algorithm

7. Results & Conclusion

Random Forest classification model held the highest accuracy of 77.34% (among other models) where ”sexism” was mostly not properly predicted in those highly imbalanced data sets Four unique cyberbullying patterns from K-means clustering silhouette score of 0.034 represented moderate clustering quality Finally, association rule mining was highly successful in discovering many strong associations from the data and high lift/confidence relationships are very important for highlighting the useful relationship. ConclusionsWe showed that our models provide a lot of valuable information for the detection of different types of cyberbullying, but their precision could be further improved and this warrants future work.

8. Limitation & Future Scope

The classification accuracy was influenced by data imbalance, especially in the ”sexism” category; silhouette scores for clustering were not so high; scalability issues appeared in association analysis. Classifier performance could be improved by adding more categories and obtaining a more balanced sample in the future. Stronger clustering techniques, i.e., Hierarchical Clustering and Sentiment Analysis can provide better identification of clusters. For handling these kinds of association rule challenges, some aspects such as scalability to a wider range (through distributed computing), and sophisticated methods like deep learning for more valuable insights can be continued efforts.

References

Al-Garadi, M. A., Varathan, K. D., and Ravana, S. D. Cybercrime detection in online communications: The experimental case of cyberbullying detection in the twitter network. *Computers in Human Behavior*, 63:433–443, 2016.

Al-Khater, W. A., Al-Maadeed, S., Ahmed, A. A., Sadiq, A. S., and Khan, M. K. Comprehensive review of cybercrime detection techniques. *IEEE access*, 8:137293–137311, 2020.

Al-Maolegi, M. and Arkok, B. An improved apriori algorithm for association rules. *arXiv preprint arXiv:1403.3948*, 2014.

Ding, C. and He, X. K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, pp. 29, 2004.

Giudici, P. and Passerone, G. Data mining of association structures to model consumer behaviour. *Computational Statistics & Data Analysis*, 38(4):533–541, 2002.

Giumetti, G. W. and Kowalski, R. M. Cyberbullying via social media and well-being. *Current Opinion in Psychology*, 45:101314, 2022. ISSN 2352-250X. doi: https://doi.org/10.1016/j.copsyc.2022.101314. URL https://www.sciencedirect.com/science/article/pii/S2352250X22000161.

Ledesma, G. R. C., Calderón-Vilca, H. D., Ibarra-Cabrera, M. J., and Cárdenas-Mariño, F. C. Architecture of cyberbullying recognizer in video game chat using deep learning model with bilstm. *International Journal of Computer Information Systems and Industrial Management Applications*, 16(1):10–10, 2024.

Liu, Y., Zavarsky, P., and Malik, Y. Non-linguistic features for cyberbullying detection on a social media platform using machine learning. pp. 391–406, 2019.

Murato, D. M., dos Santos, B. S., and Lima, R. H. P. Clustering and analysis of tweets related to petrobras. *Cadernos do IME-Série Informática*, 49(1):113–131, 2024.

Purnachandra Rao, M., Kota, N., Nidumukkala, D., Madoori, M., and Ali, D. Enhancing online safety: Cyberbullying detection with random forest classification. pp. 389–393, 2024. doi: 10.1109/ICCSP60870.2024.10543598.

Riadi, I. et al. Detection of cyberbullying on social media using data mining techniques. *International Journal of Computer Science and Information Security (IJCSIS)*, 15(3), 2017.

Romsaiyud, W., na Nakornphanom, K., Prasertsilp, P., Nurarak, P., and Konglerd, P. Automated cyberbullying detection using clustering appearance patterns. In *2017 9th International Conference on Knowledge and smart Technology (KST)*, pp. 242–247. IEEE, 2017.

Srikant, R. and Agrawal, R. Mining generalized association rules. *Future generation computer systems*, 13(2-3):161–180, 1997.

Syakur, M. A., Khotimah, B. K., Rochman, E., and Satoto, B. D. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: materials science and engineering*, volume 336, pp. 012017. IOP Publishing, 2018.

Thangarasu, G. and Alla, K. R. Detection of cyberbullying tweets in twitter media using random forest classification. pp. 113–117, 2023. doi: 10.1109/ISCAIE57739.2023.10165118.

Wang, J., Fu, K., and Lu, C.-T. Fine-grained balanced cyberbullying dataset. 2020. URL https://dx.doi.org/10.21227/kn1c-zx22.

Zhang, S. and Wu, X. Fundamentals of association rules in data mining and knowledge discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(2):97–116, 2011.