Cyberbullying Detection Report

Generated on: April 10, 2025 00:00:25

Executive Summary

This report analyzes 5 comments for potential cyberbullying content. The analysis found that **60.0%** of the comments contain language consistent with cyberbullying. A total of **3** comments were identified as potentially problematic. Key findings: • Most common terms in problematic comments: bading, hahahhaaha, manok, mcdo • Primary sentiment in flagged comments: Negative Based on these findings, immediate attention is recommended. Detailed analysis and specific recommendations are provided in the following sections.
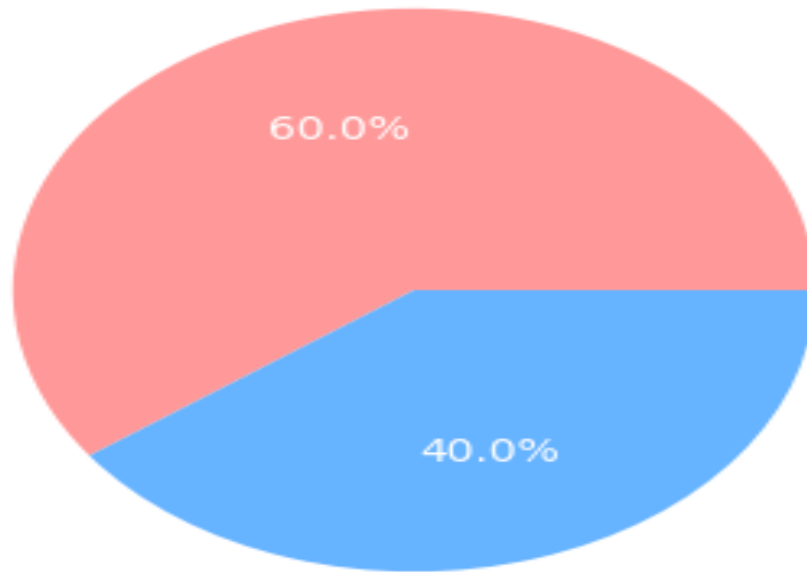
Analysis Details

Analysis Summary

Total Comments Analyzed:	5
Cyberbullying Comments:	3 (60.0%)
Normal Comments:	2 (40.0%)
High Confidence Predictions (>90%):	4 (80.0%)
Comments in Selection List:	0

The analysis shows that a majority (60.0%) of the comments are classified as cyberbullying. This indicates significant presence of harmful content that may require moderation or intervention. The model shows moderate confidence in its predictions, with 80.0% of predictions having over 90% confidence.
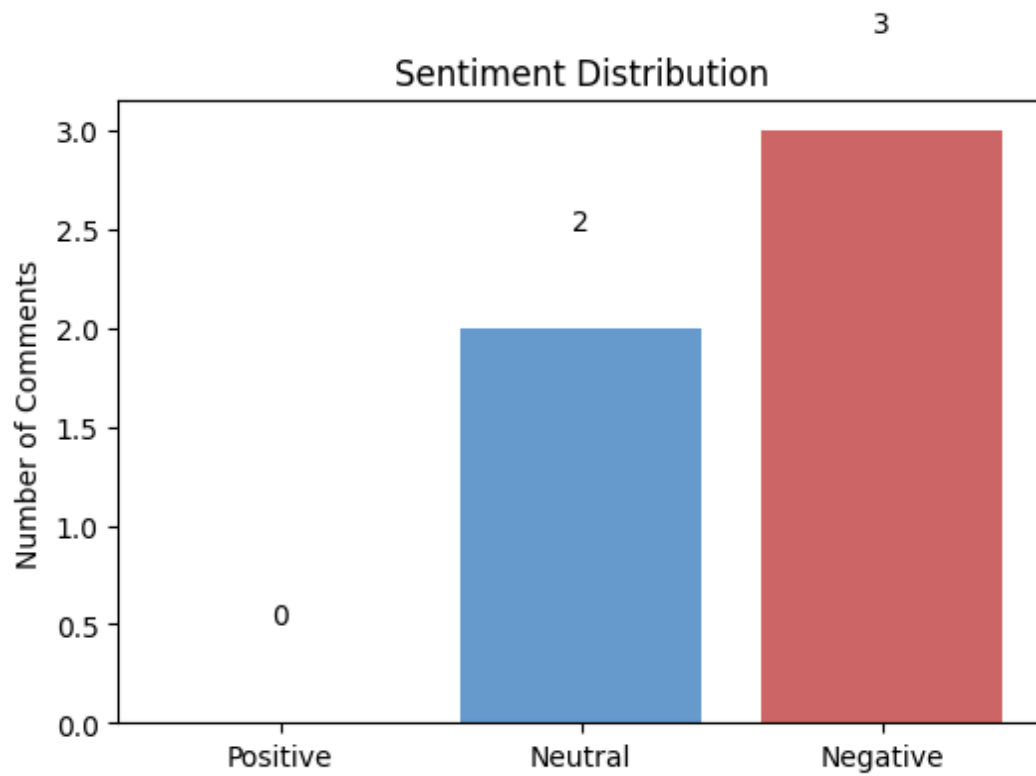
Visual Analysis

Classification Distribution

Comment Classification Distribution

60.0%

40.0%

The chart shows a concerning distribution with 60.0% of comments classified as cyberbullying. This high percentage indicates a significant moderation challenge that requires immediate attention.
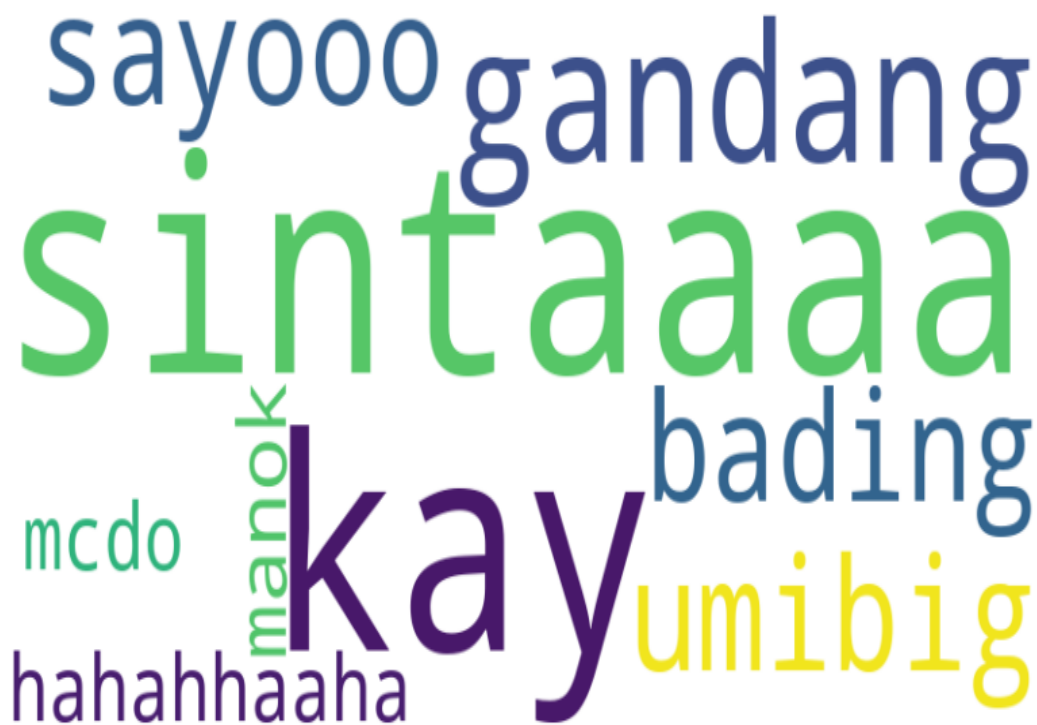
Sentiment Analysis

Sentiment Distribution

The sentiment analysis reveals a predominantly negative tone in the comments, which correlates with the high level of cyberbullying content detected. This suggests a potentially toxic communication environment that requires intervention.

Word Frequency Analysis

Most Common Words in All Comments

sayooo gandang
sintaaaa
mcdo manok kay bading
hahahhaaha umibig

The word cloud highlights the most frequent terms across all comments. Terms such as sintaaaa, kay, gandang appear frequently in the dataset, providing insight into common discussion topics.

Detailed Analysis Results

Comment	Classification	Confidence
sintaaaa	Normal	52.24%
o kay gandang umibig sayooo	Normal	96.76%
bading	Cyberbullying	97.93%
HAHAHHAAHA	Cyberbullying	97.95%
ang sarap ng manok ng mcdo	Cyberbullying	96.77%

Recommendations

High cyberbullying content detected (60.0%) 1. **Immediate Action Required:** The high percentage of cyberbullying content indicates a serious problem that requires prompt intervention. 2. **Content Moderation:** Implement stricter comment moderation measures, including: • Pre-approval of comments before they appear publicly • Automatic filtering of harmful language • Temporary comment restrictions until the situation improves 3. **Community Guidelines:** Establish or reinforce clear community guidelines that explicitly prohibit harassment, hate speech, and cyberbullying. 4. **Educational Outreach:** Consider educational initiatives to raise awareness about the impact of cyberbullying and promote positive online interaction. 5. **Regular Monitoring:** Continue to monitor the comment section closely and conduct follow-up analyses to track improvement. The predominantly negative sentiment coupled with cyberbullying content suggests an environment where negativity may be reinforcing harmful behavior.

Technical Notes

• This report was generated using an AI-powered cyberbullying detection system • Classification confidence is based on the prediction model's probability score • Word clouds exclude common stop words in both English and Tagalog • Sentiment analysis utilizes a specialized model designed for social media content • The system may not catch all instances of cyberbullying, particularly those using coded language or novel terms

For additional support or questions about this report, please contact the system administrator.

Report generated on April 10, 2025 at 12:00 AM