Comparative Summary of Existing Cyberbullying Detection Systems

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Features | Our System | Ref 1 | Ref 2 | Ref 3 | Ref 4 | Ref 5 |
| Data | | | | | | |
| Facebook | 250 posts |  |  |  |  | 4500 posts |
| Twitter | 150  tweets |  |  |  | 2477 words |  |
| YouTube | 500 comments |  | 4500 comments |  |  | 1500 comments |
| MySpace |  | 381,000 posts |  |  |  |  |
| Formspring.me |  |  |  | 3915 posts |  |  |
| Language | | | | | | |
| English | \* | \* | \* | \* | \* | \* |
| Tagalog | \* |  |  |  |  | \* |
| Dutch |  |  |  |  |  |  |
| Approach | | | | | | |
| Data Model | Unigram | Profane words,  Second pronoun,  Other pronoun | Orthony Lexicon for Negative Affect,  Profane words,  POS Bigrams | NUM and NORM dataset | Positive and Negative Sentiment | Bag of Words (BoW),  Profane word,  Word shape |
| Data Values | Frequency Count | TFIDF | TFIDF | Weighted TFIDF | TFIDF | TFIDF |
| Cyberbullying Classification | Cyberbullying,  Non-cyberbullying,  Ambiguous Cyberbullying | Bullying or Non-Bullying | Sensitive or Nonsensitive,  Race and Culture,  Sexuality,  Intelligence,  Physical Attributes | Cyberbullying,  non-cyberbullying  100, 200, 300, 400, and 500 | Bullying and Non Bullying | Bully,  Victim,  Assistants of the Bully,  Reinforcers,  Outsiders,  Defender,  N/A, |
| Algorithm | | | | | | |
| SVM | \* | \* | \* | \* | \* | \* |
| Naive Bayes |  |  | \* |  | \* | \* |
| Instance Based Algorithm |  |  |  | \* |  |  |
| Decision Trees (J48) |  |  | \* | \* |  | \* |
| JRip |  |  | \* | \* |  |  |
| K-nearest neighbor |  |  |  |  | \* |  |
| Evaluation | | | | | | |
| Accuracy  (0 - 100%) | 57.9% |  | 66.70% (SVM)  63% (NB)  61% (DT)  63% (JRip) | 81.7% (J48) | 91.31% (SVM)  87.65% (NB)  88.87% (KNN) | 59.7% |
| F-Measure | 54% | 23% |  |  |  | 57.5% (SVM)  52.4% (NB)  45.8% (DT) |
| Kappa Statistic |  |  | 65% (SVM)  44%(NB)  45%(DT)  50% (JRip) |  |  | 42.3% (SVM)  34.53% (NB)  22.54% (DT) |
| Precision | 47% |  |  |  |  | 59.7% (SVM)  53.2%(NB)  43.8%(DT) |
| Recall | 56% |  |  |  |  | 60.6% (SVM)  54.9% (NB)  50.6% (DT) |

Reference #1: Cyberbullying Detection: A Step Toward a Safer Internet Yard by Dadvar and De Jong (2012)

Reference #2: Modeling the Detection of Textual Cyberbullying by Dinakar, Reichart, and Lieberman (2011)

Reference #3: Using Machine Learning to Detect Cyberbullying by Reynolds, Kontostathis, and Edwards (2011)

Reference #4: Automatic Monitoring and Prevention of Cyberbullying by Sugandhi, Pande, Agrawalm and Bhagat (2016)

Reference #5: Automated Role Detection in Cyberbullying Incidents by Cheng and Ng (2016)