**Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network**

**Methodologies**

**Pre-Processing**: Normalization like: removing unwanted strings and correcting words.

**Features**: N-gram , TF-IDF , occurrence of pronouns , Skip-gram

**Classifiers**: Support Vector Machine (SVM) , Logistic Regression

**Results**

70.0% Precision and 77.65% Acc. using SVM

64.0% Precision and 73.76% Acc. using logistic regression

**Drawbacks:**

**N-gram:** the size tends to be large, and the performance of queries especially, long ones tends to be bad [18, 9]. These drawbacks stem from the method of extracting terms, that is, the 1-sliding technique.

**TF-IDF:**

– It computes document similarity directly in the word-count space, which may be slow for large vocabularies.

– It assumes that the counts of different words provide independent evidence of similarity.

– It makes no use of semantic similarities between words.

**Skip-gram:**

The disadvantage of skip-gram modelling is the sheer size of the training model that can be produced. This can lead to a large increase in processing time that should be leveraged against the time taken to extend the size of the training corpus.

**Logistic regression:**

logistic regression cannot predict continuous outcomes. For example, logistic regression could not be used to determine how high an influenza patient's fever will rise, because the scale of measurement -- temperature -- is continuous.

**SVM:**

the most serious problem with SVMs is the high algorithmic complexity and extensive memory requirements of the required quadratic programming in large-scale tasks.

**Sentiment Informed Cyberbullying Detection in Social Media**

-Their framework is called SICD

**Methodologies**

**Pre-Processing**: Stopwords removal and stemming.

**Features**: TF-IDF

**Classifiers**: KNN

**Results**

**F1**: 0.6105 **AUC**: 0.7539 using SICD

**Drawbacks:**

**Knn:**

The main disadvantage of this approach is that the algorithm must compute the distance and sort all the training data at each prediction, which can be slow if there are a large number of training examples. Another disadvantage of this approach is that the algorithm does not learn anything from the training data.

**Sarcastic sentiment detection in tweets streamed in real time**

In this paper the author discussed how it's hard for humans to detect sarcastic comments and the amount

of data is huge for manual detection. The author proposed a framework that recieve tweets

and process them to identify sarcastic comments. The hadoop-based framework allows user to

get and save tweets and process them. They used Twitter streaming API(twitter4j) to retieve

tweets. They used sentiment analysis to classify the tweets as negative, positive or neutral.

The tweets pass through MapReduce functions for sentiemnt classification.

they used parts-of-speach(POS) for deviding sentences into words and assign appropriate information

and they used parsing to analyze grammatical structure. Finally, they used 6 algorithms and achieved

0.97 precision when the combined 3 of them. They are: Parsed based lexicon generation algorithm(PBLGA),

Interjection word start(IWS),Postive sentiment with antonym pair(PSWAP).

**Cyberbullying Detection A Step Toward a Safer Internet Yard**

**Summary**

They used SVM model in WEKA(data mining software) for classification and they used Mypes service that allows the aggregation of users’ profiles so they can link users on multi social media platforms. The authors talked about how serious cyberbullying is, they hypothesized that the use of user’s information is essential for more precise detection like: age and gender. They calculated the ratio of profane words and the ratio of pronouns in each post then used as features along with TFIDF.

**Methodologies**

**Classifiers**: SVM

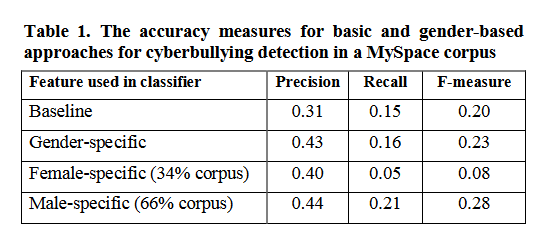
**Features**: TFIDF, Age and Gender

**Tools**: Mypes and WEKA

**Dataset**

The data-set is provided by Fundacion Barcelona Media collected from MySpace it’s size is 381,000 posts, the ground truth data-set is 2,200 labeled by three students.

**Results**



**Drawbacks**

The ground truth data-set is to small to training the classifier.

They didn’t use sentiment analysis

**Online Social Network Bullying Detection Using Intelligence Techniques**

**Summary**

They purposed a system that help governments to detect cyberbullying and take actions to prevent it. Their system uses genetic operations like crossover and mutation for optimizing the parameters before sending it to the classifier.

**Methodologies**

**Pre-Processing**: Normalization like: removing stop

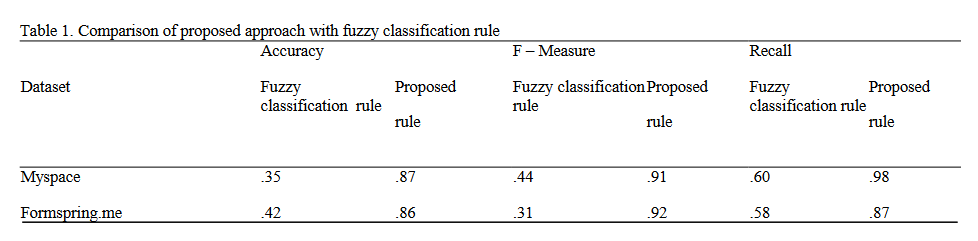
words, extra characters and hyperlinks.

**Features**: Noun, Adjective and Pronoun from the text and statistics on occurrence of word (frequency) in the text.

**Processing:** FuzGen learing algorithm.

**Classifiers**: Naive classifier technique.

**Results**

****

**Dataset**

They are available from the workshop on Content Analysis for the Web 2.0, it contains data from Myspace and Formspring.me

**Drawbacks**

**Fuzzy logic rule**: They are not robust at all. If we try to imitate the way humans reason, the minmax rule is definitely not the way.

**Stop word removal**: the stop words are important for sentiment analysis.

# **Machine learning and semantic analysis of in-game chat for cyber bullying**

**Summary**

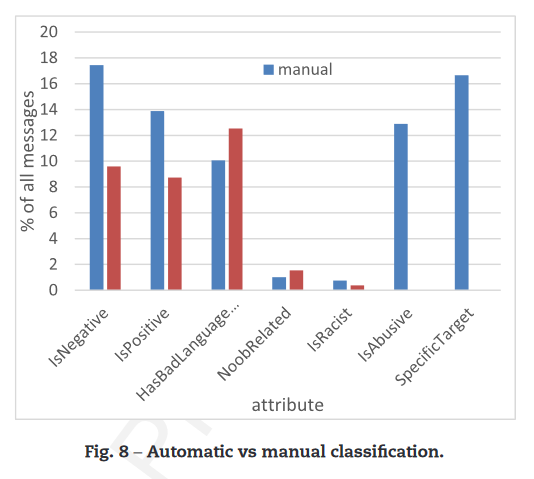
The authors discussed in this research the serious nature of cyberbullying and how to detect it. In this paper they used War Of Tanks game chat messages to help in the detection of cyberbullying. They said “The scoring schema presented here is in no way meant to be a finished product, rather it was a useful measure to help assess the utility of the data gathered and the tools employed in this paper.”[1]. They manually classified 5,000 messages into multiple categories: isNegative, isPositive, isAbusive, Bad Language, Racist and Noob related, then they made a CS(Cyberbullying Score) for each message to determine it’s level. Finally they compared it with the simple naive automatic classification.

**Methodologies**

**Features**: Sentiment analysis (Microsoft Azure and Twinword sentiment analysis API)

**Classifiers**: simple naive automatic classification

**Results**



**Dataset**

They collected the data using WOT(World Of Tanks) game api which reached 26,000 messages which 5,000 messages were manually classified to compare it to the automatic classifier

**Drawbacks**

They labeled data-set is too small and they used a simple classifier which can’t help in cyberbullying detection.

**Abusive Language Detection in Online User Content**

**Summary**

In this paper the authors used a framework called Vowpal Wabbit for classification. They developed a supervised classification methodology with NLP features that outperform a deep learning approachs.

**Methodologies**

**Pre-Processing**: Text Normalization

**Features**: N-grams, Linguistic, Syntactic and Distributional Semantics

**Classifiers**: Vowpal Wabbit (Framework): is a fast, parallel machine learning framework that was developed for distributed computing by Yahoo! Research.

**Results**

**All Features** **F-Score of news dataset**: 0.817

**F-Score of Finance dataset**: 0.795

**Dataset**

Data is sampled from comments posted on Yahoo! Finance and News during the period between October 2012 and January 2014

**Drawbacks**

use of external framework for classification can be slow, inefficient and unreliable.

**Cyberbullying Detection and Classification Using Information Retrieval Algorithm**

**Summary**

The authors used Lavenshtein distance algorithm that detects the cyberbully words present in the conversation and displays. And then they used naive bayes for classification.

**Methodologies**

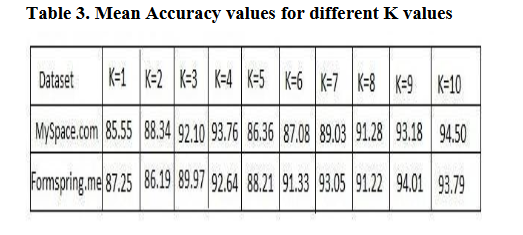
**Pre-Processing**: removing stop words, extra characters and hyperlinks.

**Features**: Noun, Adjective and Pronoun from the text and statistics on occurrence of word (frequency) in the text.

**Processing**: Lavenshtein distance algorithm.

**Classifiers**: Naïve Bayes classifier.

**Results**



**Dataset**

They are available from the workshop on Content Analysis for the Web 2.0, it contains data from Myspace and Formspring.me. 500 post from Formspring.me and 600 posts from Myspace.

**Drawbacks**

The size of data-set is small for cyberbullying classification. Also, naive bayes is a simple classifier that can be unreliable sometimes.