INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078

(Volume 8, Issue 1 - V8I1-1311)
Available online at: https://www.ijariit.com

A study of cyberbullying detection using Machine Learning techniques

Aishwary Suhas Shivarkar
shivarkarsid@gmail.com
Alamuri Ratnamala Institute of Engineering and Technology, Shahapur, Maharashtra

Vaishali Bagade
vaishali.bagade2@gmail.com
Alamuri Ratnamala Institute of Engineering and Technology, Shahapur, Maharashtra

ABSTRACT

With the widespread use of online social networks and their popularity, social networking platforms have provided us with innumerable chances not previously available, and their benefits are evident. Regardless of blessings, people can be embarrassed, ridiculed, tormented, and compelled to utilize the resource of anonymous users, strangers, or peers. In this study, we offered a cyberbullying detection method to build abilities from Twitter content material by employing a pointwise mutual facts technique. We superior a supervised tool learning solutions for cyberbullying detection and multi-beauty classification of its severity in Twitter based on the one's talents. We achieved Embedding, Sentiment, and Lexicon characteristics, as well as PMI-semantic orientation, throughout the test. The extraction of features was conducted utilizing.

Keywords: *A Cyberbullying Detection Method, Machine Learning, The Naive Bayes, KNN, Decision Tree, Random Forest, and Support Vector Machine Techniques*

1. INTRODUCTION

In this research, we use a pointwise mutual data technique to present a cyberbullying detection method for generating functions from Twitter content (tweets). Based on these features, we created a supervised device learning response in Twitter for cyberbullying detection and multi-elegance severity classification. We used Embedding, Sentiment, and Lexicon capabilities in combination with PMI-semantic orientation. The Naive Bayes, KNN, Decision Tree, Random Forest, and Support Vector Machine techniques were used to implement the extracted capabilities. In this newsletter, we will first present a quick background of the primary topics of our work. [2] defines related artworks within the nation of the artwork associated with the degree of cyberbullying. The third section

1.1 Online social network (OSN)

In this research, we present a cyberbullying detection system based on a pointwise mutual data technique for generating functions from Twitter content (tweets). Based on these characteristics, we developed a supervised device that learns Twitter answers for cyberbullying detection and multi-elegance severity categorization. We integrated Embedding, Sentiment, and Lexicon capabilities in combination with PMI-semantic orientation. To implement extracted capabilities, algorithms such as Naive Bayes, KNN, Decision Tree, Random Forest, and Support Vector Machine were employed. In this newsletter, we will first present a brief background of the primary topics of study that will be the focus of our investigation. In [2,] we define related artworks within the country of the artwork associated with the severity category of cyberbullying. [3] examples

1.2 Adverse consequences

Although the internet and social media have many good effects on society, they also have numerous negative consequences. This includes unwanted sexual publicity, cybercrime, and cyberbullying. Sexual publicity arises when criminals impersonate victims in internet advertisements and falsely claim that their victims are interested in sex[7]. Intellectual property thefts, spam, phishing cyberbullying, and various sorts of social engineering are all examples of cybercrime[8]. Because OSNS is intended to let users exchange information such as links, messages, videos, and photos[9], hackers have used it in an innovative way to perpetrate new sorts of cybercrime[10]. Bullying, including cyberbullying, has been acknowledged as a danger to public health and well-being. Although the usage of the internet and social media has increased,

1.3 Cyberbullying

Spreading false rumors based on race, gender, disability, faith, and sexuality; humiliating a person; social isolation; stalking; threatening a person online, and releasing private information about a person that was supplied in confidence are all examples of cyberbullying[21]. According to a major advocacy group in the United States[22], bullying may take various forms, including racism and sexual orientation. According to a Pew Research Center survey, there are numerous sorts of online harassment encountered by internet users. The first session includes significantly less intensive studies: swearing and humiliation because individuals who see or like it frequently claim to forget about it. However, the second sort of harassment, which is directed towards a smaller population of internet users, contains

2. RELATED WORK-CLASSIFICATION OF CYBERBULLYING

The severity level of cyberbullying has been studied[25] in OSNS using a completely language-based technique. On Form-spring. me, information was obtained from 18,554 customers. The website www.noswearing.com was used to construct a list of insult and curse words, yielding a total of 296 idioms. Each word on the list was assigned a severity rating by Reynolds[25] and his colleagues. The ranges have been 100 (for example, butt and dumb), 200 (for example, garbage and prick), 300 (for example, asshole and douchebag), 400 (for example, fuckass and pussy), and 500. (e.g. buttfucker, cuntass). They observed that while 100-degree terms are used more frequently than 500-degree phrases, they are the most predictive of cyberbullying [25].

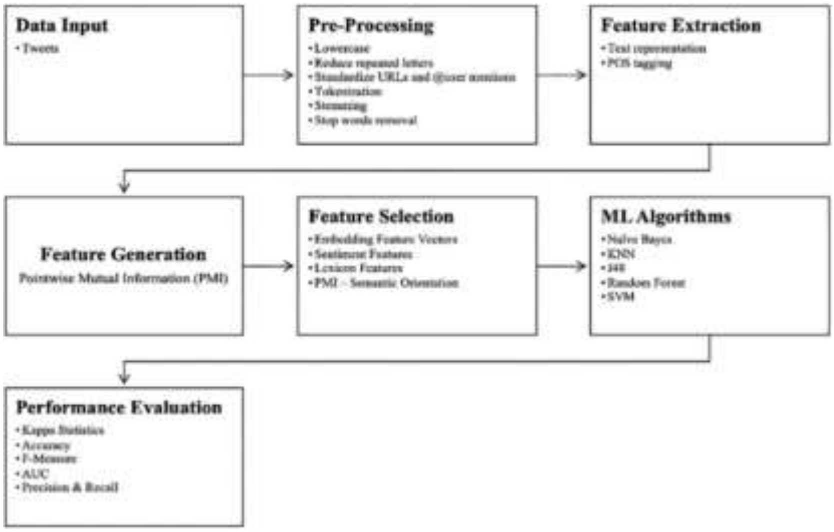
3. MATERIALS AND METHODS FOR STUDY FOR DATA ACCESSIBILITY

The major goal of putting a green cyberbullying detection device in an OSN is to prevent or, at the very least, minimize incidences of harassment and bullying[27]. These structures might be used as tools to help track online environments. Furthermore, cyberbullying detection may be more effective in assisting and recommending the victim, as well as tracking and tracking the bully. Before choosing an OSN to monitor, the following key functions must be taken into account: popularity (number of active users) and data accessibility. A significant task in cyberbullying research is to have access to relevant data that is critical to increasing fashions that reflect cyberbullying[5]. Facebook is presently the most important online social network, with over 1000000000 active users[29].

4. METHODOLOGY AND EXPERIMENTS

In summary, this step describes the study methodology we used to assess the severity of cyberbullying using the dataset described in phase 3. Figure 1 depicts all phases of our proposed framework, which are explained in the sections that follow.

Fig 1



4.1 Data collection step

For harassment research, we utilized [4]'s well-annotated corpora. The dataset has previously been classified into five types of harassing content: I sexual, II racial, III appearance-related, IV IQ, and V political (Table 1).

Table:-1 Harassment Information

Category	No	Yes	Annotated Tweets
Sexual	3616	229	3845
Racial	4273	700	4973
Intelligence	4049	810	4859
Appearance	4146	676	4822
Political	4916	698	5659
Total	21045	3113	24158

Table 1 displays the binary class of the aforementioned people in the dataset [4]. To carry out our test on severity evaluation at the harassment information set, we classified the annotated cyberbullied tweets into four levels: low, medium, high, and non-cyber bullying. We then classified sexual and appearance-related tweets as having a high level of cyberbullying severity, political and racial tweets as having a medium level of cyberbullying severity, intelligence tweets as having a low level of cyberbullying severity, and all tweets classified as 'non-cyber bullying' in each class as having a low level of cyberbullying severity. As a consequence, a dataset with verified characteristics was created (Table 2).

Table 2 Annotated Tweets

Category	Annotated Tweets
High	905
Medium	1398
Low	810
Non-Cyberbullying	21045
Combined Tota	24158

We linked sexual and look tweets together in this study mostly based exclusively on their related profane phrases used in tweets and lexicon to establish their class offered by [4], which is also our intuition for combining sexual and look-associated tweets into high-level severity tweets. Furthermore, according to the Pew Research Center, sexual harassment is a more severe kind of cyberbullying[23]. Similarly, we rate intelligence-related cyberbullied tweets as low intensity since we believe the collection of lexicons offered by [4] refers to humiliation and name-calling in this context. Furthermore, in the context of cyberbullying, the Pew Research Centre regarded name-calling and/or humiliation as significantly less extreme. It is an irritant layer, thus it is fairly unusual for individuals who see or observe it to see it.

4.2 Pre-processing step

The collected data is pre-processed before being assigned severity levels. Tweets have been updated to minimize cases to prevent sparsity difficulties, as well as to reduce repeating letters, standardized URLs, and @usermention to reduce noise inside the tweets. Tokenization was accomplished using a Twitter-specific tokenizer built solely on the CMU Tweet NLP package, and top words with a frequency of at least 10 were kept. Tokenization is the process of separating a textual content corpus into the most commonplace words, phrases, or other important components, which are then referred to as tokens. Finally, before function extraction, stop phrases and stemming procedures were finished. Stop phrases are insignificant phrases that exist in a record but aren't usually particular.

4.3 Feature extraction step

A bag of tweets has been used to represent all tweets, which is one of the most appropriate and fastest approaches. The textual content is represented as a group of phrases in this manner, and each phrase is regarded as an individual characteristic. We employed part-of-speech (POS) tagging using a Twitter-specific tagger based exclusively on the CMU Tweet NLP package for phrase experience disambiguation. The POS tagger assigns a part-of-speech tag to each phrase of the given textual material, which is represented by tuples (tag), such as nouns, verbs, adjectives, and so on.

4.4 Feature generation step

We used report-stage classification to determine the semantic orientation of each phrase in the corpus. The POS tags were utilized to extract words from the degree type of the record. After recovering the words from the dataset, their semantic orientation in terms of both cyberbullying and non-cyber bullying was determined. The idea of pointwise mutual information (PMI)[30] was used to compute the semantic orientation for each sentence in a corpus of tweets to achieve this goal. The PMI between words, word1, and word2 are:

$$PMI(word_1, word_2) = \log_2 \left[\frac{p(word_1 \& word_2)}{p(word_1)p(word_2)} \right]$$

The rating is calculated by dividing the PMI of the target phrase with a cyberbullying brilliance by the PMI of the target phrase without a cyberbullying splendor. This procedure proved to be well-suited for domain-specific lexicon production with PMI rating, therefore we built our domain-specific lexicon with PMI semantic orientation for each phrase and word using Turney's method. The semantic Orientation of a word is calculated as follows:

$$SO(phrase) = PMI(phrase, "non - cyberbullying") - PMI(phrase, "cyberbullying")$$

Turney's method is divided into three stages and is based on a professional vocabulary. First, the words are taken from the dataset. Second, emotion polarity suggested the use of each extracted word's PMI, which measures the statistical dependency between phrases. Finally, a dataset's sentiment polarity is computed by averaging the polarity of all terms in the dataset. Turney's PMI methodology no longer relies on hard-coded semantic criteria, allowing clients to apply the method to new contexts with ease.

4.5 Feature engineering and selection step

The process of producing or extracting capabilities from raw data or a corpus is known as feature engineering. The development of new functions based on current functions is referred to as function engineering. The outstanding functions given into the device when analyzing a set of rules have an immediate impact on the final results of the version prediction, not the enormous number of functions.

One of the most popular methods for enhancing cyberbullying detection is to execute function engineering, and the most common functions that increase the fine of cyberbullying detection classifier overall performance are the textual, social, person, sentiment, and phrase embeddings. We sought to reconstruct social and consumer functions because they were no longer available in the dataset supplied by [4].

4.6 Dealing with class imbalance data

Class imbalance is the state in which the number of times from one magnificence is much more than the number of times from any other magnificence. When the number of times spent on each lesson is about equal, most machine learning algorithms perform well. However, the information in many real-world programs and non-artificial datasets is imbalanced; that is, a key magnificence (also known as the minority magnificence) may have far fewer samples than the alternative magnificence (normally called the bulk magnificence). In such cases, well-known classifiers are sometimes overwhelmed by enormous grandeur and lose sight of the small dispensed moments. It generally produces a biassed classifier with better predictive accuracy over majority lessons but lower predictive accuracy over minority lessons.

Table 3 Class Distribution

Classification	Class Distribution
High	4%
Medium	6%
Low	3%
Non-Cyberbullying	87%

5. MACHINE LEARNING ALGORITHMS SELECTION STEP

The most crucial aspect of the textual content type procedure is selecting the optimal classifier. We cannot successfully select the best version for a textual content type implementation unless we have a thorough conceptual knowledge of each approach. The function derived from the tweets was used to create a model for detecting cyberbullying actions and their severity. We looked at numerous machines that study Algorithms to discover the best classifier, including Nave Bayes, Support Vector Machines (SVM), Decision Trees, Random Forests, and K-Nearest Neighbors (KNN).

5.1 Naïve Bayes

Naive Bayes is recognized as one of the most ecologically friendly and powerful inductive learning algorithms in the field of device mastering, and it has been used as a powerful classifier in different social media studies[13]. Since the 1950s, the Nave Bayes type for textual content has been widely utilized in record categorization assignments, and it may categorize any sort of data, including text, community functions, phrases, and so on. This is a generative model, which refers to the process of creating a dataset fully based on a probabilistic model. It is feasible to generate fresh statistics that are comparable to the facts on which the model is being trained by sampling from this version[22]. We used Nave's maximum primary model in our analysis.

5.2 K-Nearest Neighbours (KNN)

The K-Nearest Neighbors (KNN) approach is one of the most useful instance-based learning algorithms for multi-elegance scenarios. It is a supervised learning set of rules. In this set of criteria, distance is used to identify a wholly unique design from its neighbor. As a consequence, reveals the numerous schooling set to the K-nearby friends and sets an item into the elegance that is most popular amongst its okay nearest pals. KNN is a non-parametric lazy mastering set of rules that makes no assumptions about the distribution of underlying data.

5.3 Decision trees (J48)

A decision tree is a well-known form of algorithm in device learning, as well as one of the most commonly utilized inductive learning methodologies. It can handle school data with missing values and any continuous or discrete variable. Decision bushes are constructed utilizing classified education statistics and the idea of facts entropy. Their resilience to noisy data and capacity to explore disjunctive assertions appear to make them suitable for textual content.

5.4 Random Forest

Random forest (RF) is a set of rules used to tackle classification and regression problems. RF creates a large number of suitable wood classifiers based on a random subset of record samples and features. The categorization of the new pattern is performed by the use of majority balloting of option timber. The fundamental benefit of RF is that it works well on huge datasets, is a potent strategy for predicting missing records, and produces correct results even when a significant amount of the information is absent.

5.5 Support Vector Machine (SVM)

Based on sample popularity, SVM is a set of criteria for classifying each linear and non-linear fact. The primary SVM principle is to choose separators that may precisely differentiate the excellent lessons inside the search area. Guide vectors are information factors that partition one or more hyperplanes by using crucial education tuples. In some cases, a nonlinear SVM classifier is used when all of the information variables cannot be separated using a straight line. The kernel feature is frequently used by the nonlinear characteristic; well-known kernels include linear kernels, polynomial kernels, RBF kernels, and sigmoid kernels. When the range of functions is viewed as a whole, the Radial foundation characteristic (RBF) kernel typically outperforms others.

6. PERFORMANCE EVALUATION STEP

6.1 Candidate metrics

Performance metrics typically compare individual components of overall category performance and no longer necessarily provide comparable values. Understanding how a version operates is critical for any category algorithm. The underlying mechanics of various assessment metrics might also differ, and understanding what each of these metrics reflects and what sort of information they're seeking to portray is crucial for comparison. There are various methods for evaluating a classifier's overall performance, including don't forget, precision, accuracy, F-degree, micro-macro averaged precision, and don't forget [28]. These metrics are fully based on the "Confusion Matrix," which includes genuine excellent (TP): the number of instances accurately classified as excellent.

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where Pr(a) represents the located settlement and Pr(e) represents the hazardous settlement.

It denotes how much of an impact a better classifier has on the overall performance of a classifier that guesses at random based on the frequency of each component.

The Kappa statistic is used to assess the degree of agreement between expected and discovered categorizations in a dataset while simultaneously accounting for risk-adjusted agreement. It's just a normalized model of the proportion of accurate classifications (type accuracy), with normalization applied to the overall performance of a random classifier. At a glance, it shows how much the classifier improves over a random one.

Kappa is typically equal to or less than 1.

6.2 Chosen metrics

We were presented with a multi-magnitude problem with skewed data in our inquiry. Furthermore, because our category responsibilities are sensitive for all classes, we used the kappa statistic as our primary metric rather than the weighted f-measure highlighted while the concept of a typical place marginal distribution across raters inside a test isn't tenable, the use of Cohen's kappa is more appropriate. We also document classifier average accuracy, precision, remember genuine fantastic fee, and false-wonderful fee as reference measures. Furthermore, for comparison, we intended to test our suggested cyberbullying detection approach in a binary context utilizing the

7. DISCUSSION

The gift has progressed by identifying flaws in the current cyberbullying detection technique. In this study, we established a thorough methodology for measuring the severity of cyberbullying on Twitter, which is mostly based on earlier research from several domains. To do this, we created a device learning multi-classifier capable of categorizing cyberbullying intensity into multiple levels. We put our proposed approach for assessing cyberbullying severity to the test using a publicly accessible harassment dataset. We created a system that provides semantic orientation for each phrase in a dataset, which we then used as an input function in a variety of well-known functions including phrase embedding and sentiment analysis, as well as several phrase-stage lexicons.

8. CONCLUSION

The internet and social media have apparent societal benefits; but, their broad use may have substantial detrimental consequences. Inappropriate sexual exposure, cybercrime, and cyberbullying are all examples of this. We created a version for detecting cyberbullying activity and its severity on Twitter. Feature generation with PMI at the pre-processing level has been established as the green approach to dealing with a category imbalance in a binary and multi-elegance type, where misclassification for minority elegance (es) has a better value in terms of its effect on the detection version's reliability. The advanced version is a feature-based complete version that expands a device mastering classifier for recognizing tweets as cyberbullying or non-cyber bullying by using characteristics from tweet contents.

9. LIMITATIONS

We were unable to conduct an in-depth assessment of user behavior since the dataset we used for this study no longer supplied any information (e.g., time of the tweet, likes, followers, etc.) other than the text (tweets). Furthermore, we may have wished to end the meta-evaluation with the findings of cyberbullying severity, but this was not feasible since the study we evaluated no longer supplied essential information that would have allowed this sort of review. Despite these limitations, we feel that the current study adds to the body of knowledge by presenting a scientific framework for categorizing cyberbullying severity into various tiers to construct a device learning multi-classifier rather than a single classifier.

10. FUTURE STUDY

Online harassment and cyberbullying must become a big issue that has a large detrimental influence on people's lives. The anti-harassment coverage and criteria given by leveraging social structures, as well as the ability to report, block, or record the bully, are all excellent advances toward a more secure online society, but they are no longer enough. Popular social media sites such as Twitter, Facebook, Instagram, and others get a large volume of such highlighted information every day; therefore, reviewing key pronounced content and customers might be time-consuming and no longer possible or efficient. In such cases, it would almost certainly be useful to develop automated, data-driven systems for comparing and contrasting.

11. REFERENCES

[1] Penni J. The future of online social networks (OSN): A measurement analysis using social media tools and application. Telemat Inform. 2017;34: 498–517.
a. View Article
b. Google Scholar

[2] Lauw H, Shafer JC, Agrawal R, Ntoulas A. Homophily in the Digital World: A LiveJournal Case Study. IEEE Internet Compute. 2010;14: 15–23.
a. View Article
b. Google Scholar

[3] Hee CV, Jacobs G, Emmery C, Desmet B, Lefever E, Verhoeven B, et al. Automatic detection of cyberbullying in a social media text. PLOS ONE. 2018;13: e0203794. pmid:30296299
a. View Article
b. PubMed/NCBI
c. Google Scholar

[4] Hossein Mardi H, Shaosong Li, Zhili Yang, Qin Lv, Rafiq RI, Han R, et al. A Comparison of Common Users across Instagram and Ask. FM to Better Understand Cyberbullying. 2014 IEEE Fourth International Conference on Big Data and Cloud Computing. 2014. pp. 355–362.

[5] Citron DK. Addressing Cyber Harassment: An Overview of Hate Crimes in Cyberspace. the Internet. 2015;6: 12.
a. View Article
b. Google Scholar

[6] Wall D. What are Cybercrimes? Crim Justice Matters. 2004;58: 20–21.
a. View Article
b. Google Scholar

[7] Abu-Nimeh S, Chen T, Alzubi O. Malicious and Spam Posts in Online Social Networks. Computer. 2011;44: 23–28.
a. View Article
b. Google Scholar

[8] Ferrara P, Ianniello F, Villani A, Corsello G. Cyberbullying a modern form of bullying: let’s talk about this health and social problem. Ital J Pediatr. 2018;44. pmid:29343285
a. View Article
b. PubMed/NCBI
c. Google Scholar

[9] Volk AA, Veenstra R, Espelage DL. So you want to study bullying? Recommendations to enhance the validity, transparency, and compatibility of bullying research. Aggress Violent Behav. 2017;36: 34–43.
a. View Article
b. Google Scholar

[10] Anderson T, Sturm B. Cyberbullying: From Playground to Computer. Young Adult Libr Serv. 2007;5: 24.
a. View Article
b. Google Scholar

[11] Myers C-A, Cowie H. Cyberbullying across the Lifespan of Education: Issues and Interventions from School to University. Int J Environ Res Public Health. 2019;16. pmid:30987398
a. View Article
b. PubMed/NCBI
c. Google Scholar

[12] Zuckerberg M. One Billion People on Facebook. In: One Billion People on Facebook [Internet]. 2012 [cited 20 Oct 2019]. https://newsroom.fb.com/news/2012/10/one-billion-people-on-facebook/.

[13] Kurka DB, Godoy A, Von Zuben FJ. Online Social Network Analysis: A Survey of Research Applications in Computer Science. ArXiv150405655 Phys. 2015 [cited 24 Aug 2019]. http://arxiv.org/abs/1504.05655.

[14] Dinakar K, Reichart R, Lieberman H. Modeling the Detection of Textual Cyberbullying. 2011; 7.

[15] Ashktorab Z. A Study of Cyberbullying Detection and Mitigation on Instagram. CSCW Companion. 2016.
a. View Article
b. Google Scholar

[16] Xu J-M, Jun K-S, Zhu X, Bellmore A. Learning from Bullying Traces in Social Media. Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Montréal, Canada: Association for Computational Linguistics; 2012. pp. 656–666. https://www.aclweb.org/anthology/N12-1084.

[17] Zhao R, Zhou A, Mao K. Automatic Detection of Cyberbullying on Social Networks Based on Bullying Features. Proceedings of the 17th International Conference on Distributed Computing and Networking. New York, NY, USA: ACM; 2016. p. 43:1–43:6.

[18] Lovins JB. Development of a stemming algorithm. Mech Transl Comp Linguist. 1968;11: 22–31.
a. View Article
b. Google Scholar

[19] Nielsen FÅ. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. ArXiv11032903 Cs. 2011 [cited 17 Sep 2019]. http://arxiv.org/abs/1103.2903.

[20] Mohammad S, Kirichenko S, Zhu X. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International

Workshop on Semantic Evaluation (SemEval 2013). Atlanta, Georgia, USA: Association for Computational Linguistics; 2013. pp. 321–327. https://www.aclweb.org/anthology/S13-2053.
[21] Kiritchenko S, Zhu X, Mohammad SM. Sentiment Analysis of Short Informal Texts. J Artif Intell Res. 2014;50: 723–762.
a. View Article
b. Google Scholar
[22] Mohammad SM, Kirichenko S. Using Hashtags to Capture Fine Emotion Categories from Tweets. Compute Intell. 2013; 22.
a. View Article
b. Google Scholar
[23] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res. 2002;16: 321–357.
a. View Article
b. Google Scholar
[24] Alber M, Zimmer J, Dogan U, Kloft M. Distributed optimization of multi-class SVMs. PLOS ONE. 2017;12: e0178161. pmid:28570703
a. View Article
b. PubMed/NCBI
c. Google Scholar
[25] Kowsari K, Meimandi KJ, Heidarysafa M, Mendu S, Barnes LE, Brown DE. Text Classification Algorithms: A Survey. Information. 2019;10: 150.
a. View Article
b. Google Scholar
[26] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag. 2009;45: 427–437.
a. View Article
b. Google Scholar
[27] Huang J, Ling CX. Using AUC and Accuracy in Evaluating Learning Algorithms. IEEE Trans Knowl Data Eng. 2005;17: 299–310.
a. View Article
b. Google Scholar
[28] McHugh M. Interrater reliability: The kappa statistic. Biochem Medica Časopis Hrvat Druš Med Biokem HDMB. 2012;22: 276–82. pmid:23092060
a. View Article
b. PubMed/NCBI
c. Google Scholar
[29] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33: 159–174. pmid:843571
a. View Article
b. PubMed/NCBI
c. Google Scholar
[30] Banerjee M, Capozzoli M, McSweeney L, Sinha D. Beyond kappa: A review of interrater agreement measures. Can J Stat. 1999;27: 3–23.
a. View Article
b. Google Scholar
[31] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA Data Mining Software: An Update. SIGKDD Explore Newsl. 2009;11: 10–18.
a. View Article, b Google Scholar

