

Predicting Cyberbullying Behaviors in Big Data

Saravanan.C^{1*}, Janani.K.P², Jeevasri.P³, Karthik.M⁴

¹Associate Professor, Department of Information Technology, Velalar College of Engineering and Technology, Tamilnadu, India

²Student, Department of Information Technology, Velalar College of Engineering and Technology, Tamilnadu, India

³Student, Department of Information Technology, Velalar College of Engineering and Technology, Tamilnadu, India

⁴Student, Department of Information Technology, Velalar College of Engineering and Technology, Tamilnadu, India

ABSTRACT: Prior to the innovation of statistics verbal exchange technologies (ICT), social interactions Advanced within small cultural boundaries which include geo spatial locations. The recent trends of communiqué Technology have significantly transcended the temporal and spatial boundaries of traditional Communications. Online human networks and wealthy human behavior-related records. However, the misuse of social technologies inclusive of social media (SM) platforms, has delivered a new shape of aggression and violence that happens exclusively online. A new way of demonstrating competitive conduct in SM web sites are highlighted on this paper. The motivations for the construction of prediction models to ght aggressive conduct in SM are also outlined. We comprehensively evaluate cyberbullying prediction models and pick out the main issues associated with the creation of cyberbullying prediction models in SM. This paper affords insights on the general procedure for cyberbullying detection and most significantly overviews the methodology. Though information collection and characteristic engineering procedure has been elaborated, yet most of the emphasis is on characteristic selection algorithms and then using various system studying algorithms for prediction of cyberbullying behaviors. Finally, the issues and challenges were highlighted as well, which present new research guidelines for researchersto explore.

KEYWORDS: Index term-Big data, cyberbullying, cybercrime, human aggressive behavior, gadget learning, onlineSocial network, social media, text classification.

I. INTRODUCTION

Machine or deep studying algorithms assist researchers understand huge information. Abundant records on human beings and their societies can be obtained in this big statistics era, but this acquisition was formerly impossible. One of the principal resources of human-related data is social media (SM).By applying machine learning algorithms to SM records we can make the most historical records to expect the future of a wide range of applications. Machine gaining knowledge of algorithms offer an opportunity to successfully are expecting and discover negative forms of human behavior, such as cyberbullying. Big records analysis can find hidden understanding via deep gaining knowledge of from raw statistics. Big information analytics has progressed numerous applications, and forecasting the destiny has even end up possible via the mixture of large facts and system learning algorithms

II. RELATED WORDS

A. RISE OF AGGRESSIVE BEHAVIOR ON SM

Prior to the innovation of verbal exchange technologies, social interaction developed inside small cultural boundaries, such as locations and families. The recent development communication technology fantastically transcends the temporal and spatial limitations of traditional communicate. In the previous few years, on line verbal exchange has shifted toward user-driven technologies, which includes SM websites, blogs, online digital communities, and on line sharing platforms.

- New sorts of aggression and violence emerge exclusively on-line.
- The dramatic growth in negative human conduct on SM, with high increments in aggressive behavior, gives a brand new challenge.
- The advent of Web 2.0 technologies, consisting of SM web sites that are often accessed through cellular devices, has completely transformed functionality on the side of users.

SM web sites have turn out to be dynamic social communicate web sites for hundreds of thousands of users worldwide. Data in the shape of ideas, opinions, preferences, views, and discussions are spread among customers rapidly via on-line social communication. The on line interactions of SM users generate a big volume of statistics that may be utilized to study human behavioral patterns. SM websites also offer a fantastic opportunity to investigate patterns of social interactions among populations at a scale this is a whole lot large than before.

B. MOTIVATIONAL FOR PREDICTION AGGRESSIVE BEHAVIOR ON SM WEBSITE

SM websites contain massive amounts of text and/or non-textual content and other data related to aggressive behavior. In this painting, a content evaluation of SM web sites is performed to predict competitive behavior. Such an analysis is constrained to textual OSN content material for predicting cyberbullying behavior. Given that cyberbullying may be effortlessly committed, it's miles considered a dangerous and fast-spreading competitive behavior. Bullies handiest require willingness and a laptop or mobile smartphone with Internet connection to perform misbehavior without confronting victims. The reputation and proliferation of SM websites have improved on line bullying activities. Cyberbullying in SM websites is rampant because of the structural traits of SM web sites. Cyberbullying in traditional platforms, which includes emails or cellphone text messages, is performed on a confined range of people. SM websites allow customers to create roles for organizing friendships and communicating with other customers regardless of geographic location, therefore expanding cyberbullying beyond bodily location. Anonymous customers may exist on SM websites, and this has to be a number one motive for expanded aggressive person behavior. Developing a powerful prediction model for predicting cyberbullying is therefore of realistic significance. With a lot of these considerations, this painting performs a content-primarily based evaluation for predicting textual cyberbullying on SM websites. The motivation of this evaluation is explained inside the following section.

C. WHY CONSTRUCTING CYBERBULLYING PREDICTION MODELS IS IMPORTANT

The motivations for carrying out this evaluation for predicting cyberbullying on SM websites are discussed as follows. Cyberbullying is a prime problem and has been documented as a serious country wide health problem due to the recent growth of on-line conversation and SM web sites. Research has proven that cyberbullying exerts negative consequences at the mental and physical health and academic performance of people. Studies have also proven that cyberbullying victims incur a high threat of suicidal

D. METHODOLOGY

This segment provides the method used on this work for a literature search. Two phases were employed to retrieve posted papers on cyberbullying prediction models the rst phase included looking for official educational databases and serps. The search engines and Instructional databases used for the retrieval of relevant papers have been as follows: Scopus, Clarivate Analytics' Web of Science, DBLP Computer Science Bibliography, ACM Digital Library, Science Direct, Springer Link, and IEEE Xplore. The main keywords used for the literature search had been coined in relation to social media as follows: cyberbullying, aggressive behavior, huge data, and cyberbullying models. The second section involved looking for literature through Qatar University's digital library. The articles retrieved from the search were scrutinized to ensure that the articles met the inclusion criteria. According to the inclusion criteria, for an article to be decided on for the survey, it must file an empirical look at describing the prediction of cyberbullying on SM sites. Otherwise, the article would be excluded within the selection. Many articles had been rejected based on titles. The abstract and conclusion sections had been examined to make certain that articles satisfied the screening criteria, and those that did no longer satisfy the criteria had been excluded from the survey.

III. PREDICTING CYBERBULLYING ON SOCIAL MEDIA IN THE BIG DATA ERA USING MACHINE LEARNING ALGORITHMS

Our world is currently inside the massive facts era because 2.5 quintillion bytes of facts are generated daily. Organizations continuously generate large-scale records. These huge-scale datasets are generated from special sources, such as the World Wide Web, social networks, and sensor networks. Big statistics have nine characteristics, namely, volume, variety, variability and complexity, velocity, veracity, value, validity, verdict, and visibility. For example, Flickr generates almost 3.6 TB of records, Google is assumed to process almost 20,000 TB of facts according to day, and the Internet gathers an estimated 1.8 PB of statistics daily.

IV. PRE PROCESSING

SM is an online platform that offers customers an opportunity to create a web community, share information, and change content. SM customers and the interaction among organizations, people, and merchandise are accountable for the large amount of statistics generated on SM platforms. SM platforms, together with Facebook, YouTube, blogs, Instagram,

Wikipedia, and Twitter, are of different types. The facts generated by way of SM outlets may be based or unstructured in form. SM analytics is the evaluation of structured and unstructured records generated by SM outlets. SM analytics can be in any of the subsequent forms: hyperlink prediction, community, content, social influence, dependent, and unstructured. SM is now in the big statistics era. For example, Facebook stores 260 billion photographs in over 20 PB of garage space, and up to at least one million photos are processed in keeping with second. YouTube receives one hundred hours of downloaded motion pictures in each minute. The most not unusual means of building cyberbullying prediction fashions is to apply a textual content classification method that entails the construction of gadget studying classifiers from labeled text instances. Another way is to apply a lexicon-based totally model that involves computing orientation for a report from the semantic orientation of phrases or phrases in the file. Generally, the lexicon in lexicon-primarily based fashions may be constructed manually (much like the methods used) or automatically with the aid of the use of seed words to enlarge the listing of words. However, cyberbullying prediction using the lexicon-based totally method is uncommon in literature. The primary reason is that the texts on SM web sites are written in an unstructured manner, thus making It difficult for the lexicon-based method to come across cyberbullying based totally simplest on lexicons. However, lexicons are used to extract functions that are frequently applied as inputs to gadget getting to know algorithms. For example, lexicon based processes, inclusive of the use of a profane-based dictionary to stumble on the wide variety of profane phrases in a post, are adopted as profane functions to system learning models. The key to effective cyberbullying prediction is to have a set of capabilities which might be extracted and engineered. Features and their combos are essential within the construction of powerful cyberbullying prediction model. Most research on cyberbullying prediction used gadget studying algorithms to construct cyberbullying prediction fashions. Machine gaining knowledge of-primarily based fashions exhibit first rate performance in cyberbullying prediction. Consequently, this work opinions the construction of cyberbullying prediction models based on device mastering. The system studying specializes in the development and alertness of pc algorithms that enhance with experience. The objective of gadget studying is to identify and dene the patterns and correlations between information. The significance of analyzing massive statistics lies in discovering hidden knowledge thru deep studying from raw information. Machine getting to know can be defined as the adoption of computational models to improve device overall performance by using predicting and describing meaningful styles in training records and the purchase of information from experience.

V. FEATURE EXTRACTION AND LABELING

From an education dataset. In the checking out stage, new information are fed into the version, and times are classified to a specified class found out at some point of the schooling stage. Then, classification performance is evaluated This section opinions the most not unusual strategies in the production of cyberbullying prediction fashions for SM websites primarily based on gadget getting to know. The evaluation covers records collection, Characteristic engineering, characteristic selection, and device getting to know algorithms.

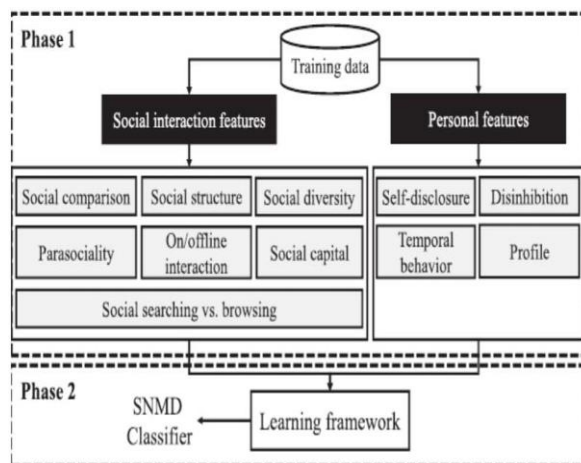


Fig.1 Extraction And Labeling

A. DATA COLLECTION

Data are critical additives of all system learning-primarily based prediction fashions. However, information (even ``Big Data'') are useless on their own until information or implications are extracted from them. Data extracted from SM web sites are used to pick out schooling and trying out datasets. Supervised prediction fashions purpose to offer computer techniques to beautify prediction performance in defined responsibilities at the basis of found instances (labeled records). Machine gaining knowledge of models for a certain challenge primarily goal to generalize; a successful version should

now not be constrained to examples in a training dataset only but should consist of unlabeled real facts. Data amount is inconsequential; what is vital is whether or no longer the extracted records represent activities on SM websites well. The main records collection strategies in preceding cyberbullying prediction studies on SM websites may be categorized into information extracted from SM websites by using either keywords, this is, words, phrases, or hash tags or by using person profiles. The problems in these information collection strategies and their results at the performance of gadget gaining knowledge of algorithms are highlighted inside the Data Collection section (related issues).

B. FEATURE ENGINEERING

Feature is a measurable property of an undertaking this is being observed. The main purpose of engineering feature vectors is to offer device getting to know algorithms with a fixed of learning vectors via which those algorithms learn the way to discriminate between one of a kind forms of classes Feature engineering is a key element at the back of the fulfillment and failure of maximum gadget gaining knowledge of fashions. The achievement and failure of prediction may be based on several elements. The most significant element is the functions used to teach the version. Most of the attempt in constructing cyberbullying

Prediction models the use of learning algorithms is dedicated to this project. In this context, the layout of the enter space (i.e., capabilities and their combinations which might be provided as an enter to the classier) is vital. Proposing a hard and fast of discriminative features, which are used as inputs to the gadget studying classier, is the main step in the direction of building an powerful classier in many applications. Feature sets can be created based on human-engineered observations, which depend on how features correlate with the occurrences of classes. Similarly, proposing a hard and fast of significant capabilities of cyberbullying engagement on SM websites is critical in growing powerful prediction models primarily based on system gaining knowledge of algorithms. State-of-the-art research has advanced features to enhance the performance of cyberbullying prediction. For example, a lexical syntactic characteristic has been proposed to deal with the prediction of offensive language; this technique is higher than conventional studying-primarily based processes in phrases of precision. Dadvar et al. examined gender information from profile statistics and developed a gender-based technique for cyberbullying prediction through the usage of datasets from my space as a basis. The gender characteristic became selected to improve the discrimination capability of a classier. Age and gender had been covered as capabilities in other research, but these features are restrained to the facts provided by customers in their on line profiles. Several researches focused on cyberbullying prediction primarily based on profane words as a feature. Similarly, a lexicon of profane phrases was constructed to indicate bullying, and these phrases were used as functions for input to device mastering algorithms. as additional features to traditional models, such as bag of phrases (n-gram n D 1). The authors claimed that adding these features improved the overall classification accuracy.

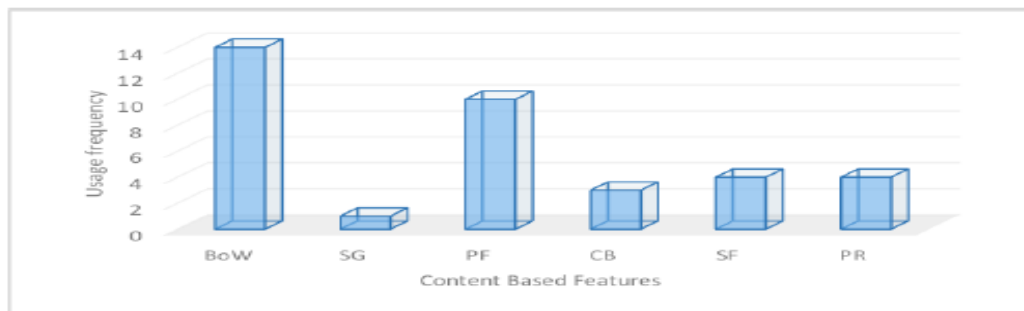


FIGURE 1. Depicting feature types used in cyberbullying prediction: Content-based features.

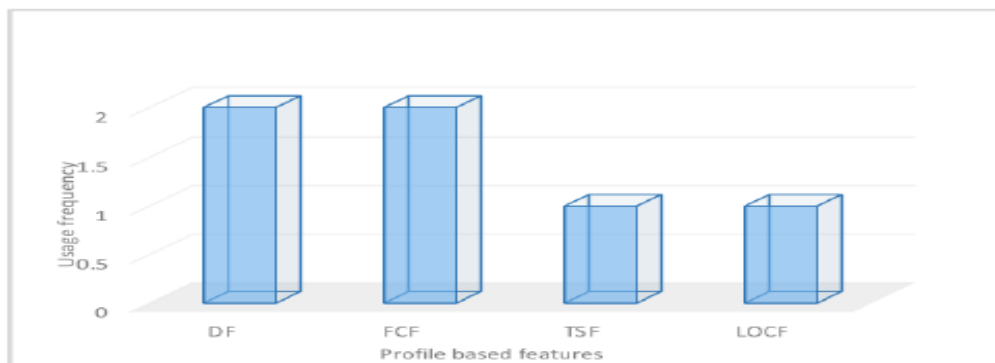


FIGURE 2. Depicting feature types used in cyberbullying prediction: Profile-based features.



VI. FEATURE SELECTION ALGORITHMS

Feature choice algorithms were hardly ever followed in state-of-the-art studies to perform cyberbullying prediction on SM web sites via system learning (all extracted features are used to train the classifiers). Most of the examined studies did not use feature selection to decide which features are essential in training gadget learning algorithms. Two studies used chi-square and PCA to select a significant feature from extracted features. These feature choice algorithms are briefly discussed in following subsections.

A. INFORMATION GAIN

Information advantage is the expected decrease in entropy produced by using isolating examples based totally on specified features Entropy is a well-known concept in facts theory; it describes the (im)purity of an arbitrary collection of examples.

Information advantage is used to calculate the power or importance of capabilities in a classification model in keeping with the class attribute. Information advantage evaluates how well a specified feature divides schooling datasets with admire to elegance labels, as explained inside the following equations. Given an education dataset (Tr), the entropy of (Tr) is defined as.

$$I(Tr) = - \sum P_n \log_2 P_n, \tag{1}$$

In which Pn is the probability that Tr belongs to magnificence n. For attribute Att datasets, the anticipated entropy is calculated as

$$I(Att) = \sum \left(\frac{Tr_{Att}}{Tr} \right) \times I(Tr_{Att}). \tag{2}$$

The facts gain of characteristic Att datasets

$$IG(Att) = I(Tr) - I(Att) \tag{3}$$

B. PEARSON CORRELATION

Correlation-based function selection is usually used in reducing characteristic dimensionality and evaluating the discrimination electricity of a characteristic in classification models. It is additionally a trustworthy version for selecting significant capabilities. Pearson correlation measures the relevance of a function by computing the Pearson correlation among it and a class. The Pearson correlation coefficient measures the linear correlation between two attributes. The subsequent value lies among -1 and 1, with -1 implying absolute negative correlation (as one characteristic increases, the other decreases) 1 denoting absolute superb correlation (as one attribute increases, the other additionally increases), and 0 denoting the absence of any linear correlation between the two attributes. For two attributes or features X and Y, the Pearson correlation Coefficient measures the correlation as follows

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n - 1) S_x S_y}, \tag{4}$$

Where x and y are the pattern means for X and Y respectively; Sx and Sy are the sample popular deviations for X and Y, respectively; and n is the size of the sample used to compute the correlation coefficient.

C. CHI-SQUARE TEST

Another not unusual function choice version is the chi-square take a look at. This take a look at is used in statistics, among other variables, to check the independence of two occurrences. In function choice, chi-square is used to test whether the occurrences of a function and sophistication are independent. Thus, the following amount is thought for each characteristic, and they're ranked by their score.

$$N = \frac{N [P(f, c_i)P(\bar{f}, \bar{c}_i) - P(f, \bar{c}_i)P(\bar{f}, c_i)]}{P(f)P(\bar{f})P(c_i)P(\bar{c}_i)} \tag{5}$$

The chi-rectangular test assesses the independence between characteristic f and sophistication ci, in which N is the whole range of documents.

D. MACHINE LEARNING ALGORITHMS: many varieties of system getting to know algorithms exist, however nearly all research on cyberbullying prediction in SM web sites used the maximum mounted and broadly used type that is supervised System gaining knowledge of algorithms. The accomplishment of device learning algorithms is decided through the degree to which the model correctly converts various types of prior observation or understanding approximately the task. Much of the practical application of machine studying considers the details of a specific problem. Then, an algorithmic version that permits for the correct encoding of the information is selected.

SVM	Support vector machine family
NB	Naïve Bayes
RF	Random forest
DT	Decision tree family
KVN	K-nearest algorithm
LR	Logistic regression
ARM	Association rule mining
RB	Rule based algorithms

Figure 3. Machine Learning Algorithm Applied In Cyberbullying Prediction

VII. CYBERBULLYING PREDICTION

A. SUPPORT VECTOR MACHINE IN CYBERBULLYING

Support vector device (SVM) is a supervised gadget mastering classier that is generally used in text classification [107]. SVM is constructed via generating a separating hyper plane in the function attributes of classes, in which the distance between the hyper plane and the adjacent records point of each magnificence is maximized. Theoretically, SVM became developed from statistical gaining knowledge of theory.

B. NB ALGORITHM

NB was used to assemble cyberbullying prediction models. NB classifiers were constructed with the aid of making use of Bayes' theorem between functions. Bayesian gaining knowledge of is typically used for textual content classification. This model assumes that the text is generated by means of a parametric model and makes use of training facts to compute Bayes-optimal estimates of the version parameters. It categorizes generated test information with these approximations. NB classifiers can address an arbitrary number of continuous or express independent features. By the use of the belief that the capabilities are independent, a high-dimensional density estimation task is reduced to one-dimensional kernel density estimation. The NB set of rules is a mastering algorithm that is grounded on the usage of Bayes theorem with strong (naive) independence assumptions.

This approach turned into mentioned in element. The NB set of rules is one of the most generally used machine studying algorithms and it's been built as a machine mastering classier in sever social media based

C. RANDOM FOREST

Random forest (RF) becomes used in the creation of cyberbullying prediction models. RF is a machine-gaining knowledge of model that combines selection bushes and ensemble mastering. This version is numerous classifications. Dataset then combines the predictions from all the timber. Therefore, RF consists of many timbers that are used randomly to pick characteristic variables for the classier input. The construction of RF is achieved within the following simplified steps.

1. The range of examples (cases) in education information is set to N, and the wide variety of attributes in the classier is M.
2. A wide variety of random selection tress is created by way of selecting attributes randomly. A schooling set is chosen for each tree with the aid of deciding on n instances from all N existing instances. The rest of the instances in the schooling set are used to approximate the mistake of the tree by way of forecasting their classes.
3. For each tree's nodes, m random variables are selected on which to base the selection at that node. The nest break up is computed the usage of those m attributes inside the schooling set. Each tree is completely built and is not pruned, as can be executed in building an everyday tree classier.

4. A large number of trees are for this reason created. This selection timber vote for the most famous elegance. These processes are known as RFs. RF constructs a version that comprises a group of trees structured classifiers, in which each tree votes for the maximum popular class.

D. DECISION TREE

Decision tree classifiers were used in construction of cyberbullying prediction models in and. Decision trees are clean to understand and interpret; hence, the decision tree algorithm may be used to analyze statistics and construct a graphic version for classification. The most normally stepped forward version of choice tree algorithms used for cyberbullying prediction is may be defined as follows. Given N wide variety of examples, first produces a preliminary tree through the divide-and-triumph over set of rules as follows.

E. K-NEAREST NEIGHBOR

K-nearest neighbor (KNN) is a nonparametric approach that comes to a decision the KNNs of X_0 and uses a majority vote to calculate the magnificence label of X_0 . The KNN classifier often makes use of Euclidean distances as the distance metric. To demonstrate a KNN classification, classifying new input posts (from a checking out set) is considered by using some of known manually categorized posts. The main venture of KNN is to classify

The unknown instance based on a nominated wide variety of its nearest neighbors, that is, to finalize the magnificence of unknown examples as both a positive or bad class. KNN classifies the elegance of unknown examples by the usage of majority votes for the nearest associates of the unknown classes. For instance, if KNN is one nearest neighbor [estimating the class of an unknown example using the one nearest neighbor vote

(K D 1)], then KNN will classify the magnificence of the unknown instance as fine (due to the fact the closest factor is high-quality). For nearest pals (estimating the magnificence of an unknown instance the use of the 2 nearest neighbor vote), KNN is unable to classify the elegance of the unknown instance because the 2d closest point is bad (positive and bad votes are equal). For four nearest friends (estimating the elegance of an unknown example the usage of the four nearest neighbor vote), KNN classifies the class of the unknown example as fantastic (due to the fact the three closest points are tremendous and best one vote is poor). The KNN set of rules is certainly one of the only classification algorithms, however regardless of its simplicity. It is able to provide aggressive outcomes. KNN become used within the creation of cyberbullying prediction fashions.

F. LOGISTIC REGRESSION CLASSIFICATION

Logistic regression is one in all the commonplace techniques imported by system studying from the statistics eld. Logistic regression is an algorithm that builds a separating hyper plane between datasets with the aid of the logistic

Function the logistic regression set of rules takes inputs (features) and generates a forecast in step with the possibility of the enter being suitable for a elegance. For instance, if the possibility is >0.5 , the classification of the instance might be a high quality elegance; otherwise, the prediction for the opposite class (bad magnificence). Logistic regression changed into used in the production of cyberbullying prediction fashions.

VIII. EVALUATION

The number one goal of constructing prediction fashions based totally on machine mastering is to generalize greater than the training dataset. When a gadget gaining knowledge of model is carried out to an actual example, it could carry out well. Accordingly, the records are divided into two parts. The rst element is the training records used to train system studying algorithms. The 2d part is the checking out information used to check device mastering algorithms. However, separately dividing data into training and checking out is not extensively employed, especially in programs wherein deriving training and checking out statistics are difficult. For example, in cyberbullying prediction, maximum state-of-art research manually labeled data. Hence, creating

Categorized records is expensive. These troubles can be decreased by move validation, that is, randomly dividing the schooling statistics into 10 subsets for example, and this process is called 10-fold pass validation. Cross validation includes the following steps: preserve a fold separate (the version does not see it) and train statistics on the version by using the final folds; take a look at each found out classifier at the fold which it did no longer see; and average the results to peer how well the unique parameter setting performs.

Evaluation metrics

Researchers measure the effectiveness of a proposed model to decide how correctly the version can distinguish cyberbullying from non-cyberbullying by the use of diverse evaluation measures. Reviewing common evaluation metrics inside the studies community is important to understand the performance of conflicting models. The most commonly used metrics in evaluation

A. ACCURACY

It became used to assess cyberbullying prediction models and its miles calculated as follows

$$Accuracy = \frac{(tp + tn)}{(tp + fp + tn + fn)} \quad (6)$$

B.PRECISION, RECALL, AND F-MEASURE

These were used to evaluate cyberbullying prediction models. They are calculated as follows

$$Precision = \frac{tp}{(tp + fp)} \quad (7)$$

$$Recall = \frac{tp}{(tp + fn)} \quad (8)$$

$$F - Measure = \frac{2 \times precision \times recall}{recision + recall} \quad (9)$$

IX. RESULT AND DISCUSSION

This observes reviewed existing literature to detect aggressive behavior on SM websites via using gadget learning approaches. We specifically reviewed four components of detecting cyber bullying messages by means of using system mastering approaches, namely, records collection, feature engineering, creation of cyberbullying detection model, and assessment of constructed cyberbullying detection models. Several varieties of discriminative capabilities that have been used to discover cyberbullying in on line social networking sites had been additionally summarized. In addition, the best supervised system learning classifiers for classifying cyberbullying messages in on-line social networking sites were identified. One of the principle contributions of modern-day paper is the definition of assessment metrics to efficiently perceive the significant parameter so the various gadgets getting to know algorithms may be evaluated against every other. Most importantly we summarized and identified the vital elements for detecting cyberbullying through device learning strategies specially supervised getting to know. For this purpose, we've used accuracy, precision recollect and f-measure which offers us the vicinity under the curve feature for modeling the behaviors in cyberbullying. Finally, the principle issues and open research challenges had been defined and discussed.

X. CONCLUSION

The researcher may also find the effects to be significantly improved, although an investigation of the way the device learning version is evaluated can also produce contradicting consequences and might not without a doubt reflect the development of performance. For example, cyberbullying posts are usually considered extraordinary cases, whereas Non-cyberbullying posts are considered normal cases. The ratio between cyberbullying and non-cyberbullying is normally big. Generally, non-cyberbullying posts comprise a huge portion. For example, one thousand posts are manually classified as cyberbullying and non-cyberbullying. The non-cyberbullying posts are 900, and the last 100 posts are cyberbullying. If a gadget learning classier classifies all 1000 posts as non-cyberbullying and is not able to classify any posts (0) as cyberbullying, then this classier is considered impractical. By contrast, if researchers use accuracy as the primary assessment metric, then the accuracy of this classier calculated as stated in the accuracy equation will yield a high accuracy percentage.

CONFLICT OF INTEREST

The authors declare here that they have no conflict of interest.

REFERENCES

1. V. Subramanian and S. Kumar, "Predicting human behaviour: The next frontiers," *Science*, vol. 355, no. 6324, p. 489, 2017.
2. H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, "Homophily in the digital world: A LiveJournal case study," *IEEE Internet Comput.*, vol. 14, no. 2, pp. 15_23, Mar./Apr. 2010.
3. M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," *Comput. Hum. Behav.*, vol. 63, pp. 433_443, Oct. 2016.
4. L. Phillips, C. Dowling, K. Shaffer, N. Hodas, and S. Volkova, "Using social media to predict the future: A systematic literature review," 2017,