Advances and Applications in Mathematical Sciences
Volume 21, Issue 9, July 2022, Pages 4973-4980

CLASSIFY BULLY TEXT WITH IMPROVED CLASSIFICATION MODEL USING GRID SEARCH WITH HYPERPARAMETER TUNING

V. INDUMATHI, S. SANTHANAMEGALA and R. PADMAPRIYA

Research Scholar Assistant Professor
School of Computer Studies
Rathnavel Subramaniam
College of Arts and Science
Coimbatore, India
E-mail: indhumathi@rvsgroup.com
santhanamegala@rvsgroup.com
padmapriya@rvsgroup.com

Abstract

With the growth of the Internet, utilize of social media has increased tremendously, and it has become the most prominent networking platform nowadays. The goal of this study is to use natural language processing (NLP) and machine learning (ML) to create and develop an effective procedure for detecting online abusive and bullying texts. From our previous work, we understand that Liner SVC and Logistic Regression are better performing than Multinomial Navies Bayes. As an enhancement in this paper, Grid Search with Hyper Parameter Tuning is done to improve the performance of classification. The simulation is used to test the usefulness of the Linear SVC and Logistic Regression models against hamming loss, accuracy, recall, and *f*-measure, among other metrics. The simulation results suggest that the Linear SVC model performs better in classification than the Logistic Regression model.

1. Introduction

Cyberbullying is a form of online harassment as given in figure 1 that occurs on social media platforms. Criminals use such networks to gather data and information that allows them to carry out their crimes, such as identifying a vulnerable victim [6]. As a result, academics have been

2020 Mathematics Subject Classification: 68T50.
Keywords: Cyber Bully, Machine Learning, Linear SVC, Logistic Regression, Grid Search.
Received January 17, 2022; Accepted May 17, 2022

attempting to drive with various ways and approaches for detecting and preventing cyberbullying. Cyberbullying monitoring systems have recently gotten a lot of attention, and their purpose is to quickly identify cyberbullying situations [7]. The main idea behind such systems is to extract specific features from social media messages and then create classifier algorithms based on those collected features to detect cyberbullying. Users, material, emotions, and/or social networks could all be used to create such features [2].

Figure 1. Cyber bullying in Social Media.

Youths were amid the first to use social media in [3], and continue to use it at high levels; also, elderly persons have begun to utilize it. We are focusing on creating a classifier from a small collection of positive training examples, while all other data stays unlabeled and no negative samples are available from training, in contrast to existing cyberbullying detection approaches that require both positive and negative training. The paper is structured as follows: Section 2 explores Related Works. In Section 3, Research Methodology is defined with better-performed models based on our previous work further tuned using Grid search in this paper. In Section 4, the Result and analysis are presented. In Section 5, the paper is concluded.

2. Related Work

In [7], created and implemented a new cyberbullying annotation scheme that identifies the presence and intensity of cyberbullying, the role of the post author (harasser, victim, or bystander), and several fine-grained cyberbullying categories such as abuse and threats. They describe their findings on the automatic detection of cyberbullying in online posts, as well as the capability of recognizing more fine-grained cyberbullying categories. An F-score of 55.39 percent is earned for the first job.

In [8], proposed an integrated model that includes the feature extraction engine and the classification engine is suggested using raw text datasets from a social media engine as input. The simulation results suggest that the ANN-DRL provides better classification results than conventional machine learning classifiers.

In [9], present a session-based structure for detecting cyberbullying automatically from massive amounts of unlabeled streaming material. They use an ensemble of 1 class classifiers in the session-based structure since the gathered data from Social Networks receives enormous volumes to the server system. Their key contribution is to detect cyberbullying automatically in real-world scenarios where labeled data is not easily available.

In [11], develop a method for detecting cyberbullying and identifying the most active predators and victims. We use a weighted TFIDF to improve classification performance, in which bullying-like features are scaled by a factor of two. Weighted TFIDF surpassed other approaches in terms of overall performance. This encapsulates our plan to increase the use of inductive language in detrimental posts.

In [12], present a technique for tracking the phenomenon of cyberbullying on social media. Through a simple, easyto-use interface, the system seeks to assist supervising persons (e.g., educators) in recognizing probable incidents of cyberbullying. This shows the results of a hate speech detection system as well as the network through which the messages are passed around.

3. Research Methodology

Grid search with Hyperparameter Tuning was performed to improve the efficiency of the classification models used in our previous work [1]. With this optimization, we can able to identify bully text in a better way.

3.1. Dataset. This model was built using a dataset of the Talk page of Wikipedia comments. In this dataset, for Training 80% of the data was streamed and tested remaining 20% data. There are 159,571 observations with 8 columns in the training data and 153,164 observations with 2 columns in the test data.

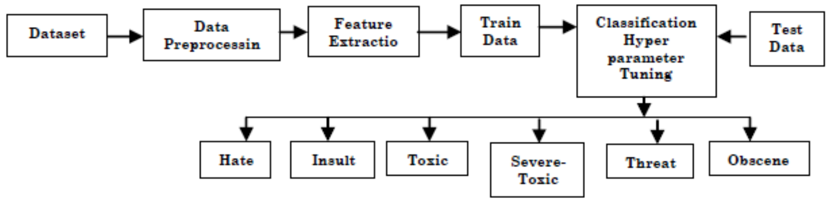

Figure 2. Proposed Model.

3.2. Data Preprocessing

Before processing the data from the dataset it has to be cleaned and prepared for the model. Data went through a series of steps, including tokenization, normalization, lemmatization, and finally, reducing the word length to less than 5. Tokenization is accomplished with the Tokenize() function, which removes punctuation and special characters [9]. So to overcome such issues spell-check tool was used to correct the misspelled words because in previous work lots of bully text was not recognized due to wrong spelling.

3.3. Feature Extraction

TF-IDF (Term Frequency-Inverse Document Frequency) was used to reduce the facts of tokens that occur frequently in the corpus and are thus empirically less illuminating than features [10] that rise in a short period in the training model. In comparison to other vectorizers, Bag of Words is used to define manually the bully text which is all around identified in this dataset. Developed a function to construct word clouds to gain an understanding of which words contribute the most to certain categories. A parameter label [12] is passed to the function (i.e., toxic, insult, threat, etc). While testing the model it has been identified that 1347 words were misclassified as non-bully when it was harmful.

3.4. Modeling

3.4.1. Pipelines

A pipeline is a set of data preparation options, modeling operations, and prediction transform processes organized in a logical order. A machine learning pipeline is a method of codifying and automating the process of creating a machine learning model. The pipeline aims to integrate numerous processes to cross-validated while adjusting various parameters. It does this by allowing you to set the parameters of the several phases using their names and separated parameter name as '__', as seen below.

"name = pipe.steps[-1][1].__class__.__name__.split('.')[-1]"

3.4.2. Grid Search with Hyperparameter Tuning

Hyper-parameters are variables that aren't learned directly by estimators. Searching the hyper-parameter space for the best cross-validation score is achievable and encouraged. Grid Search calculates the performance for each combination of all the supplied hyperparameters and their values and then chooses the optimum value for the hyperparameters. Based on the number of hyperparameters involved, this makes the processing time-consuming and costly. GridSearchCV's grid search exhaustively creates candidates from a grid of parameter values defined by the param grid argument.

4. Result and Analysis

4.1.1. Hamming Loss

The Hamming loss is the percentage of labels that are incorrect compared to the total number of labels. The hamming distance between y true and y pred is used to calculate hamming loss (as per equ. 1) in multi-class classification. Hamming loss penalizes only the individual labels in multi-label categorization. Since Hamming loss is defined as

$$HL = \frac{1}{m}\sum_{i=1}^{m}\frac{|Y_i \cap Z_i|}{|L|} \quad (1)$$

where $\cap$ denotes intersect, Y_i and Z_i stands for Boolean that the i^{th} datum (prediction) contains the l^{th} label, it equals to (1 - accuracy) for binary case $(L = 1)$. Mislabeling is no longer a hard wrong or right in multi-label classification. A prediction that includes a selection of the exact classes is preferable to one that does not include any of them. From the figure 3, it is clear not Logistic Regression is doing a great job since it has the lowest rate of wrong labeling.

Figure 3. Hamming Loss of Classification Models.

4.1.2. Evaluation Metrics

Previous results [1] will concentrate on Logistic Regression and Linear SVM because they perform better. For comparative purposes Training Time, average F1 score, Recall will be displayed. Both models F1 Score increased after Grid Search with Hyper Parameter Tuning as shown in figure 4 compare to the previous score. Logistic Regression was (0.917) before tuning and after tuning it raised to (0.972) same way Linear SVC was (0.918) before tuning and after tuning it raised to (0.918).

Figure 4. F1 Score.

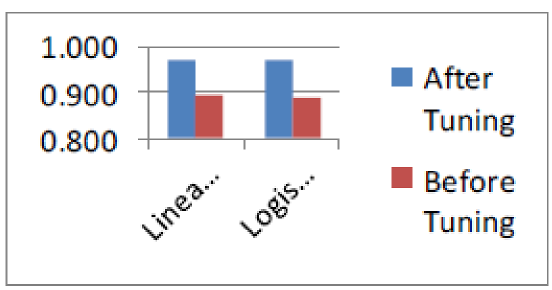

Figure 5. Recall.

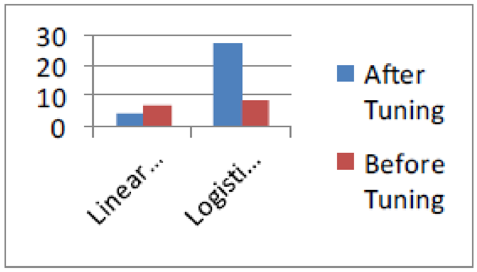

Figure 6. Training Time.

Training Time for Linear SVC decreased to (4) after Grid Search with Hyper Parameter Tuning as shown in figure 5 and 6 compare to the previous score (7). For Logistic Regression increased to (27) after tuning compared to its previous score (8).

Conclusion

Linear SVC efficient than Logistic regression. Notice much better results after modifying class weight compared with the basic models. LinearSVC outperforms Logistic Regression by a factor of about 1. At the point of misclassification, Logistic Regression has a lesser rate of loss compared to Linear SVC. So further work wants to investigate if ensembling could help us obtain better outcomes because ensemble learning improves machine learning results by integrating numerous models and allows for higher prediction performance than a single model.

References

[1] V. Indumathi and S. Santhanamegala, Enhanced Classification Model on CyberBullying detection with Multi-Label, 2021.

[2] V. Indumathi, S. Santhana Megala, R. Padmapriya, M. Suganya, and B. Jayanthi, Prediction and analysis of plant growth promoting bacteria using machine learning for millet crops, Annals of R.S.C.B., ISSN:1583-6258, 2021.

[3] T. Pratheebhda, V. Indhumathi and S. Santhana Megala, An empirical study on data mining techniques and its applications, 2021.

[4] V. Indumathi, P. Vijayakumar An efficient dimension reduction in-text categorization using clustering technique, 2018.

[5] Jalal Omer Atom, Cyberbullying detection through sentiment analysis, 2020.

[6] Cynthia Van Hee, Els Lefever, Ben Verhoeven, Julie Mennes, Bart Desmet Guy De Pauw, Walter Daelemans and Veronique Hoste, Detection and fine-grained classification of cyberbullying events, Proceedings of Recent Advances in Natural Language Processing, Hissar, Bulgaria 7(9) (2015), 672-680.

[7] N. Yuvaraj, K. Srihari, Gaurav Dhiman, K. Somasundaram, Ashutosh Sharma, S. Rajeskannan, Mukesh Soni, Gurjot Singh Gaba, Mohammed A. Alain and Mehedi Masud, Nature-Inspired-Based approach for automated cyberbullying classification on multimedia social networking, 2021.

[8] Vinita Nahar, Xue Li, Chaoyi Pang and Yang Zhang, Cyberbullying Detection based on Text-Stream Classification, 2013.

[9] B. Sri Nandhini, J. I. Sheeba and Cyberbullying, Detection and Classification Using Information Retrieval Algorithm, 2015.

[10] K. Nalini and L. Jaba Sheela, Classification of tweets using text classifier to detect cyber bullying, 2015.

[11] K. Nalini, L. Jaba Sheela, classification of tweets using text classifier to detect cyberbullying, Advances in Intelligent Systems and Computing, 2015.

[12] Stefano Menini, Giovanni Moretti, Michele Corazza, Elena Cabrio, Sara Tonelli and Serena Villata, A system to monitor cyberbullying based on message classification and social network analysis (2019), 105-110.