International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

A HYBRID DEEP LEARNING APPROACH FOR CYBERBULLYING DETECTION ON TWITTER: THE DEA-RNN MODEL

[1]AKKAPALLI SRINIVAS, [2]SHEELAM BHARGAVI
[1]Assistant Professor,[2]Student
Department of CSE
Sree Chaitanya College of Engineering, Karimnagar

Abstract : The prevalence of cyberbullying (CB) on social media sites has grown. Due to the fact that social media is so widely used by people of all ages, it is imperative that social media platforms be made safer from cyberbullying. The DEA-RNN hybrid deep learning model is presented in this study as a way to identify CB on the social media network Twitter. To reduce training time and fine-tune the parameters of the Elman RNN, the Dolphin Echolocation Algorithm (DEA) is combined with Elman type Recurrent Neural Networks (RNN) in the proposed DEA-RNN model. Using a dataset of 10,000 tweets, we conducted a detailed evaluation of DEA-RNN and compared its results to those of cutting-edge algorithms like Random Forests (RF), RNN, SVM, Bi-directional long short term memory (Bi-LSTM), and SVM. Based on the testing data, it was determined that DEA-RNN performed better in every case. It fared better than the well thought-out current methods for Twitter platform CB detection. In scenario 3, where it attained an average of 90.45% accuracy, 89.52% precision, 88.98% recall, 89.25% F1-score, and 90.94% specificity, DEA-RNN demonstrated greater efficiency.

I. INTRODUCTION

People of all ages now choose social media networks like Facebook, Instagram, Flickr, Twitter, and Flickr for online communication and socialisation. These platforms have made it possible for individuals to engage and communicate in previously unimaginable ways, but they have also given rise to evil practices like cyberbullying. One kind of psychological abuse that has a big effect on society is cyberbullying. Events involving cyberbullying are on the rise, namely among young individuals who spend a lot of time switching between social networking sites. Because of their widespread usage and the anonymity that the Internet affords to those who misuse it, social media platforms like Facebook and Twitter are especially vulnerable to cyberbullying. For instance, 14% of abuse in India happens on Facebook and Twitter, and 37% of these instances include children [1]. Furthermore, cyberbullying may result in major mental health problems and detrimental impacts on mental health. Anxiety, sadness, stress, and social and emotional challenges brought on by cyberbullying incidents account for the majority of suicides [2]_[4]. This highlights the need for a method to spot cyberbullying in communications on social media (such as tweets, posts, and comments).

The primary emphasis of this essay is the issue of detecting cyberbullying on the Twitter network. The main objectives in combating cyberbullying risks are to identify instances of cyberbullying from tweets and

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

provide preventative measures, since cyberbullying is becoming a common issue on Twitter [5]. As a result, further study on social network-based CB is required in order to gain deeper understanding and support the creation of useful tools and strategies to address the issue of cyberbullying [6]. It is almost hard to manually monitor and regulate cyberbullying on the Twitter platform [7]. Furthermore, it might be challenging to mine social media posts for signs of cyberbullying. For instance, communications on Twitter are often succinct, rife with slang, and may include emoticons and gifs. As such, it is hard to infer someone's intents and meanings only from posts on social media. Furthermore, bullying may be hard to see if the perpetrator hides it with techniques like sarcasm or passive-aggressiveness. Cyberbullying detection on social media is an open and active study issue, despite the difficulties that social media communications provide. The Twitter network has mostly used tweet categorisation and, to a lesser degree, subject modelling techniques for cyberbullying detection. Tweets are often classified as either bullying or non-bullying using text categorisation based on supervised machine learning (ML) models [8]_[17]. Tweets have also been classified as either bullying or non-bullying using deep learning (DL) based classifiers [7], [18]_[22]. When class labels are fixed and irrelevant to newly occurring events, supervised classifiers perform poorly [23]. Furthermore, it could only work with a preset set of events; tweets that change on the _y cannot be handled by it. For a long time, topic modelling techniques have been used as a means of identifying the key subjects from a dataset in order to create the classes or patterns across the whole dataset. Specialised unsupervised short text topic models were used since, despite their comparable concepts, generic unsupervised topic models are ineffective for short texts [24]. These methods extract the hot themes from tweets and efficiently identify them for further processing. These approaches facilitate the extraction of significant themes by using bidirectional processing. But in order to have enough previous information, these unsupervised models need a lot of training, which isn't always enough [25]. Given these constraints, a very effective tweet classification strategy has to be created to close the gap between the topic model and the classifier, resulting in much improved flexibility.

In this paper, we present DEA-RNN, a hybrid deep learning-based method that automatically identifies bullying from tweets. The enhanced Dolphin Echolocation Algorithm (DEA) is combined with Elman type Recurrent Neural Networks (RNN) in the DEA-RNN technique to fine-tune the parameters of the Elman RNN. The dynamic character of brief texts and subject models for the successful extraction of trending themes are handled by DEA-RNN. When it came to identifying cyberbullying on the Twitter platform across all scenarios and assessment criteria, DEA-RNN fared better than the techniques that were taken into consideration.

The following is a summary of this article's contributions:

_ Create a better DEA optimisation model to be used to automatically adjust the RNN parameters to improve performance; _ Propose DEA-RNN to combine the Elman type RNN and the improved DEA for the best tweet classification; _ Gather a new Twitter dataset based on cyberbullying keywords to compare the effectiveness of DEA-RNN and the current approaches; and _ Use Twitter datasets to evaluate the effectiveness of DEA-RNN in identifying and categorising cyberbullying

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

tweets. The comprehensive experimental findings show that in terms of recall, precision, accuracy, F1 score, and specificity, DEA-RNN performs better than other competing models.

The remainder of this piece is organised as follows: Section II reviews and analyses recent relevant research. In Section III, the suggested DEA-RNN model is explained. Performance metrics, outcomes analysis, and experimental analysis are included in Section IV. Section V provides an introduction to the topic. Section VI provides a conclusion and some avenues for further research.

II. LITERATURE SURVEY

Reviewing the state-of-the-art in CB detection and classification on Twitter datasets is the primary goal of this section. The categorisation of tweets including cyberbullying often makes use of machine learning (ML) based techniques using various feature selection techniques. The SVM and Information Gain (IG) based feature selection approach was used by Purnamasari et al. [26] to identify instances of cyberbullying in tweets. Muneer and Fati [11] used a variety of classifiers for the detection of cyberbullying incidents in tweets, including AdaBoost (ADB), Light Gradient Boosting Machine (LGBM), SVM, RF, Stochastic Gradient Descent (SGD), Logistic Regression (LR), and MNB. TF-IDF and Word2Vec were the approaches used in this investigation to extract features. SVM and Random Forests (RF) models with TF-IDF were used by Dalvi et al. [12] [27] for feature extraction in order to identify cyberbullying in tweets. Even though SVM performed well in these models, adding more class labels causes the model's complexity to rise. Al-garadi et al. [28] used a variety of machine learning classifiers, including RF, Naïve Bayes (NB), and SVM, to study the detection of cyberbullying based on variables that were retrieved from Twitter, including (tweet content, activity, network, and user). An strategy that incorporated textual content elements with social media features was proposed by Huang et al. [29] for the identification of CB from social media. The IG technique is used to rank the characteristics. There is use of well-known classifications like NB, J48, and Bagging and Dagging. The results suggested that social attributes could help improve the detection accuracy of cyberbullying. Squicciarini et al. [30] identified cyberbullying and predicted cyberbullying on social networks such as spring.me and MySpace by using a decision tree (C4.5) classifier with a social network, personal, and linguistic attributes. In order to identify cyberbullying incidents from tweets and categorise tweets into several cyberbullying classes, such as aggressors, spammers, bullies, and normal, Balakrishnan et al. [31] used a variety of machine learning techniques, including RF, NB, and J48. The investigation came to the conclusion that the detection rate is unaffected by the emotional characteristic. This model works well, however it can only handle a short dataset with few class labels. Using the single and double ensemble-based voting model, Alam et al. [32] presented an ensemble-based categorisation method. These ensemble-based voting models used unigram TF-IDF and mutual information bigrams as feature extraction models and decision tree, logistic regression, and bagging ensemble model classifiers for the classification. The Bagging ensemble model yielded the greatest accuracy when analysed across the Twitter dataset, but it took other factors into account. Despite the fact that these ensemble models shortened the training and execution times for classification, they had a significant drawback when applied to sarcastic tweets and acronyms with many meanings. Chia

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

et al. [8] classified irony and sarcasm from cyberbullying tweets by using several machine learning and feature engineering techniques. While this strategy considerably identifies the sarcasm and irony phrases among cyber-bullying tweets, the detection rate is still quite low [33].

Several classifiers and feature selection approaches were evaluated in this approach.

Similar to this, Rafiq et al. [17] used a Vine dataset to detect the occurrences of cyberbullying using decision trees, AdaBoost, NB, and Randon Forest classifiers. The Vine media dataset was gathered by the authors and tagged using Crowd-Sourced and CrowdFlower web tools. As features, they made use of the comments, unigrams, media details, and profile. In order to identify CB in social media, Nahar et al. [34] proposed a semi-supervised learning approach in which training data samples are enhanced and a fuzzy SVM algorithm is used. The training set is automatically expanded and extracted using the augmented training technique from the unclassified streaming text. As an initial input, a tiny, constrained training set is used to accomplish the learning. The proposed approach gets beyond streaming data's complicated and dynamic nature. Numerous off-the-shelf techniques, such as Bagof-Words models and LDA and LSA-based modelling, were offered by Xu et al. [35] for predicting bullying traces on Twitter. Cheng et al. [36] proposed PI-Bully, a personalised cyberbullying detection system, to identify cyberbullying from the Twitter dataset. Three components make up PI-Bully: a global component that finds the traits that all users share, a personalised component that records the unique qualities of each user, and a peer influence component that can measure the varied influences of other users. In the literature, deep learning (DL) based methods for identifying cyberbullying in tweets have also been put out. Tweets on cyberbullying were categorised using Artificial Neural Network (ANN) and Deep Reinforcement Learning (DRL) by N. Yuvaraj et al. [9]. This method's computational complexity is larger, however. To improve the sentiment analysis task performance, Chen et al. [37] developed a text classification model based on CNN and 2-D TF-IDF features. In comparison to the baseline LR and SVM models, the experimental findings demonstrated that the CNN model produced the best outcomes.

Agrawal [16] used Transfer Learning in conjunction with LSTM to identify cyberbullying across a number of social media platforms. Zhao et al. [38] proposed a novel representation learning method for cyberbullying detection called smSDA (Semantic-Enhanced Marginalised Denoising Autoencoder). Robust and discriminative representations were generated using smSDA. After that, SVM may be used to process the learnt numerical representations. Zhang [39] proposed a novel model for detecting hate speech that combines CNN layers and the Gated Recurrent unit Network GRU layers. In order to identify hate speech related to cyberbullying in Arabic tweets, Al-Hassan and Al-Dossari [19] used SVM as the baseline classifier and evaluated it against four DL models: CNN + LTSM, LTSM, CNN + GRU, and GRU. But because of their increased complexity, CNN+LSTM and CNN+GRU may not be able to handle bigger datasets well. A brand-new classification approach for CB identification from Twitter data was put out by Natarajan Yuvaraj et al. [18]. For tweet classification, deep decision-tree classification using multi-feature based AI was utilised. By combining the decision tree classifier and the hidden layers of deep neural networks, the deep decision tree

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

classifier was created. Three feature selection techniques were also used in this method: IG, Pearson Correlation, and Chi-Square. But it is not accurate enough to handle high-dimensional data. To identify cyberbullying in tweets, Fang et al. [20] developed a classification model that incorporates a bi-directional Gated Recurrent Unit (Bi-GRU) with a self-attention mechanism. This model improved the categorisation process for cyberbullying tweets by using merit to understand the underlying associations between words using BI-GRU in conjunction with a self-attention mechanism. Nevertheless, understanding every association between the tweets is limited by the attention network's context-independent behaviour.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Purnamasari et al. [26] utilized the SVM and Information Gain(IG) based feature selection method for detecting cyberbullying events in tweets. Muneer and
Fati [11] used various classifiers, namely AdaBoost(ADB), Light Gradient Boosting Machine (LGBM), SVM, RF, Stochastic Gradient Descent (SGD), Logistic Regression (LR), and MNB, and for cyberbullying events identification in tweets. This study extracted features using Word2Vec and TF-IDF methods.

Dalvi et al. [12] [27] used SVM and Random Forests (RF) models with TF-IDF for feature extraction for detecting cyberbullying in tweets. Although SVM in these models achieved high performance, the model complexity increases when the class labels are increased. Al-garadi et al. [28] investigated cyberbullying identification using different ML classifiers such as RF, Naïve Bayes (NB), and SVM based on various extracted features from Twitter such as (tweet content, activity, network, and user). Huang et al. [29] suggested an approach for identifying CB from social media, which integrated the social media features and textual content features.

The features are ranked using IG method. Well-known classifies such as NB, J48, and Bagging and Dagging are utilized. The findings implied that social characteristics could aid in increasing the accuracy of cyberbullying detection. Squicciarini et al. [30] utilized a decision tree (C4.5) classifier with a social network,
personal and textual features to identify Cyberbullying and cyberbullying prediction on social networks like spring.me, and MySpace. Balakrishnan et al. [31] utilized different ML algorithms such as RF, NB, and J48 to detect cyberbullying events from tweets and classify tweets to different cyberbullying classes such as aggressors, spammer, bully, and normal. The study concluded that the emotional feature does
not impact the detection rate. Despite its efficiency, this model is limited to a small dataset with fewer class labels.
Alam et al. [32] proposed an ensemble-based classification approach using the single and double ensemble-based voting model. These ensemble-based voting models utilized decision tree, LR, and Bagging ensemble model classifiers for the classification while utilizing mutual information bigrams and unigram TF-IDF as feature extraction models. On analysis over the Twitter dataset, the Bagging ensemble model provided the best precision but considered other parameters. Although, these ensemble models reduced the training and execution time for classification, the major limitation comes when utilized sarcasm tweets and multiple-meaning acronym terms. Chia et al. [8] also utilized different ML and feature

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

engineering-based approaches to classify irony and sarcasm from cyber-bullying tweets. In this approach, many classifiers and feature selection methods were tested; while this approach greatly detects the sarcasm and irony terms among cyber-bullying tweets, the detection rate is still very low [33].

Disadvantages

- The system is not implemented cyberbullying detection due to absence of an effective ML classifiers.
- The system is not implemented DEA-RNN techniques which lead very less prediction.

Proposed System

In this article, we propose a hybrid deep learning-based approach, called DEA-RNN, which automatically detects bullying from tweets. The DEA-RNN approach combines Elman type Recurrent Neural Networks (RNN) with an improved Dolphin Echolocation Algorithm (DEA) for fine tuning the Elman RNN's parameters. DEA-RNN can handle the dynamic nature of short texts and can cope with the topic models for the effective extraction of trending topics. DEA-RNN outperformed the considered existing approaches in detecting cyberbullying on the Twitter platform in all scenarios and with various evaluation metrics. The contributions of this article can be summarized as the following:

_ Develop an improved optimization model of DEA for use to automatically tune the RNN parameters to enhance the performance;

_ Propose DEA-RNN by combining the Elman type RNN and the improved DEA for optimal classification of tweets;

_ A new Twitter dataset is collected based on cyberbullying keywords for evaluating the performance of DEA-RNN and the existing methods; and

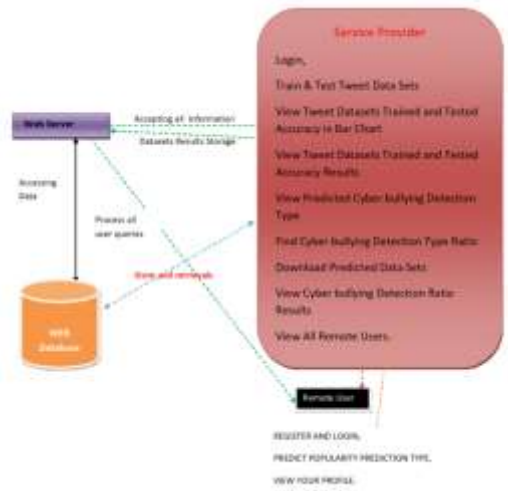
_ The efficiency of DEA-RNN in recognizing and classifying cyberbullying tweets is assessed using Twitter datasets. The thorough experimental results reveal that DEA-RNN outperforms other competing models in terms of recall, precision, accuracy, F1 score, and specificity.

Advantages

- The proposed system effectively identifies the trending topics from tweets and extracts them for further processing. An effective models help in leveraging the bidirectional processing to extract meaningful topics.

An effective system which is mainly tested and trained by SVM, Multinomial Naive Bayes (MNB),Random Forests (RF) classifiers.

IV. SYSTEM ARCHITECTURE



V. SYSTEM IMPEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train & Test Tweet Data Sets, View Tweet Datasets Trained and Tested Accuracy in Bar Chart, View Tweet Datasets Trained and

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Tested Accuracy Results, View Predicted Cyber bullying Detection Type, Find Cyber bullying Detection Type Ratio, Download Predicted Data Sets, View Cyber bullying Detection Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT CYBERBULLYING TYPE, VIEW YOUR PROFILE.

VI. CONCLUSION

In order to improve the efficacy of topic models for the identification of cyber-bullying occurrences, this article built an efficient model for tweet categorisation. For effective parameter tweaking, the DEA RNN was created by fusing the Elman type RNN with the DEA optimisation. Additionally, it was evaluated against the current Bi-LSTM, RNN, SVM, RF, and MNB techniques using a freshly generated Twitter dataset that was obtained via the use of CB keywords. According to the experimental research, the DEA-RNN outperformed the other approaches in use in every situation when measured using a variety of metrics, including accuracy, recall, F-measure, precision, and specificity. This illustrates how DEA affects RNN performance. The hybrid suggested model outperformed the other evaluated existing models in terms of performance rates; however, when input data increases beyond the initial input, DEA-RNN's feature compatibility decreases. The present investigation was restricted to the Twitter dataset alone; other Social Media Platforms (SMP) including Instagram, Flickr, YouTube, Facebook, and others need to be examined to identify any patterns of cyberbullying. Subsequently, future research will examine the feasibility of using different source data for cyber-bullying identification. Additionally, we were only able to analyse the content of tweets; we were unable to analyse the behaviour of the individuals. Future efforts will include this. The textual content of tweets is used by the suggested model to identify cyberbullying, whereas additional media, such as pictures, videos, and audio, are still up for investigation and may be the subject of future studies. Additionally, our goal is to identify and categorise CB tweets in a live stream.

REFERENCES

1. F. Mishna, M. Khoury-Kassabri, T. Gadalla, and J. Daciuk, ``Risk factors for involvement in cyber bullying: Victims, bullies and bully_victims,"Children Youth Services Rev., vol. 34, no. 1, pp. 63_70, Jan. 2012, doi:10.1016/j.childyouth.2011.08.032.
2. K. Miller, ``Cyberbullying and its consequences: How cyberbullying is contorting the minds of victims and bullies alike, and the law's limited available redress," Southern California Interdiscipl. Law J., vol. 26, no. 2, p. 379, 2016.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

3. A. M. Vivolo-Kantor, B. N. Martell, K. M. Holland, and R. Westby, ``A systematic review and content analysis of bullying and cyber-bullying measurement strategies," Aggression Violent Behav., vol. 19, no. 4,pp. 423_434, Jul. 2014, doi: 10.1016/j.avb.2014.06.008.
4. H. Sampasa-Kanyinga, P. Roumeliotis, and H. Xu, ``Associations between cyberbullying and school bullying victimization and suicidal ideation, plans and attempts among Canadian schoolchildren," PLoS ONE, vol. 9, no. 7, Jul. 2014, Art. no. e102145, doi: 10.1371/journal.pone.0102145.
5. M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, ``Improving cyberbullying detection with user context," in Proc. Eur. Conf. Inf. Retr., in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artifcial Intelligence and Lecture Notes in Bioinformatics, vol. 7814,2013, pp. 693_696.
6. A. S. Srinath, H. Johnson, G. G. Dagher, and M. Long, ``BullyNet:Unmasking cyberbullies on social networks," IEEE Trans.Computat. Social Syst., vol. 8, no. 2, pp. 332_344, Apr. 2021, doi:10.1109/TCSS.2021.3049232.
7. A. Agarwal, A. S. Chivukula, M. H. Bhuyan, T. Jan, B. Narayan,and M. Prasad, ``Identi_cation and classi_cation of cyberbullying posts:A recurrent neural network approach using under-sampling and class weighting," in Neural Information Processing (Communications in Computer and Information Science), vol. 1333, H. Yang, K. Pasupa, C.-S. Leung, J. T. Kwok, J. H. Chan, and I. King, Eds. Cham, Switzerland: Springer, 2020, pp. 113_120.
8. Z. L. Chia, M. Ptaszynski, F. Masui, G. Leliwa, and M. Wroczynski, ``Machine learning and feature engineering-based study into sarcasm and irony classi_cation with application to cyberbullying detection," Inf. Process. Manage., vol. 58, no. 4, Jul. 2021, Art. no. 102600, doi: 10.1016/j.ipm.2021.102600.
9. N. Yuvaraj, K. Srihari, G. Dhiman, K. Somasundaram, A. Sharma, S. Rajeskannan, M. Soni, G. S. Gaba, M. A. AlZain, and M. Masud, ``Nature-inspired-based approach for automated cyberbullying classi_cation on multimedia social networking," Math. Problems Eng., vol. 2021, pp. 1_12, Feb. 2021, doi: 10.1155/2021/6644652.
10. B. A. Talpur and D. O'Sullivan, ``Multi-class imbalance in text classifcation: A feature engineering approach to detect cyberbullying in Twitter," Informatics, vol. 7, no. 4, p. 52, Nov. 2020, doi: 10.3390/informatics7040052.
11. A. Muneer and S. M. Fati, ``A comparative analysis of machine learning techniques for cyberbullying detection on Twitter," Futur. Internet, vol. 12, no. 11, pp. 1_21, 2020, doi: 10.3390/_12110187.
12. R. R. Dalvi, S. B. Chavan, and A. Halbe, ``Detecting a Twitter cyberbullying using machine learning," Ann. Romanian Soc. Cell Biol., vol. 25, no. 4,pp. 16307_16315, 2021.
13. R. Zhao, A. Zhou, and K. Mao, ``Automatic detection of cyberbullying on social networks based on bullying features," in Proc. 17th Int. Conf. Distrib.

Comput. Netw., Jan. 2016, pp. 1_6, doi: 10.1145/2833312.2849567.

14. L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, ``XBully: Cyberbullying detection within a multi-modal context," in Proc. 12th ACM Int. Conf. Web Search Data Mining, Jan. 2019, pp. 339_347, doi: 10.1145/3289600.3291037.
15. K. Reynolds, A. Kontostathis, and L. Edwards, ``Using machine learning to detect cyberbullying," in Proc. 10th Int. Conf. Mach. Learn. Appl. Workshops (ICMLA), vol. 2, Dec. 2011, pp. 241_244, doi: 10.1109/ICMLA.2011.152.