JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

Comparative Analysis and Assesment on Different Hate Speech Detection Learning Techniques

Dr.Dharmveer Yadav,
Assistant Professor, Department of Computer Science,
St.Xavier's College,Jaipur,Rajasthan,India
Email: dharmveerya21@gmail.com

Dr.Mahaveer Kumar Sain,
Professor,Department of Computer Science & Informatics, Maharishi Arvind University, Jaipur,Rajasthan,India Email:mahaveersain@gmail.com

Abraham Amal Raj B,
Departmentof Computer Science & Informatics, Maharishi Arvind University,Jaipur,Rajasthan,India
Email:amalrajsj@gmail.com

Received 2022 October 20; **Revised** 2022 November 18; **Accepted** 2022 December 26.

Abstract—Increased social networking services have modified the way and scale of cyberspace communication. Even so, owing to the anonymity and mobility of these services, the no. of online hate speech is growing. In recent years, several groups, scholars, and social media networks have worked to mitigate the detrimental effect of hate speech on social media. Despite these attempts, social media users continue to be subjected to hate speech. The issue is significantly more visible in social organizations that foster public discourse. Because it is costly and time-consuming to automatically identify hate speech by human annotators, an algorithm is needed for automatic recognition. It has proved effective to transfer knowledge with the finished of a pretrained language model for various downstream tasks in the area of natural language processing (NLP). This paper presents a deep analysis of how hate speech is detected by state-of-art speech detection methods. There are various methods to perform multitask learning in deep learning.The deep multi-task learning system to leverage valuable knowledge from these multiple tasks in numerous benchmark data sets showthe effectiveness of the method which is convincing about the most advanced models. Such DL techniques are used to detect hate speeches from the benchmark datasets. Also, a comparative evaluation has also done using these different types of machine learning techniques that also included deep learning with word embedding methods for hate speech detection. From this comparative analysis and discussion found that machine learning techniques achieved good outcomes but Deep learning techniques beat them. However, deep learning with word embedding techniques has achieved more than 92% performance using Bi-GRU-GloVebut deep learning without word embedding methods outperforms and achieved more than 95% performance efficiency results using LSTM over state-of-art techniques.

Keywords—*Social Media, Emotions Recognition, Hate Speech Detection, Detection Methods, Multitask Learning, Deep Learning.*

I. INTRODUCTION

The move to online platforms of human communication is a double-edged sword. Social advantagescontain sharing views & experiences, receiving immediate feedback & discussing top topics. From an economic perspective, online communications informationenhances businesses' information, improves services, and increases their business performance. [1].Online analysis, social media data provide lots of customer information including dislikes about the product or service [2].

Remarkable progress has been made in the next generation of auxiliary attributes, including subjects and sentiments. However, for the broad corpus, high-quality labeled information is hard to get. The purpose of this paper is therefore to detect a ranking of social emotions that can be of use for more controlled text generation, with different intensities evoked by Online Documents. Existing studies also see each text as a document that does not catch the internal relationship between sentences in a document [3].

Real-time microblogging helps users to share brief digital content including email, associations, images, or videos. Social mediais a microblog contraband word. The generation of major metadata for data mining & simulated modeling by users in social media like Twitter, Facebook, YouTube, Instagram, WhatsApp, Snapshot, LinkedIn, so on. [4][5]. Microblogging is a modern means of communications compared to traditional media amongstoperators, establishments, & scholars in numerousarenas[6][7][8][9]. Microblogging is attractive because of its unique message feature, including portability, instant message, and simple use that allows real-time communiqué with little to no content restrictions.

In the daily lives of individuals, emotions drive global decision-making. Also, emotional factors can adversely move human communication, attention & personal knowledge recall. Although emotional recognition and interpretation always come effortlessly to humans, these tasks face serious computational routine challenges.The term effectual computing denotesmethods to detect, identify & predict human emotions to adapt the computational systems to these conditions, such as joy, anger, sadness, trust, surprise, anticipation. The subsequent computer systems can not only show empathy but can also offer decision-making support for an individual's emotional state. [10].

Hate speech is typically described as any message that ignores or differentiates against a person or group associated with particular criteria including color, ethnicity, sexuality, gender identity, ethnicity, language, or another factor. [11][12]. The no. of HS is growing increasingly owing to increased Twitter & resulting big data from content created by a user. [13]. However, the difficulties in algorithms for incident detection limit most approaches to hate speech detection. A dynamic research focus continues to focus on classifying hate speech with social media information but there have been few research attempts to create a general metadata architecture.Twitter hate speech detection is based on such techniques as Machine learning (ML), NLP, data extraction, content extraction, & text mining. But there are a lot of tweets, contaminated content [14], and rumors [15] on Twitter streams that adversely affected the performance of the classification algorithm.

This analysis offers a series of HS data sets that researchers can use to test any learning model on the classification of HS. The study also presented the advantages and drawbacks in the Table III classification for single and hybrid MLapproaches. Also presented was the summary of the efficiency assessment for MLapproaches. This study provides a detailed review of recent and past hate speech classification techniques to guide other investigators into novel directions for research.

A. Motivation

The hate speech classifiers are based on annotation methods of questionable reliability that are very problematic to describe. And a Manual, like the ones Facebook utilizes, is a difficult task. Censorship is a potential risk if automated text classification systems are appliedto deal with that problem. All options should also be taken into consideration. Action has been taken to censor and block posts deemed hateful and/or threatening to the online digital community and society, with adverse consequences. Our work aims to find easy and successful ways to enhance existing research in the area of classification of HS.

The paper content is structured accordingly. Section II takes an overview ofHate-speech detection in social networking and VLSP2019's Hatespeech dataset. Section III presents a state of arts HSDtechniquesfor solving the HSD problem on a dataset. Section IV showsthe concept of multitask learning in deep learning in detail.Section V discusses the various deep learning techniques used to Hate speech detection. Section VI lists related works on Hate speech detection. Then, a comparative analysis and deep discussion also provides the illustration and comparison on different Hate speech detection algorithms using machine learning and deep learning techniques in Section VII. Finally, Section VIII concludes the paper.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

II. HATE SPEECH DETECTION (HSD)

The rapid growth and interactions among people from different countries, cultures & ethnic communities are enabled by Web technology and social networks. Although this global communication has clear benefits and positive effects, invisibility, anonymity, & accessibility have made remarks xenophobic, racial & sexist easy then unpunished to express oneself.The so-called inhibition outcome on the Web [16] allows users to conduct themselves face to face. The rapid spread of online content, in particular in social networks, has also created such harassment behavior highly harmful, to the damage produced by hateful propaganda on the Internet can effectively reduce.

A. Hate Speech

Hate speech may be called a language to show hate to a targeted group or to be derogatory, humiliating, or insulting group members [17]. Internet distributors like Facebook recently announced that they are stepping up their exertions to censor & fight hateful speech [18] to prevent hateful language and harm their customers. But they recognized that they did not detect any such material. [19]. another corporation, similarto Twitter, is actively reviewing policies to contain more offensive behaviors and create new ways of combating hateful content, Warnings to delete harmful tweets or even permanent user suspension. However, while they have made considerable efforts and engaged many human resources, a huge quantity of data produced by the users is a challenge for them [20].

B. Hate Speech Detection

The spread of social networks e.g. Facebook and Twitter have made HSD of social-network language is a most recent area of study. The role of detecting hate speech in online communities is to detect hateful content. Expanding the Internet and expanding online interactions can harass many users of social media, blogs & forums.The online experience and the general population may be affected by this kind of abuse. Sentences like "such niggers work for locals" or "All people who are #muslim should be killed. The story's end. The target minority groups #islam #IslamicState are viewed online everyday [21].

In recent years, the movement for the protection of minority groups has been promoted on the internet [22]. Artificial Intelligence for the Social Good (AI4SG). Sexual harassment [22], sexual discrimination detection [24], cyberbullying & trolls have been discussed in some papers. [25].

More recent word has discussed the suicide detection of ideas to address certain actual Internet consequences of hatred [26]. The first study on detecting hate speech was based on bag-of-words (BOW)approaches [27][28]. This is the basis of this research. In the year 2012, in contrast to pattern-based approaches, we can find one of the previous studies using classifiers based on computers for the identification of hate speech [29][30]. Traditional MLmethods have been commonly used to detect hate speech, including LR [27], SVM &DT. Non-linguistic features such as the author's gender or ethnicity can help improve the classification of hate speech but this data are often unavailable or dishonest in social media [28]. Any types of sentiment details are available as features. Hate, thus, is a negative emotion, it is safe to say that messages that convey negative emotions are more likely to display hate than messages that are neutral or positive. Sentiment analysis & methods of polarity detection also are typically used to detect hate. [31].

Also, the usage of external lexical tools has been used in HSD, inspired by sentiment analysis& affective computing [32]. A hate verb lexicon is developed in [33] that endorses or promotes performances of violence. Subsequently, lexicon-based classification is highly dependent on the consistency of outside resources available, another initiative mergesadvantages of ML with classification methods based on lexicons to HSD. [34].

In recent years, the focus has been paid to the use of neural patterns in the identification of hate speech. These models typically use DL practices, e.g. CNN's & LSTM Networks [35], which display an important outcome for many NLP tasks. Because of the tools available, the commonHS testing studies are based on English texts.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

C. Hate-Speech Dataset

The shared VLSP task [36] provided the Hate-speech dataset. the dataset comprises 20,345 Facebook comments or posts. Any post is mentioned with the CLEAN, OFFENSIVE, and HATE labels of the same name. Table I defines dataset sample data:

TABLE I. Three class sample data from the Dataset

No.	Comments / Posts content	Label
1	English: Send me music	CLEAN
2	English: I fucking sulky this	OFFENSIVE
3	English: See your damn fool, this guy's comment!	HATE

The table below explains the number of messages, average terms, and vocabulary for each label:

TABLE II. VLSP-HSD Dataset Statistics

Label	Clean	Hate	Offensive
Amount of comments	18,614	709	1,022
Word length average	18.69	20.46	9.35
Size of vocabulary	347,949	14,513	9,556

The vocabulary size for each level is determined in Table II by the total number of terms. Dividing the vocabulary by the commentary number on each label determines the average terms.

III. STATE-OF-THE-ART OF HSDTECHNIQUES

These techniques aim to minimize the error of new comments or posts between the predicted labels with a real label to comments or posts. Also, in comparing their efficiency, both traditional model ML& DN models will be used.

A. Word Embedding (WE)

WEare NLP feature-learning method that translates words or phrases from word collections to a true number vector. The distributed images of words that have been namedWE lead to the accuracy of foreign language models [37]. As inputs obtained the first rank in a joint NER challenge ordered by VLSP, Thai-Hoang Pham & Phuong LeHong use RNN models with pretrained terms [37].

B. Traditional Models

Logistic regression (LR): This is a key & well-known classification algorithm. The manual feature extractions are needed in the text classification. The logistic regression with 2 factors is used in this paper:

- TF-IDF is (1.3) and its maximum functionality is 20,000 with word analyzer, stopwords2 & NGR.
- TF-IDF range of n-gram is (3,6), the highest char analyzer used stop-words is 40.000 & with the char analysis tool.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

Support Vector Machine (SVM): In ML, classification, regression, and other tasks of learning are common. SVC Kernel, which may include2 classes & multiclass classification, is included in a LIBSVM [38]. SVM uses the same extractor as LR with a linear classifier:

- TF-IDF Analyzer of words for &
- TF-IDF N-gram range is with char n-gram analyser (3,6).

A. Deep Neural Network Models (DNN)

Text-CNN:CNNis a multi-stage classification NNarchitecture. It can detect mixture features by using fully convolutional layers. [39].

GRU:GRU is an RNN-type and is an LSTM-model variation. In comparison, Gated Recurrent Units are just 2 gates: Reset & Update Gates. This makes training less complex than LSTM, which means it's faster than LSTM, & more complicated than the LSTM model. [40].

B. Single & Hybrid ML Methods

In most classification activities, ML requires mapping some output vectors to certain input vectors. ML algorithms can essentially be separated intoa hybrid & single strategy[41].

1) Single Methods

This technique is the detector and categorizes hate speech into Twitter data using a single classification of ML. Also, ML is utilized for extracting & preprocessing Twitter data in different capacities. ML is commonly considered to be an algorithmic and statistical approach for solving problems. In the following stages, the key steps in HS classification models:

- Step of preprocessing
- Step of Data representation
- Step of detection
- Step of classification

2) Hybrid Methods

This methodcombines various methods of ML to increase the efficiency of simple human methods. This is viewed as an improvement to single ML approaches to produce improved outcomes with Twitter hate speech designation. Hybrid approaches are very stable, and the high volume of metadata generated on social media platforms is believed to be better processed. Hybrid models are more computer-efficient and yield better performance relative to their single form.

TABLE III. The Benefits and Drawbacksof hybrid & single approaches

Main type	Benefits	Drawbacks	Approaches
Single method	Stability, Adaptable, Extensible	Low precision, The problem of fragmentation, Poor class imbalance data performance	1) Fuzzy logic (FRB, FML, AR) 2) ANN (RNN, CNN, MLP) 3) DL (LSTM, CNN-1D) 4) BN (Naïve Bayes) 5) GA (GP) 6) Kernel method (SVM) 7) LR 8) DT (J-48graft, RF)
Hybrid	Consisten	Higher	1) FL & NLP

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

method	cy, Efficient, Flexible Adaptability to the size of data	accuracy, More difficulty of time	2) BN & RNN 3) LSTM & NN 4) Embedding & DL 5) Bi- LSTM & MLP 6) Cat swarm optimization & NN 7) K-means & cuckoo search 8) NB & feedforward NN 9) NB & LR 10) NB & PSO 11) PSO, GA & DT 12) DL & LR 13) N-gram & ANN 14) N-gram & SVM 15) Embedding, LSTM & gradient boosted trees 16) SVM, LR & DT

IV. MULTI-TASK LEARNING

Multi-Task Learning (MTL) is an ML subfield in which a common model performs several tasks at the same time. Such methods provide benefits such as increased data efficiency, reduced overfit by common representations & rapid learning by using auxiliary data. However simultaneous learning of several tasks presents new challenges for design & optimization & it is a non-trivial problem to choose which tasks to be learned together.

A. Problem Associated with Learning

Learning definitions for different tasks lead to challenges that do not occur in one task. In specific, there could be conflicting needs in different tasks. In this case, increasing a model's performance in one task would harm a task with various requirements, called negative transfer or destructive interference [42].

B. Multi-Task Learning

MTL is a learning paradigm where ML models are simultaneously trained using data from different tasks to learn the shared insights amid a series of similar tasks. These typical representations improve data quality & can potentially speed up learning in connection with similar tasks or the future, thus contributing to alleviating well-known limitations in DL: large-scale data requirements & computational demand. Yet, it has not been easy to produce these results and is now an active field of study. MTL represents more accurately than just a task in learning the learning process for people since information integration across domains is a key aspect of human intelligence.

C. Multi-Task Architectures

When designing a shared architecture, several common variables must be taken into consideration: the share of model parameters shared among roles and how task-specific and shared modules can be designed and integrated. More variables occur when seeing architectures for a particular problem area, such as how-to convolutional filters for certain vision tasks are partitioned into general and task-specific classes. Numerous planned MTL architectures play a balance with the extent to which knowledge is exchanged among tasks: To lot sharing leads to a negative transfer which can allow joint multi-task models to perform worse for each task than for individual models, whereas too little sharing does not allow the model to leverage knowledge efficiently among tasks. MTL's most efficient architectures are those that share well. We divide into four groups the MTL architecture [44]:

- The task-domain architectures,
- Multi-modal architectural structures,
- Architectures that have been learned and

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

- Architecture conditional.

We consider computer vision domains, NLP & improving learning for single-domain architectures. Multimodal architectures perform input tasks in more than just one mode, for example, with visual and language elements answering visual queries. It is important to note that we only accept multiple-task multimodal architectures. Architectures learned are organized between architectural learning steps such that the same calculation is done with the same task input. The architecture applied to a certain piece of data relies on the data itself in conditional architectures.

D. Two MTL Methods for Deep Learning

We now look at the two most commonly used methods of performing multiple tasks learning in deep neural networks to make MTL ideas more concrete. MTL usually takes place in the sense of DL through a common hard or soft parameter of hidden layers.

1) Hard Parameter Sharing

It is widely utilized for the MTL method in NNs is hard parameter sharing and it goes back to [45]. It is typically used to divide hidden layers between tasks while retaining multiple task-based output layers.

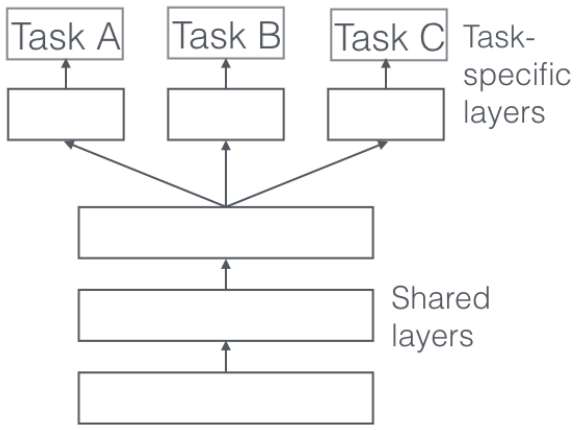

Fig. 1. Hard parameter sharing for MTL in DNNs

The use of Sharing hard parameters greatly decreases the chance of overfitting. [46] it has been shown that N, where N is no. tasks, can be less than the no. of tasks, i.e. output layers, if shared parameters may be overfitting. This has an intuitive effect: the more tasks we learn, the more our model gets an image of all tasks, the less chance we have to solve our original task.

2) Soft Parameter Sharing

The specific roles have their model with their very own parameters., instead, for soft parameter sharing. To encourage similar parameters, the distance among model parameters is then regulated. For example, [47] uses the regularization l2l2 model while [48] uses the norm trace requirement.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

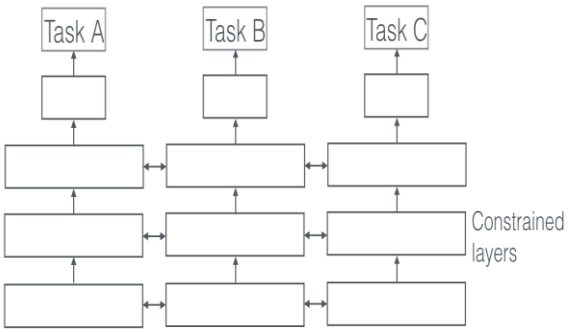

Fig. 2. Soft parameter sharing for MTL in DNN

The limitations used in deep neural networks for the exchange of soft parameters were largely driven by MTL regularization techniques developed for other models.

V. DEEP LEARNING TECHNIQUES

Deep Learning is an ML platform that has a connection to the ability to learn unsupervised from unlabeled data under the umbrella of AI. DL tools for estimation and classification of incidents e.g. HSD & opinion classification are used in various fields of data mining and text classification.

Here we should analyze essential types of deep-learning architecture:

A. RBM

RBM is an undirected two-layer, visible layer, and hidden layer neural network. No connections are available within each layer, but links are visible to hidden. The expected data log probability is maximized. It is known. The inputs are binary vectors since each input is learned from the Bernoulli distributions. The activation function is similarly calculated as the regular logistic function is used in a typical neural network from 0-1. Any neuron started if activation is higher than the random variable is treated as a chance. Visible units are taken as inputs for the hidden layer neurons. Seen neurons are the original input of binary inputs and the probabilities for the hidden layer. [49].

B. AE

AE is traditional feedforward NN that aims to learn a compressed and distributed data set representation. An AE is a network of three layers that are trained to rebuild its inputs using them as output. They need to learn features that capture data variance in terms of creating it. If linear activation functions are used that can be used for dimensional reduction, they are similar to PCA. Since the training, the activations of the hidden layer are utilized as learned features and the top layer can be dismissed [49].

C. CNN

CNN is on the verge of providing neural networks with neural learning weights & biases. With OpenCV, this has been employed for signal processing including picture identification in the realm of vision. Modular NNs give a self-contained collection of various networks that contributes to the output. NN has inputs that are distinct from the build-up as well as performance subtasks of many other networks. These networks do not interact or alert each other while performing activities. A modular NN has the advantage of breaking down a large computational function into smaller parts [49].

D. RNN

RNN works by storing layer output and feeding it back to an origin to better anticipate layer output. Even by the output of the sum of weights and features, the first layer is equivalent to the nerve distribution networks in this scenario. If this is calculated, the neural network cycle repeats, ensuring that every neuron may retain any information from one phase to

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

the next. Each neuron, like such a brain cell, functions in measures. In this manner, we will support the neural network to perform front-end propagation while keeping in mind what information it needs for future usage [50].

E. GRU Network

GRU contains two gates, which are called update & reset gates. An update gate's responsibility is to show the content of the last maintenance cell. reset gate demonstrates the transport process with the new input of previous cells. By initiating Reset Gate 1 and Updated Gate 0, the GRU represents a standard RNN. GRU model workability in comparison with LSTM is simple. It may be trained in a short period & is known to be best performed [50].

F. DSN

The deep-convex networks are also known as DSN. DSN differs from most traditional structures for DL. It's called deep because it comprises a vast number of deep networks with their secret layers in each network. The DSN claims that teaching is not a simple and isolated issue, but incorporates individual training issues. DSN contains a mixture of network modules present in the architecture. The DSN is composed of three modules. There is an input zone, secret zone & and output zone in each model module. Subroutines are located on top of each other with each module input as effects of the previous layer & authentic input vector [50].

G. LSTM

LSTM [51] has been developed and used for many applications with the efforts of Hochreiter&Schimdhuber. For speech recognition, IBM designated LSTMs mostly. The LSTM usages a memory device named a cell that is long enough just to retain its value and treat it as its input function. This allows the unit to save the last value designed. Memory unit or cell comprised of 3 gates that regulate information movement in a unit in & out of the cell.

- The port of input or gate regulates the new memory flow of data.
- The second gate that is named for forgets port controls & is applied to forget an existing data piece & to help a cell store new data.
- The output gate task is again to monitor information in a cell that is applied as cell output.

For control purposes, the weight of the cell may be applied. training method widely referred to as time backpropagation that enhances weight is important. The approach requires an optimization network performance error.

VI. LITERATURE SURVEY

DL methods have revolutionized computer vision & NLP fields in just a few years and are now a quasi-standard. In the literature on decision support recently, several DL-based approaches emerged. We analyze the following studies and differentiate between approaches to decision-making.

N. Albadi et al. (2018) Discussed extensively in Arabic Twitter the issue of detection of religious hate speech. Their work describes how they created the first public Arabic data set which was annotated for religious HSD & 1st Arabic lexicon which included terms widely used in religious discussions along with statistics that reflect their polarity & strength. They found new models of classification based upon lexicon, n-gram & DL. A complete comparison of diverse models is then seen in a brand-new unseen dataset. With 0.84 regions below receiver Operating characteristic curve, single RNN design and Gated Recurrent Units (GRUs) and performed presentations can correctly detect religious hate speech [52].

Zewdie Mossie et al. (2019) The suggested methodology explores the introduction of a vulnerable group recognition hate speech detection approach. With the Amharic text data example on Facebook, they recognize a potentially vulnerable group for social media hatred. They composed & interpreted Amharic data for the role of HSD, associated with inclusive communities like Ethiopia. They have used the distributed Apache Spark data preprocessing and functional extraction platform since social media are very noisy and big and require powerful processing resources. [53].

E. Sazany and I. Budi et al. (2019) As adaptive a DL paradigm are if it is tested for various domains on diverse text data set. This study suggests several methods of DLto recognize HS on texts from Twitter using variants of the RNN to simulate other sets of test data obtained from Facebook & Twitter. The research was performed to assess the variance in model performance among the training & test process. Experimental data revealed that both in the training phase, the GRU algorithm (85.37 percent F1) LSTM algo (76.30 percent F1) during the test period, the proposed approach outperformed the ML technique. Then proposed approach produced comparable results with the baseline method concerning the adaptability of model output [54].

H. Sohn and H. Lee (2019) The multi-channel model suggested for detecting hate speech. They compared our model to previous techniques using three non-English database sets: the 2018 SemEvalHatEval Spanish dataset, the 2018 GermEval shared goal of discovering offensive language datasets, as well as the EvalIta Ha SpeED Italian dataset. Lastly, with detailed studies, They were up to obtain new or comparable findings with these datasets. [55].

Prashant Kapil and Asif Ekbal (2020) The present paper proposes valuable data from several correlated classification tasks to enhance the efficiency of each task through a deep MTL system. This MTL model is built on a communal private system that allocates shared & private layers with 5 classification tasks, to capture shared roles & task-specific features. Experiments on the 5 data sets show that the proposed system achieves positive macro-F1 and weighted-F1 results. [56].

SafaAlsafari et al. (2020) this work is to establish an appropriate system for Arabic hate and offensive speech detection to deal with this important problem. First, by crawling Twiter data with 4 robust extraction strategies which were dependent on 4 hate forms, they generated a reliable Arabic textual corpus. First, the corpus is labeled using a three-hierarchic annotation scheme to evaluate inter-annotation agreement at each level. Based on ML and DL techniques, they have established multiple classification models in 2, 3, and 6 classes, mixing them with diverse techniques of attribute extraction, e.g. word semantic embeddings. Lastly, a detailed experiment was performed to determine the outcomes and misclassification of various learned models. Compared to previous hatred and aggressive speech research in Arabic and other languages, the results are very good [57].

PolychronisCharitidis et al. (2020) Intended toward hate speech targeted at social news accounts from reporters. This has been achieved by a group of journalists describing HS in connection with the journalist's views and the types of hate speech characteristic of journalists. Then build a huge pool of journalism tweets in a large number of languages. We are following a brief annotation technique that requires direct learning annotation processes to label a pool of unlabeled tweets according to the description. The findings of this paper are a recent series of 5 different languages open to the public for Twitter datasets. They are testing state-of-the-art hate speech analysis architectures by using our annotated data sets to train and analyze them. Finally, a detection algorithm that fits any individual model was proposed. [58].

Y. Zhou et al. (2020) An important attempt is a common method to improve overall classification results by integrating different classifier results. The emphasis is on several popular machine learning approaches to identify texts, including Embeddings from Language Model (ELMo), Bidirectional Encoder Representation from Transformer (BERT), & CNN, & these refer to SemEval2019 data sets Challenge 5. Then, to merge classifications to increase the overall classification efficiency, follow some fusion techniques. The findings indicate that classification accuracy & F1-score are significantly improved [59].

S. W. A. M. D. Samarasinghe et al. (2020) offered a DL process that makes use of two CNNs that would first categorize a given text content as hateful or not before learning how to recognize hostile content. To make choices, authorities may utilize the hate level of the text corpus to classify any material that includes hateful language. In this work, they employed FastText word embedding to turn the text data into numerical vectors for further analysis. Findings demonstrate that hate speech categorization and hate level classifications are accurate 83 percent and 60 percent, correspondingly [60].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

VII. COMPARATIVE ANALYSIS AND DISCUSSION

The comparative experiments conducted in this research are presented in this section. The following are the results of the hate speech detection using ML and DL techniques in sentiment analysis.

A. Performance metrics

Performance metrics are required to compute the experiment result for the classification models. There may be many performance measures to reflect the result of the classification.

1) Accuracy

An important performance metric is accuracy, which is no. of correctly categorized items separated by the total no.of observations.

$$\text{Accuracy} = \frac{TP + TN}{\text{N}} \dots (1)$$

2) Precision

Precision is the percentage of entries that are accurately projected positive to the total no.of positive entries. Low false positive rates are associated with a better accuracy rating.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \dots\dots (2)$$

3) Recall

The recall is the proportion of accurately anticipated positive entries to the total no.of positive entries. In essence, it shows how many positive observations were categorized properly.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \dots\dots (3)$$

4) F1 Score

False negatives and false positives are included in the F1 score since accuracy and recall are weighted together. F1 score becomes a superior performance metric than accuracy when dealing with an issue of class imbalance.

$$\text{F}_\beta = \frac{(1 + \beta^2)(\text{Precision} - \text{Recall})}{\beta^2 * (\text{Precision} + \text{Recall})} \dots\dots (4)$$

B. Comparative Results

For each model, we computed the accuracy, precision, recall, or F1 score. The computed values are shown in the following tables, as well as graphical illustrations have been utilized to demonstrate the comparison of state-of-the-art methodologies.

1) Using ML Model BoW with n-grams along with Doc2Vec and Sentiment analysis

Table IV shows how three distinct feature engineering strategies in NLP were used in this research. The top 1000 vectorized n-grams are utilized for this, and the BoW vectorizer is employed. Gensim's Doc2Vec technique was also used to build 500 feature embeddings for each text. Lastly, TextBlob was used to do sentiment analysis. All of these features are combined to generate a feature vector of 1501 characters. The 1501 features that were created are employed with Logistic Regression, Random Forest, Decision Tree, or Naive Bayes.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

TABLE IV: Classification Using Doc2vec, Bow & Sentiment Analysis [61]

Models	Accuracy	Precision	Recall	F1 Score
LR	0.9564	0.9378	0.9621	0.9498
DT	0.9502	0.9307	0.9549	0.9426
RF	0.9531	0.9231	0.9687	0.9453
NB	0.9445	0.9282	0.9447	0.9363

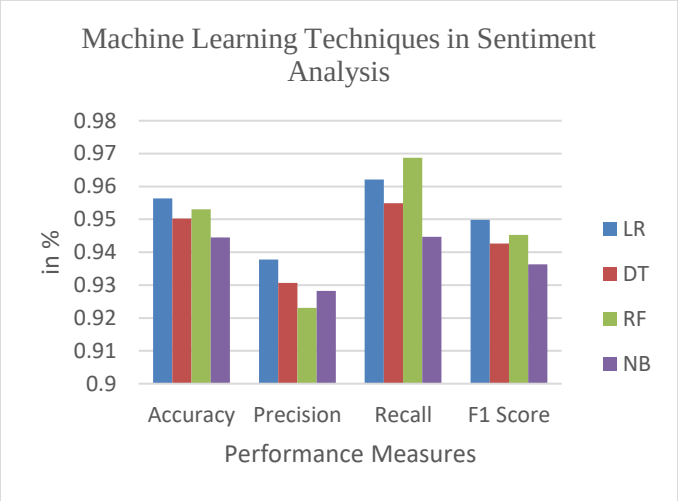

Fig. 3. Comparison graph of classification using Doc2vec, Bow, and Sentiment Analysis

Figure 3 illustrated the comparative graph to show the classification performance using different performance measures that are displayed at the x-axis and the y-axis represents the results that are measured in %. The methods are used in this for HTD in Sentiment Analysis is Doc2vec, and Bow using four different machine learning techniques. This plot shows that since the Random forest with Doc2vec, and Bow provides higher accuracy (95.31%), recall (96.87%), and f1-score (94.53%) except precision (92.31%) which is less than all other techniques like naïve Bayes (92.82%), decision tree (93.07%) and logistic regression (93.78%). But, logistic regression with Doc2vec, and Bow provides the highest accuracy (95.64%), precision (93.78%), and f1-score (94.98%) except recall (96.21%) which is less than Random forest with Doc2vec, and Bow recalls value (96.87%).

This comparative evaluation revealed that logistic regression outperforms other state-of-arts machine learning techniques.

2) Using DL Model with word-embeddings

In addition, we have employed Deep Learning-based models associated with a wide variety of word embedding methods. In the experiment described in Table V, we employed GloVe pretrained embeddings to get the results. We also used the dataset as a corpus to train our embeddings using Gensim's Word2vec [62] technique, which produced excellent results. Both of these embeddings are used in the training of an LSTM and a GRU model, respectively.

TABLE V: Classification Using DL Models With Word Embeddings [61]

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

Models	Embeddings Used	Accuracy	Precision	Recall	F1 Score
Bi-LSTM	GloVe	0.9291	0.9062	0.8992	0.9014
Bi-GRU	GloVe	0.9320	0.8881	0.9195	0.9022
Bi-LSTM	Word2vec	0.9222	0.9009	0.8978	0.8972
Bi-GRU	Word2vec	0.9209	0.8983	0.9004	0.8966

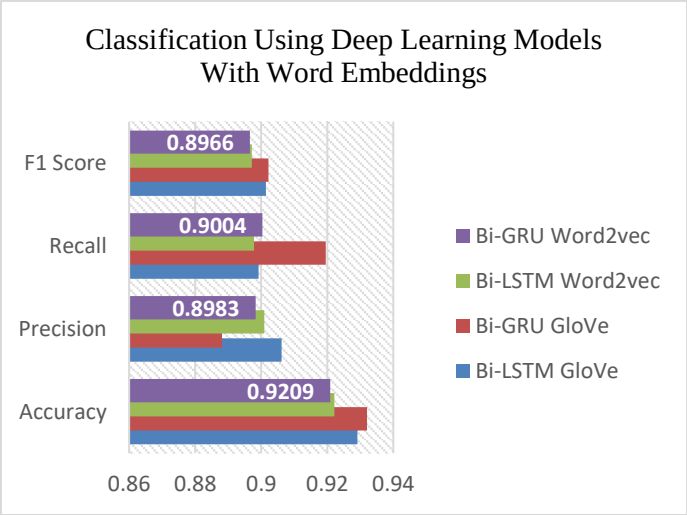

Fig. 4. Comparison graph of classification using DL methods with word embedding

Figure 4 illustrated the comparative bar graph to show the classification results using DL models with word embedding techniques. Here Bi-LSTM and Bi-GRU classification techniques are considered as DL techniques while Word2vec and GloVe methods are used as word embedding techniques. From this illustration, we can see that the Bi-GRU with GloVe provides higher accuracy (92.09%), recall (91.95%), and f1-score (90.22%) except precision (88.88%) which is less than Bi-LSTM with GloVe precision (90.62%). Similarly, Bi-LSTM with Word2vec provides higher accuracy (92.22%), precision (90.09%), and f1-score (89.72%) except recall (89.78%) which is less than Bi-GRU with Word2vec recall value (90.04%).

This observation found that however GloVe based deep learning techniques perform better than Word2vec based techniques. But Bi-GRU with GloVe outperforms these two GloVe based deep learning techniques and among all techniques.

3) Hate Speech Detection using Deep Learning techniques

For HSD, the experimental findings produced by using several DL models are examined in this following section. The CNN, DCNN, DNNs, LSTM, or Bi-LSTM models are all utilized in the DL model. We determined the numbers for accuracy, precision, recall, or F1 score for both models using the same formulas and assumptions. Table VI presents the results of the calculations.

TABLE VI: Performance Measure Scores for deep learning techniques

Models	Accuracy	Precision	Recall	F1 Score

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

LSTM [63]	0.9785	0.9598	0.9986	0.9785
Bi-LSTM [63]	0.9781	0.9582	0.9990	0.9781
CNN [64]	0.71	0.74	0.82	0.86
DCNN with k-fold [65]	0.95	0.97	0.88	0.92
DNNs	0.848	0.839	0.840	0.839

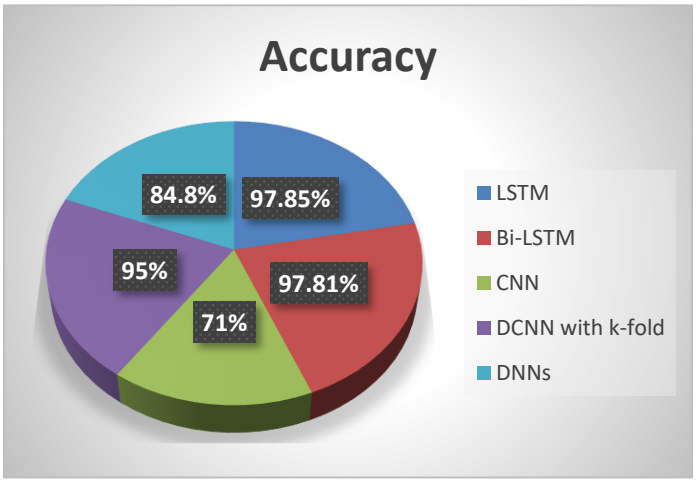

Fig. 5. Accuracy chart of deep learning techniques

Figure 5 displayed the pie chart plot to show the accuracy for hate speech detection data set using LSTM, Bi-LSTM, CNN, DCNN with k-fold and DNNs. from this chart, we can see that the least accuracy was achieved by CNN model and DNN, while others deep learning techniques achieved more than 94% accuracy. Among all of these deep learning techniques, LSTM is superior and obtained the highest accuracy which is 97.85%, and the least accuracy achieved by CNN which is 71%.

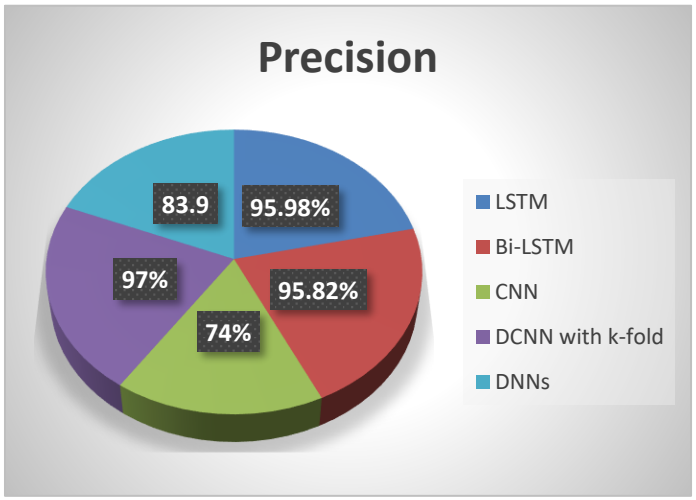

Fig. 6. Precision chart of deep learning techniques

Figure 6 displayed the pie chart plot to show the precision value for hate speech detection using five different deep learning techniques, namely, LSTM, Bi-LSTM, CNN, DCNN with k-fold, and DNNs. This pie chart illustrated that the

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

CNN has obtained the least precision which is 74% while the DNN model has achieved 83.9% precision. In comparison to these both techniques, others LSTM, Bi-LSTM, DCNN with k-fold deep learning techniques achieved the highest precision value which is more than 95%. Among all of these five deep learning techniques, DCNN with k-fold is superior and obtained the highest precision which is 97%, and the least precision achieved by CNN which is 74%.

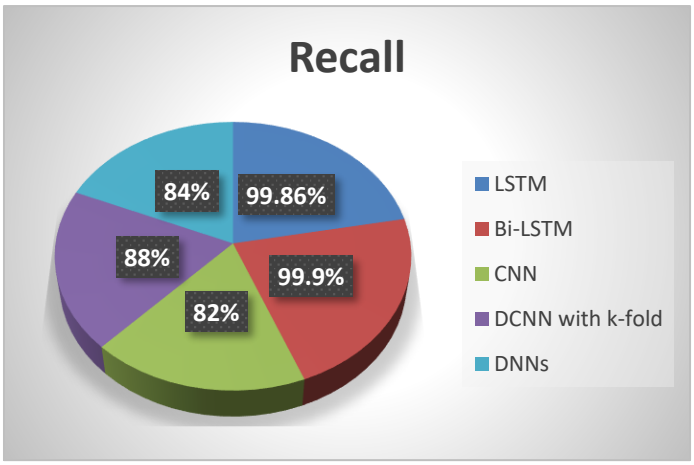

Fig. 7. Recall chart of deep learning techniques

Figure 7 visualized the pie chart plot to show the best recall value for hate speech detection from LSTM, Bi-LSTM, CNN, DCNN with k-fold and DNNs. From this pie, chart representation found that the DCNN with k-fold achieved high recall value (88%) than DNN (84%), and CNN (82%) while two remaining LSTM and Bi-LSTM deep learning techniques achieved more than 99% recall value. Among all of these deep learning techniques, Bi-LSTM is superior and obtained the highest recall value which is 99.9% and the least recall value achieved by CNN which is 82%.

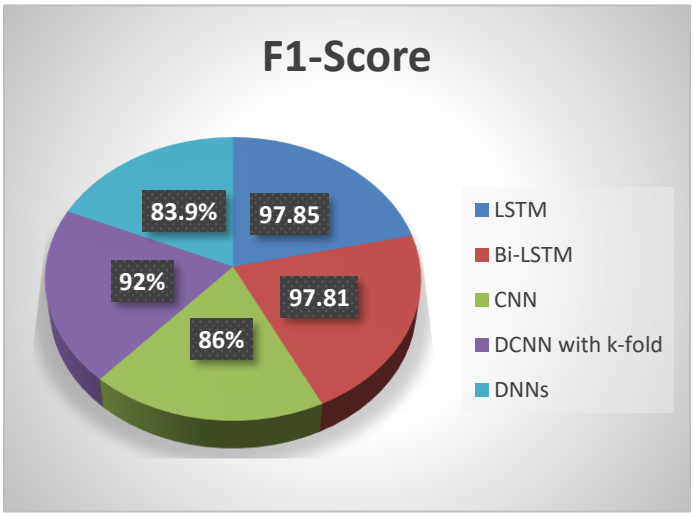

Fig. 8. Recall chart of deep learning techniques

Figure 8 displayed the pie chart plot to show the f1-score for detecting hate speech using five different deep learning techniques. This pie chart demonstrated that the DNN has obtained the least f1-score which is 83.9% while the CNN model has achieved an 86% f1-score. In comparison to these both techniques, the other three deep learning techniques LSTM, Bi-LSTM, DCNN with k-fold achieved the highest f1-score which is more than 92%. Among all of these five deep learning techniques, LSTM is superior and obtained the highest f1-score which is 97.85%, and the least precision achieved by DNN which is 83.9%.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

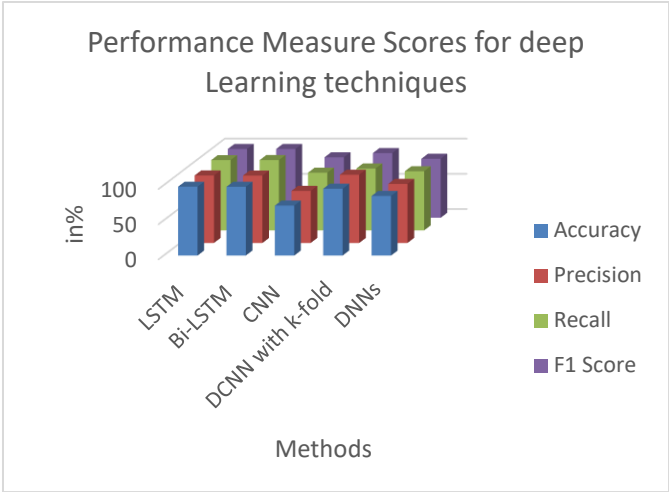

Fig. 9. Comparison graph of performance metrics with deep learning techniques

Figure 9 displayed a comparative 2D plot for HDSwith DL techniques to the calculated performance measures. The accuracy for the CNN model is 0.71, precision is 0.74, recall is 0.82 & f1-score is 0.86, respectively, which is very minimal whereas the accuracy for the DNNs model is 0.84, precision is 0.83, recall is 0.84 and f1-score is 0.83, respectively. We may conclude from findings acquired using DL-based models that the model with a fixed split of train & test samples failed to achieve sufficient performance in detecting HS tweets. Subsequently, the k-fold cross-validation procedure is performed, with k set to 10. For HS tweet predictions, the DCNN model with a filter size of 4g produced accuracy, recall, & F1-score values of 0.97, 0.88, & 0.92, respectively. The estimated accuracy, precision, & f1 values indicate that LSTM outperformed Bi-LSTM. However, Bi-LSTM has a higher recall score than LSTM. The recall is defined as the proportion of positive classification to total positive classification. In this research, we classified hate speech as positive. That implies the model detects hate speech with less error. Bi-LSTM offers a minor advantage over LSTM in this scenario. Despite this, the difference in scores is too minor to draw any comparison between the two models.

VIII. CONCLUSION

The growing prevalence of HD in social media in recent years has been seen as a significant worldwide issue. Numerous governments & organizations, who have attracted the attention of the science world, have made important developments in methods for hate speech detection. Although there is a great deal of literature on this issue, the performance of each method is still hard to evaluate, as each has its advantages and drawbacks.es. The lack of resources to identify online hate material effectively is one of the biggest obstacles to stopping this crime. The techniques for DL including MTL are discussed in this work.DL presents the 2 most common MTL methods. There are different DL techniques have discussed that are used to detect hate speech. Also, a comparative analysis has been made using different machine learning and deep learning techniques using word embedding methods for hate speech detection. This comparison is useful for further detecting hate and offensive speech and implementing a new model.

This analysis is useful to perform learning with multitask using deep learning techniques to accurately detect hate speeches from social media or microblogging data.

References

[1] Elizaveta Zinovyeva, Wolfgang Karl Härdle, Stefan Lessmann, Antisocial online behavior detection using deep learning, Decision Support Systems, Volume 138, 2020, 113362, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2020.113362.

[2] Sujata Khedkar, Subhash Shinde, Deep Learning and Ensemble Approach for Praise or Complaint Classification, Procedia Computer Science, Volume 167, 2020, Pages 449-458, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.03.254.
[3] Deyu Zhou, Meng Zhang, Yang Yang, Yulan He, Hierarchical state recurrent neural network for social emotion ranking, Computer Speech & Language, Volume 68, 2021, 101177, ISSN 0885-2308, https://doi.org/10.1016/j.csl.2020.101177.
[4] N.A. Ghani, S. Hamid, I.A.T. Hashem, E. Ahmed, Social media big data analytics: A survey, Comput. Hum. Behav. 101 (2019) 417–428.
[5] J.Q. Dong, C.H. Yang, Business value of big data analytics: A systemstheoretic approach and empirical test, Inf. Manage. 57 (1) (2020) 103124.
[6] F. Atefeh, W. Khreich, A survey of techniques for event detection in twitter, Comput. Intell. 31 (1) (2015) 132–164.
[7] J. Earl, R.K. Garrett, The new information frontier: toward a more nuanced view of social movement communication, in: Social Movement Studies, Taylor & Francis, 2016, pp. 1–15.
[8] R.Z. Medina, J.C.L. Diaz, Social media use in crisis communication management: An opportunity for local communities? in: Social Media and Local Governments, Springer International Publishing, 2016, pp. 321–335.
[9] X. Lu, Online communication behavior at the onset of a catastrophe: an exploratory study of the (2008) wenchan earthquake in China, Nat. Hazards 91 (2) (2018) 785–802.
[10] Bernhard Kratzwald, SuzanaIlić, Mathias Kraus, Stefan Feuerriegel, Helmut Prendinger, Deep learning for affective computing: Text-based emotion recognition in decision support, Decision Support Systems, Volume 115, 2018, Pages 24-35, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2018.09.002.
[11] B. Vidgen, T. Yasseri, "Detecting weak and strong islamophobic hate speech on social media", J. Inf. Technol. Polit. 17 (1) (2020) 66–78.
[12] J.T. Nockleby, Hate speech, in: Leonard W. Levy, Kenneth L. Karst, et al. (Eds.), Encyclopedia of the American Constitution, vol. 3, second ed., Macmillan, New York, 2000, pp. 1277–1279.
[13] N.D. Gitari, Z. Zuping, H. Damien, J. Long, A lexicon-based approach for hate speech detection, Int. J. Multimedia Ubiquit. Eng. 10 (4) (2015) 215–230.
[14] K. Lee, B.D. Eoff, J. Caverlee, Seven months with the devils: A long-term study of content polluters on Twitter, in: International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 2011.
[15] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in: Proceedings of the 20th International Conference on World Wide Web, WWW '11, ACM, New York, NY, 2011, pp. 675–684
[16] M. F. Wright, B. D. Harper, and S. Wachs, ''The associations between cyberbullying and callous-unemotional traits among adolescents: The moderating effect of online disinhibition,'' J. Personality Individual Differences, vol. 140, pp. 41–45, Apr. 2019.
[17] T. Davidson, D. Warmsley, M. Macy, and I. Weber, ''Automated hate speech detection and the problem of offensive language,'' in Proc. ICWSM, 2017, pp. 1–4.
[18] D. Gershgorn and M. Murphy. (2017). Facebook is Hiring More People to Moderate Content than Twitter has at Its Entire Company Quartz. Accessed: Jun. 20, 2019. [Online]. Available: https://bit.ly/2ZbhsHu
[19] T. Vega, ''Facebook says it failed to bar posts with hate speech,'' The New York Times, 2013. Accessed: Jun. 10, 2019. [Online]. Available: https://nyti.ms/2VXy9Ex
[20] R. Meyer, ''Twitter's famous racist problem,'' The Atlantic, 2016. Accessed: Jul. 5, 2019. [Online]. Available: https://bit.ly/38EnFPw.
[21] F. Rodríguez-Sánchez, J. Carrillo-de-Albornoz and L. Plaza, "Automatic Classification of Sexism in Social Networks: An Empirical Study on Twitter Data," in IEEE Access, vol. 8, pp. 219563-219576, 2020, doi: 10.1109/ACCESS.2020.3042604.
[22] Z. Ryan Shi, C. Wang, and F. Fang, ''Artificial intelligence for social good: A survey,'' 2020, arXiv:2001.01818. [Online]. Available: http://arxiv.org/abs/2001.01818
[23] A. Khatua, E. Cambria, and A. Khatua, ''Sounds of silence breakers: Exploring sexual violence on Twitter,'' in Proc. ASONAM, 2018, pp. 397–400.

[24] A. Khatua, E. Cambria, K. Ghosh, N. Chaki, and A. Khatua, ‘‘Tweeting in support of LGBT? A deep learning approach,’’ in Proc. ACM India Joint Int. Conf. Data Sci. Manage. Data, 2019, pp. 342–345.
[25] E. Cambria, P. Chandra, and A. Hussain, ‘‘Do not feel the trolls,’’ in Proc. SDoW Workshop 9th Int. Semantic Web Conf., 2010, pp. 1–12.
[26] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, ‘‘Suicidal ideation detection: A review of machine learning methods and applications,’’ IEEE Trans. Comput. Social Syst., early access, Sep. 17, 2020, doi: 10.1109/TCSS.2020.3021467.
[27] Z. Waseem, ‘‘Are you a racist or am i seeing things? Annotator influence on hate speech detection on Twitter,’’ in Proc. 1st Workshop NLP Comput. Social Sci., 2016, pp. 138–142.
[28] Z. Waseem and D. Hovy, ‘‘Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter,’’ in Proc. NAACL Student Res. Workshop, 2016, pp. 88–93.
[29] G. Xiang, B. Fan, L. Wang, J. Hong, and C. Rose, ‘‘Detecting offensive tweets via topical feature discovery over a large scale Twitter corpus,’’ in Proc. 21st ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2012, pp. 1980–1984.
[30] P. Gianfortoni, D. Adamson, and C. Rose, ‘‘Modeling of stylistic variation in social media with stretchy patterns,’’ in Proc. EMNLP, 2011, pp. 49–59.
[31] E. Cambria, ‘‘Affective computing and sentiment analysis,’’ IEEE Intell. Syst., vol. 31, no. 2, pp. 102–107, Mar. 2016
[32] F. Del Vigna, A. Cimino, F. Dell’Orletta, M. Petrocchi, and M. Tesconi, ‘‘Hate me, hate me not: Hate speech detection on facebook,’’ in Proc. ITASEC, 2017, pp. 86–95.
[33] N. D. Gitari, Z. Zhang, H. Damien, and J. Long, ‘‘A lexicon-based approach for hate speech detection,’’ Int. J. Multimedia Ubiquitous Eng., vol. 10, no. 4, pp. 215–230, Apr. 2015.
[34] Y. Tang and N. Dalzell, ‘‘Classifying hate speech using a two-layer model,’’ Statist. Public Policy, vol. 6, no. 1, pp. 80–86, Jan. 2019, doi: 10.1080/2330443X.2019.1660285.
[35] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, ‘‘Deep learning for hate speech detection in tweets,’’ in Proc. ACM WWW, 2017, pp. 759–760.
[36] X.-S. Vu, T. Vu, M.-V. Tran, T. Le-Cong, and H. T. M. Nguyen, “HSD shared task in VLSP campaign 2019: Hate speech detection for social good,” in Proceedings of VLSP 2019, 2019.
[37] H. Pham and L.-H. Phuong, “End-to-end recurrent neural network models for vietnamese named entity recognition: Word-level vs. characterlevel,” in The 15th International Conference of the Pacific Association for Computational Linguistics, 2017.
[38] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, May 2011.
[39] V. A. Ho, D. H.-C. Nguyen, D. H. Nguyen, L. T.-V. Pham, D.-V. Nguyen, K. Van Nguyen, and N. L.-T. Nguyen, “Emotion recognition for vietnamese social media text,” arXiv preprint arXiv:1911.09339, 2019.
[40] Z. Zhang and L. Luo, “Hate speech detection: A solved problem? the challenging case of long tail on twitter,” Semantic Web, 2018.
[41] Femi Emmanuel Ayo, Olusegun Folorunso, Friday Thomas Ibharalu, Idowu Ademola Osinuga, Machine learning techniques for hate speech classification of twitter data: State-of-the-art, future challenges and research directions, Computer Science Review, Volume 38, 2020, 100311, ISSN 1574-0137, https://doi.org/10.1016/j.cosrev.2020.10031
[42] Evgeniou, T., Micchelli, C. A., and Pontil, M. (2005). Learning multiple tasks with kernel methods. Journal of Machine Learning Research, 6:615–637.
[43] Sebastian Ruder, “An Overview of Multi-Task Learning in Deep Neural Networks”, arXiv:1706.05098v1 [cs.LG] 15 Jun 2017, pp. 1-14.
[44] Michael Crawshaw, “Multi-Task Learning with Deep Neural Networks: A Survey”, arXiv:2009.09796v1 [cs.LG] 10 Sep 2020, pp. 1-43.
[45] Caruana, R. "Multitask learning: A knowledge-based source of inductive bias." Proceedings of the Tenth International Conference on Machine Learning. 1993.
[46] Baxter, J. (1997). A Bayesian/information theoretic model of learning to learn via multiple task sampling. Machine Learning, 28, 7–39. Retrieved from http://link.springer.com/article/10.1023/A:1007327622663

[47] Duong, L., Cohn, T., Bird, S., & Cook, P. (2015). Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers), 845–850.

[48] Yang, Y., &Hospedales, T. M. (2017). Trace Norm Regularised Deep Multi-Task Learning. In Workshop track - ICLR 2017. Retrieved from http://arxiv.org/abs/1606.04038

[49] V. Pream Sudha and R. Kowsalya, "A Survey on Deep Learning Techniques, Applications and Challenges", International Journal of Advance Research in Science and Engineering (IJARSE), Vol. No.4, Issue 03, March 2015, pp. 311-317.

[50] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari and Gulshan Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning", Archives of Computational Methods in Engineering, 2019, pp. 1-23. https://doi.org/10.1007/s11831-019-09344-w.

[51] D. Goularas and S. Kamis, "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 2019, pp. 12-17, doi: 10.1109/Deep-ML.2019.00011.

[52] N. Albadi, M. Kurdi and S. Mishra, "Are they Our Brothers? Analysis and Detection of Religious Hate Speech in the Arabic Twittersphere," 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018, pp. 69-76, doi: 10.1109/ASONAM.2018.8508247.

[53] Zewdie Mossie, Jenq-Haur Wang, Vulnerable community identification using hate speech detection on social media, Information Processing & Management, Volume 57, Issue 3, 2020, 102087, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2019.102087.

[54] E. Sazany and I. Budi, "Hate Speech Identification in Text Written in Indonesian with Recurrent Neural Network," 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS), Bali, Indonesia, 2019, pp. 211-216, doi: 10.1109/ICACSIS47736.2019.8979959.

[55] H. Sohn and H. Lee, "MC-BERT4HATE: Hate Speech Detection using Multi-channel BERT for Different Languages and Translations," 2019 International Conference on Data Mining Workshops (ICDMW), Beijing, China, 2019, pp. 551-559, doi: 10.1109/ICDMW.2019.00084.

[56] Prashant Kapil, Asif Ekbal, A deep neural network based multi-task learning approach to hate speech detection, Knowledge-Based Systems, Volume 210, 2020, 106458, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2020.106458.

[57] SafaAlsafari, Samira Sadaoui, Malek Mouhoub, Hate and offensive speech detection on Arabic social media, Online Social Networks and Media, Volume 19, 2020, 100096, ISSN 2468-6964, https://doi.org/10.1016/j.osnem.2020.100096.

[58] PolychronisCharitidis, Stavros Doropoulos, Stavros Vologiannidis, IoannisPapastergiou, Sophia Karakeva, Towards countering hate speech against journalists on social media, Online Social Networks and Media, Volume 17, 2020, 100071, ISSN 2468-6964, https://doi.org/10.1016/j.osnem.2020.100071.

[59] Y. Zhou, Y. Yang, H. Liu, X. Liu and N. Savage, "Deep Learning Based Fusion Approach for Hate Speech Detection," in IEEE Access, vol. 8, pp. 128923-128929, 2020, doi: 10.1109/ACCESS.2020.3009244.

[60] S. W. A. M. D. Samarasinghe, R. G. N. Meegama and M. Punchimudiyanse, "Machine Learning Approach for the Detection of Hate Speech in Sinhala Unicode Text," 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 2020, pp. 65-70, doi: 10.1109/ICTer51097.2020.9325493.

[61] T. Dhamija, Anjum, and R. Katarya, "Comparative Analysis of Machine Learning and Deep Learning Algorithms for Detection of Online Hate Speech," 2021, doi: 10.1007/978-981-16-0942-8_48.

[62] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013.

[63] C. Paul and P. Bora, "Detecting Hate Speech using Deep Learning Techniques," Int. J. Adv. Comput. Sci. Appl., 2021, doi: 10.14569/IJACSA.2021.0120278.

Volume 14, No. 1, 2023, p. 29-48
https://publishoa.com
ISSN: 1309-3452

[64] P. Rani, S. Suryawanshi, K. Goswami, B. R. Chakravarthi, T. Fransen, and J. P. McCrae, "A Comparative Study of Different State-of-the-Art Hate Speech Detection Methods in {H}indi-{E}nglish Code-Mixed Data," Proc. Second Work. Trolling, Aggress. Cyberbullying, 2020.

[65] P. K. Roy, A. K. Tripathy, T. K. Das, and X. Z. Gao, "A framework for hate speech detection using deep convolutional neural network," IEEE Access, 2020, doi: 10.1109/ACCESS.2020.3037073.