Bangla Interrogative Sentence Identification from Transliterated Bangla Sentences

Md Montaser Hamid
Computer Science and Engineering
Shahjalal University of Science and Technology
Sylhet, Bangladesh
montaserhamid13@gmail.com

Tanvir Alam
Computer Science and Engineering
Shahjalal University of Science and Technology
Sylhet, Bangladesh
tanviralam997@gmail.com

Sabir Ismail
Computer Science and Engineering
Stony Brook University
New York, United States
sabir.ismail@stonybrook.edu

Md Forhad Rabbi
Computer Science and Engineering
Shahjalal University of Science and Technology
Sylhet, Bangladesh
frabbi-cse@sust.edu

***Abstract*—In this paper, we propose a method to identify Bangla interrogative sentences from transliterated Bangla sentences. All over the internet, we generate a huge number of Bangla interrogative sentences and they are mostly written using transliteration. In transliterated Bangla, identifying interrogative sentences possesses great challenges. The question marks at the end of the interrogative sentences are not used in transliterated Bangla profoundly especially on social media. Moreover, people often make interrogative sentences without using Bangla question words when writing in transliterated form. To find the solution, we discuss the rule-based approach, supervised learning approach and a deep learning approach. In the rule-based approach, we design a set of rules based on grammar and data analysis. For employing supervised learning, we use machine learning techniques such as Support Vector Machine, k-Nearest Neighbors, Multilayer Perceptron and Logistic Regression and achieved accuracies of 91.43%, 75.98%, 92.11% and 91.68% respectively. To apply deep learning, we implement a Convolutional Neural Network. This approach provides a decent result with an accuracy of 84.64% and dignifies the scope of Convolutional Neural Network as an ideal model for Bangla natural language processing.**

Keywords*—*Interrogative sentence identification, transliterated Bangla, Convolutional Neural Network

I. Introduction

A transliterated form of language is defined as the form in which alphabet of one language is used for representing another language [1]. For example, *"Ami bari jabo"* is a Bangla transliterated sentence using the Latin alphabet. The use of transliteration is huge in many languages. The Latin alphabet is used in western and eastern European languages extensively. The Latin alphabet is also used in Turkey, Vietnam, Somalia and in east African Swahili Language [2].

We use the transliterated form to write Bangla in various sectors. The maximum use of the transliterated Bangla sentences is in social media, like Facebook, Instagram and twitter. Another sector where this transliterated form is used extensively is in chatting applications, like Messenger, WhatsApp, Viber etc. We also use this form to write in online blogs and web portals.

In all these sectors mentioned earlier, we are generating a considerable number of interrogative sentences every day. The identification of interrogative sentences is very much important for various reasons. In data analytics, this identification can play a significant role. Service providers can study the client behavior, expectations, demands, queries from the interrogative sentences posted by the clients in the form of Facebook status, comments, tweets, online blog posts, live chats etc. This identification is also important for the development of smart assist applications, medical applications, question-answering based applications, user-interactive applications, chatbot programs, 24/7 question-answering services and so on.

In transliterated Bangla, the interrogative sentences come in various forms. In many cases these sentences differ to great extent from the traditional grammatical format of the interrogative sentences. Many people do not use question mark at the end of the interrogative sentences. This common practice makes the challenge of identification very hard. It is found that almost 30% of the online questions do not contain any question mark at the end [3]. One important thing is that the question mark at the end of the sentence does not necessarily express questioning property all the time. For instance, *Tomar eto boro spordha?*. This sentence expresses exclamation rather than questioning property. Another example is that people often omit the Bangla question words (*ke, kokhon, kar* etc.) while writing interrogative sentences. For instances, *Apni jacchen tahole?, Apni shotti chole jaben?* etc. There are some sentences where the presence of question word does not indicate the questioning property rather the question words may act as a linker. Example of such sentences taken from our dataset are given bellow:

- Ami kokhon bari jabo ta ekhono janina
- ki kori tate karo kichu ashe jay na
- Tini janena tara ki karone asheni
- khela koto tarikhe hobe ta ekhono jana jay ni

In this paper, for the identification of interrogative sentences, we have the rule based approach, supervised learning based approach and the deep learning approach. In the rule based approach we have followed the rules of grammar and some rules which we have made by observing and analyzing the dataset. The drawback of this approach is that due to variations in sen-

978-1-5386-8207-4/18/$31.00 ©2018 IEEE

tences no set of rules can identify the interrogative sentences accurately and efficiently. For better accuracy and efficiency, we have used supervised learning. In supervised learning, We have used the State Vector Machine(SVM), Logistic Regression, Multilayer Perceptron(MLP) and k-Nearest Neighbors(k-NN) algorithm for the classification and identification. Lastly, we have used Convolutional Neural Network(CNN) for the deep learning approach.

II. Related Works

Though, the topic of text classification and identification is common for English language, related works regarding this topic is sparse in Bangla language. In English, the work on detection and retrieval of questions is done based on the contents generated from blogs, web portals, twitter and emails. A good number of work has been done for question detection from Community Question Answering(CQA). Deep learning for the classification of sentences is also highly practiced.

The question and query detection are mostly used and analyzed for developing online question answering Services. The most famous services are: Apple's Siri, Microsoft's Cortana, IBM Watson, Wolfram Alpha and Google search engine. They detect and extract answer by following IR (Information Retrieval)based approach, knowledge-based approach and hybrid approach [4].

Rule based approach is the most time-worn process of detecting questions. Efron et al. [5] implemented a rule-based approach which is able to detect and analyze the questions asked in microblogging environment like twitter. The authors have made a taxonomy of questions from huge collections of questions taken from twitter.

Wang et al. [6] applied a learning-based approach based on lexical and syntactic features that are employed to detect questions retrieved from Community Based Question Answering Services (CQA). Sequential patterns of the sentences are mined to detect the questions. The sentences are decomposed into a stream of tokens. In sequential pattern extraction, the Part-of-Speech (POS) tag of all the tokens of a sentence are used. Support Vector Machine (SVM) algorithm is used for the classification of the questions and non-questions. Li et al. [7] used another learning based approach that is employed to identify questions from twitter. The prefix span algorithm is used for mining the frequent question patterns and SVM is also employed to differentiate the questions.

For designing question answering systems, deep learning is a very aspiring and efficient way. With the formation of Recurrent Neural Networks(RNN), it is possible to analyze longer text [8]. Here, RNN models and end-to-end memory networks are used for designing question answering system. By forming Convolutional Neural Network of one layer of Convolution with the help of word2vec and static vectors, it is possible to design a state-of-the-art language independent sentence classifier and question classifier [9], which performs better than most of the classifiers with the help of tuning the hyperparameters.

Razzaghi et al. [10] employed machine learning techniques SVM and Naive Bayes to detect Frequently Asked Question(FAQ). Information Gain (IG), ChiSquared Attribute Evaluation(Cfs), and CfsSubset (Chi) methods are used for feature selections. Syntactic features, question words, semantic features, bag of words are used for forming the set of features.

Banerjee et al. [11] used multiple models such as Naive Bayes, Kernel Naive Bayes, rules induction and decision tree classifiers to classify Bangla questions. They have achieved an accuracy of 91.65%.

Wang et al. [12] used Naive Bayes and Support vector Machines for sentiment classification. They have employed different variants, methods and features varying from the traditional approaches to improve the sentiment and topic classification.

Yin et al. [13] compared CNN and RNN for different natural language processing tasks and depicted the basis of selection between these two neural networks.

Kalchbrenner et al. [14] made a dynamic CNN for language independent sentence modelling without any dependence of parse tree. They tested their model on various tasks like sentiment predictions, question classification.

Liu et al. [15] designed multi-task systems and associated multi-task learning concepts with the RNN. They have used RNN with multi task frameworks to classify texts.

III. Methodology

In the transliterated form of Bangla sentences, a lot of things are taken under consideration. In this form, concrete grammatical rules or hand-picked rules for the identification will not be good enough for obtaining better accuracy and efficiency. The variations and informalities in the sentences are extremely high therefore a learning-based approach is a must for addressing the problem in a greater extent. For this purpose, we have used supervised learning and deep learning.

A. Rule Based Approach

Rule based approach is the conventional way to address the solution of the problem. With the combination of grammar and in-depth analysis of the dataset, we have designed a set of rules for the identification purpose. We have done some feature extraction to design the rules. The analysis behind the design of these rules are discussed in section IV. The position of the Bangla question words serves great importance in our rule-based approach. Therefore, we have made a Bangla question word list which contains 20 words. These question words are listed in Figure 1.

Ki, Keno, Kivabe, Kothay, Koto, Kar, Kon, Kobe, Kisher, Kokhon, Ke, Kemon, Koy, Ke Ke, Kake, Kara, Kader, Koi, Koyta, Kotha.

Fig. 1. Bangla Question Words

As the question mark is not that much significant and often gets omitted, we have excluded the presence of question mark as a rule. The rules are as follows:

- **Rule 1:** A Bangla question word is present as the first or last word of a sentence.
- **Rule 2:** A Bangla question word is present as the second word of the sentence and the first word is the subject or object of the sentence.
- **Rule 3:** A Bangla question word is present just before the last word of the sentence.
- **Rule 4:** The word "Naki" is considered as a question word and the sentence with this word follows the previous rules.

The findings and results of this approach have given us the insights that the designing and implementation of rule-based approach is not pragmatic. Thus we have deviated ourselves from this approach and given our focus on the learning-based approaches. The findings of this approach are discussed in section V.

B. Supervised Learning Approach

As mentioned earlier the identification can't be done with accuracy and efficiency by only following a set of hand-picked rules. For better performance a learning-based approach is mandatory. We have employed the following supervised machine learning techniques for the identification of interrogative and non-interrogative sentences:

- Support Vector Machine (SVM)
- Logistic Regression(LR)
- Multilayer Perceptron(MLP)
- k-Nearest Neighbors (k-NN)

These machine learning algorithms are very common for text classification and identification problems in Natural Language Processing(NLP). Our main challenge was to employ them in a dataset for transliterated Bangla sentences and extracting the ideal features. We have used lexical features from the datasets for all the machine learning techniques employed here. The results of this approach are described in section V.

C. Deep Learning Approach

For classification and identification, using deep learning is a modern and dynamic approach. For employing deep learning we have used Convolutional Neural Network (CNN) for classification. We have tried to replicate the model of CNN described in [9]. We have followed this tutorial [16] for implementing the CNN.

We need to embed the words for CNN. The first layer of the network is used to embed the words into low-dimensional vectors. This layer is a learning step. The embedded words act as a lookup table. The second layer is used for convolution. Embedded words are used for performing convolutions. The result of this convolution layer is max-pooled to get the feature vector and dropout regularization.

In the implemented network the value of the dimensional embedding is 128. The filter size is taken as 3,4 and 5 which means that the convolutional filter will cover 3,4 and 5 words respectively. All the sentences are padded to have the same length which is 59. The number of filters per filter size in the network is 128. The batch size is 64 and dropout is 0.5. There are two classes (interrogative and non-interrogative) in the output layer of this network and the hyperparameters of the CNN are not tuned for the test set.

IV. Data Analysis

A. Making Corpora

The dataset we have worked on has been formed by extracting 44,538 comments from cricket-based Facebook public groups using a web application [17].

1) Making of Primary Corpus for Rule-Based Approach: The comments we have extracted from facebook groups contain various types of sentences. We have excluded the comments written in English language and in Bangla alphabet and made a dataset and named it as the *Primary Corpus*. From this corpus we have excluded the sentences with the question mark and made the *Interrogative Mega Corpus*. From this corpus we have taken the unique distinct sentences without any similarities between each other and formed the *Interrogative Corpus* and the rest of the primary corpus is named as the *Other Mega Corpus*. We have used these corpora for the rule-based approach. The information about the corpora is reported in TABLE I.

2) Making Corpus for Learning Based Approaches: For employing supervised learning, we have scrutinized the primary corpus and omitted a large number of sentences which were very raw in nature and had anomalous content. We have manually picked out the interrogative and non-interrogative Bangla transliterated sentences from the primary corpus to form our *Cricket Domain Corpus*. Implementing SVM, k-NN, MLP and logistic regression classifiers on this corpus, we have obtained good results. In this experiment, the test dataset and the training dataset have the same type of contents which are mainly cricket related comments. From the result, we have realized that the result was influenced by the common domain of the training dataset and test dataset. Therefore, we tried to introduce data from another domain in our dataset. We have collected data for designing a chatbot for university admission

TABLE I
MAKING CORPORA

Corpus	**Number of Sentences**	**Total Number of Words**
Primary Corpus	145,009	4,29,883
Interrogative Mega Corpus	4,624	23,785
Other Mega Corpus	25,259	121,224
Interrogative Corpus	700	3,073

test for our another project. This university admission test dataset contains queries, questions and the corresponding answers regarding university admission test which are frequently asked by the applicants. This dataset follows proper and standard form of transliterated sentences. We have taken the queries and the questions as the interrogative sentences and the answers as the non-interrogative sentences and formed a new dataset. We have named this dataset as *University Admission Corpus*. After adding this corpus with the Cricket Domain Corpus we have got our *Mixed Domain Corpus*. The information about these three corpora is reported in TABLE II.

B. Analyzing Corpora

1) Interrogative Corpus Analysis: This corpus is the most important corpus for our experiments in rule-based approach. In this corpus, We have calculated that the average number of words is 4.7, the average length of words is 4 and average number of letters is 19 per sentences.

We have also calculated the position of the question words of Figure 1 in the interrogative corpus. This calculation is reported in TABLE III. We have also seen that the word *"Naki"* is used as the question word in 24 sentences. It is to be noted that 149 sentences out of 700 sentences from this corpus do not contain any Bangla question words.

2) Other Mega Corpus Analysis: In this corpus all the sentences without question mark remain present. Like the interrogative corpus, We have calculated that the average number of words is 5, the average length of words is 4 and average number of letters is 11 per sentences.

V. Experiments & Results

By applying 3 different approaches we have achieved distinct results. The findings of our experiments are discussed in this section.

A. Evaluation of Rule Based Approach

From the set of rules prescribed in section III.A, we have tested the interrogative corpus. The main basis of this approach is the position of the Bangla question words in the sentence as written in TABLE III. We exclude the number of occurrence of question words as the last word of the sentences from the other positions. Then we find the mean position of the Bangla question words in the sentence which is 1.90. At first, we have tested the corpus according to the rule 1, then gradually we integrated the other rules and observe the combined effects of the other rules in the improvement of accuracy. The evaluation is given in TABLE IV.

From this level of accuracies, from a known small dataset we can say the rule-based approach is not satisfactory. As the number of rules can't be made concrete due to the diversity and varieties of sentences. From TABLE III, we have seen that the maximum likely position of Bangla question word is the last word of the sentence. In the sentence *Tader jete bollam kothay ar tara gelo kothay*, the question word *Kothay* is at the last position but the sentence is not an interrogative sentence whatsoever. But according to our rule-based approach, this sentence will be regarded as an interrogative sentence as the question word *Kothay* is present as the last word of the sentence.

From this, we say that the rule-based approach is not

TABLE II
ALL CORPORA

Corpus	**Total Sentences**	**Interrogative Sentences**	**Non-Interrogative Sentences**
Cricket Domain Corpus	8797	1704	7093
University Admission Corpus	2993	2434	559
Mixed Domain Corpus	11790	4138	7652

TABLE III
QUESTION WORD POSITION

Position of Bangla Question Word in Sentences	**Number of Sentences**	**Percentage of Sentences with the Question Word**
1st Word	112	16%
Last word	171	24.43%
2nd Word	167	23.86%
3rd Word	37	5.29%
4th Word	4	0.57%
5th Word	4	0.57%
6th Word	2	0.29%
7th Word	1	0.14%
8th Word	1	0.14%
Just Before the Last Word	52	7.43%

TABLE IV
RULE BASED APPROACH EVALUATION

Method	**Accuracy %**
Rule 1	40.42
Rule 1 + Rule 2	64.29
Rule 1 + Rule 2 + Rule 3	71.71
Rule 1 + Rule 2 + Rule 3 + Rule 4	75.14

pragmatic and there should be a learning based approach for addressing the identification and classification.

B. Evaluation of Supervised Learning Approach

We have analyzed the Cricket Domain Corpus and the Mixed Domain Corpus with the help of SVM, k-NN, MLP and logistic regression classifiers. At first, we have used the Cricket Domain Corpus where the training and test set have come from the same domain. Then we have taken the Mixed Domain Corpus where the training and test set differs in terms of subject domain.

1) Result of Cricket Domain Corpus: To evaluate the Cricket Domain Corpus, we have taken 30% of the data as the test set and the rest as the training set. The distribution of the training and test dataset is represented in Table V.

Though the domain is same, it has huge variations as it is taken from real people. We have labeled the test set and the training set then we have found out the accuracy for SVM, k-NN, MLP and logistic regression classifiers. The training time was really small. The accuracy level we have observed is described in Table VI.

2) Result of the Mixed Domain Corpus: The Mixed Domain Corpus contains data from two domains, one is cricket another is hand picked data for university admission test related queries. We have employed the same four classifiers for this corpus. This time admission test based dataset was the training set and the cricket based dataset was the test set. The result we have found from this corpus is stated in Table VI.

3) Assessment of the results: We observe that the accuracy of the Cricket Domain Corpus differs from the accuracy of the Mixed Domain Corpus. It happened due to the change of the domain of the training dataset in the Mixed Domain Corpus. As the training dataset(admission dataset) contains formal sentences and sentences with standard form of transliteration, it can not handle all the real time variations of the sentences of the Cricket Domain Corpus, as a result the accuracy label drops. With a standard form of test dataset our model of classification will work swiftly but with the standard test dataset the real time variations of the sentences will be ignored.

TABLE V
DISTRIBUTION OF THE CRICKET DOMAIN CORPUS FOR SUPERVISED LEARNING

Dataset	**Number of Interrogative Sentences**	**Number of Non-Interrogative Sentences**
Training Dataset	1193	4965
Test Dataset	511	2128

TABLE VI
RESULT OF SUPERVISED LEARNING APPROACH

Classifier	**Accuracy for Cricket Domain Corpus**	**Accuracy for Mixed Domain Corpus**
SVM	90.36%	82.64%
k-NN	80.66%	80.59%
MLP	90.36%	74.66%
Logistic Regression	90.18%	79.79%

C. Evaluation of the Deep Learning Approach

To implement CNN, we have used our Mixed Domain Corpus as this corpus contains the maximum number of sentences. 10% of the data is taken as the test dataset and other 10% of the data is taken as the validation set. 80% of the corpus is taken for the training dataset. The distribution of the test set, validation set and training set is described in TABLE VII.

Interrogative and *Non-Interrogative* are the two output classes of our network. For the validation dataset we have observed an accuracy of **85.77%**. For the test dataset, the accuracy of CNN is **84.64%**. If we use pretrained wor2vec vectors for the embedding matrices and a training dataset with more sentences we can gain more accuracy from this approach.

D. Classifying Sentences Using Our CNN Model

Using CNN, We have become successful in classifying the following sentences. I indicates interrogative class and NI indicates the non-interrogative class.

- ajke khela kokhon hobe bolte parben -I
- apnar desher bari kothay -I
- apni koto din dhore ei kaj korchen -I
- amra shobai besh valo achi -NI
- apni ki koren ta diye amader kichu ashe jay na -NI
- shobaike diye ki ar shobkichu korano jay -NI
- apnar naam ta bole jabe ki -NI
- tader bashay kothay sheta ki tumi jano -I

We can even classify sentences like *"ajka kala kokhan"*, *"ame ajka kala dekba"* which are extreme cases of transliterated form of Bangla considering the popular standard way of

TABLE VII
DISTRIBUTION OF THE MIXED DOMAIN CORPUS FOR CNN

Dataset	**Number of Interrogative Sentences**	**Number of Non-Interrogative Sentences**
Training Dataset	3312	6122
Validation Dataset	413	765
Test Dataset	413	765

TABLE VIII
ACCURACY COMPARISON BETWEEN THE SUPERVISED LEARNING AND THE DEEP LEARNING APPROACHES

Approach	**Accuracy for the Test Dataset**
SVM	91.43%
k-NN	75.98%
MLP	92.11%
Logistic Regression	91.68%
CNN	84.64%

spelling. From this result of classification, we can say that CNN can identify transliterated Bangla sentences efficiently.

E. Evaluation of All Approaches

All our three approaches show significant and insightful results. To compare the accuracies of the supervised learning approach to the deep learning approach, we have tested all the four classifiers of the supervised learning approach with the test dataset of the deep learning approach. For training the classifiers, we have merged the validation and the training dataset of the Mixed Domain Corpus of the CNN. This experiment gives proper comparison between all the approaches as the test dataset is same for all the approaches. We have excluded the rule-based approach from this experiment as it has no futuristic aspect. The result of this experiment is shown in Table VIII.

The results of the supervised learning and the deep learning approaches dignifies the scope of machine learning techniques and the deep neural network as the potential solution for identifying transliterated Bangla interrogative sentences.

VI. CONCLUSION

In this paper, we have discussed the challenges and difficulties in the process of designing a system to identify Bangla interrogative sentences from transliterated Bangla sentences. As less works have been done in this area the problem has remained complex and vast. We have introduced three approaches to address the solution of this problem. The results of the experiments demonstrate that the rule based approach is not suitable for identification purpose. Supervised learning approach have given us insightful results which gives the idea that the applied classifiers can do the work of identification of the sentences with decent accuracy. But this process has also shown sensitivity with the domain of the training and test dataset. With the implementation of deep learning approach, we have successfully demonstrated the efficiency of Convolutional Neural Network (CNN) in identification and classification of Bangla transliterated interrogative sentences. We are currently working on employing Recurrent Neural Network(RNN), another state-of- the-art artificial neural network model for the identification purpose. We are also trying to develop a more diversified dataset with various and critical examples of both interrogative and other form of transliterated sentences. We are looking forward to designing question-answering system for transliterated Bangla using the insight of this paper. The datasets and the codes are available at: https://goo.gl/wa1PqY .

REFERENCES

[1] "What is transliteration?," [Online]. Available: http://www.translitteration.com/what-is-transliteration/en/. [Accessed: June 8, 2018].

[2] "Index of languages by writing system," [Online]. Available: http://www.omniglot.com/writing/langalph.htm. [Accessed: June 8, 2018].

[3] C. Gao, C.-Y. Lin, Y.-I. Song, and Y. Sun, "Finding question-answer pairs from online forums," in The 31st Annual International ACM SIGIR Conference 20-24 July 2008, Singapore. ACM, 2008, pp.467-474.

[4] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 2nd ed., New Jersey: Prentice Hall, 2009.

[5] M. Efron and M. Winget, "Questions are content: a taxonomy of questions in a microblogging environment," in Proceedings of the American Society for Information Science and Technology 22 - 27 October, 2010, Pittsburgh, Pennsylvania, vol. 47. ACM, 2010.

[6] K. Wang and T.-S. Chua, "Exploiting salient patterns for question detection and question retrieval in community-based question answering," in Proceedings of the 23rd International Conference on Computational Linguistics 23 - 27 August, 2010, Beijing, China. ACM, 2010, pp. 1155-1163.

[7] B. Li, X. Si, M. R. Lyu, I. king, and E. Y.Chang, "Question identification on twitter," in Proceedings of the 20th ACM international conference on Information and knowledge management 24 - 28 October, 2011, Glasgow, Scotland, UK. ACM, 2011, pp. 2477-2480.

[8] "Question Answering Using deep learning," 2016. [Online]. Available: https://cs224d.stanford.edu/reports/StrohMathur.pdf. [Accessed: May 13, 2018].

[9] Y. Kim, "Convolutional neural networks for sentence classification," arXiv:1408.5882, 2014.

[10] F. Razzaghi, H. Minaee, and A. A. Ghorbani, "Context free frequently asked questions detection using machine learning techniques," in International Conference on Web Intelligence 13-16 October, 2016, USA. IEEE, 2016.

[11] S. Banerjee and S. Bandyopadhyay, "An empirical study of combining multiple models in bengali question classification," in Proceedings of International Joint Conference on Natural Language Processing (IJCNLP), Japan, 2013, pp. 892-896.

[12] S. Wang and C.D.Manning, "Baselines and bigrams: simple, good sentiment and topic classification," in proceedings of ACL, 2012, 90-94.

[13] W. Yin, K. Kann, M. Yu, and H. Schtze, "Comparative study of CNN and RNN for natural language processing," arXiv:1702.01923v1, 2017.

[14] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," arXiv:1404.2188, 2014.

[15] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," arXiv:1605.05101v1, 2016.

[16] "Implementing a CNN for Text Classification in TensorFlow," [Online]. Available: http://www.wildml.com/2015/12/implementing-a-cnn-for-text-classification-in-tensorflow/. [Accessed: August 8, 2018].

[17] "Facebook Groups Analytics & Management Tools," [Online]. Available: https://grytics.com/. [Accessed: June 9, 2018].