A New Technique to Classification of Bengali News Grounded on ML and DL Models

Tamim Al Mahmud
Department of Computer Science and Engineering
Green University of Bangladesh
Dhaka, Bangladesh

Sazeda Sultana
Department of Computer Science and Engineering
Green University of Bangladesh
Dhaka, Bangladesh

Antara Mondal
Department of Computer Science and Engineering
Green University of Bangladesh
Dhaka, Bangladesh

ABSTRACT

News classification is the process of categorizing news articles into predefined categories based on their content. It involves utilizing techniques such as machine learning and NLP to analyze the textual features of news articles and assign them to relevant categories. This enables efficient organization and retrieval of news articles, aiding users in accessing information based on their interests. Many topics are covered by news, including local, national, and international news, politics, business, sports, and entertainment.

In our native Bengali language, there are numerous classifications for news. The motivation for working in this research is because of the limited research in Bengali text and the scarcity of resources. This study proposed a new technique for the categorization of the news based on a comparison among different machine learning and deep learning model results. First, we have collected more than 20K data. Afterwards, applied several machine learning and deep learning models to pre-processed and cleaned data. Then apply our proposed technique on evaluated results produced by the machine learning and deep learning models to make a final decision to identify the news category.

In our proposed technique, we counted the predicted categories from three types of Machine Learning models and two types of Deep Learning models. Finally, we computed the maximum predicted class and assigned the category for the specific news.

General Terms

Sentiment, Machine Learning, Natural Language Processing, Artificial Intelligence, Deep Learning

Keywords

Newspaper, Classification, Pre-Processing, Data Visualization, Splitting Dataset, Tokenization, Support Vector Machine, Gated Recurrent Unit, Singular Vector Decomposition, Long Short-Term Memory, Bag of Word, Linear Support Vector Classification, Random Forest

1. INTRODUCTION

News classification is necessary for internet users to get their desired news in real-time. This process saves the user's valuable time. In this modern era, all newspapers maintain the news category to attract their target customers to find their desired news easily. Further, every newspaper has an online portal where they post some news at intervals. So, manually identifying the category of specific news is time-consuming. However, machine learning techniques come with a blessing to identify the news category easily. Our research aims to explore various machine learning algorithms and techniques for news classification in Bengali text. We will analyze the performance of these algorithms and compare them to identify the most effective one. Additionally, we will also investigate the impact of different feature extraction methods on the accuracy of news classification. This research will not only provide insights into the application of machine learning in Bengali text but also contribute to the development of automated news classification systems that can aid in information retrieval and analysis. The results of this study can be useful for journalists, researchers, and organizations that deal with large volumes of news articles in Bengali text. Overall, our research will pave the way for future developments in Bengali text analysis using machine learning and natural language processing.

This research looks at how different Machine Learning and Deep Learning classifier techniques can help with discoveries. In the classifications task, the machine learning classification algorithm assists in quickly locating the appropriate category of news and removing distracting and harmful content news. There is some work done in this area and all the researchers used either machine learning techniques or deep learning techniques. Our proposed technique is the use of maximum count category decision-making can greatly enhance the accuracy of machine learning and deep learning models. This technique can be applied to a wide range of industries, including finance, healthcare, and marketing. By accurately categorizing data, businesses can make more informed decisions and improve their overall performance. Additionally, this technique can help to reduce the risk of errors and inaccuracies in data analysis. As technology continues to advance, it is important that we explore new methods for improving the accuracy and efficiency of machine learning models. The use of maximum count category decision making is just one example of how we can achieve this goal and continue to push the boundaries of what is possible with AI.

As a result, our proposed system received an improved accuracy of 8.5% compared to each of the machine learning and deep learning models. These results provide us encouragement to apply this method to another research in this field. The following list outlines the main contributions made by this research represent to achieving the goals:

(1) Conducting an in-depth investigation in the researched field to find the latest works.
(2) Identified a new technique that can classify Bengali text easily with improved accuracy.
(3) Collected more than 20K data and structured these data for processing.
(4) Make a comparison and select the best model among available machine learning and deep learning models.

The remaining sections are arranged as follows: section 2 discusses the previous works in this field, section 3 illustrates and discusses the working methodology, section 4 result analysis and evaluation and section 5 contains the conclusion.

2. RELATED WROK

The state-of-the-art efforts in the research field are summarized in this section. We evaluated recent research and made an effort to determine its advantages as well as its drawbacks.

They proposed combining supervised learning algorithms with NLP methods to categorize the forms of the Bengali language. The authors used a decision tree classifier and the SVM with a linear kernel. Their task does not provide much accuracy and the task was not valid enough as they only used two models [1]. The author proposed Bengali News classification using Graph Convolutional Networks. The authors modified graph convolution networks (GCN) and applied BiLSTM, GRU-LSTM, and LSTM. The work process took too much time for classifying the Bengali documents [2]. The researcher is proposed a way that can classify Bengali articles using a new comprehensive Dataset. In this work, the authors used a deep learning model, and support vector machine classifier algorithms. Their task does not work well on unstructured data [3]. The author used an RSS crawler for each site and generated Full-Text-RSS. They used only the Naive Bayes algorithm. Thus, their work was not good enough as they did not use KNN, SVM and other traditional Machine Learning models [4]. They proposed a technique based on the comparative analysis of different machine learning models. They used a scheme to enlarge the dataset which can be used to increase accuracy [5]. The author's proposed the classification of Bengali News articles using Bi-LSTM. They used ANN LSTM, BiLSTM model. They did not use the tag generation from the headline. As a result, they received lower accuracy [6]. The authors used features with SVM, LR, KNN, DT, and NN. However, they did not add the sequence of words and words to the vectors [7]. In this paper, the authors tried to show the Bangla crime type Classification in the Bangla language. They used CNN and RN classifiers. However, they were unable to show specific crime information [8]. In order to identify Bangla false news, the authors employed the supervised machine learning techniques SVM and MNB classifiers [9]. Authors used deep learning, Bi-LSTM, and RF for their works. They choose the best classifiers for Bangla text's false news identification [10]. In this paper, the authors calculated the distance or not similarity between the pairs of newspapers. They found the news with the most unique information [11]. Authors prepared an automatic word generator and did three different modified techniques of hierarchical clustering on the same cluster. However, they did not use any traditional machine learning models [12]. To identify false Bangla News, they suggested a deep hybrid learning approach. The Authors found successfully differences between fake news and real news [13]. They proposed an aspect-based sentiment analysis for Bangla News headlines. They used the small training classifiers dataset for this analysis in logistic regression, SGD classifier, SVM, and Random Forest [14]. The authors used different classification models. They also added predictors. However, they were unable to find original fake news [15]. They used different types of classifiers and found the model performance in an acceptable stage [16]. The author's trying to identify the Bangla satirical news spreading between the different online news. In this paper, they used only the hybrid feature [17]. In this paper authors, used the TF-IDF technique with different term frequency thresholds as their feature selection and they got acceptable results [18]. The authors show here the differences between applied algorithms, and which one is performing better. They used only text classification [19]. In this paper, the authors showed Naïve Bayes better than the other three algorithms. Here they did not use any regression [20]. The authors used only CNN BiLSTM for sentiment analysis and received good accuracy [21]. Several research articles [22]-[26] have been conducted on the Bengali language across diverse fields.

3. METHODOLOGY

All the language processing systems should follow some algorithm rules and methodology. After analyzing many ideas and methods we have prepared our solving system where we build and carry out experiments in our system. Our proposed methodology architecture is displayed in fig-1.
This work mostly focused on the headlines and major news. Additionally, we divided the news into four groups, including entertainment, sports, politics, and health. Out of the 20000 data, we obtained 16213 trained data. We used Deep Learning and Machine Learning algorithms in our system architecture, shown in fig-1.

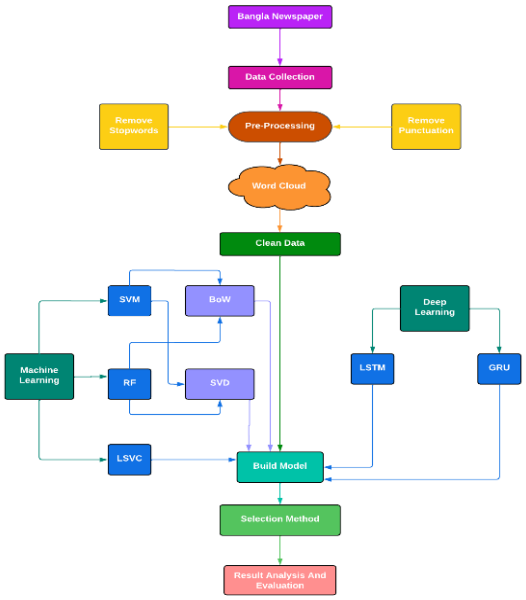

Figure 1: Our Proposed Architecture

3.1 Data Collection

We need a huge amount of data. So that we select some of the daily newspapers and collect more news from the database. we pick randomly news from the newspaper. We choose headlines and main news. We collect four types of news. They are politics, sports, entertainment, and health. We collect more than 20K data from different newspapers. When we complete our full data set, we prepare this set in CSV format and then put it in the drive where we locate it. The percentage of collecting data from 18 different newspapers is displayed in the pie chart in fig-2. We accumulated and made a dataset containing a total of 20K data. The dataset has column serial, Headlines, Main News, Category, and Source. Based on the news category we labelled each data into one of four categories Sports, Politics, Entertainment and Health. fig-3, displayed the Bar chart of each category of data.

3.2 Data Processing

After collecting the dataset then we make a system to process the whole data set. The pre-processing has been done with the following steps:
(1) **Text Cleaning**: After setting the CSV file then we clean our dataset We apply the function [27] to the data frame then we

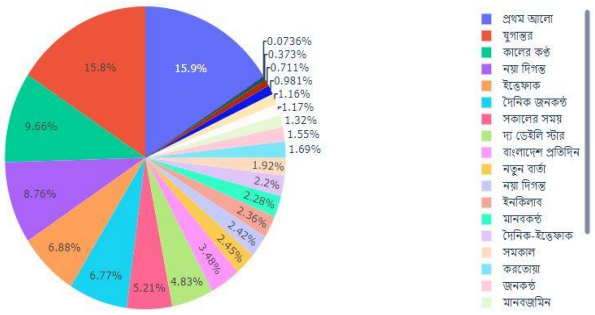

Figure 2: Different Sources in a Pie Chart

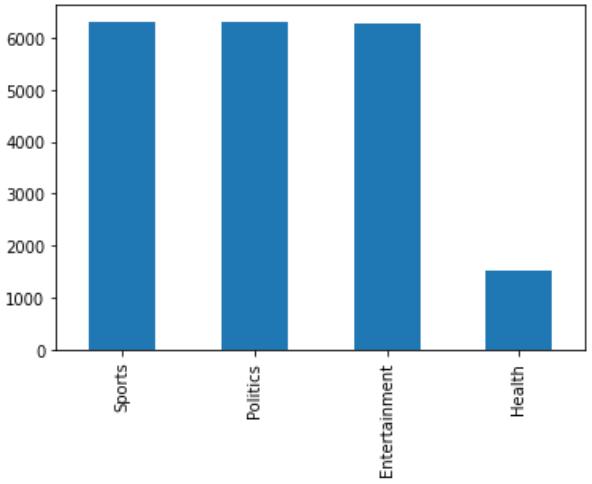

Figure 3: Bar Chart of different News Category

remove unnecessary symbols and characters. Punctuation marks can be irrelevant or unnecessary for certain text analysis tasks. Removing them can help simplify the text and reduce noise, making it easier to process and analyze.

(2) **Tokenization**: In Natural Language Processing, tokenization involves splitting text into individual units or tokens, such as words or phrases. Removing punctuation before tokenization can help ensure that punctuation marks are not considered as separate tokens, leading to more accurate and meaningful tokenization results. We must eliminate the dataset of unnecessary punctuation. Such as: (, !, , #, , .., ..., ..., (,), , , [,], /, etc).

(3) **Stop-words**: Stop-words are common words that are often considered insignificant and do not carry much meaning in the context of NLP and text analysis tasks. These are the terms that frequently appear in words but don't really mean much. The words must be eliminated because they are no longer necessary in the context of sentiment analysis. Words include: ও", " এবং", "েকন", "িকন্তু", "অথবা", " করেবা" etc. Afterwards, we remove low- length data which we do not want to carry. Pre-processing in natural language processing (NLP) offers advantages such as improved accuracy by reducing noise, enhanced efficiency through data standardization, and better generalization by eliminating biases, resulting in more accurate and efficient analysis and modeling.

Splitting Dataset: Data splitting, also known as dataset splitting, is a common technique used in machine learning and data analysis to divide a dataset into multiple subsets for different purposes. The main goal of data splitting is to create distinct subsets that can be used for training, validation, and testing. We split the dataset into training and test categories. The training data contains 80% of the original data and the test dataset contains the remaining 20%. Finally, we filter out the data outliers.

3.3 Processing Method

We used singular value decomposition and bag of words processing techniques with the machine learning models random forest and support vector machine.

Singular Value Decomposition: Singular Value Decomposition is a matrix factorization technique widely used in linear algebra and numerical analysis. It decomposes a given matrices, allowing for efficient analysis and manipulation of the original matrix. It is also powerful matrix factorization technique that finds application in various fields, including data analysis, signal processing, image compression and machine learning.

Bag of Words: The Bag of Words model is an easy and popular method for representing text data in natural language processing, information retrieval, and other related fields. It disregards the order and structure of words in a document and focuses solely on the occurrence or frequency of individual words, treating each document as an unordered collection.

3.4 Data Visualization

In our data analysis part, we analyze the whole dataset, and we count words, unique words, and most frequent words. Then we visualization full of data statics. We visualized the statistics for all four categories Politics, Entertainment Sports and Health. Here we illustrate the results in fig-4 only for the Entertainment category. Figure 4 illustrates the Entertainment category's total number of news counts. Then the maximum, min and mean length among all news. The figure also illustrates the standard deviation of the news length.

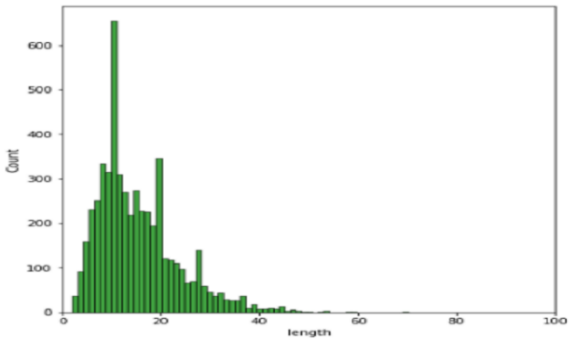

Figure 4: Distribution of Text Lengths

Word Cloud: The size of each word in a word cloud, sometimes referred to as a tag cloud or wordle, is based on how frequently or how important it occurs in the text. It is a popular data visualization technique used to display textual information in an aesthetically appealing and easily digestible manner. We also generate the word clouds for each category. Word-cloud is the image where it represents the most frequently occurring words in the text are displayed with larger font sizes, while less frequent words appear smaller fig-5 displayed the word clouds for the Entertainment category.

Here in this Figure 5, the most frequent words are "ছিব ", "গান ", "কথা" etc. These words are very common in Bengali movies and any kind of entertainment. However, we could not represent the exact Bengali word for some cases in the word cloud. We are still working on producing the exact word cloud for Bengali text.

Figure 7: Visualizing Data into Word-Clouds

3.5 Training Models

In this part, we train our dataset with different Machine Learning and Deep Learning models. Three machine learning models and two deep learning models were used. In respect of deep learning, we use the Gated Recurrent Units and Long Short-Term Memory. Here we use Keras with the gated recurrent units and long short-term memory model for training data. We used machine learning models such as linear support vector classification, support vector machine and random forest.

Support Vector Machine: Support Vector Machine (SVM) is a supervised machine learning technique used for classification and regression applications called the vector machine. By locating an ideal hyperplane that optimally divides several classes in the feature space, it establishes a decision boundary. By decreasing classification errors and increasing the distance between data points and the decision border, SVM seeks to achieve good generalization.

Random Forest: Random Forest is a machine learning combination learning method that combines various decision trees to generate predictions. To increase accuracy and decrease overfitting, a set of decision trees is built, and their outputs are combined. Random Forest utilizes random feature selection and bootstrapped data samples to enhance diversity and robustness in the ensemble model.

Linear Support Vector Classification: For binary classification problems, the Support Vector Machine (SVM) variation known as Linear Support Vector Machine (SVM) employs a linear decision boundary. It looks for the ideal surface that divides the classes by the greatest margin. Because of its ease of use and strong generalization capabilities, linear SVM is useful, efficient, and frequently utilized in a variety of applications.

We used two types of algorithms in deep learning models. One is long short-term memory (LSTM) which can retain information for a long time in the memory. Another model is gated recurrent unit (GRU).

Long Short-Term Memory: A recurrent neural network (RNN) architecture called Long Short-Term Memory is created to get over the problem of vanishing gradients and identify long-term dependencies in sequential input. The basic goal of LSTM is to create a memory cell that can selectively recall or forget information across a range of sequential cycles. As a result, the network can retain information for a very long time, making it particularly useful for work involving lengthy data sequences.

Gated Recurrent Unit (GRU): For modeling sequential data, such as time series or natural language processing tasks, recurrent neural network architectures such as the gated recurrent unit (GRU) are used. Its structure is similar to that of the more well-known long short-term memory network, although it uses a more basic filtering system. The reset gate and update gate are its two primary parts. By limiting the information flow within the network, these gates enable the network to selectively modify and restore the hidden state.

Our classifiers would then be generated the results using the gated recurrent unit, long short-term memory, linear support vector classification, support vector machine (bag of word, singular vector decomposition), and random forest (bag of word, singular vector decomposition). Then we find out the accuracy of different classifiers. Initially, we applied the class's default parameters. In the later stage, we tuned the parameters to get better results for each model.

The parameters are calculated based on the classification data. Then we got accuracy, and we find out which one provides the best accuracy.

3.6 Proposed Model

Our research target was to beat the benchmark result. So, we proposed a new technique that can improve accuracy by about 8.5% compared to each of the machine learning and deep learning models we used in our task. The idea is to count the predicted class for each news produced by the machine learning and deep learning models. Then, make a decision for the class based on the maximum count. The proposed methodology's algorithm overflow is displayed in Algorithm 1.

Algorithm 1: Our Proposed Election System Algorithm

INPUT: News Data N
OUTPUT: A_c = (P, E, H, S) where A_c = Assigned Class,
P = Politics, E= Entertainment, H= Health, and S =Sports
Initialization: $Sum_P = 0$, $Sum_E = 0$, $Sum_H = 0$, $Sum_S = 0$
For each category C_i in results
if C_i is Politics then
$Sum_P \leftarrow Sum_P + 1$
else if C_i is Entertainment then
$Sum_E \leftarrow Sum_E + 1$
else if C_i is Health then
$Sum_H \leftarrow Sum_H + 1$
else if C_i is Sports then
$Sum_S \leftarrow Sum_S + 1$
end if
For each News N_i
if $Sum_P = MAX(Sum_P , Sum_E , Sum_H , Sum_S)$ then
$A_c \leftarrow$ Politics
else if $Sum_E = MAX(Sum_P , Sum_E , Sum_H , Sum_S)$ then
$A_c \leftarrow$ Entertainment
else if $Sum_H = MAX(Sum_P , Sum_E , Sum_H , Sum_S)$ then
$A_c \leftarrow$ Health
else if $Sum_S = MAX(Sum_P , Sum_E , Sum_H , Sum_S)$ then
$A_c \leftarrow$ Sports
end if $Com_i \leftarrow$ Negative

end if

4. RESULT ANALYSIS & EVULATION

Result evaluation is the main characteristic of every research observation. We have taken four metrics "Precision", "Recall", "Accuracy" and "F- measure for the performance evaluation of different machine learning and deep learning models.

Precision- A binary classification performance statistic called precision assesses how accurately positive predictions are made. This ratio is derived as the sum of true positives and false positives divided by the true positive prediction rate. Greater accuracy reflects the model's dependability in positively

categorizing situations and a lower rate of false positive prediction.

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

Recall- Recall is an indicator of performance that measures how well a model can identify all relevant instances of a given class when performing classification tasks. In other words, it is the percentage of true positives to the total of true positives plus false negatives. Less missed class instances are indicated by higher recall.

$$Recall = \frac{TP}{(TP+FN)}$$

Accuracy- The total accuracy of a model's prediction can be evaluated using accuracy, a performance indicator used in classification tasks. It measures the proportion of predictions that were accurate compared to those that were made as a whole. A model's total performance is frequently assessed using its accuracy, which reflects the percentage of accurate predictions.

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)}$$

F-measure- A statistic used to evaluate the effectiveness of binary classification models is the F-measure. It creates a single score by combining precision (accuracy of positive predictions) and recall (ability to discover positive occurrences). By considering both false positives and false negatives, it offers an accurate evaluation of a model's efficiency.

$$F\text{-}measure = 2 * \frac{Precision \times Recall}{(Precision + Recall)}$$

where
(1) The amount of accurately categorized positive events is referred to as TP (True Positives).
(2) The amount of accurately categorized negative events is referred to as TN (True Negatives).
(3) The amount of incorrectly categorized positive events is referred to as FP (False Positives).
(4) The amount of incorrectly categorized negative events is referred to as FN (False Negatives).

In this task, we used support vector machine (bag of words), support vector machine (singular value decomposition), random forest (bag of words), random forest (singular value decomposition), linear support vector classifier, gated recurrent unit, and long short-term memory. We calculated the accuracy, precision, recall, and f-measures. The models' results are shown in Table 1.

Table 1: raining Results of SVM, Random Forest, Linear SVC, LSTM and GRU on Dataset

Training Results							
Statistic Measure	SVM (BoW)	SVM (SVD)	RF (BoW)	RF (SVD)	Linear SVC	LSTM	GRU
Accuracy	87.12	85.46	84.63	83.52	85.95	83.39	84.51
Precision	85.02	83.26	82.59	82.24	83.57	81.37	83.45
Recall	82.39	80.86	80.31	77.44	82.50	82.39	82.77
F-Measure	83.23	81.57	81.22	79.03	82.66	82.44	82.56
Algorithm for our proposed technique as a result : **95.45**							

Result Analysis: In this article, we used a support vector machine with bag of word, support vector machine with singular value decomposition, random forest with bag of word, random forest with singular value decomposition, linear support vector classifier. And find we find accuracy, precision, recall, f-measures. For support vector machine (singular value decomposition), we found accuracy is 87.12, precision is 85.02, recall is 82.39, f-measure is 83.23. Which is our best accuracy in machine learning algorithm. For SVM(SVD), we found accuracy is 85.46, precision is 83.26, recall is 80.86, f-measure is 81.57. For random forest (bag-of- word), we found accuracy is 84.63, precision is 82.59, recall is 80.31, and f-measure is 81.22. For random forest (singular value decomposition), we found accuracy is 83.52, precision is 82.24, recall is 77.44, f-measure is 79.03. Linear support vector classifier (LSVC), we found accuracy is 85.95, precision is 83.57, recall is 82.50, and f-measure is 82.66. For Long short-term memory (LSTM), we found accuracy is 83.39, precision is 81.37, recall is 82.39, and f-measure is 82.44 in deep learning. For gated recurrent unit (GRU), we found accuracy is 84.51, precision is 83.45, recall is 82.77, and f-measure is 82.56 in deep learning. In our proposed technique we achieved the maximum accuracy which is 95.45.

5. CONCLUSION

The main target of our research investigation is to enhance the research on Bengali text as well as increase the resources in this domain. About 300 million people are talking in Bengali language and it is the most spoken native language number fifth. However, there is not much research on this language as compared to other languages. We used a unique election technique to classify the different news categories and the technique accuracy topped compared to three Machine learning techniques Support Vector Machine (Bag of Words), Support Vector Machine (Singular Value Decomposition), Random Forest (Bag of Words), Random Forest (Singular Value Decomposition) and Linear Support Vector Classifier and two Deep learning techniques Gated Recurrent Unit, and Long Short-Term Memory. This work will surely enrich the research in Bangla texts. Thus, it motivates further research in this field. Therefore, it will be a little contribution from us in this regard.

6. ACKNOWLEDGMENTS

Our thanks to the reviewers for their scholarly review and help to enhance the quality of the article. Furthermore, thanks to the experts who have contributed towards development of the template.

7. REFERENCES

[1] Abdul Bari Parves, Abdullah Al Imran, and Md. Riazur Rahman. 2020. Incorporating Supervised Learning Algorithms with NLP Techniques to Classify BengaliLanguage Forms. In Proceedings of the International Conference on Computing Advancements (ICCA 2020). Association for Computing Machinery, New York, NY, USA, Article 62,1–7. https://doi.org/10.1145/3377049.3377110

[2] M. M. Rahman, M. A. Z. Khan and A. A. Biswas, "Bangla News Classification using Graph Convolutional Networks," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, https://doi.org/10.1109/ICCCI50826.2021.9402567

[3] M. Tanvir Alam and M. Mofijul Islam, "BARD:Bangla Article Classification Using a New Comprehensive Dataset," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018, pp. 1-5, https://doi.org/10.1109/ICBSLP.2018.8554382

[4] A. N. Chy, M. H. Seddiqui and S. Das, "Bangla news classification using naive Bayes classifier," 16th Int'l Conf. Computer and Information Technology, 2014, pp. 366-371, https://doi.org/10.1109/ICCITechn.2014.6997369.354–355

[5] Rifat Rahman. 2020. A Benchmark Study on Machine Learning Methods using Several Feature Extraction Techniques for News Genre Detection from Bangla News Articles & Titles. In Proceedings of the 7th International Conference on Networking, Systems and Security (NSysS '20). Association for Computing Machinery, New York, NY, USA, 25–35. https://doi.org/10.1145/3428363.3428373

[6] M. M. H. Shahin, T. Ahmmed, S. H. Piyal and M.Shopon, "Classification of Bangla News Articles Using Bidirectional Long Short Term Memory," 2020 IEEE Region 10 Symposium (TENSYMP), 2020, pp. 1547-1551, https://doi.org/10.1109/TENSYMP50017.2020.9230737

[7] S. Rahman, S. K. Mithila, A. Akther and K.M.Alam, "An Empirical Study of Machine Learning-based Bangla News Classification Methods," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021,pp. 1-6, https://doi.org/10.1109/ICCCNT51525.2021.9579655

[8] S. Tabashum, M. M. Hossain, A. Islam, M. Y. Mahafi Taz Zahara and F. N. Fami, "Performance Analysis of Most Prominent Machine Learning and Deep Learning Algorithms In Classifying Bangla Crime News Articles," 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 2020, pp. 1273-1277, https://doi.org/10.1109/TENSYMP50017.2020.9230785

[9] M. G. Hussain, M. Rashidul Hasan, M. Rahman,J.Protim and S. Al Hasan, "Detection of Bangla Fake News using MNB and SVM Classifier," 2020 International Conference on Computing, Electronics Communications Engineering (iCCECE), 2020, pp. 81-85, https://doi.org/10.1109/iCCECE49321.2020.9231167

[10] Rachna Jain, Deepak Kumar Jain, Dharana, and Ntika Sharma. 2021. Fake News Classification: A Quantitative Research Description. ACM Trans. Asian Low-Resour. Lang.Inf. Process. 21, 1, Article 3 (January2022), 17 pages. https://doi.org/10.1145/3447650

[11] A. A. Babu, A. Hasan and S. Ismail, "Suggesting an Informative Order of Bangla Follow up News on a Specific Issue from Several Days Newspapers," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018, pp. 1-5, https://doi.org/10.1109/ICBSLP.2018.8554538

[12] K. Wohiduzzaman and S. Ismail, 'Recommendation System for Bangla News Article with Anaphora Resolution," 2018 4th International Conference on Electrical Engineering and Information Communication Technology (iCEEiCT), 2018, pp. 467-472, https://doi.org/10.1109/CEEICT.2018.8628075

[13] Q. A. R. Adib, M. H. K. Mehedi, M. S. Sakib, K. K. Patwary, M. S. Hossain and A. A. Rasel, "A Deep Hybrid Learning Approach to Detect Bangla Fake News," 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2021, pp. 442-447, https://doi.org/10.1109/ISMSIT52890.2021.9604712

[14] M. N. H. Hridoy, M. M. Islam and A. Khatun, "Aspect Based Sentiment Analysis for Bangla Newspaper Headlines," 2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI), 2021, pp. 1-4, https://doi.org/10.1109/STI53101.2021.9732611

[15] A. Motaleb et al., "Analyzing Human Abilities to Detect Fake Bangla News with Respect to Different Features,"2021 IEEE 12th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), 2021, pp.0337-0346, https://doi.org/10.1109/UEM-CON53757.2021.9666493

[16] R. Amin, N. S. Sworna, M. N. K. Liton and N. Hossain, "Abstractive Headline Generation from Bangla News Articles Using Seq2Seq RNNs with Global Attention," 2021 International Conference on Science Contemporary Technologies (ICSCT), 2021, pp. 1-5, https://doi.org/10.1109/ICSCT53883.2021.9642642

[17] A. S. Sharma, M. A. Mridul and M. S. Islam, "Automatic Detection of Satire in Bangla Documents: A CNN Approach Based on Hybrid Feature Extraction Model," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), 2019, pp. 1-5, https://doi.org/10.1109/ICBSLP47725.2019.201517

[18] Q. Ishtiaque Mahmud, N. Islam Chowdhury and M.Masum, "Reducing Feature Space and Analyzing Effects of Using Non Linear Kernels in SVM for Bangla News Categorization," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018, pp. 1-6, https://doi.org/10.1109/ICB-SLP.2018.8554844

[19] K. Salehin, M. K. Alam,M. A. Nabi, F. Ahmed and F. B. Ashraf, "A Comparative Study of Different Text Classification Approaches for Bangla News Classification," 2021 24th International Conference on Computer and Information Technology (ICCIT), 2021, pp.1-6, https://doi.org/10.1109/ICCIT54785.2021.9689843

[20] L. Nahar, Z. Sultana, N. Jahan, and U. Jannat, "Filtering Bengali Political and Sports News of Social Media from Textual Information," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-6, https://doi.org/10.1109/ICASERT.2019.8934605

[21] O. Sarkar, M. F. Ahamed, T. T. Khan, M. K. Ghosh and M. R. Islam, "An Experimental Framework of Bangla Text Classification for Analyzing Sentiment Applying CNNBiLSTM," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp.1-6, https://doi.org/10.1109/INCET51464.2021.945639

[22] T. A. Mahmud, S. Sultana, T. I. Chowdhury and F.R. Anando, "A New Approach to Analysis of Public Sentiment on Padma Bridge in Bangla Text," 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2022, pp.1-6, https://doi.org/10.1109/STI56238.2022.10103315

[23] M. G. Hussain, T. A. Mahmud and W. Akthar, "An Approach to Detect Abusive Bangla Text," 2018 International Conference on Innovation in Engineering

and Technology (ICIET), Dhaka, Bangladesh, 2018, pp. 1-5, https://doi.org/10.1109/CIET.2018.8660863

[24] M. G. Hussain, S. Kabir, T. A. Mahmud, A. Khatun and M. J. Islam, "Assessment of Bangla Descriptive Answer Script Digitally," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019, pp. 1-4, https://doi.org/10.1109/ICBSLP47725.2019.202042

[25] T. Ahmed, S. F. Mukta, T. Al Mahmud, S. A.Hasan and M. Gulzar Hussain, "Bangla Text Emotion Classification using LR, MNB and MLP with TF-IDF CountVectorizer," 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, 2022, pp. 275-280, https://doi.org/10.1109/ICSEC56337.2022.10049341

[26] Md Gulzar Hussain, Tamim Al Mahmud. (2019). A Technique For Perceiving Abusive Bangla Comments. GREEN UNIVERSITY OF BANGLADESH JOURNAL OF SCIENCE AND ENGINEERING, 04(01). https://doi.org/10.5281/zenodo.3544583

[27] Tamim Al Mahmud, Md Gulzar Hussain, Sumaiya Kabir, Hasnain Ahmad, and Mahmudus Sobhan. 2020.A Keyword Based Technique to Evaluate Broad Question Answer Script. In Proceedings of the 2020 9th International Conference on Software and Computer Applications (ICSCA 2020). Association for Computing Machinery, New York, NY, USA, 167–171 https://doi.org/10.1145/3384544.3384604