BAIUST Academic Journal (BAJ), Volume 4, Issue 1, 2023: 26–44
https://doi.org/10.63307/BAJ.4.1.S3

Bengali News Headline Categorization: A Comprehensive Analysis of Machine Learning and Deep Learning Approach

Ovi Chowdhury, Mamun Ahmed*, Mst. Tasmim Ara, Saha Reno, and Ankhi Alam

Department of Computer Science and Engineering, Bangladesh Army International University of Science and Technology (BAIUST), Cumilla, Bangladesh.
Email: mamun@baiust.edu.bd

*Corresponding Author

Abstract: Text classification, a prominent application of natural language processing, is gaining popularity in the Bengali language, much like in many other languages. A significant effort in this domain involves categorizing various unlabeled news items, spanning topics like national, international, and IT news. Bengali news platforms are experiencing growth, supported by effortless internet accessibility, leading to widespread engagement in online news consumption. These platforms commonly encompass a broad spectrum of news genres. This article presents an approach for classifying news headlines sourced from websites or news portals through the utilization of a machine learning algorithm. The acquired data underwent thorough evaluation and training, which encompassed preprocessing procedures such as tokenization, numeric character elimination, exclamation mark removal, symbol removal, and the exclusion of stop words. Additionally, we compiled a list of stop phrases to further enhance performance, recognizing the importance of effective stop word elimination in feature selection. Rather than scrutinizing news articles from diverse online sources, our research exclusively concentrates on categorizing Bengali news headlines. We consider eight distinct news categories, and our model is trained to categorize input data accordingly. Remarkably, our comprehensive model attained its peak performance employing the GRU technique, resulting in an accuracy rate of 84% in this specific case.

Keywords: *N*ews portal, Bengali news headline categorizing, online publication, text classification, word elimination

1. Introduction

In the realm of Natural Language Processing (NLP) systems, techniques are employed to enhance text comprehension and symbol interpretation. One valuable method is text clarification, which involves the assignment of specific terms to a given text (Yuslee & Abdullah, 2021). Another significant approach is text categorization or classification, which involves sorting articles into predetermined groups (Shahin et al., 2020). Although this poses a challenge in managing unstructured text within predefined categories, it offers both theoretical insights into document collections and practical utility (Yang & Joachims, 2008). By categorizing content, users can efficiently locate information within specific sections, rather than navigating through the entire dataset. As the volume of information escalates, the necessity for effective text categorization becomes increasingly evident.

While numerous studies have addressed news headline classification across various languages, limited attention has been directed towards Bangla newspapers. To support the creation of an automated system that integrates categorization algorithms based on machine learning, we formulated a method for classifying news content from Bengali newspapers. Our approach involves the creation and training of classifiers using a corpus of training documents. These classifiers are then utilized to allocate documents to relevant categories. We opted for the domain of online news due to the absence of robust search capabilities and visualization tools on existing news websites. Our observation revealed a gap in comprehending and analyzing data trends, exacerbated by the continuous influx of news content and references. To address this, we designed a system catering to two user groups: news readers seeking categorized news stories and stakeholders or analysts interested in statistical analysis to identify historical and current news patterns. Furthermore, several news organizations aspire to systematically categorize their publications.

2. Historical Background

Scholars dealing with real-world data derive advantages from classification methodologies. In an era of limited technological resources, investigators undertook remarkably audacious research endeavors. Certain scholars achieved favorable outcomes utilizing machine learning classifiers, whereas others gained entry to RNN. This segment deliberates pertinent research that attains a notable precision level akin to the classifiers we have implemented, serving as a wellspring of inspiration.

2.1 Earlier Research

To classify concise content, Yang Li introduced an innovative SVM KNN approach (Shahin et al., 2020). A range of machine learning classifiers including CNN, SVM, NB, RNN, and LSTM were employed. Ultimately, the SVM + CNN (SVMCNN) hybrid classifier demonstrated the most superior performance. The utilization of the SVM KNN methodology yielded approximately 90% accuracy in their findings (Zia et al., 2015). Another pertinent research Shahi and Pant (2018) ventured into predictive analysis for automatic classification of Nepalese articles. The work of Shahi and Pant (2018) encompassed the selection of an appropriate classifier model and artificial neural networks. While machine learning classifiers such as SVM and Naive Bayes employed multi-layer connectivity to achieve 74.65% accuracy in Nepali news text categorization, the neural network performed at a slightly lower efficiency of 73%. The dataset consisted of 4964 instances of Nepali news text, categorized into 20 different types.

For the classification of Bangla news headlines, Pranshengit Dhar and Md. Zainal Abedin harnessed prominent machine learning techniques (Omidvar et al., 2018). Their methodology incorporated SVM, Naive Bayes, and Adaboost classifiers, yielding an

https://doi.org/10.63307/BAJ.4.1.S3

accuracy level of approximately 81%. Similarly, Sheikh Abujar introduced a system for multi-labeling Bengali news, utilizing neural networks, achieving comparable performance (Cai et al., 2018). Their dataset comprised over 86 thousand news headlines, and employing SVM, NB, Random Forest, Logistic Regression, and Neural Networks as machine learning techniques, they reached an impressive accuracy of about 90% through Neural Network methods. In the domain of text categorization pertaining to hate speech, Bjorn Gamback focused on harnessing convolutional neural networks (Shahi & Pant, 2018). Their adoption of CNN led to an achievement of 86.68 percent accuracy. By implementing a distinct word embedding method along with the softmax function and max pooling, an increase of 7.3 percent was observed. Word embedding proved crucial in data preparation, aligning with Roger Alan Stein’s assertion that it mitigates the system’s weakest performance (Dhar & Abedin, 2021). Amin Omidvar’s research involved utilizing clickbait web data from media sources, subjecting it to analysis via both machine learning classifiers and neural networks (Khushbu et al., 2020). This comprehensive approach further enriched the body of knowledge in the field.

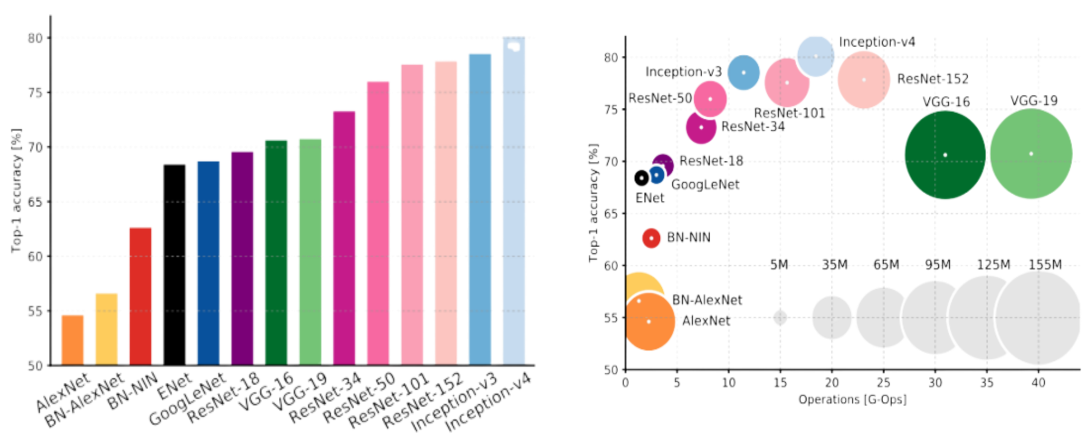
2.2 Recent Research

The central focus of emotion research centers on classifying tasks using affirmative or adverse metrics, with the intention of discerning maternal perspectives or particulars. This study is driven by the goal of enhancing customer engagement, revenue generation, and brand recognition. The techniques employed have diverse applications spanning fields such as finance, economics, spam detection, stock exchange activities, online transactions, and various other business domains. With the capacity to swiftly respond and facilitate informed decision-making, effective sentiment analyses can yield substantial benefits across areas including policy-making, governance, organizational strategies, campaigns, and enterprises. The cost-efficient acquisition of neural networks is feasible. A wealth of data including numerous emails, comments, and evaluations, collectively numbering in the thousands, can be harnessed. Text classification methodologies need to be adaptable to cater to businesses of all scales. Organizations must be well-prepared to address critical scenarios with prompt and efficient actions. Computer information retrieval should emulate designer labeling consistently and in real-time to promptly identify essential attributes. Although the notion of document categorization is not groundbreaking in the field of computational linguistics (Stein et al., 2019), it remains an ongoing focus in the context of Bangla text. Categorizing online news holds significant importance in an age heavily reliant on digital news sources. This proposed study, centered on the Bangla language, is dedicated to achieving this categorization goal. The study materials from our literature survey section draw upon certain Bangla datasets. In comparison to alternative machine learning methods, our hybrid modeling approach exhibits higher efficacy.

2.3 State-of-the-Art Technology

Recent research endeavors have been directed towards reducing the computational burden of deep learning networks in common applications, all while maintaining exceptional predictive precision. This pursuit holds particular significance for embedded Internet of Things (IoT) devices. These devices often benefit from large Convolutional Neural Networks (CNNs) boasting intricate, closely interconnected layers, known for their superior performance. However, their effectiveness can be severely hampered by their resource-intensive computational demands, rendering them impractically sluggish. In the context of agricultural systems, we delved into the application of cutting-edge CNN architectures. Our study encompassed an evaluation of both accuracy and computational requisites, drawing insights from relevant literature and incorporating the latest advancements in network designs, as depicted in Figure 1.

Figure 1
Accuracy vs Computational Cost



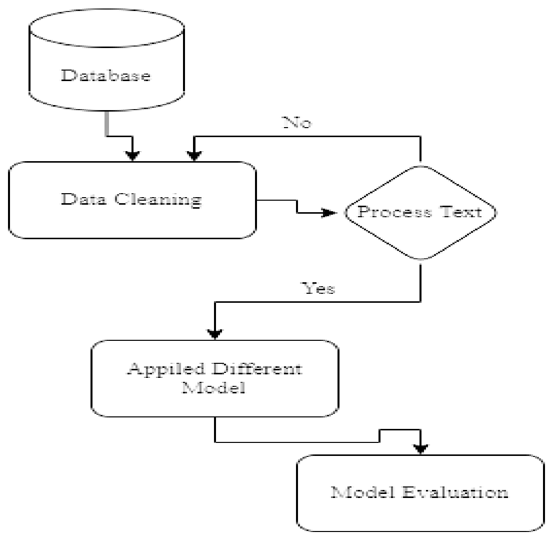
3. Proposed Method

We collected data from multiple Bangla newspapers, employing the Python module BeautifulSoup for web scraping. Once data retrieval and compilation were accomplished, we proceeded to eliminate unnecessary symbols from the datasets. Within this section, essential metrics such as word count, document count, and the count of unique words per class are detailed. Subsequently, we constructed the length frequency distribution based on the refined datasets. The subsequent phase involved data preparation for model training. An 80% portion of the data was employed for training, while the remaining 20% was allocated for testing. This data was then labeled using an encoded sequence. The model training was executed with 10 epochs and a batch size of 64, culminating in a fully prepared model. To facilitate the forecasting of news headlines, we deployed two distinct deep learning systems: Long

https://doi.org/10.63307/BAJ.4.1.S3

Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). These models were trained using a batch size of 64 and 10 epochs. Post-training, we evaluated the performance of these models by deriving metrics such as accuracy, precision, recall, and the F1 score. The subsequent phase entails a comparison between the outcomes generated by these two deep learning algorithms. This comparative analysis will enable us to gauge the efficacy of each algorithm in forecasting news headlines.

Figure 2
Model Architecture



3.1 Deep CNN

Deep learning stands as a pivotal machine learning methodology, emulating human learning patterns, thereby playing a pivotal role in the expansive realm of data science. It seamlessly integrates statistics and predictive modeling, streamlining the intricate journey undertaken by data scientists in assimilating, dissecting, and comprehending extensive datasets. These algorithms in the domain of deep learning are structured in an ascending hierarchy of complexity and abstraction. This hierarchy constitutes a foundational facet of predictive analytics automation. Notably, this contrasts with the linear nature observed in machine learning algorithms.

3.2 ReLu

In a neural network, the activation function serves to transform the aggregated weighted input of a node into the node's activation or output. Among these functions, the rectified linear activation function, often abbreviated as ReLU, is a prevalent choice. ReLU operates by directly outputting the input if it's positive and returning zero if it's negative. This straightforward behavior, combined with its improved performance during training, has established ReLU as the go-to activation function across various neural network architectures.

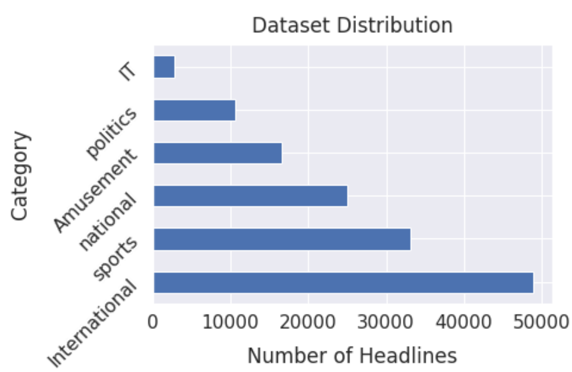
3.3 Sigmoid

A distinctive real function known as a sigmoid possesses several specific characteristics: a non-negative derivative at every point, applicability to all real input values, and a singular inflection point. The concepts encapsulated by the terms “sigmoid function” and “sigmoid curve” encapsulate the same underlying principle.”

4. Data Collection

To amass data from a variety of Bangla newspapers, a scraping methodology was employed. This curated collection encompasses an extensive one million plus entries. Prominent publications, including Bangladesh Protidin (BD Protidin, 2023), Dainik Juganttor (Jugantor, 2023), Daily Inqilab (Daily Inqilab, 2023), and Kaler Kantho (Kaler Kantho, 2023), among others, contribute to this dataset. These newspapers enjoy widespread readership within Bangladesh. This data compilation serves a crucial role in discerning the prevalent types of content sought by readers. The extraction of data from websites was streamlined by employing the Chrome Web Scraper tool in tandem with Python utilities. Our dataset comprises three principal columns: headlines, their associated categories, and the names of the respective newspapers. It’s noteworthy that this dataset is accessible to the public, fostering transparency and openness.

Figure 3
Dataset Distribution



4.1 Data Cleaning

Information can be acquired from diverse origins, spanning various sources. Newspaper headlines are systematically categorized into numerous sections. Real-time data retrieval is carried out from several online publications originating from Bangladesh. The process of data collection entails the utilization of scraping tools and modern technologies.

https://doi.org/10.63307/BAJ.4.1.S3

Figure 4
Data Preprocessing

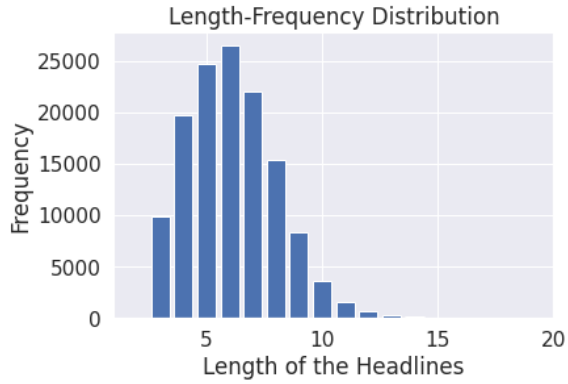
4.2 Data Preprocessing

Given the concise nature of headlines, the exclusion of stopwords is often unnecessary. Employing regular expressions aids in removing extraneous data from our dataset. After undergoing the cleaning process, the data sample appears refined and streamlined.

Original: ক্ষমা চেয়েও মুক্তি পেলেন না পরিচালক গাজী মাহবুব
Cleaned: ক্ষমা চেয়েও মুক্তি পেলেন না পরিচালক গাজী মাহবুব
Original: ব্র্যান্ডউইথের ব্যবহার ৮০০ জিবিপিএস ছাড়িয়ে
Cleaned: ব্র্যান্ডউইথের ব্যবহার ৮০০ জিবিপিএস ছাড়িয়ে
Original: জামিনে মুক্তি পেলেন ছাত্রদল সভাপতি
Cleaned: জামিনে মুক্তি পেলেন ছাত্রদল সভাপতি
Original: দ. কোরিয়ায় ১০০টি খালি কফিন পাঠিয়েছে যুক্তরাষ্ট্র
Cleaned: দ কোরিয়ায় ১০০টি খালি কফিন পাঠিয়েছে যুক্তরাষ্ট্র
Original: ফ্লোরিডায় হামলাকারী ‘মানসিকভাবে অসুস্থ’: ট্রাম্প
Cleaned: ফ্লোরিডায় হামলাকারী মানসিকভাবে অসুস্থ ট্রাম্প
Original: সেরাটা দিতে পারলে সিরিজ জিতবে বাংলাদেশ: মাশরাফি
Cleaned: সেরাটা দিতে পারলে সিরিজ জিতবে বাংলাদেশ মাশরাফি

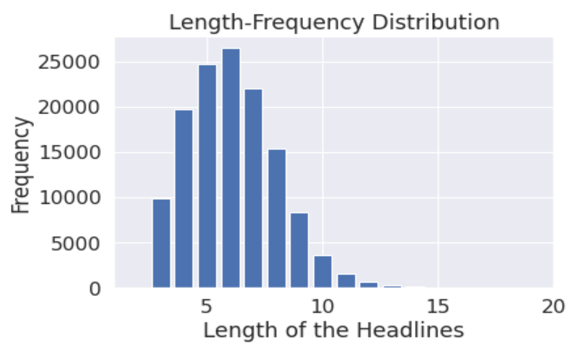
Following data cleaning, it is essential to determine the ideal headline length for standardization, ensuring consistency among all headlines. Figure 5 provides insights into the maximum, minimum, and average headline lengths.

Figure 5
Length Frequency Distribution



Furthermore, within each category, a substantial volume of terms exists. We meticulously select words from each category that encompass both uniqueness and relevance. This process is commonly referred to as data statistics, as illustrated in Figure 6.

Figure 6
Dataset statistics



4.3 Data Encoding and Model Building

After the preparation of the headlines, the text data is represented by encoded sequences, where each sequence is a vector of an index number containing words in each headline. Numeric values are also assigned to the categories.

====== Encoded Sequences ======
মোদির পাশে তৈমুর!
[4172, 2216, 6238, 301, 2629, 5925]
====== Padded Sequences ======
মোদির পাশে তৈমুর!
[4172 2216 6238 301 2629 5925 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0]

We add labels to each category after padding the sequence.

https://doi.org/10.63307/BAJ.4.1.S3

5. Model Description

5.1 Logistic Regression

Logistic Regression (LR) stands as a supervised classification algorithm within the realm of machine learning. The central function of an LR model is to calculate the total of input characteristics and then determine the sigmoid of the result. Employed primarily for probability prediction concerning specific classes, LR's application revolves around target or dependent variables. Notably, such variables inherently encompass just two potential classes. In mathematical terms, a logistic regression function of X predicts P(Y=1). This signifies that the dependent variable entails two exclusive possibilities: either 0 or 1 (Al-Tahrawi, 2015).

If model = X and predicted given input = yT, then the equation is:

$$yT = model(X) \tag{1}$$

In a model, a team of parameters is called a coefficient or beta. If we take inputs x1, x2,x3,......,xn and its coefficients is b0, b1,b2,b3,.....,bn then the equation is:

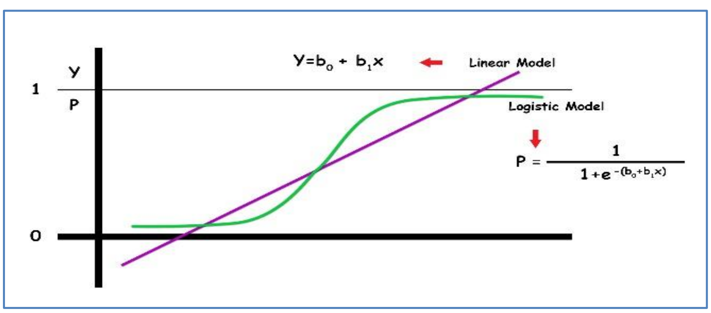
$$yT = b0 + b1 * x1 + b3 * x2 + \ldots\ldots + bn * xn \tag{2}$$

$$y = X * beta \tag{3}$$

It is an identical linear regression equation. It gives a real value instead of a class label.

Figure 7

Logistic Regression Model Architecture



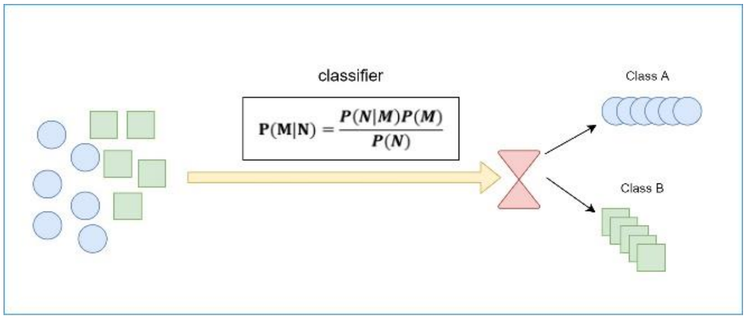
5.2 Multinomial Naive Bayes

Multinomial Naive Bayes (MNB) emerges as a swift and user-friendly algorithm designed for dataset classification, operating through a conditional probabilistic learning approach. Its versatility extends to both binary and multi-class classification, with notable efficacy demonstrated in the latter. MNB's foundation lies in Bayes Theorem, underpinned by the concept of conditional probability. The equation of the Bayes theorem is:

$$\frac{P(A \vee B) = P(B \vee A) P(A) P(B)}{P(A)} \tag{4}$$

Within this equation, the symbol A signifies the probability associated with the hypothesis, while B represents the event under observation. The elements P(A) and P(B) stand for the prior probability and the marginal probability correspondingly. Additionally, P(A|B) and P(B|A) respectively denote the posterior probability and the likelihood probability.

Figure 8
Logistic Regression Model Architecture



5.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression challenges. SVM works by creating a high-dimensional feature space for categorizing data points, even when the data are not linearly separable. Mathematically, if we have the hyperplane, we can make predictions easily. For this, we need the like equation. The equation of the line is:

y = ax + b (5)

Let, x = x1 and y = x2 we get,

ax1 - x2 + b = 0 (6)

If, x = (x1, x2) and w = (a,-1) then we get,

w.x + b=0 (7)

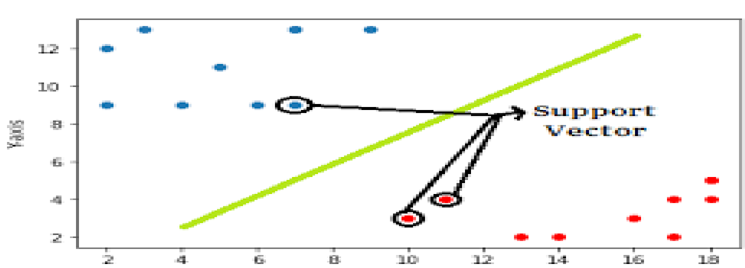
The hyw.x+bpothesis function is,

h(Xi)= {+1 if w.x+b>=0}"

{-1 if w.x+b<0} (8)

By finding a hyperplane we can separate the data accurately by using the SMV learning algorithm. Optimal hyperplanes can be found from these best accurate data.

Figure 9
Support Vector Machine Model Architecture



https://doi.org/10.63307/BAJ.4.1.S3

5.4 Random Forest

Random Forest (RF) operates as a supervised machine learning algorithm suitable for both classification and regression tasks. It operates by aggregating the results of multiple decision trees, yielding a reliable and predictable output. Notably, RF demonstrates proficiency in managing large-scale datasets, offering robust predictive capabilities in conjunction with interpretability. In addressing regression challenges via RF, a pivotal role is played by the Mean Squared Error (MSE) equation. This equation is crucial in understanding how data partitions occur at each node:

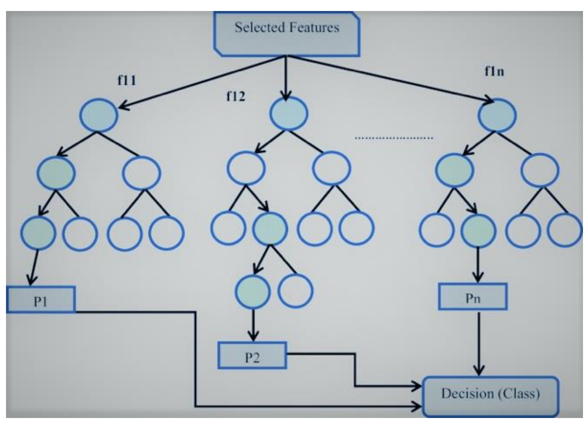
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2 \qquad (9)$$

Here, number of data points = N, returning value of model = f1, actual value of data point = yi.Mathematically, to solve classification problems by using RF we have to use the equation of the Gini index. The equation of the Gini index is,

$$Gini = 1 - \sum_{i=1}^{C}(p_i)^2 \qquad (10)$$

Here, observing the dataset's relative frequency = pi and c is the representation of the number of classes.

Figure 10
Random Forest Model Architecture

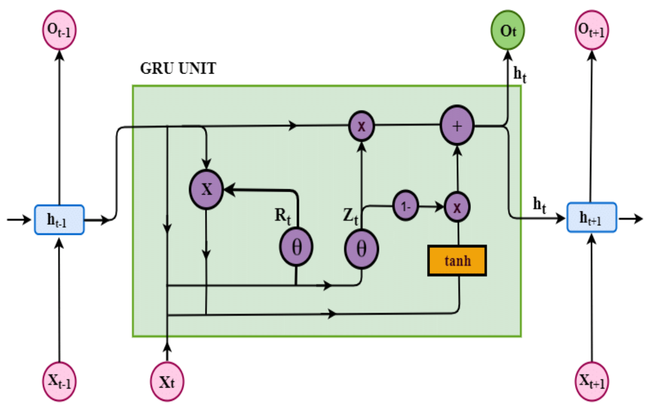


5.5 GRU

Gated Recurrent Neural Networks (RNN) have demonstrated their efficacy across a spectrum of applications reliant on sequential or temporal data. They have garnered significant adoption in domains like speech recognition, natural language processing, and machine translation. Specifically, both Long Short-Term Memory (LSTM) RNN and the more recently introduced Gated Recurrent Unit (GRU) RNN have showcased their prowess in managing prolonged sequences. The triumphant performance of these networks largely stems from their incorporation of gating signals that orchestrate the utilization of present inputs and memory for updating current activation and forming the ongoing state.

Throughout the learning process encompassing training and evaluation, these gating mechanisms exhibit unique sets of weights that are dynamically adjusted. While these models substantially enhance learning within RNN frameworks, they concurrently introduce an augmented level of parameterization owing to their gate networks. This, in turn, results in an increased computational overhead compared to the conventional RNN model. It is noteworthy that the GRU RNN operates with two gate networks, whereas the LSTM RNN employs three distinct gate networks. To optimize efficiency, there's a suggestion to potentially reduce the number of external gates to a minimum of one, with a preliminary assessment of its long-term efficacy.

Figure 11
GRU Model Architecture

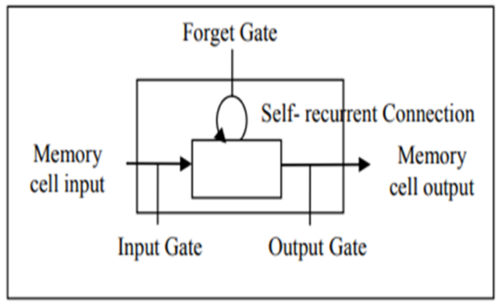


5.6 LSTM

Recurrent Neural Networks (RNNs) have evolved to manage sequential data based on their inherent recurrent architecture. However, a limitation arises when there is a substantial temporal gap between the relevant information-containing unit and the unit where RNNs struggle to effectively establish connections. This struggle is attributed to issues like gradient vanishing or exploding, particularly when the volume of required information escalates. To overcome these challenges, an enhancement to the initial RNN architecture has been introduced in the form of Long Short-Term Memory (LSTM) networks. These LSTM networks incorporate three fundamental gates: the input gate, the forget gate, and the output gate. This strategic addition aims to address the complications posed by the aforementioned gradient issues. Through these gates, LSTM networks are better equipped to capture the relationship between distant information units, mitigating problems related to vanishing or exploding gradients. Both RNNs and LSTM networks have recently showcased significant accomplishments in various domains involving time sequence data, such as human action understanding, spoken language interpretation, cross-lingual communication, image description generation, and visual content analysis.

https://doi.org/10.63307/BAJ.4.1.S3

Figure 12
LSTM Model Architecture



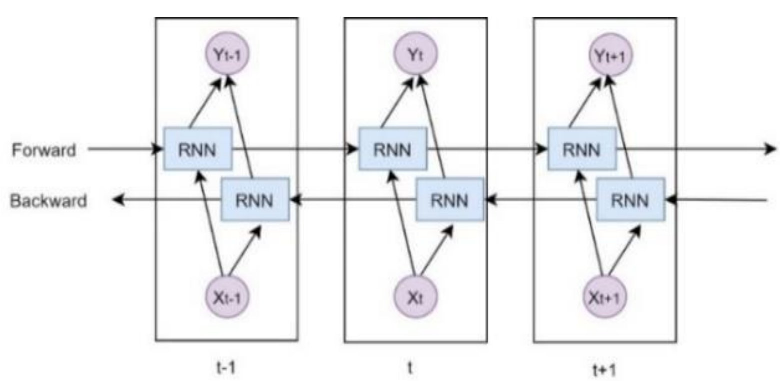
5.7 BI-LSTM

Bidirectional Long-Short Term Memory (Bi-LSTM) is a sequence processing model that incorporates information from both forward and backward directions. This inherent bidirectionality allows it to leverage insights from both sides of the sequence. Mathematically, Bi-LSTM integrates the outcomes of both forward and backward LSTMs at each time step, covering information from past to future and vice versa. This produces two distinct outputs from the cell: the activation (a) and the new candidate (c). So the equation is:

By using LSTM we can control 3 different gates,
$\Gamma_u = \sigma\left(W_u\left[a^{<t-1>}, x^{<t>}\right] + b_u\right)$ for update gate.
$\Gamma_f = \sigma\left(W_f\left[a^{<t-1>}, x^{<t>}\right] + b_f\right)$ for forget gate.
$\Gamma_0 = \sigma\left(W_o\left[a^{<t-1>}, x^{<t>}\right] + b_o\right)$ for output gate. Now, considering both activations (forward, backward) the calculated output $\hat{y}$ at time t is

$$\hat{y}^{<t-1>} = g\left(W_y[a^{<t>} \rightarrow, a^{<t>} \leftarrow] + b_y\right) \quad (11)$$

Figure 13
BI-LSTM Model Architecture



6. Result Analysis

6.1 ML-based Result Analysis

Measurement serves as a foundational aspect of our comprehension of the world, a knowledge accumulated over countless years of experience. Scientists require tools to quantify measurements, yet inherent to any measurement is an element of uncertainty attributed to errors. In this context, two crucial concepts come to the fore: accuracy and precision. Accuracy pertains to the proximity of a measurement to an established or acknowledged value. On the other hand, precision revolves around the closeness of multiple repetitions of the same measurement to one another. In essence, accuracy gauges how faithfully a measurement aligns with a known value, while precision evaluates the consistency of repeated measurements of a particular quantity.

Within our machine learning methodology, the core algorithmic foundation was established on the framework of the random forest classifier. The Random Forest (RF) constitutes an ensemble of individual decision trees. The decision-making process within each branch is informed by the "Gini index," aiding in the selection of the optimal decision path. The Gini index is derived through the following equation:

$$GINI = 1 - \sum_{i=1}^{c} Pi^2 \tag{12}$$

In this context, the variable c symbolizes the overall count of class labels, while pi denotes the probability associated with a particular subclass. For our strategy, we chose a forest consisting of 100 trees, where the 'Gini' criterion is employed to evaluate the degree of division. The process involves dividing nodes when a minimum of 2 internal nodes exists, and each internal node takes into account all pertinent system attributes.

For forecasting news headlines, we employed Machine learning, LSTM, and GRU. These two separate models had varied outcomes. The accuracy of the models is discussed in Table 1. The GRU Model is more accurate. GRU produces superior results. A bidirectional model and soft-max activation function are both employed. The higher the score, the more data is tightly categorized. We used the GRU Model to properly classify the categories.

Table 1
Model Accuracy

Model	Accuracy
LR	64.45%
MNB	61.33%
SVM	65.29%
RF	65.42%
GRU	84.01%
LSTM	82.74%
Bi-LSTM	83.42%

In the consistency of our model's evaluation, the highest accuracy and performance were acquired by the GRU which was 84.01%. In cases of ML approaches the Logistic Regression got 64%, the multinomial naïve bias achieved 61%, the Support Vector Machine got around 65.29% and the Random forest was able to get the highest

https://doi.org/10.63307/BAJ.4.1.S3

among the ML approaches consisting of the 65.42%. And then if we focus on the deep learning approaches we can contemplate the variation of the GRU which stands at the highest of 84% of total accuracy. As the LSTM has a more subtle and linear approach it was able to acquire around 82% and the Bidirectional LSTM approach was withered a little better with 83% than the traditional LSTM approach.

Table 2

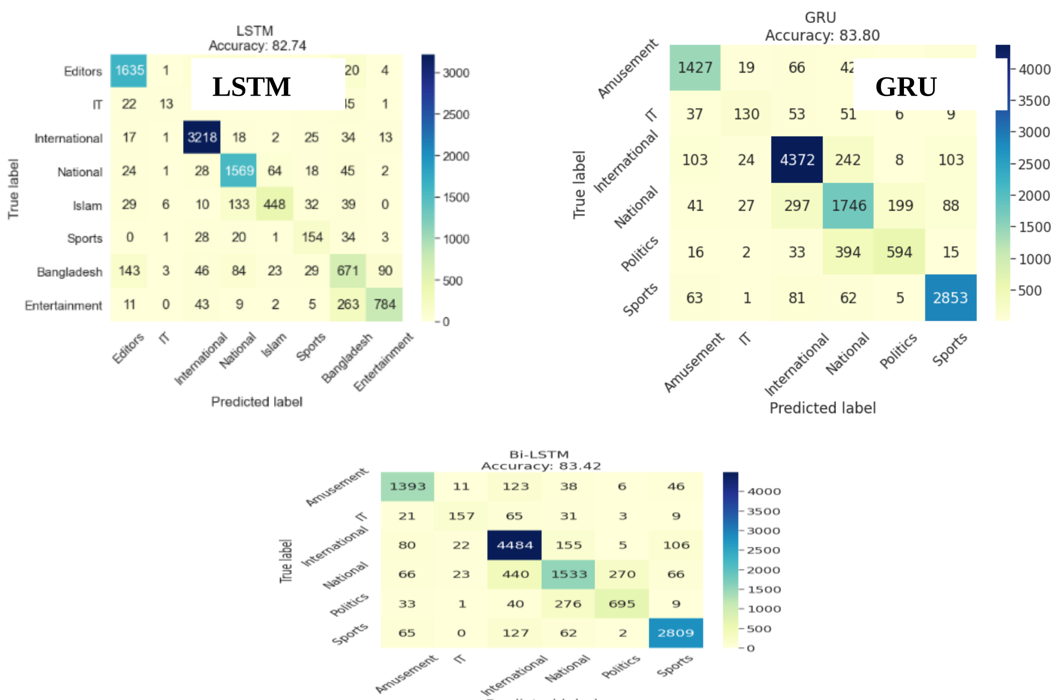
Precision, Recall, F1 Score of the Models

Models Name	Precision	Recall	F1 Score
LR	0.64	0.63	0.63
MNB	0.63	0.61	0.58
RF	0.65	0.64	0.64
SVM	0.65	0.64	0.64
LSTM	82.91	82.74	82.37
GRU	83.71	84.01	83.78
Bi-LSTM	83.02	83.42	83.07

The confusion matrix existed harnessed for the interpretation estimates of the forms after the classification. For the classification the filmfalm of the confusion matrix is described in Figure 14.

Figure 14

LSTM Model Architecture



Bi-LSTM

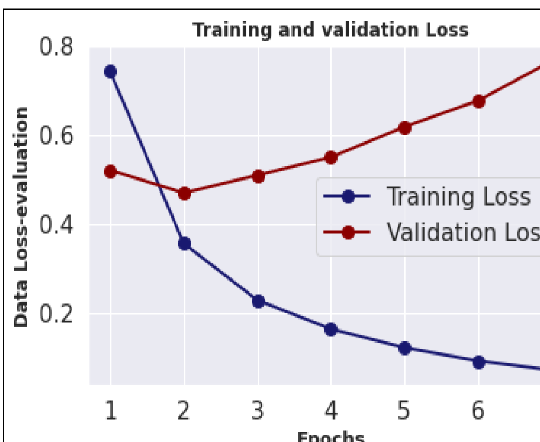
In the current research, we leveraged a combination of neural network and machine learning classifiers, encompassing Support Vector Machine (SVM), Random Forest (RF), Logistic Regression, and Naive Bayes (NB), to effectively categorize our intricate Bengali dataset. Classification hinges on a kernel system that empowers the anticipation of outcomes based on distinct datasets. This approach enables the system to make informed decisions tailored to the specific nature of the dataset at hand.

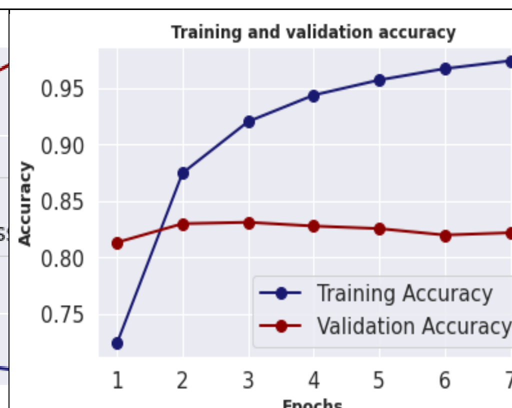
Consequently, the utilization of classifiers yields substantial output through the neural network. In contrast, the outcomes generated by other classifiers fall below the average mark. The achievement of a successful outcome is underpinned by the provision of a substantial dataset. Preceding the application of classifiers, data preprocessing was essential, involving the implementation of an enhanced tokenizer and the removal of stop words from around 30,000 entries in the dataset.

The previous studies utilized models that required a substantial number of parameters and were notably more intricate. In contrast, our present endeavor focuses on crafting a considerably simpler model. Additionally, our approach demonstrates the capacity to swiftly and accurately yield favorable outcomes, even when handling datasets with irregular distributions. As we advance through epochs, denoting the number of iterations, we assess the loss and validation loss for each model, aiding in predicting the class of News Headline categorization. Loss, reflecting the error encountered during each iteration on the training dataset, and validation loss (val-loss), computed for the testing dataset, stand as critical metrics. Remarkably, Figure 15 showcases a congruent trend in both loss and validation loss across all models, signifying a notable alignment of the data with the models. This coherence underscores a successful adaptation of the data to the models, indicating a commendable model fit.

Figure 15

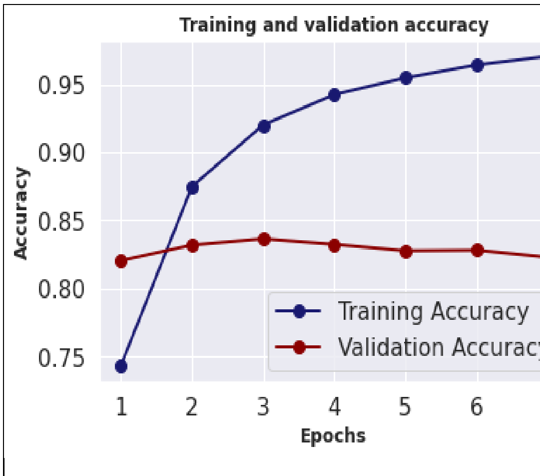
Train-validation Accuracy and Loss-validation Graph



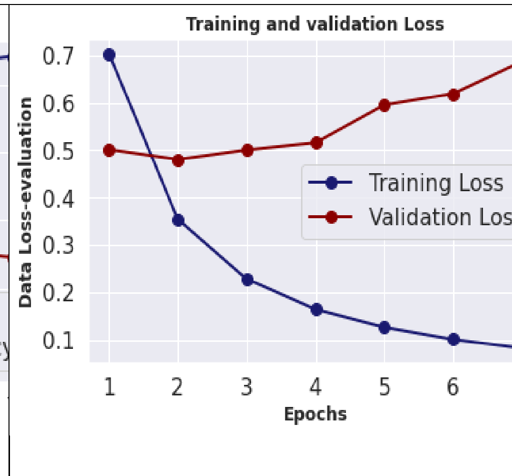


Bi-LSTM	Bi-LSTM

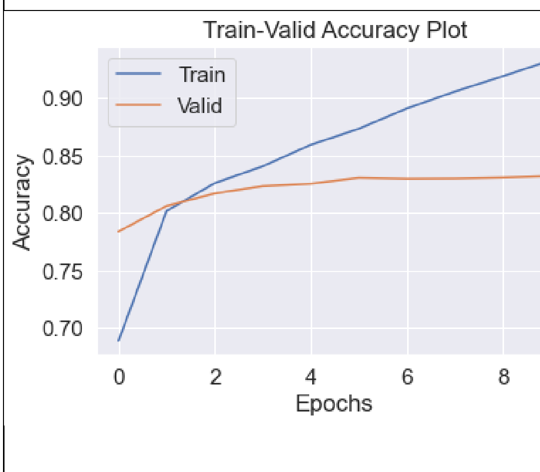
https://doi.org/10.63307/BAJ.4.1.S3



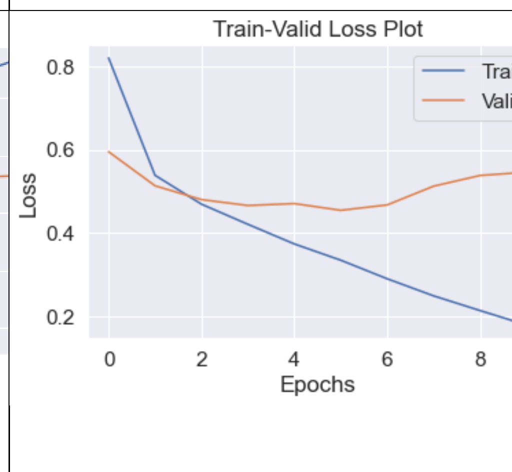
GRU



GRU



LSTM



LSTM

7. Future Work

As exemplified by our efforts above, we can affirm that our work stands as an amalgamation of diverse contradictions. The models we have utilized and the approaches we have followed can establish the foundation for the forthcoming aspects of our exploration into natural language processing, machine learning, and the experimentation with deep learning frameworks. Overcoming the challenges posed by Bengali linguistic-based datasets has proven to be one of the most formidable tasks within this realm, and we are committed to providing unwavering support to achieve new heights in research. This pursuit holds the potential to usher in a novel era of social and counterproductive value, making meaningful strides within the realm of computer science progress.

8. Conclusion

In this study, a machine learning-based model was meticulously devised to classify news headlines. Existing research within the literature primarily considers divergent linguistic publications. Within the realm of classification techniques, the GRU and LSTM algorithms emerge as stalwart contenders, contributing to the creation of robust models. The results of the categorization undertaken herein align harmoniously with antecedent investigations. It's imperative to note that due to the utilization of two distinct methods for categorization, outcomes may exhibit variability across different models. The ambit of news categorization was delimited to eight distinct categories, with the results being generalizable across these categories. This approach demonstrates its efficacy in furnishing accurate outcomes when confronted with extensive and diverse datasets. Diverse facets of information drive corporations to classify news contingent on the content published in various periodicals, thus aligning the outcomes with their intended objectives. In culmination, several pertinent insights surface. There exists a call for further research to delve deeper into the subject. The current dataset at our disposal is modest, and consequently, amplifying the dataset's scale could potentially yield superior results. Altering the model's attributes engenders diverse outcomes, and the results are subject to variation when epochs are modified. The activation function's omission from the models emerges as a pivotal factor influencing outcomes. The realm of machine learning harbors a rich landscape of models, each offering distinct results and insights, underscoring the significance of model selection in rendering outcomes.

9. Acknowledgement

We extend our heartfelt gratitude to the newspapers Jungantor, Jay Jay Din, Dainik Ittefaq, and Dainik Inqilab for generously providing us with invaluable data that proved instrumental in the execution of our research.

References

Al-Tahrawi, M. M. (2015). Arabic text categorization using logistic regression. *International Journal of Intelligent Systems and Applications*, 7(6), 71–78.

Bangladesh protidin, (2021). https://www.bd-protidin.com

Cai, J., Li, J., Li, W., & Wang, J. (2018). Deeplearning model used in text classification. In 2018 15th international computer conference on wavelet active media technology and information processing (ICCWAMTIP) (pp. 123–126). IEEE. https://ieeexplore.ieee.org/document/8632592

Daily Inqilab. (2021). https://www.dailyinqilab.com

Dhar, P., & Abedin, M. (2021). Bengali news headline categorization using optimized machine learning pipeline. *International Journal of Information Engineering & Electronic Business*, 13(1), 15–24.

https://doi.org/10.63307/BAJ.4.1.S3

Doinik Jugantor, (2021). https://www.jugantor.com

Elgabry, O. (2019). The ultimate guide to data cleaning. Towards to data science. https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4

Gambäck, B., & Sikdar, U. K. (2017). Using convolutional neural networks to classify hate-speech. In Proceedings of the first workshop on abusive language online (pp. 85-90). https://aclanthology.org/W17-3013/

Khushbu, S. A., Masum, A. K. M., Abujar, S., & Hossain, S. A. (2020). Neural network based Bengali news headline multi classification system: Selection of features describes comparative performance. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE. https://ieeexplore.ieee.org/document/9225611

Kostadinov, S. (2017). Understanding GRU networks. Towards Data Science. Towards Data Science, Towards Data Science, 16. https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be

Omidvar, A., Jiang, H., & An, A. (2018). Using the neural network for identifying click baits in online news media. In Annual International Symposium on Information Management and Big Data (pp. 220–232). Springer, Cham. https://www.semanticscholar.org/paper/Using-Neural-Network-for-Identifying-Clickbaits-in-Omidvar-Jiang/b0c63783250a0d4c3fad34e2a72f5c3dea459132

Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015). Convolutional, long short-term memory, fully connected deep neural networks. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4580-4584). IEEE. https://ieeexplore.ieee.org/document/7178838

Shahi, T. B., & Pant, A. K. (2018). Nepali news classification using naïve bayes, support vector machines and neural networks. In 2018 International Conference on Communication Information and Computing Technology (ICCICT) (pp. 1–5). IEEE. https://ieeexplore.ieee.org/document/8325883

Shahin, M. M. H., Ahmmed, T., Piyal, S. H., & Shopon, M. (2020). Classification of Bangla news articles using bidirectional long short-term memory. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1547-1551). IEEE. https://ieeexplore.ieee.org/document/9230737

Stein, R. A., Jaques, P. A., & Valiati, J. F. (2019). An analysis of hierarchical text classification using word embeddings. *Information Sciences*, 471, 216–232. https://www.sciencedirect.com/science/article/abs/pii/S0020025518306935

Yang, Y., & Joachims, T. (2008). Text categorization. Scholarpedia, 3(5), 4242. http://www.scholarpedia.org/article/Text_categorization

Yuslee, N. S., & Abdullah, N. A. S. (2021). Fake News Detection using Naive Bayes. In 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET) (pp. 112-117). IEEE. https://ieeexplore.ieee.org/abstract/document/9612540

Zia, T., Abbas, Q., & Akhtar, M. P. (2015). Evaluation of feature selection approaches for Urdu text categorization. *International Journal of Intelligent Systems & Applications*, 7(6), 33–40.