Different Machine Learning based Approaches of Baseline and Deep Learning Models for Bengali News Categorization

Mohammad Rabib Hossain
Department of Electrical & Electronic Engineering, Shahjalal University of Science & Technology, Sylhet

Soikot Sarkar
Department of Electrical & Electronic Engineering, Shahjalal University of Science & Technology, Sylhet

Moqsadur Rahman
Department of Computer Science & Engineering, Shahjalal University of Science & Technology, Sylhet

ABSTRACT
Today's universe is the type of world where everyone thrives to live in virtual life. According to the perspective of the present time, the online news portal holds a major door to that gradually increasing greedy life. So around the globe, the various platform has been developed to fulfill the requirement of mankind. A heavy load of work has been carried out for making this platform autonomous in the English language. That's why the machine learning approach is quite a fully developed field in English in news classification. But it can't be said the same for Bangla language. These put in the inspiration to do a research on this topic. So, here Bangla news which has been collected from newspapers and gathered around to make a Bengali Corpus. After preprocessing the news text, different sorts of procedures to classify the news text using baseline and deep learning models of Machine Learning are applied.

Keywords
Sentiment Analysis, Bangla, News Categorization, Confusion Matrix, CNN, BiLSTM.

1. INTRODUCTION
The Web is rapidly moving towards a platform for mass collaboration in content production and consumption. The increasing number of people are turning to online sources for daily news. Bangla online newspaper started appearing at about the same time that the internet became public in Bangladesh. Every day, the vast amount of Bangla news articles are created by several news sites that exist in the World-Wide-Web and its rate increases exponentially. An online newspaper has many forms [1]. One form is the electronic edition of the printed newspaper. The user can read the online edition similar to a paper edition; there is no categorization, neither with respect to content nor with respect to layout. Another form of online newspaper is a news website, which enables the user browsing in menus that are organized in subject categories and subcategories. Common to most of the above forms of online newspapers is that the user is assumed to read the news from a computer screen, while connected via the Internet to a certain news provider. However, these services might not be sufficient for many readers and reading situations. A lot of newspaper readers like to view and analyze news from various news sources. Many times, readers are interested only in the news articles of their categories of interest [2]. Therefore, the users have to scan through all the news articles of several news sites in order to get the articles of his/her interest. For example, a user interested in sports-related news has to go through all the news articles from various news sites and their time spent in analyzing news from multiple sources which is tedious. So a reader would prefer a system that would gather news articles from various sources; which is accessible to all the time and anywhere, also from a mobile reading device. A reader would like to visit a unique newspaper that includes the articles from various favorite sources, arranged and presented in an order that best fits her interests and reading habits.

Currently, a wide range of news is being distributed through the news entries. That implies the distribution of all-out news in these news entrances. Whatever it may be, not all individuals are inclined towards a wide array of news. Some people are more interested in perusing sports news than political news. A few people get a kick out of the chance to peruse political news than alternate news. A few people get a kick out of the chance to peruse amusement news. It really relies on one's decision. In any case, now and then, it has turned out to be such a great amount of exhausting to see the news that, really, isn't favored by the client. The news entry turns into the most proficient in the event that it demonstrates the news as indicated by the particular client's decision. In any case, for this, the main errand is to distinguish the news class. There are bunches of undertakings on news orders in English. In any case, there is extremely poor work on Bangla. In the event that Bangla news order gets some exploration chips away at it, it can be utilized as a part of such numerous genuine applications. Other than this, the present world is such a great amount of concentrating on the proposed framework. Clients expect everything that the better things will be prescribed to them by the framework. To make a framework to be proposal skilled must be able to take choice without anyone else's input. To take a choice independent from anyone else must need the information mining ability. The research-based work was intrigued by this. So, here the work is completely related to machine learning methods and has a few information mining strategies as well.

2. RELATED WORKS
It is the matter of sorrow that very few works on this field have accomplished by this time though in the present time, working on this field is increasing day by day. There are enough resources for the English language [3] as there has been done many works in this field for English. Recently, not only in Bangla language but also in other languages as like Chinese[4], Indonesian[5][6], Hindi[7][8], Urdu[8], Arabic[9], English-Hindi[10] and so on, are being included on Natural Language Processing related works. There are being enriched with resources day by day after doing more research works on this field. The desired objective for any order is only

to build an arrangement of models by utilizing preprocessed datasets. These datasets are the core of such an undertaking. At that point, the datasets are being partitioned into two sections preparing dataset and testing dataset. At last, these two sub-datasets are utilized to manufacture the model. The intention of building such a sort of model is to anticipate the class of various s. An exploration group from BRAC University chipped away at such a point. They essentially work in view of N-Gram based order [11]. In 2019, a group from CUET proposed a framework for classifying Bengali text documents using deep convolution nets. The proposed framework consists of word embedding and document classifier models. Experiments with more than 1 million Bengali text documents reveal that the proposed system worthy of classifying documents [12]. A Group from United International University presented a supervised learning-based Bangla content classification method. They have created a large Bangla content dataset and made it available for use publicly. This dataset was tested using several machine learning algorithms using text-based features. Their experiments showed logistic regression worked best compared to other algorithms. They have developed an online tool and made it available at: http://samspark1-001-site1.etempurl.com[13]. A research group of Chittagong University proposed an approach that provides a user to find out news articles that are related to a specific classification. Through a web crawler, they extract useful text from HTML pages of news article contents to construct a Full-Text-RSS. Each news article content is tokenized with a modified light-weight Bangla Stemmer. In order to achieve better classification results, they removed the less significant words i.e. stop −word from the document. They applied the naive Bayes classifier for classification of Bangla news article contents based on the news code of IPTC. Their experimental result showed the effectiveness of the classification systems [14].

3. METHODOLOGY

3.1 Dataset Collection

Any data set creation step first begins with ample data collection. To ensure that the diversity of collected data set, twelve categories were selected to extract: accident, art, economics, education, entertainment, environment, international, opinion, politics, crime, sports, science and technology. For these topics, the used dataset was from Open Source Bengali Corpus from a thesis project of Dept of Computer Science and Engineering, SUST. The dataset had been stored as titled বাংলা ডাটাসেট (কপরাস) in online. For each of the above categories, around 250 news from the dataset was got. The size of the whole sample was 3000.

3.2 Balanced Dataset

The source of the data doesn't have a balanced dataset. All the category has a various amount of news. So, the categories won't be balanced if all the news for different categories is taken. If the balance among the different categories isn't maintained then any of the above 12 categories can make the trained dataset a biased one. So that is not suitable for machine learning purposes. As the target was to classify news through its extracted feature so training dataset was kept as unbiased as possible. So, 250 news (Table 1) for each of the news categories is taken to keep the data clean of any kind of bias tendency which can be seen in Figure 1.

Table 1: Category wise total data

news type	Quantity
accident	250
art	250
crime	250
economics	250
education	250
entertainment	250
environment	250
international	250
opinion	250
politics	250
science_tech	250
sports	250

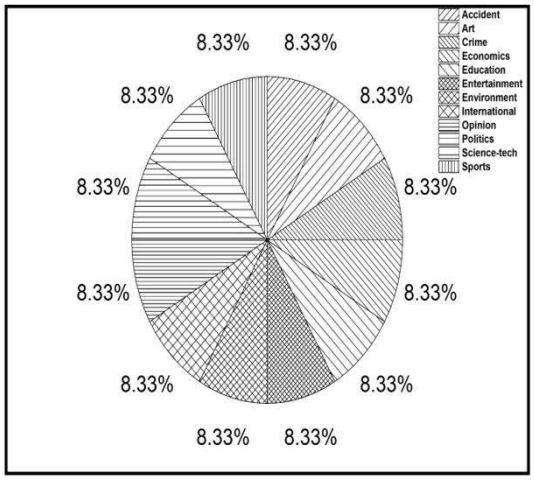

Fig 1: Category wise total data

Table 2: Raw data before pre-processing

Raw News without Pre-processing	Category
নওগাঁর মহাদেবপুর উপজেলার নওগাঁ-রাজশাহী মহাসড়কে একটি বাস ও ট্রাকের সংঘর্ষের ঘটনা ঘটেছে। এতে চার জন নিহত ও অন্তত ২০ জন আহত হয়েছেন।	Accident
সাহিত্যের সঙ্গে ক্রিকেটের সম্পর্ক হয়তো একটু দূরের। তবে যোগাযোগ একেবারে ক্ষীণ নয়, দেশ-বিদেশের সাহিত্যে ক্রিকেট স্থান পেয়েছে নানাভাবে।	Art
উচ্চ আদালতের নির্দেশ অমান্য করে পটুয়াখালীর বাউফল উপজেলার কাছিপাড়া মাধ্যমিক বিদ্যালয়ের শিক্ষক ও কর্মচারীদের ৩ লাখ ৭০ হাজার টাকা বেতন-ভাতা তোলা হয়েছে।	Crime
দশ পায়ের কাঁকড়া, রপ্তানির সম্ভাবনাময় এক পণ্য। দেশের উপকূলীয় এলাকায়, বিশেষ করে সুন্দরবন অঞ্চল থেকে প্রচুর কাঁকড়া ধরা হয়।	Economics

Table 3: Dataset after using custom made tokenizer function

Raw News without Pre-processing	Category
'নওগাঁর', 'মহাদেবপুর', 'উপজেলার', 'নওগাঁ' 'রাজশাহী', 'মহাসড়কে', 'একটি', 'বাস', 'ও', 'ট্রাকের', 'সংঘর্ষের', 'ঘটনা', 'ঘটেছে', 'এতে', 'চার', 'জন', 'নিহত', 'ও', 'অন্তত', 'জন', 'আহত', 'হয়েছেন'	Accident
'সাহিত্যের', 'সঙ্গে', 'ক্রিকেটের', 'সম্পর্ক', 'হয়তো', 'একটু', 'দূরের', ' তবে', 'যোগাযোগ', 'একেবারে', 'ক্ষীণ', 'নয়', 'দেশ', 'বিদেশের', 'সাহিত্যে', 'ক্রিকেট', 'স্থান', 'পেয়েছে', 'নানাভাবে'	Art
'উচ্চ', 'আদালতের', 'নির্দেশ', 'অমান্য', 'করে', 'পটুয়াখালীর', 'বাউফল', 'উপজেলার', 'কাছিপাড়া', 'মাধ্যমিক', 'বিদ্যালয়ের', 'শিক্ষক', 'ও', 'কর্মচারীদের', ''লাখ', 'হাজার', 'টাকা', 'বেতন', 'ভাতা', 'তোলা', 'হয়েছে'	Crime
'দশ', 'পায়ের', 'কাঁকড়া', 'রপ্তানির', 'সম্ভাবনাময়', 'এক', 'পণ্য', 'দেশের', 'উপকূলীয়', 'এলাকায়', 'বিশেষ', 'করে', 'সুন্দরবন', 'অঞ্চল', 'থেকে', 'প্রচুর', 'কাঁকড়া', 'ধরা', 'হয়'	Economics

3.3 Data Preprocessing

Table 2 shows the condition of various class's news conditions before preprocessing. English words are omitted here. All symbols, English numbers are removed from this data set also. There are no punctuation marks left as it does not hold any meaning for the purpose. During the preprocessing, the Bengali numbers also have been removed. Table 3 shows a sample of the dataset after doing the preprocessing steps. Each news is labeled according to its category. This makes it possible for anyone to look for additional data from the data set. Table 3 shows the tokenized data which has been tokenized by using a custom made tokenizer.

3.4 Model Overview

Here SVM, Naive-Bayes, Random Forrest, Logistic Regression for baseline evaluation and BiLSTM and CNN for fine-tuned predictions are used.

3.4.1 Support Vector Machine (SVM)

Support Vector Machine is a classification algorithm that divides samples into multiple groups in a hyper plane [15]. It essentially looks for support vectors, the most extreme cases of each group, to draw a distance between all groups. SVM algorithm repeatedly learns from the data set until it can draw a satisfactory margin between different classes [16]. In that way, it can distinctively differentiate and identify various classes. Figure 2 is a representation of SVM classifying two classes by finding a sufficient margin. Here SVM is used for mainly baseline evaluation. So any parameters are not tuned. Here linear SVM is used for measuring the baseline performance of the SVM model.

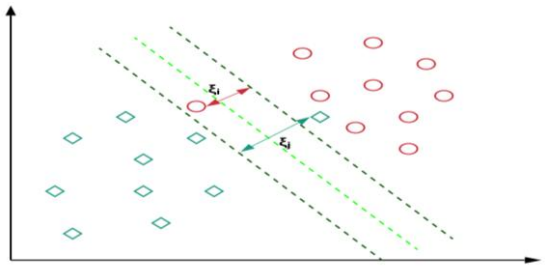

Fig 2: Support Vector Machine [17]

3.4.2 Bidirectional Long Short Term Memory (BiLSTM)

LSTM had one obvious problem: it looks at a sentence from only one direction. But human language is very complicated and the relationship between words in a sentence can be two folds [18]. For example, in the sentences "the bank is situated by the river" and "the riverbank is filled with people", the relationship between bank and river is different. If the language model only reads the sentence from either left to right or right to left, it will not be able to grasp the full meaning of the sentences. To solve this, BiLSTM was proposed. This model is similar to LSTM except for a very crucial point: BiLSTM reads a sentence from both directions. This enables the model to fully grasp the relationship between words in a sentence [19]. Figure 3 shows the inner working of traditional BiLSTM architecture.

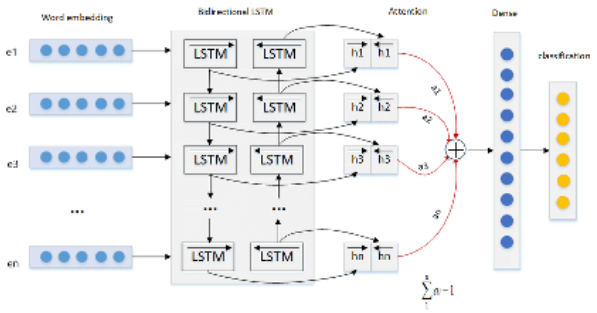

Fig 3: Bidirectional Long Short Term Memory Architecture

3.4.3 Convolutional Neural Network (CNN)

Convolutional Neural Network is mainly used for image recognition. But it has also been used successfully in Natural Language Processing tasks. CNN is another deep learning model where activation functions like ReLu [20] or tanh [20] scan over two-dimensional representation of data and pull out information from them [21]. This in turns becomes a series of tensors that were fed into hidden layers for learning. Figure 4 describes the working principle of a CNN architecture.

Here 128X5 convolution layers are used on top of 100 embedded layers. The activation function for convolution layers was 'ReLU'. 10 layers of dense function are used with activation function set as 'ReLU' and finally a 2 layer dense neuron for classification with activation function set as 'sigmoid'. Finally, the optimizer was set as 'adam'.

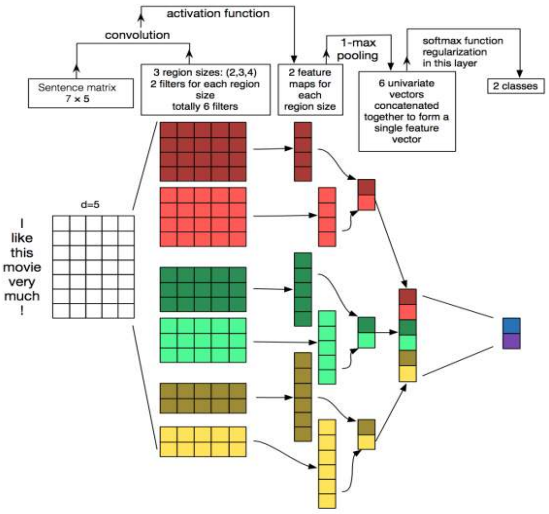

Fig 4: Convolutional Neural Network [3]

3.4.4 Naïve Bayes

In machine learning, the naïve Bayes classifiers Research are a family of basic ”probabilistic classifiers” focused on applying a theorem of Bayes with clear (naïve) assumptions of independence between features. They are among the easiest versions of Bayesian networks. Naïve Bayes has been studied extensively since the 1960s. It was introduced (though not under that name) into the text retrieval community in the early 1960s,[22] and remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines [23]. It also finds application in automatic medical diagnosis [24]. Below diagram shows how naive Bayes works:

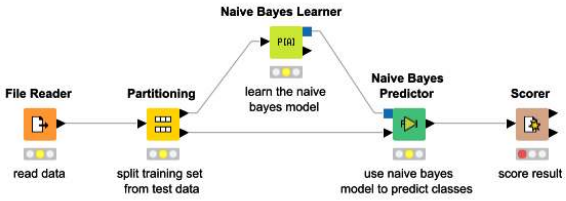

Fig 5: Working Naïve Bayes

3.4.5 Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees[25][26]. Random decision forests correct for decision trees habit of overfitting to their training set [27]:587–588. The first algorithm for random decision forests was created by Tin Kam Ho[1] using the random subspace method,[26] which, in Ho’s formulation, is a way to implement the ”stochastic discrimination” approach to the classification proposed by Eugene Kleinberg[28][29]. Figure 6 demonstrates how random forest work.

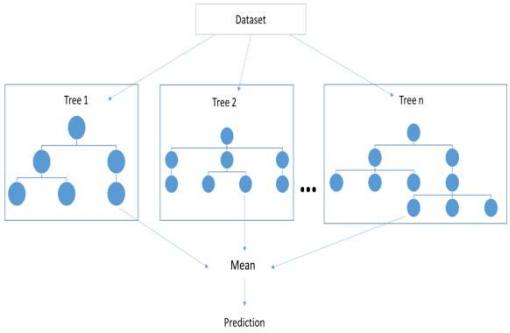

Fig 6: Working of Random Forest

3.4.6 Logistic Regression

The logistic model (or logit model) is a statistical model that is usually taken to apply to a binary dependent variable. In regression analysis, logistic regression or logit regression is estimating the parameters of a logistic model. A logistic model is one where the log-odds of the probability of an event is a linear combination of independent or predictor variables. The two possible dependent variable values are often labeled as “0” and “1”, which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick. So, the binary logistic regression model can be generalized to more than two levels of the dependent. Logistic regression measures the relationship between the dependent variables and one or more independent variables. It is done so by estimating probabilities using the logistic function. The probabilities must then be transformed into binary values in order to actually make a prediction. The logistic function used for this purpose is called sigmoid function. The sigmoid function takes any real value input and maps it to 0 or 1. “-1 and 1” can also be the choice.

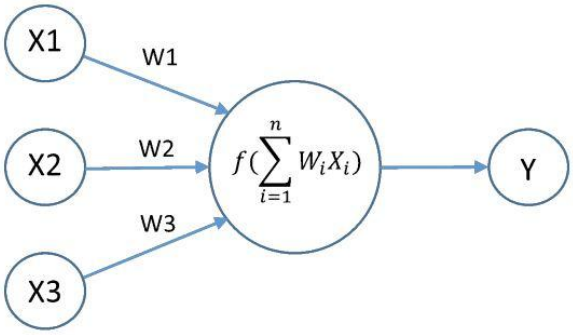

Fig7: Working of Logistic Regression

4. RESULTS AND DISCUSSION

In the experiment, after feature extraction and normalization, the mentioned Bengali corpus has been used as training datasets for obtaining models, and then, using these models testing data sets are predicted. But the whole dataset has not been used from each category as the training model. So each and every type of news has a percentage in the training model and also in the testing sample. But the percentage for each category is not the same. More to say that though it is quite random those categories don’t have an unequal quantity of samples. All the class of news has the same amount of samples to be fed. But 12 types have been chosen simultaneously so for the randomness of the election process of the training model the number of each type of class is not the same. In Figure 8 it is shown that the number of training samples for every category is varied with each type of news.

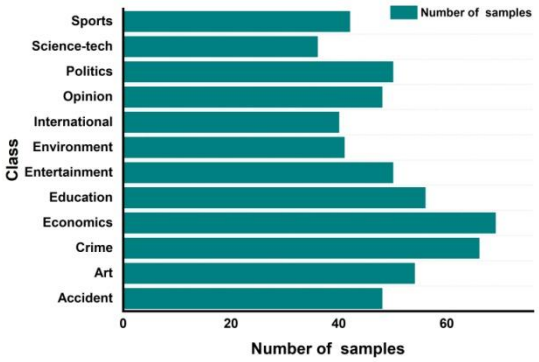

Fig8: Different number of samples for different classes in learning process

4.1 Accuracy of Model

The confusion matrix is a technique for summarizing the performance of a classification algorithm. Classification accuracy alone can be misleading if you have an unequal number of observations in each class or if you have more than two classes in the dataset. So the above approach has been used to give out the result for better understanding. A confusion matrix shows how many positive and negative cases a model actually predicted. This gives us even insight to how well the model the performing.

Table 4: Confusion Matrix

True positive	False positive
False negative	True negative

4.2 Base Model

4.2.1 Naive Bayes Classifier

The dataset has 12 different categories and it is clear there are different accuracy rates for different types of news. If all of these above categories are combined to get the accuracy rate of the naive-bayes model then the accuracy is 79.83%. Here entertainment has the lowest rate of accuracy in the prediction model.

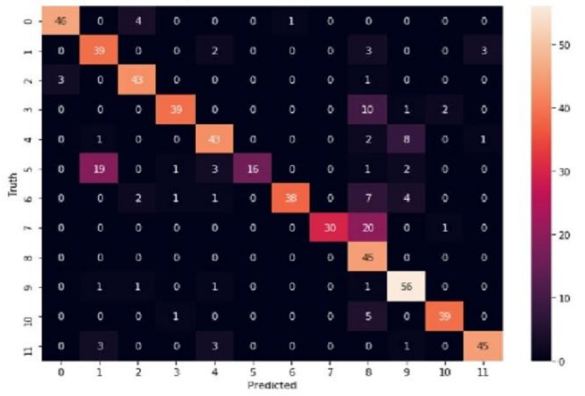

Fig9: Confusion Matrix for Naive-Bayes

4.2.2 Logistic Regression

The accuracy of this model stands at 87.83%. Again, entertainment has the lowest and science has the highest rate of accuracy in the prediction model.

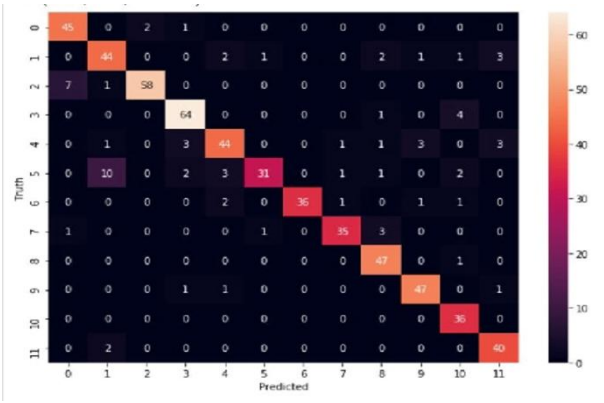

Fig10: Confusion Matrix for Logistic Regression

4.2.3 Random Forest Classifier

If all the categories are combined to get the accuracy rate of the random forest model then the accuracy is 82.5%. In the prediction model the lowest rate of accuracy is found in education and the accident category shows better accuracy for this model. A confusion matrix is shown which can show which class has better results. So, it can be considered as heatmap.

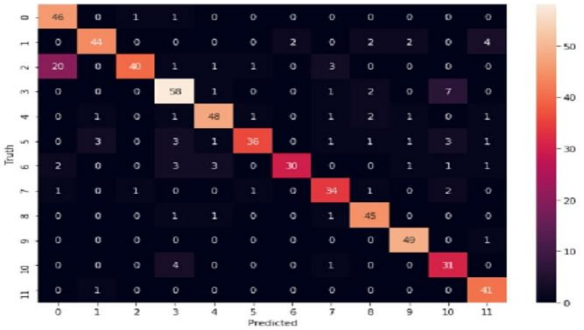

Fig11: Confusion Matrix for Random Forest Model

4.2.4 Linear SVM

There are different types of kernels available for SVM model. Among those kernels, the linear kernel is used here. If all of these above categories are combined to get the accuracy rate of the model then the accuracy is 91%. Education shows the lowest rate of accuracy here again in the prediction model and the accident category shows better accuracy for this model. The confusion matrix is shown below which can show which class has a better result. So, it can be considered as a heatmap.

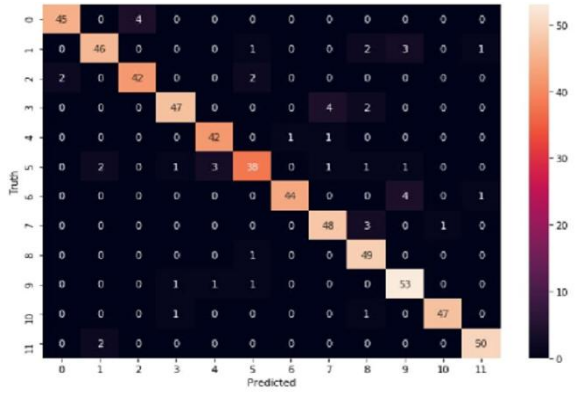

Fig12: Confusion Matrix for SVM Model

4.3 Deep Learning model

4.3.1 BiLSTM

In this case, international type news has the lowest rate of accuracy in the prediction model. For the base model such as naive-bayes, Logistic regression entertainment has the lowest accuracy but not in here. On the contrary, it is clear that the accuracy rate is quite impressive. The accident category has the highest rate of accuracy here. Again this model has better accuracy than the previously given base models. After calculation, the accuracy rate was 92.6%.

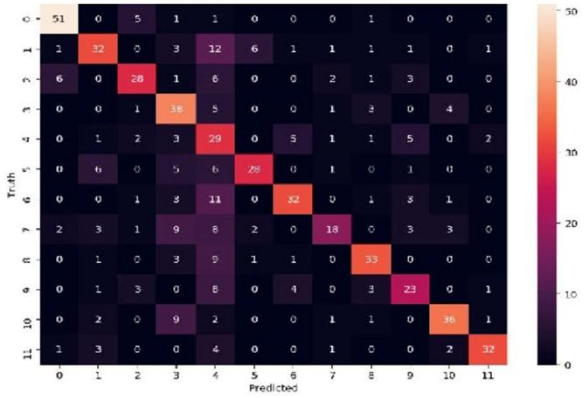

Fig12: Confusion Matrix for BiLSTM Model

4.3.2 CNN

This is the last model from the deep learning algorithm but here a better rate of accuracy is found in this prediction model. For the base model such as naive-bayes, Logistic regression and also in Bi-LSTM lower accuracy was found than this model. On the contrary, its accuracy rate is quite impressive. The same quantity of samples has been fed in this approach which is 80% of the whole dataset and after calculation, the accuracy rate is 93.43%.

4.4 Comparison of Algorithms

Table 5 shows that the highest result comes from the Support Vector Machine in the base model and CNN in deep learning. Here CNN gives the best performance. To increase the accuracy level, the dataset must need to preprocess properly. All categorical news should be equally numbered. At that, to increase the accuracy level, data cleaning has no alternative. The more data are preprocessed, the more accurate prediction will be shown by the classifier.

Table 5: Accuracy of various Models

Model	Accuracy (%)
Naive Bayes	79.8
Random Forrest	82.5
Logistic Regression	87.8
Linear SVM	91
Bi-LSTM	92.6
CNN	93.43

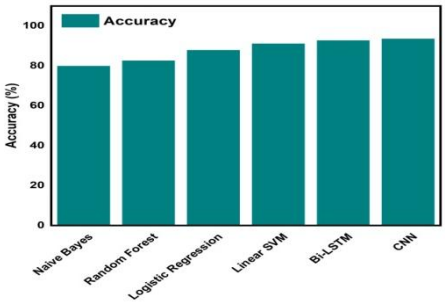

Fig13: Accuracy of various Models

5. CONCLUSION

Through CNN the maximum accuracy of 93.43% is found, which is quite good for a Sentiment Analysis System. But, by a large amount of balanced data, it is possible to build a more reliable System. Again the stop words have not been removed from the whole dataset so the prediction of the models has been hampered. The irrelevant letters that normally get attached to the actual words to express the person, time or you can say the grammatical terms that are necessary for the Bengali language don't always hold a meaningful concept. So the performance has been degraded for the above applied models. As also there are a lot of specific words that are exclusively specific for a specific category but also available in other category samples so there is a built-in confusion arise in the training time. If specific words are been filtered that are relevant to the specific class of news then it would really uphold the accuracy score for the applied models.

6. REFERENCES

[1] Tenenboim, L., Shapira, B. and Shoval, P. 2008. Ontology-based classification of news in an electronic newspaper.

[2] Pendharkar, B., Ambekar, P., Godbole, P., Joshi, S. and Abhyankar, S. 2007. Topic categorization of rss news feeds, Group.

[3] Carreira, R., Crato, J. M., Gonçalves, D. and Jorge, J. A. 2004. Evaluating adaptive user profiles for news classification. 9th international conference on Intelligent user interfaces, pp. 206–212.

[4] Xu, J., Ding, Y.-X. and Wang, X.-L. 2007. Sentiment classification for chinese news using machine learning methods. Journal of Chinese Information Processing, vol. 21, no. 6, pp. 95–100.

[5] Asy'arie, A. D. and Pribadi, A. W. 2009. Automatic news articles classification in Indonesian language by using naive bayes classifier method. 11th International Conference on Information Integration and Web-based Applications & Services, pp. 658–662.

[6] Buana, P. W., Jannet, S. and Putra I. 2012. Combination of k-nearest neighbor and k-means based on term re-weighting for classify indonesian news. International Journal of Computer Applications, vol. 50, no. 11, pp. 37–42.

[7] Dutta, K., Kaushik, S. and Prakash, N. 2011. Machine learning approach for the classification of demonstrative pronouns for indirect anaphora in hindi news items. The Prague Bulletin of Mathematical Linguistics, vol. 95, pp. 33–50.

[8] Kanan, T. and Fox, E. A. 2016. Automated arabic text classification with p-s temmer, machine learning, and a tailored news article taxonomy. Journal of the Association for Information Science and Technology, vol. 67, no. 11, pp. 2667–2683.

[9] El-Barbary, O. 2016. Arabic news classification using field association words. Advances in Research, pp. 1–9.

[10] Haque, R., Dandapat, S., Srivastava, A. K., Naskar, S. K. and Way, A. 2009. English-hindi transliteration using context-informed pb-smt: the dcu system for news 2009. Named Entities Workshop: Shared Task on Transliteration. Association for Computational Linguistics, pp. 104–107.

[11] Mansur, M. 2006. Analysis of n-gram based text categorization for bangla in a newspaper corpus. Ph.D. dissertation, BRAC University.

[12] Hossain, M. R. and Hoque, M. M. 2019. Automatic bengali document categorization based on deep convolution nets. Emerging Research in Computing, Information, Communication and Applications. Springer, pp. 513–525.

[13] Mostakim, S. Al., Ehsan, F., Hasan, S. M., Islam, S. and Shatabda, S. 2018. Bangla content categorization using text based supervised learning methods. International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE, pp. 1–6.

[14] Chy, A. N., Seddiqui, M. H. and Das, S. 2014. Bangla news classification using naive bayes classifier. 16th Int'l Conf. Computer and Information Technology, IEEE, pp. 366–371.

[15] Ladwani, V. M. 2018. Support vector machines and applications. Computer Vision: Concepts, Methodologies, Tools, and Applications. IGI Global, pp. 1381–1390.

[16] Cauwenberghs, G. and Poggio, T. 2001. Incremental and decremental support vector machine learning. Advances in neural information processing systems, pp. 409–415.

[17] Huang, X., Maier, A., Hornegger, J. and Suykens, J. A. 2017. Indefinite kernels in least squares support vector machines and principal component analysis. Applied and Computational Harmonic Analysis, vol. 43, no. 1, pp. 162–172.

[18] Plank, B., Søgaard, A. and Goldberg, Y. 2016. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. arXiv preprint arXiv:1604.05529.

[19] Zhou, P. Shi, W. Tian, J., Qi, Z., Li, B. Hao, H. and Xu, B. 2016. Attention-based bidirectional long short-term memory networks for relation classification. 54th annual meeting of the association for computational linguistics (volume 2: Short papers), pp. 207–212.

[20] Lawrence, S., Giles, C. L., Tsoi, A. C. and Back, A. D. 1997. Face recognition: A convolutional neural-network approach. IEEE transactions on neural networks, vol. 8, no. 1, pp. 98–113.

[21] Kalchbrenner, N., Grefenstette, E. and Blunsom, P. A. 2014. Convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188.

[22] Maron, M. E. 1961. Automatic indexing: an experimental inquiry. Journal of the ACM (JACM), vol. 8, no. 3, pp. 404–417.

[23] Rennie, J., Shih, L., Teevan, J. and Karger, D. 2003. Tackling the poor assumptions of naive bayes classifiers. ICML [Accessed: 10-Feb-2017].

[24] I. Rish et al. 2001. An empirical study of the naive bayes classifier. IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22, pp. 41–46.

[25] Ho, T. K. 1995. Random decision forests. 3rd international conference on document analysis and recognition, vol. 1. IEEE, pp. 278–282.

[26] Technol, L. and Hill, M. 1998. The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence, vol. 20, no. 8, pp. 832–844.

[27] Friedman, J., Hastie, T. and Tibshirani, R. 2001. The elements of statistical learning. Springer series in statistics New York, vol. 1, no. 10.

[28] Kleinberg, E. 1990. Stochastic discrimination. Annals of Mathematics and Artificial intelligence, vol. 1, no. 1-4, pp. 207–239.

[29] Kleinberg E. et al. 1996. An overtraining-resistant stochastic modeling method for pattern recognition. The annals of statistics, vol. 24, no. 6, pp. 2319–2349.