Multi-label News Classification Using Traditional ML Models and BERT

Mohammad Ashekur Rahman
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
mohammad.ashekur.rahman@g.bracu.ac.bd

Asma Ul Hussna
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
asma.ul.hussna@g.bracu.ac.bd

Sanjida Ali Shusmita
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
sanjida.ali.shusmita@g.bracu.ac.bd

Samiha Khan
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
samiha.khan@g.bracu.ac.bd

Iffat Immami Trisha
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
iffat.immami.trisha@g.bracu.ac.bd

Md Humaion Kabir Mehedi
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
humaion.kabir.mehedi@g.bracu.ac.bd

Md Motahar Mahtab
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
md.motahar.mahtab@g.bracu.ac.bd

Annajiat Alim Rasel
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
annajiat@gmail.com

***Abstract*—Now a days tremendous increase in web-based online information forces firms to embrace new resources and strategies that may aid in boosts the ability of high-dimensional data processing and handling 90 percent of the content on the Web is encrypted, unorganized and there are various ways to turn this data into something beneficial.News Classification is one technique to get meaningful, structured data.It's important and required to put together a good collection of groupings. Autonomous news analytics is becoming increasingly important as the amount of machine readable texts grows.Although since emergence of digital information, automatic multiple news classification has been a major application and development area. News and Texting has become very popular in recent years because of the vast amount of data categorization required.We have a large number of news articles to deal with everyday.In this paper,we collected the dataset from HuffPost which has almost 200k news headlines between the years of 2012 to 2018.We used seven different types of Machine Learning and Deep learning algorithms to classify various types of news and found that Bert has the highest effective model for classifying multiple news as it has more than 70 percent accuracy while other algorithms such as Multinomial Naive Bayes,Support Vector Classifier,Logistic Regression,Random Forest,Decision Tree classifier and Gaussian Naive Bayes has 46.66 percent,46.46 percent, 44.77 percent and 41.19 percent,25.62 percent and 12.94 percent accuracy scores subsequently.**

***Index Terms*—Bert, Multinomial Naive Bayes, Support Vector Classifier, Logistic Regression, Random Forest, Decision Tree classifier, Gaussian Naive Bayes**

I. Introduction

A news article is a form of reporting that covers current or recent events of wide interest such as national papers or on a specialized topic such as political or trade news magazines, club newsletters or technology news websites. An expert view of a current occurrence might be included in a news report. When we visit a news portal, we must have noticed that the information is separated into sections. Tech, entertainment, sports as well as other prominent topics may be found on practically every news website. Before releasing a newspaper story, every news portal classifies it so that viewers may quickly find the sort of news that attracts them the next time they visit the news portal. For instance, If someone enjoys reading about the newest technical innovations, someone will always go to the technology section of a news website. However, someone might not even be willing to read about technology, instead someone may be interested in politics, business, entertainment or sports. Nowadays, news content is sorted manually by news web page management. Furthermore, they can save time by implementing a machine learning model

in the platforms that reads the news title or text and identifies the news by type [1].

News classification datasets are now used to sort machine translation texts into different categories based on their content. Consider categorizing news stories by subject or categorizing textbooks by beneficial or harmful opinion. News categorization is indeed useful for detecting languages, managing client feedback, and eliminating risk. Machine learning models can automate this procedure, which is time-consuming when done by hand. News categorization is a multi-label text classification challenge. The purpose is to allocate a news report to one or more categories. A group of binary classifiers is a common strategy for multi-label text classification.

In our paper, we used a 200k HuffPost news dataset to classify the category of news. Since we use a very large dataset it requires gpu's and tpu's to train quickly.Then,the news dataset is trained on using seven different types of Machine Learning and Deep Learning Algorithm such as BERT,Logistic Regression,Gaussian Naive Bayes,Random Forest,Multinomial Naive Bayes,Decision tree classifier and Support Vector classifier model.Excluding the BERT, the accuracy scores for the other six algorithms were generated first. Multinomial Naive Bayes had the highest accuracy of 46.66 percent between them.But on the other hand Support Vector Classifier, Logistic Regression, and Random Forest have 46.46 percent, 44.77 percent and 41.19 percent accuracy scores correspondingly. The accuracy scores of the Decision Tree classifier and Gaussian Naive Bayes classifier were 25.62 percent and 12.94 percent correspondingly and were the lowest among all the others.Finally,we found that Bert has the highest accuracy which is 70.45 percent among all the seven machine learning and deep learning models we used for multiple news classification in our project.

II. Literature Review

In this section, we will briefly discuss some existing news classification techniques which were proposed by other authors. Various news classification methods have been implemented by different authors to achieve better classification results.

Nugroho et al. [2] proposed a Spark NLP approach for large scale news classification. MapReduce improves text processing efficiency by ensuring parallelization of large computations. Authors used a transformer architecture based BERT model to classify large amounts of news data. To build news classification models in Spark NLP, text processing and word embedding used from the BERT pre-trained model. A multi class text classifier named ClassifierDL used in Spark NLP to create a classifier. The authors performed the classification task using the BERT model with and without Spark NLP.

G. Kaur and K. Bajaj [3] established a neural network technique for news classification. The authors discussed Naïve Bayes, Support Vector Machine, Artificial Neural Network, Decision Tree, K-nearest neighbor to classify the news dataset. An overall review of news classification methods briefed in this paper. News classification helps users to get desired news in real time.

S. Divya et al. [4] demonstrated machine learning techniques for audio based news classification. In this study news videos analyzed and classified based on audio content with the help of machine learning methods. Before watching the video content this method helps the user to find the genre of news. NLP applied to get unigrams and bigrams. The classifying ability of different classifiers were analyzed in this study. This study ensures that the Multinomial naïve bayes classifier gives the highest performance compared to other classifiers.

J. Ahmed and M. Ahmed [5] proposed machine learning techniques for online news classification. Authors mentioned that a classification approach is required to convert unstructured data into structured form of data. Automatic text classification methodology helps to use the data in the correct way. Text labeling which is a supervised learning technique used to categorize unlabeled texts into labeled texts. Bayesian classifier outperforms in this study and gave highest accuracy in automatic classification of online news articles.

Wei et al. [6] discussed a Feature Extraction based news-comment relevance classification algorithm. Irrelevant news comments are found in different news articles. This study demonstrates news-comment relevance analysis issues. BERT feature extraction strategy proposed here for this type of classification problem. For the extraction of feature vectors of news, an extensive semantic database trained model is used by this algorithm. This algorithm takes the extracted feature vectors as input for classification tasks.

W. Jing and Y. Bailong [7] proposed a Wide and Deep Bert Model based technology for news text classification. Sparse feature information creates problems with text classification. This study demonstrated a deep learning method for text classification and users personal recommendation. Multiple English text data used here for learning text features. BERT pre-trained model along with Wide and Deep BERT model discussed in this study. In the verification step Tensorflow deep learning framework used here.

A. Marium and G S Mamatha [8] conducts a study on fake news classification using Cloud Processing Capacity. The authors used Bi-directional Encoder Representation for Transformers (BERT) for detecting fake news. It is tested on Google Cloud GPU capacity. For training purposes tokenized corpus transferred into tensors. Least validation loss granted for accuracy. This study demonstrates an appropriate cloud platform for hosting this type of model.

A. Singh and G. Jain [9] proposed a simple transformers technique for sentiment analysis of news headlines. Simple transformers are interrelated with natural language processing. Four different transformers models are used here to perform sentiment analysis. Fine tuned and pre trained models used in this study. BERT, RoBERT, Distilled BERT and XLNet base case used here to improve the efficiency of the model. The model outperforms in Bert based cases.

M. Varasteh and A. Kazemi [10] proposed a Persian News Classification method by using ParsBert on Augmented Data.

Machine learning based methods like Naïve Bayes, Support Vector Machine require feature engineering and they are straight forward. They truly depend on selected features. The authors augmented the data and then used a pre-trained model named ParsBERT for classification tasks. This study also presents a comparison with other traditional machine learning methods.

D. Lorga et al. [11] discussed a method for detecting the vulnerabilities from news websites. They proposed a machine learning model for the early detection of the vulnerabilities. This study is a part of automatic detection of early cybernetic threats using Open Source Intelligence (OSINT). Support Vector Machine, Multinomial Naïve Bayes classifier and fine tuned BERT model demonstrated in this model. The authors ensure that NLP models are easily applicable and well fitted for extracting relevant information from cyber security news articles.

III. Dataset Description

The dataset in this paper is collected from HuffPost and contains almost 200k news headlines between the year 2012 to 2018. For having a huge amount of data, it almost covers every news category with an ample number of data. It has 5 attributes for each data which are ”category”, ”headline”, ”authors”, ”link”, ”short_description” and ”date”. Here, ”category” was taken as class or label and ”headline”, ”short_description” were used as features to predict and train the model.

TABLE I
FREQUENCY OF EACH CATEGORY IN DATASET

Category	Count	Category	Count
POLITICS	32,739	CRIME	3,405
WELLNESS	17,827	MEDIA	2,815
ENTERTAINMENT	16,058	WEIRD NEWS	2,670
TRAVEL	9,887	GREEN	2,622
STYLE & BEAUTY	9,649	RELIGION	2,556
PARENTING	8,677	STYLE	2,254
HEALTHY LIVING	6,694	SCIENCE	2,178
QUEER VOICES	6,314	WORLD NEWS	2,177
WORLDPOST	6,243	TASTE	2,096
FOOD & DRINK	6,226	TECH	2,082
BUSINESS	5,937	MONEY	1,707
COMEDY	5,175	ARTS	1,509
SPORTS	4,884	FIFTY	1,401
BLACK VOICES	4,528	GOOD NEWS	1,398
HOME & LIVING	4,195	ARTS & CULTURE	1,339
PARENTS	3,955	ENVIRONMENT	1,323
WEDDINGS	3,651	COLLEGE	1,144
WOMEN	3,490	LATINO VOICES	1,129
IMPACT	3,459	CULTURE & ARTS	1,030
DIVORCE	3,426	EDUCATION	1,004

According to Table-1, this dataset has 40 unique news categories containing 32,739 data for ”Politics” from 200k data. Likewise, the ”Wellness” and ”Entertainment” categories have the 2nd and 3rd highest amount of data (17,827 and 16,058 respectively). However, ”Culture Arts” and ”Education” have the least amount of data among all other 40 categories.

In figure 1, illustrations of word frequencies taken from “short_description” and ‘headline” represent three different categories (“Politics”, ”Entertainment” and “Wellness”) among 40 categories are shown. The highlighted and most occurred word from the ”Entertainment” class is star, movie show,

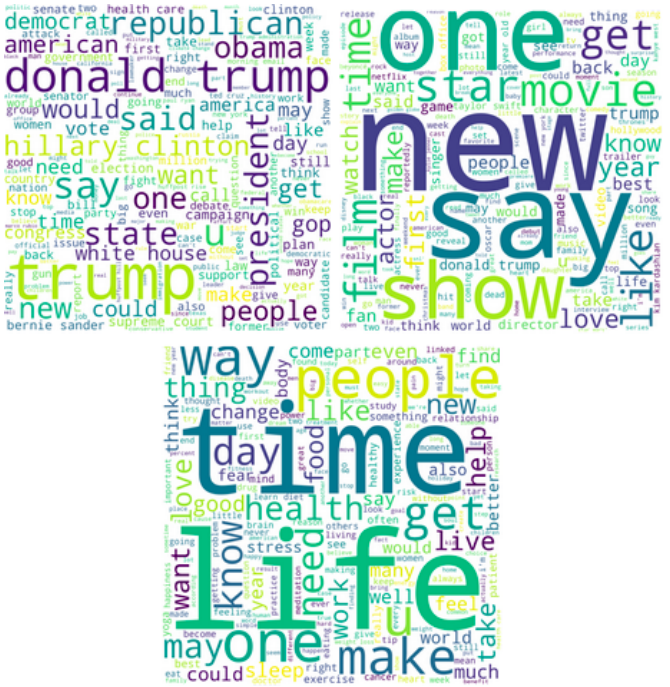

Fig. 1. The frequent words of ”politics”, ”entertainment”, ”wellness” category

film, actor, game and many more. The ”Politics” category has frequent words like donald, trump, hillary, obama, democrat, republican and so on. Similarly, life, time, people, health words are quite common in the ”Wellness” category according to this dataset.

IV. Proposed Model

Logistic regression, Gaussian Naive Bayes, Random Forest, multinomial Naive Bayes, decision tree, support vector machine algorithms along the BERT model have been used to train the news classifier in this paper.

SVM is a supervised machine learning technique that is being used to assist with classification and regression difficulties. Its goal is to find the optimum solution among the probable outputs. Text classification is one of the most popular relevant topics for multiclass classification using SVM. When the data set contains more noise, such as overlapping target classes, SVM does not function well. The SVM will be very slow if the number of features for each data point exceeds the number of training data samples. For this paper, the huge dataset of news classification with multiclass was the most challenging thing in training and testing period.

Random Forest is one of the most user-friendly among all of the machine learning algorithms. It involves a significant amount of computational power along with resources because it constructs several trees and combines their outcomes. It also takes longer to train because it uses several decision trees to decide the class.

Naive Bayes classifiers are usually applied in text classification have a greater success rate in resulting in better prediction for the multi-class problem. Here Multinomial Naive Bayes

and Gaussian naive bayes were used for news classification. However, this algorithm doesn't care about the word sequence and ends up predicting wrong. As we had used a large dataset, it is not very wise to rely on the result of only this classifier.

A decision tree classifier is a method for systematically categorizing multiple classes. It asks the dataset a series of questions about its attributes and features. Because of the complexity and time required, decision tree is relatively costly.

SVM is a supervised machine learning technique that is being used to assist with classification and regression difficulties. The goal of this classifier is to find the optimum solution among the probable outputs. Text classification is one of the most popular relevant topics for multiclass classification using SVM. When the data set contains more noise, such as overlapping target classes, SVM does not function well. The SVM will be very slow if the number of features for each data point exceeds the number of training data samples. For this paper, the huge dataset with multiclass was the most challenging thing in training and testing period.

Logistic Regression is not a great choice for multiclass classification but we have used this one to compare with all other classifiers accuracy scores. Bert-large-uncased model, a masked language modeling (MLM) objective was used to train a model on the English language. This model is uncased, meaning it does not distinguish between Bengali and English or capital and small letter. When given a sentence, the model masks 15% of the words in the input randomly, then runs the entire masked sentence through the model, which must predict the masked words. This is distinct to classic recurrent neural networks (RNNs), which often see words sequentially, which masks potential tokens or words. It enables the model to learn a sentence's bidirectional representation.

As 6 traditional ML algorithms has been used in this paper apart from bert-large-uncased model, TPU was only configured to build BERT model. As Bert is a huge model, it requires tpu's to train fast. Tensor Processing Unit (TPU) is an integrated circuit and open-source Framework by google which is customized with high speed network for machine learning and TensorFlow.After the TPU configuration is done, the dataset needs to be cleaned and transformed to apply different classifier. First, the label "THE WORLDPOST" was replaced by "WORLDPOST" for being in the same category. In this dataset, we have 40 unique categories of news. The categories were encoded using the Label Encoder function which assigns a value between 0 and no of unique categories -1. This is how the categories were mapped to an integer.

Here we have used 6 other ML models and a bert-large-uncased model where all features need to be in the same case, so the features' data were transformed into lower case. Length of the features headlines and description played an important role. From Table 2, lots of the minimum length of headlines and descriptions is 0 which means it doesn't have any words in many of them. The figure 3 is indicating that the description feature had no words in almost 3% cases but the headline had few words in most of the cases. As a result, the headlines were concatenated before a short description field to make those data usable.

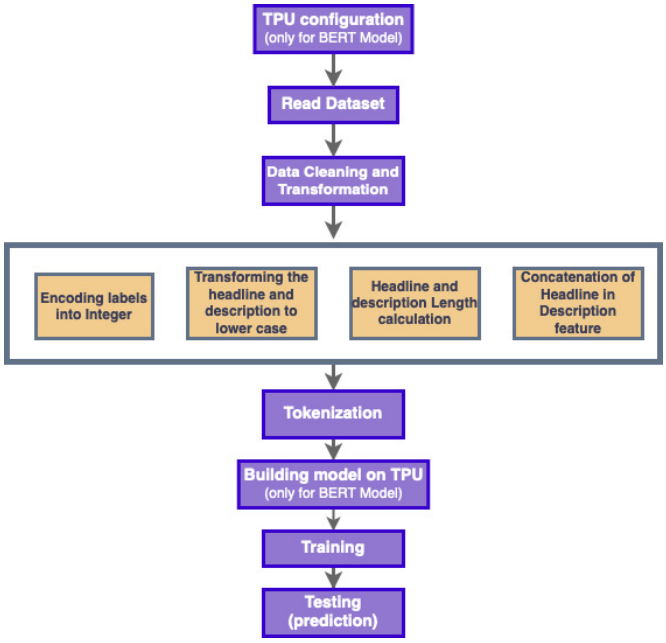

Fig. 2. The Work Flow diagram of our model

TABLE II
SUMMARY OF DESCRIPTION AND HEADLINE LENGTH

	categoryEncoded	descr_len	headline_len
count	200853.000000	200853.000000	200853.000000
mean	22.024366	19.728289	9.538563
std	10.753991	14.409189	3.089320
min	0.000000	0.000000	0.000000
25%	13.000000	10.000000	7.000000
50%	24.000000	19.000000	10.000000
75%	30.000000	24.000000	12.000000
max	39.000000	243.000000	44.000000

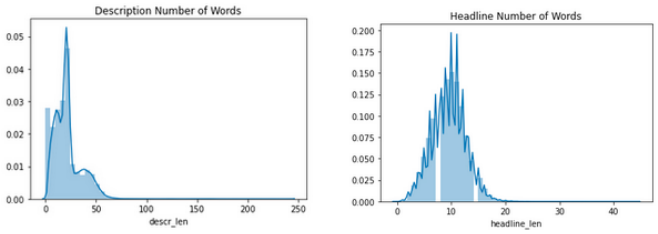

Fig. 3. The length of description and headline

After data cleaning and transformation, the HuggingFace tokenizer has tokenized the data using pretrained data for bert-large-uncased. The dataset has been split into train and test datasets with a 7:3 ratio. Also, tf.keras.utils.to_categorical has been used to tokenize the news descriptions and to convert the categories into one-hot vectors.

Here, To build the BERT model and other algorithms, categorical cross entropy is used which is a loss function for multi-class classification models with two or more output labels. A single-hot category encoding value in the form of

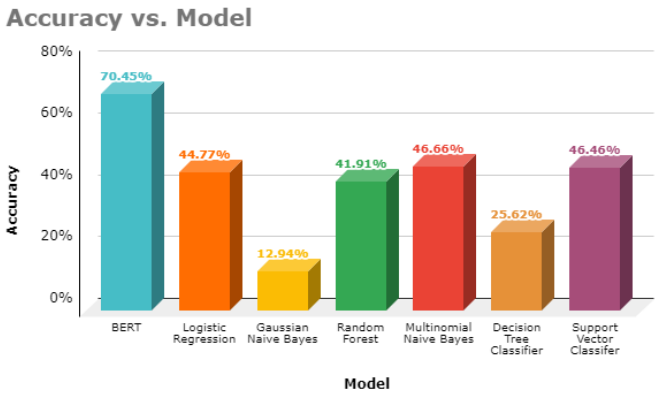

Fig. 4. The accuracy of the different classifiers

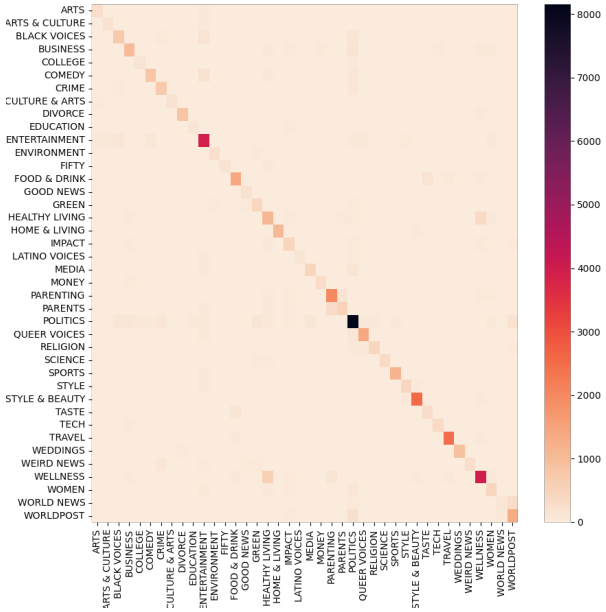

Fig. 5. Heatmap of the BERT classification

0s and 1s is allocated to the output label. If the output label is in integer form, Keras is used to convert it to categorical encoding. Again, the same pretrained method is used to build the bert-large-uncased model on TPU(other classifiers were run without using TPU).

As building the model on TPU is completed, AUTOTUNE allocated the CPU budget to assign the value dynamically at runtime. The data was trained 10 times to improve the accuracy level and to decrease the loss value. After the training dataset, the test dataset was used to predict the class. As the dataset was encoded while data preprocessing, encoding was performed to transform them back into actual categories. Lastly, Accuracy score and confusion matrix will determine if Bert method is suitable for this multiclass news classification or not. A Confusion matrix is an N x N matrix to evaluate a classification model's performance, where N is the number of unique classes.

V. Result Analysis

As discussed in the method and from figure 4, accuracy scores for the other 6 algorithms except BERT were calculated first. Among them, multinomial Naive Bayes had the highest accuracy of 46.66%. Then support vector classifiers, logistic regression and random forest have 46.46%, 44.77% and 41.91% respectively. 25.62% and 12.94% were the accuracy score of the decision tree classifier and Gaussian Naive Bayes classifier, which were the least among the mentioned ones.

Because of having such low accuracies(not more than 46.66%), the BERT model has been used in this paper to achieve a higher accuracy score. The accuracy of the BERT Model was surprisingly high(70.45%) for the same dataset. As a result, the BERT model is the most reliable algorithm for this kind of language classifier project. Figure 5 illustrates a confusion matrix heatmap (exact deviation) for multi-class news classification using the BERT model; the "Politics' ' category was detected with more than 8,000 occurances. However, this classification was also confused by the classifiers "Black Voices' ', "Business", "Comedy", "Crime", "WorldPost' ' and so on. "Wellness" and "Entertainment" was classified correctly for more than 4,000 data each. On the contrary, "Wellness" was also mistaken with Healthy living and other classes. Similarly, "Entertainment" was misunderstood by "Black Voices" and "Comedy". "World News' ' has one of the least occurrences in the dataset and was not misled much while predicting the class with the BERT model.

In short, this confusion matrix heatmap has most of its value in diagonal cubes which emphasize that the BERT model has predicted most of the classes correctly compared to the wrong prediction.

TABLE III
Classification Report of BERT Classifier

Category	Precision	Recall	F1-score	Support
ARTS	0.46	0.65	0.54	435
ARTS & CULTURE	0.53	0.6	0.56	356
BLACK VOICES	0.58	0.51	0.54	1384
BUSINESS	0.6	0.57	0.59	1758
COLLEGE	0.48	0.53	0.5	353
COMEDY	0.64	0.55	0.59	1504
CRIME	0.59	0.7	0.64	1020
CULTURE & ARTS	0.66	0.6	0.62	335
DIVORCE	0.83	0.82	0.82	1036
EDUCATION	0.52	0.47	0.49	318
ENTERTAINMENT	0.76	0.81	0.78	4871
ENVIRONMENT	0.59	0.6	0.6	393
FIFTY	0.64	0.41	0.5	449
FOOD & DRINK	0.8	0.76	0.78	1850
GOOD NEWS	0.39	0.48	0.43	452
GREEN	0.47	0.54	0.5	784
HEALTHY LIVING	0.44	0.57	0.49	1926
HOME & LIVING	0.86	0.83	0.84	1273
IMPACT	0.39	0.44	0.42	1053
LATINO VOICES	0.66	0.43	0.52	348

Category	Precision	Recall	F1-score	Support
MEDIA	0.68	0.55	0.61	835
MONEY	0.63	0.61	0.62	508
PARENTING	0.69	0.75	0.72	2640
PARENTS	0.56	0.46	0.5	1179
POLITICS	0.8	0.83	0.82	9843
QUEER VOICES	0.78	0.76	0.77	1880
RELIGION	0.65	0.55	0.6	796
SCIENCE	0.72	0.57	0.64	613
SPORTS	0.79	0.79	0.79	1461
STYLE	0.62	0.63	0.63	680
STYLE & BEAUTY	0.88	0.89	0.88	2881
TASTE	0.51	0.51	0.51	617
TECH	0.6	0.59	0.6	606
TRAVEL	0.85	0.86	0.85	2937
WEDDINGS	0.89	0.84	0.86	1097
WEIRD NEWS	0.6	0.33	0.43	817
WELLNESS	0.78	0.75	0.76	5320
WOMEN	0.43	0.42	0.43	1074
WORLD NEWS	0.54	0.24	0.33	657
WORLDPOST	0.61	0.7	0.65	1887

The classification report(Table - 3) was generalised to find out the quality of the prediction from the BERT classification method. Precision and recall are two most important models of evaluation metrics. Precision value refers to the relevance percentage of the predicted value. "Weddings' ' class has a precision value of 0.89 and the most relevant prediction

compared to other classes. Even though the "Impact" category has a precision score of as low as 0.39, Bert model didn't evaluate or justify data from each category equally.

Recall Validates the percentage of total relevant accurate results classified by the model. The recall value of "Style and Beauty" is 0.89 which means it was predicted correctly most of the time but "World News" was predicted wrong most of the times with a recall score of 0.24.

"Style and Beauty" has the greatest F1-score of 0.88 which makes the prediction almost accurate. Whereas, 0.33 is the F1-score of "World News" and the least accurate among all the categories. By analyzing the score of F1-score, not all the categories were equally predicted accurately with this BERT model.

In this multiclass classifier, support values from the table indicate that "Politics" has the highest number of actual occurrences, which is 9843. On the other hand, "Education" has the least occurrence(318) among all the 40 categories. This is clear that it is an unbalanced dataset.

From the above classification report, the Recall and F1-score has shown mostly the same behaviour for most of the categories. In general, these four performance evaluation metrics represent that this BERT model can predict almost 90% of the time, but unbalanced data is making it harder to predict each category correctly.

VI. Conclusion

A text classifier system was used in this study to categorize online news posts. Approximately 200k HuffPost news datasets were used to classify the news category from seven different websites using their tags. Besides, the dataset analysis was demonstrated by incorporating seven distinct types of classifiers. In the paper, we used BERT, Logistic Regression, Gaussian Naive Bayes, Random Forest, Multinomial Naive Bayes, Decision tree classifier, and Support Vector classification algorithms for classification and performed a comparison of different classifiers' outcomes from seven different sites on the same dataset, and the findings confirm that BERT performed better than most classifiers; when working with various news datasets, it provides adequate classification accuracy. Moreover, The BERT model is a good NLP model for large-scale work.

A. Future work

We intend to improve our framework in the future by analyzing additional architectures and pre-trained models to improve classification performance and computational resources. Furthermore, we try to look into the effects of text preprocessing prior to training. Our long-term goal is to create and implement a classification methodology in multiple regional languages.

References

[1] A. F. Akyürek, L. Guo, R. Elanwar, P. Ishwar, M. Betke, and D. T. Wijaya, "Multi-label and multilingual news framing analysis," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.

[2] K. S. Nugroho, A. Y. Sukmadewa, and N. Yudistira, "Large-scale news classification using bert language model: Spark nlp approach," in *6th International Conference on Sustainable Information Engineering and Technology 2021*, 2021, pp. 240–246.

[3] G. Kaur and K. Bajaj, "News classification using neural networks," *Commun. Appl. Electron*, vol. 5, no. 1, pp. 42–45, 2016.

[4] S. Divya, R. Raghavi, N. Sripriya, S. Mohanavalli, and S. Poornima, "Analysis of audio-based news classification using machine learning techniques," in *International Conference on Mathematical Analysis and Computing*, Springer, 2019, pp. 429–441.

[5] J. Ahmed and M. Ahmed, "Online news classification using machine learning techniques," *IIUM Engineering Journal*, vol. 22, no. 2, pp. 210–225, 2021.

[6] H. Wei, W. Zheng, Y. Xiao, and C. Dong, "News-comment relevance classification algorithm based on feature extraction," in *2021 International Conference on Big Data Analysis and Computer Science (BDACS)*, IEEE, 2021, pp. 149–152.

[7] W. Jing and Y. Bailong, "News text classification and recommendation technology based on wide & deep-bert model," in *2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE)*, IEEE, 2021, pp. 209–216.

[8] A. Marium and G. Mamatha, "Bert model for classification of fake news using the cloud processing capacity," in *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)*, IEEE, pp. 1–6.

[9] A. Singh and G. Jain, "Sentiment analysis of news headlines using simple transformers," in *2021 Asian Conference on Innovation in Technology (ASIANCON)*, IEEE, 2021, pp. 1–6.

[10] M. Varasteh and A. Kazemi, "Using parsbert on augmented data for persian news classification," in *2021 7th International Conference on Web Research (ICWR)*, IEEE, 2021, pp. 78–81.

[11] D. Iorga, D. Corlătescu, O. Grigorescu, C. Săndescu, M. Dascălu, and R. Rughiniş, "Early detection of vulnerabilities from news websites using machine learning models," in *2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, IEEE, 2020, pp. 1–6.