BlackOps at CheckThat! 2021: User Profiles Analyze of Intelligent Detection on Fake Tweets Notebook for PAN

S. M. Sohan [1], Sharun Akter Khusbu[1], Md. Sanzidul Islam[1] and Md. Arid Hasan[1]

[1] *Daffodil International University, Dhaka, Bangladesh*

Abstract

An expensive task is fake news detection for recent trends among the concept of misinformation or rumors. In everywhere most of the times information lead or play emergent preface but forthwith misinformation also in everywhere to mislead the peoples mind and activity. Therefore, detecting fake content in any system can be a weapon over fictitious news. In any language cross over the exponential growth of fake news in social sites. Hence, it is the real time process to produce online fake news so that it has been needed to implement an automated technique whenever detect true from false. According to the solution of this approach made a research

On English language textual inputs as twitter news from user profiles. At this point, due to accurate analysis for social media we experimented with supervised learning such as Decision tree, Random forest and gradient boosting. In between all the ML classifiers outperformed with 88% detection accuracy that mention the research of detection is more accurate.

Keywords

Decision Tree, Random Forest, Gradient boosting, ML, Fake news detection.

1. Introduction

Quotidian information or opinions are paving the both way of positive and negative as a text version. Thus, a vast amount of text and news has been split around the world from person to person online. Tropically, in any language the public used to make comments, news, gossip, debate and individual opinions for their own conception. In that way, miscellaneous comments are produced by the daily activities so that people use abusive or wrong concepts over actual comments. Among these occurrences, fake news reaches the common people moreover people are getting confused between fake news and raw news. Therefore, authentication of any news is to be difficult or doubtful to be identified. This unnecessary situation is responsible for producing more information and each news mixes up with the fake news.

Rumors spread by this fake news which is interpreted also make the purpose of manipulation in different concepts [5]. Within milliseconds all over the world spreading misinformation. However, now is the time to stop spreading rumors and news, wrong concepts. Therefore, there has been a necessity for proper tools to enhance and solve these issues. Due to the emergence of web tools may reduce the maximum number of fake news or misconceptions. Many fake news detection had been completed among different languages such as similar works on Urdu language augmentation over fake data [6] and several works over relational features on social media, entities and facts in fake mining text [7]. Fake opinion detection in social networks [8] and text messages in online tools [9] are the areas of communication activity where people used to text on any news. Fake news detection with gradient boosting [10] outperformed in large datasets whenever text input as a sentence.

[1]CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania
EMAIL: sohan15-10466@diu.edu.bd (A. 1); sharun.cse@diu.edu.bd (A. 2); sanzid.cse@diu.edu.bd (A. 3); arid.cse0325.c@diu.edu.bd (A. 4)
ORCID: 0000-0001-9567-7628 (A. 1); 0000-0002-8900-8580 (A. 2); 0000-0002-9563-2419 (A. 3);0000-0001-7916-614X (A. 4)

CEUR Workshop Proceedings (CEUR-WS.org)

Hence, news detection aims to solve people's anxiety with a new invention system which is detecting deceptive news. Basically, there needs to be an involvement of truthful and fraudulent prior reviewed news. In [11-15] authors are highlighted on hybrid features in fake data, different self-made corpora dataset in Bangla news detection, benchmark work on Urdu fake news, fake news on covid-19 issues and another analysis on UrduFake2020 900 annotated data analysis with ML, CNN, BERT. Many authors covered fake news in three categories such as knowledge based, context based and style based. Each individual who works there has challenges with difficulties hence available resources and dataset is limited. To the best of our knowledge work procedure summarized as follows:

1. In our proposed work, we started a competition in CLEF! Check That Lab 2021 with Task3 dataset corpus on misinformation over English language dataset.
2. After collecting the csv started the analysis along with columns and data volume and completed preprocessing method.
3. Proper experiments for a new system on fake news detection using Bidirectional LSTM which is motivated by linguistic features.

In every section described precisely as follows. Brief background is related research given in Section 2, Section 3 is the methodology on overall work procedure and Section 4 describes the final results and outcomes discussion at last the necessary Section 5 for conclusion and future works scope.

2. Literature Review

CLEF Check that! workshop provides the dataset on different task [35-36] after applying the train and test set that number of research work present in this workshop [37-38]. It's very common that, different kinds of news are split by social media. All that news is not real. So, fake news is a very critical mass in this era. So, here we mentioned the two major components. Which are false news predictions on social platforms and author profiling.

False news prediction on social platforms: Basically, false news on social platforms are generated by two major perspectives. One is social context based and the another is news context based. News based contents are visual and textual and those fake news are split based one these visual and textual contexts with an incorporated approach. For example, in this paper the writer compares both news. Where he finds emotional concern in the language of fake news. Because, people prioritize the emotion one more. The approach is to collect news from different sources like newspapers or someone's user profile in social media and mix up those news and split them again with an incorporated approach [1]. Arthur is actually trying to say that there are two kinds of users on social platforms. First type believes in false news and shares false news. The second type is to believe in real news and split it. These two groups of people are used to perform the fake news classification task [2]. This paper is very relative to the topic of fake news splitting. Here, Arthur describes how it actually works [3]. In an publication, here it describes that commonly fake news split by social bots [4]. In another research paper, here it discusses tweet distributions. Classification via features, such as the account age and similar was also shown to work well [10]. In a recent survey, results showed that fact checking is a very important step to maintain social platform news quality. By employing automated systems, capable of prioritizing potentially interesting users, less time is spent on manual curation, which can be an expensive and time-consuming process [25]. In the recent work, a new thing has been proposed, where it builds a model which uses user context like a temp text data against plain context and fuse the context information [26]. After conducting a thorough study of 83 classes from fact-checkers, we define four fake news classes in the following article [31]. Fake article classification(Benchmark classification) is defined in [32]. We used the unique of collecing the data, we put human in the loop to get the high quality data, Steps used in Data collection defined in [33]. Domain Categorization [34].

Author Profiling: Author profiling is a very important fact in this regard. Since 2013, PAN has annually taken author profiling as a shared task. In the time between 2013 to 2020, many facts of author profiling have been covered in different media platforms. For example, social bots, age detection and gender detection. At PAN task in different volumes, it shows that variations of analysis perform best for textual classification [27]. Here the participants use a Support Vector Machine classifier, with many types of word and character like n-grams as attributes [29]. Many used stylistic analysis in the research as authors acceptance statistics in order to use term occurrences [28]. Emotions impact on author profiling has also been justified before, use of other attributes like subjective measurement has been neglected in this task [30].

3. Methodology

This work has been completed through four steps. The general discussion of those steps is illustrated here. The first step is selecting the acceptable pretend news dataset from Conference and Labs of the analysis Forum and preprocessing the dataset. Once that, Classify the dataset using (Decision Tree, Random Forest, Gradient Boosting) classifiers and measure model performance exploitation using totally different metrics like (accuracy, recall, and precision) as represented in Figure 1.

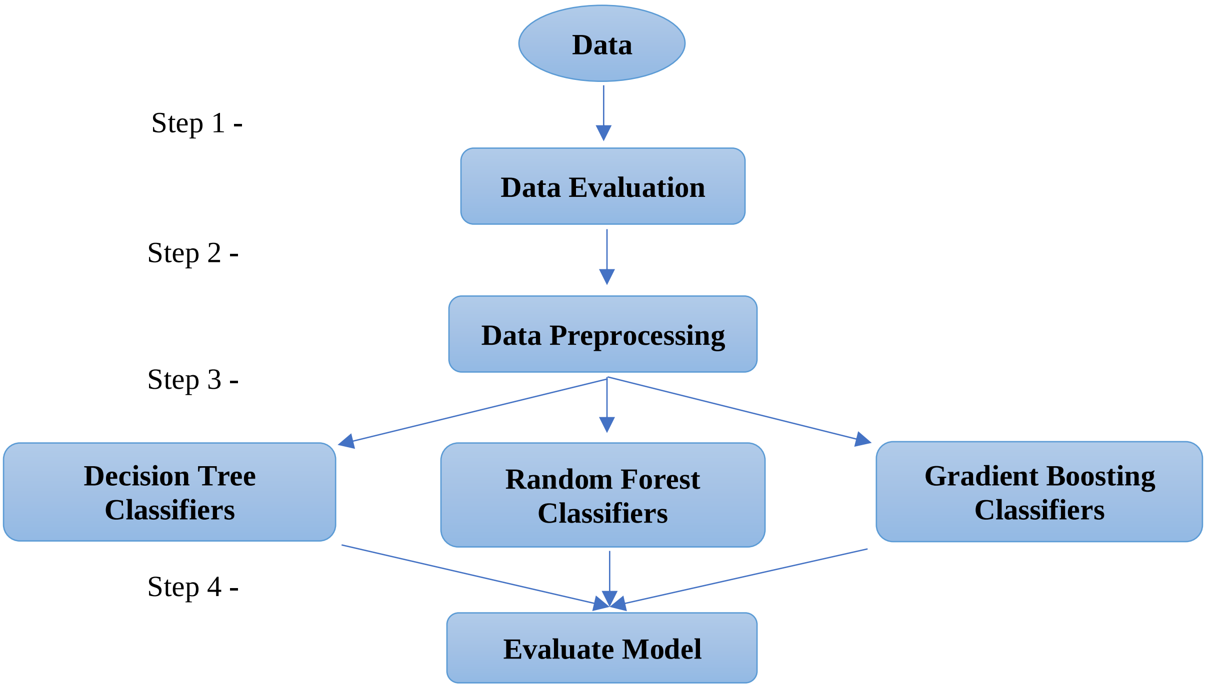

Figure 1: Work design step of fake news detection.

3.1. Data description

We collected the data from a competition named "CLEF2021". Thays arrange workshops every year. The coaching knowledge is discharged in batches and roughly comprises 900 articles with the various labels and add more 60,752 data from the same place then total 61,652. Given the text of a news article, verify whether or not the most claim created within the article is 23,727 of True News, 237 of Partially False News, 33,610 of False News and 4,078 of Other News, as shown in Figure 2.

Our definitions for the classes are as follows:

- **False -** the most claim made in a commentary is untrue.
- **Partially False -** the most claim of an article could be a mixture of true and false info. The article contains partially true and partially false information however can't be thought of one hundred pc true. It includes all articles in categories like partially false, partially true, mostly true, miscaptioned, deceptive etc., as outlined by completely different fact-checking services.

- **True -** This rating indicates that the first parts of the most claim are incontrovertibly true.
- **Other-** a commentary that can't be classified as true, false, or part false because of lack of proof concerning its claims. This class includes articles relevant and unverified articles.

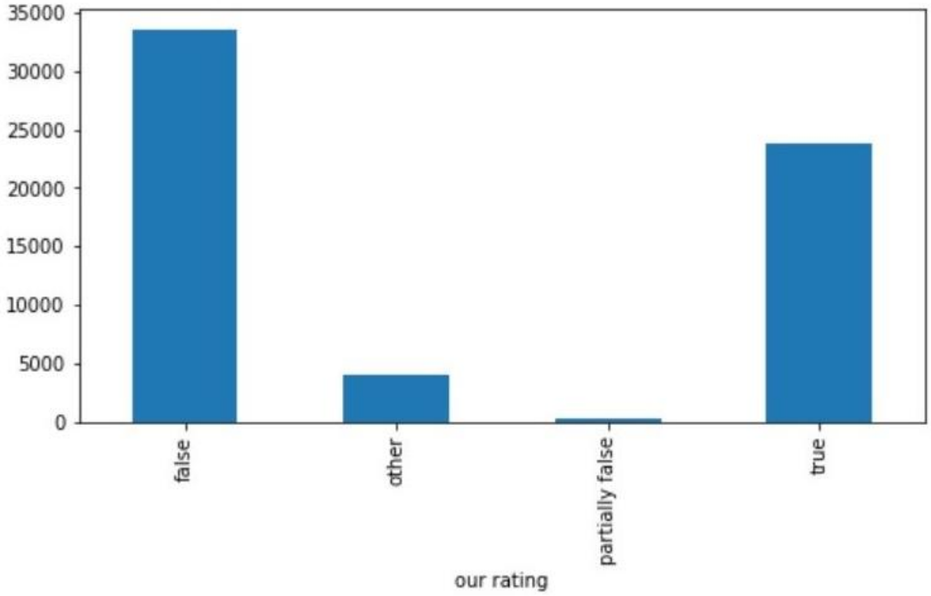

Figure 2: Count Label.

3.2. Data preprocessing

The data must be subjected to certain filtering and cleansing processes, e.g. B. Removing stop words, punctuation marks, removing upper and lower case letters and removing special characters, numbers, spaces and adding class columns Where True is 1, False is 2,Partially False is 3 also Other is 4[16]. By removing the intangible information contained in the data, the size of the data set is reduced and only the valuable information remains in the data set [17-18]. Table 1 shows an example of the data set used, which represents the raw data collected. With no preprocessing step, while Table 2 shows the data after the preprocessing step.

Table 1
Before the preprocessing step

	text	our rating
0	Distracted driving causes more deaths in Canada	FALSE
1	Missouri politicians have made statements after	partially false
2	Home Alone 2: Lost in New York is full of viol	partially false
3	But things took a turn for the worse when riot	FALSE
4	It’s no secret that Epstein and Schiff share a	FALSE

Table 2
After the preprocessing step

	text	our rating	class
0	distracted driving causes more deaths in canada	FALSE	2
1	missouri politicians have made statements after	partially false	3
2	home alone lost in new york is full of viol	partially false	3
3	but things took a turn for the worse when riot	FALSE	2
4	It s no secret that epstein and schiff share a	FALSE	2

3.3. Model learning

(i) **Decision Tree -** Decision tree builds classification or regression models at intervals the vary of a tree structure. It breaks down a dataset into smaller associate degreed smaller subsets whereas at the same time Associate in nursing associated decision tree is incrementally developed. The last word ends up in a tree with decision nodes and leaf nodes. Associate degree alternate node has a pair of or a decent deal of branches. Leaf node represents a classification or decision. The easiest decision node throughout a passing tree that corresponds to the foremost effective predictor mentioned as root node. Decision trees can handle every categorical and numerical data [19-20].

Algorithm 1
Decision Tree

Input: Predefined classes
Output: Built decision tree Num of features 17000 Max –depth 2
Begin
Step1: Create a root node for the tree
Step 2: If all examples are positive, return leaf node 'positive.' Else if all examples are negative, return leaf node 'negative.'
Step 3: Calculate the entropy of current state H(S)
Step 4: For each attribute, calculate the entropy concerning the attribute 'x' denoted by H(S, x)
Step 5: Select the attribute which has a maximum value of IG(S, x)
Step 6: Remove the attribute that offers the highest IG from the set of attributes
Step 7: Repeat until we run out of all attributes or the decision tree has all leaf nodes.
End

(ii) **Random Forest -** The core unit of random forest classifiers is the choice tree. The choice tree could also be a data structure that's designed to take advantage of the alternatives of the Associate in nursing information set. Every node of the choice tree is choppy with a live tree involving a bunch of the features. The nodes are split and support the entropy of a selected set of the features. The random forest is an associate assortment of decision trees that are relating to a group of bootstrap samples that are generated from the primary information set. As we all know that a forest is formed from trees and additional trees means that more strong forest. Similarly, random forest algorithm creates call trees on data samples so gets the prediction from every of them and at last selects the most effective resolution by means that of voting. Thorough info on random forest classifiers is found at intervals inside the papers by Breiman. At intervals victimization the quality random forest approach, the bootstrapping technique helps the event of random forest with a group of required vary of decision trees thus enhancing classification accuracy through the conception of overlap dilution as mentioned in Suthaharan 2015. In many cases, the performance of a random forest is like growth, making it easier to train and optimize. Therefore, random forest is a general algorithm suitable for multiple packets [21-22].

(iii) **Gradient Boosting -** Gradient boosting classifiers are a gaggle of machine learning algorithms that blend many weak learning models on to form a sturdy revelatory model. Decision trees are sometimes used once doing gradient boosting. Gradient boosting models became common as a results of their effectiveness at classifying powerful data sets, and have recently been accustomed win several Kaggle informatics competitions. The key plan is to line the target outcomes for this next model so as to reduce the error. However are the targets calculated? The Python machine learning library, Scikit-Learn, supports entirely altogether completely totally different implementations of gradient boosting classifiers, at the aspect of XGBoost this text well appraise the speculation behind gradient boosting models classifiers, and look at two alternative ways in which of closing classification with gradient boosting classifiers in Scikit-Learn [23-24].

Algorithm 2
Random Forest

Input: Predefined classes
Output: Built Gradient Boosting Num of features 17000 Num of estimators (num of tree in the forest) 100
Begin
Step 1: extract features from texts (X1, X2, …, Xn: float number)
Step 2: Compute the best splinter point between the n features For the node d.
Step 3: Utilize the optimal splinter point to split the node into two child nodes.
Step 4: Repeat steps 1, 2 to n number of nodes was reached.
Step 5: Build the forest through the repetition of steps 2- 4 for D time
End

Algorithm 3
Gradient Boosting

Input: Predefined classes
Output: Built Forest trees Num of features 5000 Max –depth 7
Begin
Step 1: Compute the negative gradient.

$$\overline{y_i} = -[\frac{\partial L(y_i, F(x_i)}{\partial F_{x_i}}]$$

Step 2: Fit a model.

$$\alpha_m = \arg min_{\alpha,\beta} \sum_{i=1}^{N} [\bar{}y - \beta h(x\_i \ ; \ \alpha\_m)]^2$$

Step 3: Choose a gradient descent step size as.

$$\rho_m = \arg min_{\rho} \sum_{m-1}^{N} L(y_i, \ Fm - 1(x_i) + \ \rho h(x_i; \alpha))$$

Step 4: Update the estimation of F(x).

$$F_m(x) = F_{m-1}(x) + \ \rho_m h(x; \ \alpha_m)$$

End

4. Experiment Result

The classification results showed that the accuracy of the choice tree, random forest and Gradient Boosting classifier is 85%, 87% and 88%, respectively. Figure 3, Figure 4 and Figure 5 represent the resulting confusion matrix with T-true, T-false, T-partially-false T-other, F-true, F-false, F-partially-false and F-other values. Table 3, Table 4 and Table 5 illustrate all results of used analysis metrics applied to classify the fake news accurately.

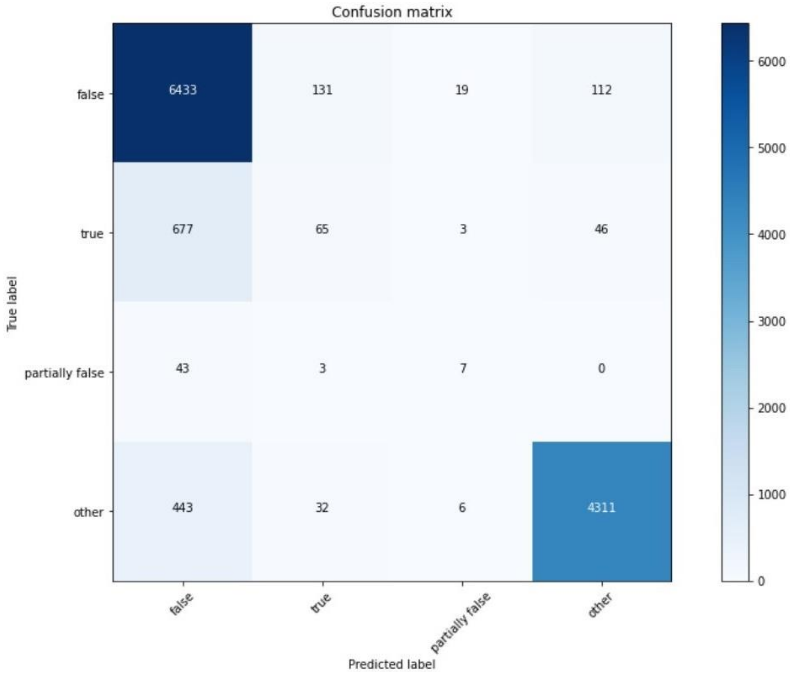

Figure 3: Confusion matrix of Decision Tree.

Table 3
Results of Decision Tree.

Pointer	Result
True	7877
False	11010
Partially False	63
Other	1396
Precision	55%
Recall	54%
Accuracy	85%

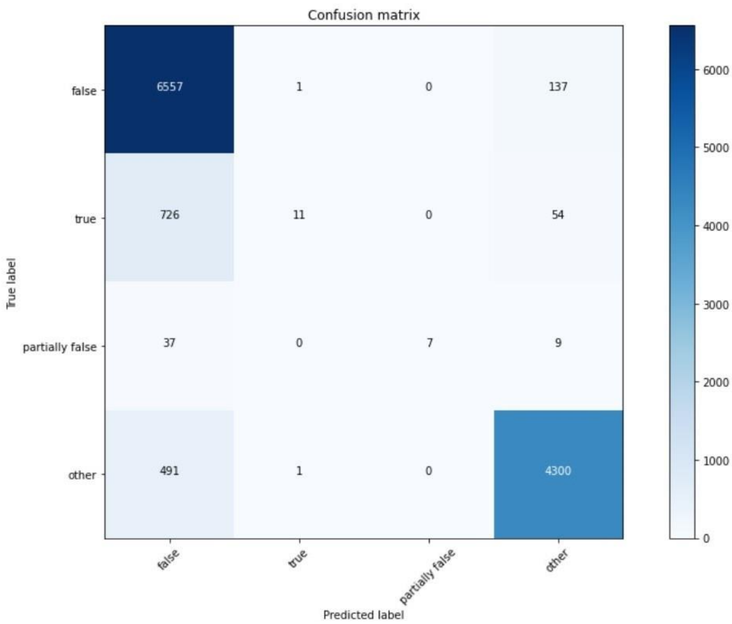

Figure 4: Confusion matrix of Random Forest.

Table 4
Results of Random Forest.

Pointer	Result
True	7877
False	11010
Partially False	63
Other	1396
Precision	80%
Recall	49%
Accuracy	87%

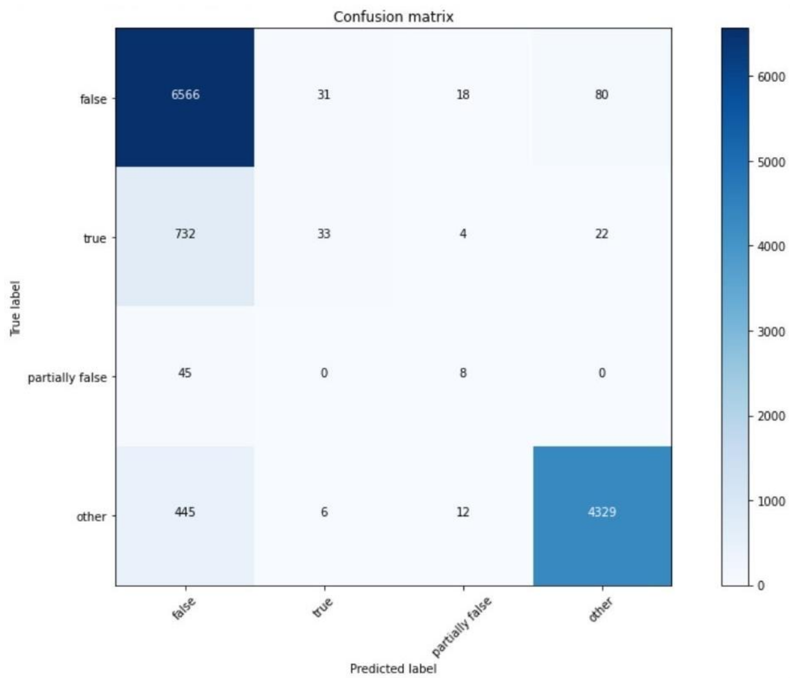

Figure 5: Confusion matrix of Gradient Boosting.

Table 5
Results of Gradient Boosting.

Pointer	Result
True	7877
False	11010
Partially False	63
Other	1396
Precision	61%
Recall	50%
Accuracy	88%

From the results shown above, it seems that the Gradient Boosting outperforms better than a random forest and in call Tree terms of accuracy, wherever the accuracy of gradient boosting equals 88% whereas in random forest equals 87% and decision tree equals 85%. This can be thanks to the characteristics and behavior of every algorithm and its impact on the dataset used. supported our dataset, the options used impotence plays a vital role in classification accuracy since the gradient boosting algorithm offers high importance to some features over others. For these reasons, the gradient boosting with this kind of fake news dataset offers a higher result than the decision tree and random forest within the classifying process.

Additionally, in our results, the random forest prediction takes an extended time than the decision tree, wherever the time of running random forest is (2m 9s) and decision tree is (2m 8s), whereas the Gradient Boosting is (21m 18s). Besides, Internal processes can be checked and thus permit the replica of work. After that, we compared our classification methods' accuracy with the accuracy of alternative connected works.

Table 6
CLEF2021 CheckThat! Lab - Task 3 Results.

Team/Participant Name	Score
SaifuddinSohan	0.38
nomanashraf712	0.38
NLytics	0.38
Ninko	0.35
talhaanwar	0.35
abaruah	0.34

We got a good score in a contest made from CLEF2021 CheckThat. If you look at the Table 6 you will understand that our score has been much better than other participatory score. They used Neural network based Bi-LSTM, LSTM and Other model but our model scored better than them.

5. Conclusion

Detecting fake news spreaders is an important step to control the spread of fake news through social platforms. In our work, we used three kinds of classification algorithms to detect fake news spreaders.Here, we use Decision Tree, Random Forest and Gradient Boosting classification algorithms. We tested the data set with those three classification algorithms. After testing all the data sets with those algorithms, we got the best score from the Gradient Boosting algorithm. Our model obtained an accuracy score of 0.88 in the test data using the Gradient Boosting classification algorithm. It performed better than the Random Forest and Decision Tree algorithm.

There is more opportunity to improve the model. If some use different algorithms like Bi-LSTM or LSTM, the result could be different or even better. Fake news detection is a very versatile topic to research. There is more to do in further research from our opinion.

6. Acknowledgment

This is unbelievable support we’ve got from some of the faculty members and seniors of DIU NLP and ML Research Lab to continue our whole research flow from the beginning. We acknowledge Dr. Firoj Alam for guiding us and informing the secretes of research workshops. Also, we’re thankful to Daffodil International University for the workplace support and the academic collaboration in some cases. Dr. Touhid Bhuiyan and Dr. Sheak Rashed Haider Noori also supported us with guidance, motivation, and advocating in institutional supports. Lastly, for sure we’re thankful to our God always for every fruitful work with our given knowledge.

7. References

[1] Ghanem B., Rosso P., Rangel F., “An Emotional Analysis of False Information in Social News Articles”, ACM Transactions, 20(2), 2019.

[2] Shu K., Zhou X., Wang S., Zafarani R., Liu H., “The Role of User Profiles for Fake News Detection”, ASONAM, pp. 436-439, 2019.

[3] Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology Computer Science (2016)

[4] Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A., Menczer, F.: The spread of fake news by social bots. arXiv preprint arXiv:1707.07592 96, 104 (2017)

[5] K. Shu, A. Sliva, S. Wang, J. Tang and H. Liu, “Fake News Detection on Social Media: A Data Mining Perspective”, 19(1) , 22–36, 2017.

[6] Maaz Amjad, Grigori Sidorov, Alisa Zhila, “Data Augmentation using Machine Translation for Fake News Detection in the Urdu Language”, CIC, pp: 2537- 2542, 2020.

[7] Adrian M.P. Bra, sadoveanu, Razvan Andonie, “Semantic Fake News Detection: A Machine Learning Perspective", IWANN, 2019.

[8] A. Balali, M. Asadpour and H. Faili, “A Supervised Method to Predict the Popularity of News Articles”, 703–716, 1405-5546, 2017.

[9] R. Satapathy, I. Chaturvedi, E. Cambria, S.S. Ho and J.C. Na, “Subjectivity Detection in Nuclear Energy Tweets, 657–664, 1405-5546, 2017.

[10] Gilani Z., Kochmar E., Crowcroft J., “Classification of Twitter Accounts into Automated Agents and Human Users”, ACM, pp. 489 -496, 2017.

[11] Mohammed K. Elhadad, Kin Fun Li, Fayez Gebali, “ A Novel Approach for Selecting Hybrid Features from Online News Textual Metadata for Fake News Detection”, pp 914-925, 2019.

[12] Md Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam, Sudipta Kar, “BanFakeNews: A Dataset for Detecting Fake News in Bangla”, 2020.

[13] Maaz Amjad, Grigori Sidorov, Alisa Zhila, Helena Gomez-Adorno, Ilia Voronkov, Alexander Gelbukh, “Bend The Truth: Benchmark Dataset for Fake News Detection in Urdu Language and its Evaluation”, 2020.

[14] Caio V. Meneses Silva, Raphael Silva Fontes, Methanias Colaco Junior, “Intelligent Fake News Detection: A Systematic Mapping”, pp 168 - 189, 2020.

[15] Maaz Amjad, Grigori Sidorov, Alisa Zhila, Alexander Gelbukh, Paolo Rosso, “Overview of the Shared Task on Fake News Detection in Urdu at FIRE 2020”, CEUR workshop, 2020.

[16] Hadeer Ahmed, Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques", 2017.

[17] Suhad A. Yousif, Venus W. Samawi, "Utilizing Arabic WordNet Relations in ArabicText Classification:NewFeature SelectionMethods", 2019.

[18] Suhad A. Yousif, Islam Elkabani, "Arabic Text Classification: The Effect of the AWN Relations Weighting Scheme". 2017.

[19] Suhad A. Yousif, Hussam Y. Abdul-Wahed, Nadia M. G. Al-Saidi, "Extracting a new fractal and semivariance attributes for texture images", 2019.

[20] Sam F., "Decision Tree Classification with Differential Privacy: A Survey", 2019.

[21] Jehad Ali, R. K., Nasir Ahmad, Imran Maqsood, "Random Forests and Decision Trees", 2019.

[22] Suhad A.Yousif, Venus W. Samawi, Islam Elkaban, Rached Zantout, "Enhancement of Arabic text classification using semantic relations of Arabic WordNet", 2015.

[23] Pritika Bahad, Preeti Saxena,"Study of AdaBoost and Gradient Boosting Algorithms for Predictive Analytics", 2019.

[24] Navoneel Chakrabarty, Tuhin Kundu, Sudipta Dandapat, Apurba Sarkar, Dipak Kumar Kole,"Flight Arrival Delay Prediction Using Gradient Boosting Classifier", 2018.

[25] Zhou, X., Zafarani, R.: Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315 (2018).

[26] Cai C., Li L., Zengi D., “Behavior Enhanced Deep Bot Detection in Social Media”, ISI, pp. 128 – 130, 2017.

[27] Daneshvar S., Inkpen D., Bellot P., Trabelsi C., Mothe J., Murtagh F., Nie J.Y, Soulier L., Sanjuan E., Cappellato L., Ferro N., "Gender Identification in Twitter using N-grams and LSA Notebook for Pan at CLEF2018", 2018.
[28] Johansson F., Cappellato L., Ferro N., Lasada D. E., Muller H., "Supervised Classification of Twitter Accounts Based on Textual Content of Tweets Notebook for PAN CLEF 2019", 2019.
[29] Rangel, F., Rosso, P., Montes-y Gomez, M., Potthast, M., Stein, B., "Overview of the 6th author profiling task at pan 2018: multimodal gender identification in twitter", Working Notes Papers of the CLEF (2018)
[30] Rangel F., Rosso P., "On the impact of emotions on author profiling. Information Processing and Management" 52(1), 73–92, https://doi.org/10.1016/j.ipm, 2015.
[31] Shahi., Gautam Kishore., Dirkson, Anne,. Majchrzak., Tim A, , "An exploratory study of covid-19 misinformation on twitter" 100104, 2021.
[32] Shahi., Gautam Kishore., Nandini, Durgesh., "Fake{C}ovid -- A Multilingual Cross-domain Fact Check News Dataset for COVID-19" http://workshop-proceedings.icwsm.org/pdf/2020_14.pdf, 2020.
[33] Shahi., Gautam Kishore., "AMUSED: An Annotation Framework of Multi-modal Social Media Data" arXiv preprint arXiv:2010.00502, 2020.
[34] Shahi., Gautam Kishore., "A multilingual domain identification using fact-checked articles: A case study on COVID-19 misinformation" arXiv preprint, 2021.
[35] Preslav Nakov., Da San Martino., Tamer., Alberto Barron Cedeno., Ruben Miguez., Shaden Shaar., Firoj Alam., Fatima Haouari., Maram Hasanain., Nikolay Babulkov., Alex Nikolov., Shahi, Gautam Kishore., Struß, Julia Maria., Thomas Mandl., "The {CLEF}-2021 {CheckThat}! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News" ECIR~'21, 639—649, https://link.springer.com/chapter/10.1007/978-3-030-72240-1_75 ,2021.
[36] Preslav Nakov., Da San Martino., Giovanni., Tamer Elsayed., Alberto Barron Cedeno., Ruben Miguez., Shaden Shaar., Firoj Alam., Fatima Haouari., Maram Hasanain., Nikolay Babulkov., Alex Nikolov., Shahi, Gautam Kishore., Struß, Julia Maria., Thomas Mandl., Sandip Modha., Mucahid Kutlu., Yavuz Selim Kartal., "Overview of the CLEF-2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News" CLEF~'2021 ,2021.
[37] Shahi, Gautam Kishore., Struß, Julia Maria., Thomas Mandl., "Overview of the CLEF-2021 CheckThat! Lab Task 3 on Fake News Detection" CLEF~'2021, 2021.
[38] Gautam Kishore Shahi., Julia Maria Struß., Thomas Mandl Task 3: Fake News Detection at CLEF-2021 CheckThat!" 10.5281/zenodo.4714517, https://doi.org/10.5281/zenodo.4714517 , 2021.