

Urdu Fake News Detection Using Ensemble of Machine Learning Models

Asha Hegde, Hosahalli Lakshmaiah Shashirekha

Department of Computer Science, Mangalore University, Mangalore, India

Abstract

False information or fake news has the potential to mislead the public, damage the social order, undermine government legitimacy, and pose a major threat to societal stability. Fake news is not new, but the spread of fake news is only accelerated to a wider audience through social media causing more damage to the society. As a result, early detection of fake news on social media platforms is gaining importance day by day. Concurrently, the difficulty of swiftly identifying the fake news in various languages on social media is becoming more significant as the global use of the internet grows with the influence of the availability of ambiguous information. The majority of the fake news detection models focus on resource-rich languages like English and Spanish. Due to lack of bench marked corpus, fake news detection in languages like Urdu and many Indian languages have garnered very little attention. However, few workshops and shared tasks are being organized to promote fake news detection in resource-poor languages. One among them is UrduFake 2021 - a shared task at FIRE (Forum for Information Retrieval Evaluation) 2021 that encourages researchers to develop working models for detecting fake news in Urdu - a resource-poor language. In this paper, we - team MUCS, describe the ensemble of four Machine Learning (ML) classifiers, namely: Random Forest (RF), Multilayer Perceptron (MLP), Gradient Boosting (GB) and Adaptive Boosting (AB), submitted to UrduFake 2021. Using word uni-grams, character n-grams and fastText Urdu word vectors as features to train the ensembled classifier, the proposed model obtained macro F1-score of 0.552, accuracy of 0.713, and 12th rank in the shared task.

Keywords

Word2Vec, n-grams, Fake news, Machine Learning, Ensemble Learning

1. Introduction

In the era of internet, social media and blogs have become the key sources of news, be it real or fake. With the widespread use of smartphones and the ease of access and freedom to share the content on social media, people are no longer restricted only to be the consumers of news, but are also playing a major role as news producers. As a result, several real/ cooked up/ fake news could be instantly and initially reported on social media even without checking the facts and figures, before they could appear on traditional media [1]. Miscreants/ culprits are taking undue advantage of the technology and spreading fake news on social media platforms. The anonymity of users on online platforms provide a significant chance for fake news spreaders to


Forum for Information Retrieval Evaluation, December 13-17, 2021, India

✉ hegdekasha@gmail.com (A. Hegde); hlsrekha@gmail.com (H. L. Shashirekha)

🌐 <https://mangaloreuniversity.ac.in/dr-h-l-shashirekha> (H. L. Shashirekha)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

manipulate peoples' thoughts, social trust and generate wrong opinions [2]. Fake news also has the potential to undermine society's unity while also having a detrimental influence. For example, Russia developed phoney accounts and social bots in order to promote misleading tales [3]. When public opinion is required for any event, fake news thrives and becomes widespread on social media. Following the 2016 presidential election in the United States, false news and its consequences are widely addressed [4].

Manually detecting the ever-increasing amount of fake news is time-consuming and error-prone. Further, social media text is often noisy and does not adhere to the syntax of any one language. Hence, there is a huge need for the techniques to detect the fake news efficiently and automatically [5]. Majority of the proposed fake news detection tasks have focused on resource-rich languages such as English and Spanish. Due to lack of annotated data, resource-poor languages such as Urdu and many Indian native languages have got very less attention. Off late, few shared tasks and workshops are being organized to promote fake news detection in resource-poor languages like Urdu [6]. In order to boost up the text processing activities in Urdu language, UrduFake 2021 - a shared task at FIRE 2021 aims at distinguishing the real news from the fake news in Urdu [7]. This shared task can be modeled as a binary Text Classification (TC) task as there are only two labels, Fake and Real.

Researchers have developed a slew of tools and algorithms for TC in general. However, an algorithm that performs well for one dataset may not exhibit the similar performance on other datasets. Therefore, claiming that an algorithm or a model performs well on all the datasets is illogical. In this paper, we - team MUCS, describe the model submitted to UrduFake 2021 shared task to detect Urdu fake news. The proposed model is an ensemble of four ML algorithms, namely: RF, MLP, GB and AB, trained with word uni-grams, character n-grams and fastText Urdu word vectors.

Rest of the paper is organized as follows: Section 2 highlights the recent work in detecting fake news and Section 3 details the proposed methodology. While Section 4 spreads light on experimental results, the paper concludes with future work in Section 5.

2. Related Work

Detecting fake news is challenging, especially for languages with limited resources. Due to lack of or limited availability of bench marked corpus, several researchers have constructed their own dataset and developed various techniques to detect fake news. UrduFake 2020¹ is a shared task at FIRE 2020 to detect fake news in Urdu language. Several models were submitted by the researchers to this shared task and the description of some of the good performing models are given below:

Fazlourrahman et al. [8] proposed an ensemble of ML classifiers to detect Urdu fake news using an ensemble of three ML classifiers, namely: Multinomial Naïve Bayes (MNB), Logistic Regression (LR), and MLP with hard voting. The classifiers were trained using the vectors generated by CountVectorizer of word n-grams in the range (1, 2) and character n-grams in the range (1, 5). Their model obtained a macro F1-score of 0.770 and 5th rank in the shared task. A model to detect fake news proposed by Kumar et al. [9] uses Dense Neural Network (DNN) based

¹<https://www.urdufake2020.cicling.org/>

classifier to predict the Fake or Real tag for the given Urdu article. Term Frequency–Inverse Document Frequency (TF-IDF) vectors of character n-grams in the range (1, 3) are fed as features into the dense layers of DNN which is having four dense layers containing 2,048, 512, 128, and 2 neurons in first, second, third, and fourth layers respectively. To reduce over-fitting, *dropout* with a rate of 0.4 was used between each pair of dense layers. Further, the authors used *Rectified Linear Unit (ReLU)* activation² in each of the dense layers and *softmax* activation³ at the output layer. This DNN based model obtained a macro F1-score of 0.810 and 3rd rank in the shared task.

Balaji et al. [10] proposed MLP based classifier to detect fake news in Urdu language. To construct their model, the authors used fastText pre-trained word embeddings to map words to vectors and character n-grams in the range (1, 4) and achieved a macro F1-score of 0.7881 and placed 6th rank in the shared task. Lina et al. [11] proposed a Urdu fake news detection model based on the combination of word and character embedding. They extracted word based features using pre-trained Urdu Robustly Optimized Bidirectional Encoder Representations from Transformers (ROBERTa)⁴ and character based features by adopting character level Convolutional Neural Network (CNN) [12]. Further, sentence embeddings created separately using word and character embeddings was concatenated and fed to the *softmax* layer which predicts the Fake or Real tag. They have also used label smoothing technique to reduce over-fitting. Their model obtained 1st rank with a macro F1-score of 0.907 and outperformed all other models submitted to the shared task.

Reddy et al. [13] proposed a Bidirectional Gated Recurrent Unit (Bi-GRU) based model to detect fake news in Urdu articles. They tokenized the dataset using Keras tokenizer⁵ and used the skip-gram model to create Urdu word embeddings to represent text as vector values [14] which were fed to max and average pooling layers. These layers were concatenated to reflect the spatial relationship of features generated by Bi-GRU. The final feature representation was created by concatenating these two pooling representations and the feature vectors were fed to a sigmoid layer to classify the input text into Fake and Real classes. This model obtained a macro F1-score of 0.807 and obtained 4th rank in the shared task.

Researchers have explored different features like character n-grams, character level embedding, word based features and pre-trained word vectors and different classifiers like ML and Deep Learning (DL) to detect fake news in Urdu news articles. But none of the methods promise 100% results which gives scope for further studies.

3. Methodology

Several experiments were conducted using various features and various learning models to detect Urdu fake news. Based on the results obtained on the development (Dev.) set provided by the organizers of the shared task to detect Urdu fake news, three best performing models were submitted to the shared task. Out of the three models, an ensemble of RF, MLP, GB and AB classifiers with soft voting obtained good results in the final evaluation of the models by the

²<https://numpy-ml.readthedocs.io/en/latest/numpy-ml.neural-nets.activations.html>

³<https://docs.scipy.org/doc/scipy/reference/generated/scipy.special.softmax.html>

⁴<https://huggingface.co/urduhack/urdu-roberta>

⁵https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/text/Tokenizer

organizers and the same is described here. The structure of the proposed ensemble model is depicted in Figure 1. Pre-processing, feature extraction and model construction steps carried out in the proposed methodology are discussed in the following sub-sections:

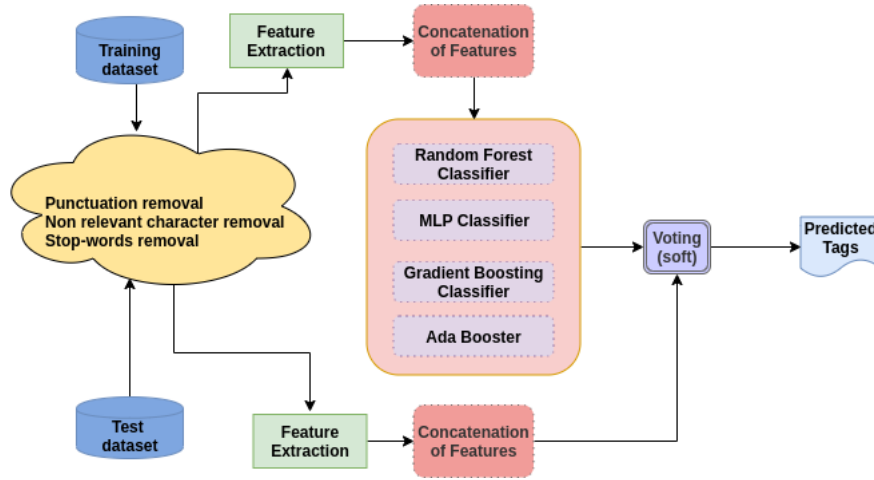


Figure 1: Structure of the proposed ensemble model

3.1. Pre-processing and Feature Extraction

Pre-processing is a crucial step in cleaning up the textual content and transforming it into a format that ML algorithms can understand. As the given dataset is already pre-processed, only punctuation, non-relevant characters and stopwords⁶ are removed as they do not contribute to the TC task.

Feature Extraction is the process of extracting features from the text which are used to train and evaluate the learning models. A combination of TF-IDF of words, char n-grams and pre-trained Urdu word embeddings from fastText are used as features to train and evaluate the classifiers. Feature extraction approaches are given below:

- **TF-IDF vectors** provide a better normalized representation of the text document by removing the impact of overly repeated words. It expresses the relative importance of the word in a document versus the entire corpus. Word uni-grams (17,793) and character n-grams (18,068) in the range (2, 3) are extracted using TfidfVectorizer⁷ and the range of char n-grams is fixed based on performing various experiments.
- **Word vectors** is a type of mapping that allows words with similar meanings to have similar vector representations. The concept behind word vectors is that the surrounding words are used to represent target words. fastText⁸ pre-trained Urdu embeddings available in gensim library⁹ with a window size 5 and 300-dimensional space are used to build word

⁶<https://www.kaggle.com/ratman/urdu-stopwords-list>

⁷https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

⁸<https://fasttext.cc/docs/en/crawl-vectors.html>

⁹<https://pypi.org/project/gensim/>

vectors for the words present in the dataset. Each document in the dataset is represented as a sum of the vector values of the words present in that document.

3.2. Model Construction

The performance of the classifier heavily relies on the dataset and the features extracted from the dataset. Further, the performance of the classifiers is not the same for all the datasets [15]. This has motivated to propose an ensemble of classifiers where the weakness of one classifier will be compensated by the strength of another. Ensemble learning is also a method of generating a new classifier from multiple base classifiers, which outperforms any constituent classifier in the ensemble. It may be noted that any number of classifiers can be ensembled with compatible parameters. An ensemble of four ML classifiers, namely: RF, MLP, GB and AB classifiers with soft voting is used to identify Urdu fake news.

RF classifier is made up of a large number of individual decision trees which work together as an ensemble by itself. Each individual tree in the RF classifier produces a class prediction and the majority vote of all the classes in the forest is used to predict the final tag. This classifier which resolves the over-fitting issues is also well-suited to deal with noisy high-dimensional data [16]. Unlike other classification algorithms such as RF or MNB, MLP classifier consists of three layers namely: input layer, hidden layers and output layer [17] and performs classification using an underlying Neural Network (NN). MLP classifier is a simple feed forward NN classifier compared to other NN classifiers such as recurrent NN or CNN and it is widely used in TC tasks.

In recent years boosting algorithms have grown in popularity in building ML models. Boosting algorithm is a type of ensembled ML algorithm that combines the predictions of many weak learners. The advantage of employing boosting algorithm is that it generates a robust classifier by converting a group of learners with poor classification performance. In GB, regularisation methods penalize different parts of the algorithm and improve overall performance by reducing over-fitting [18]. AB classifier is a meta-estimator in which weights are assigned to each instance and higher weights are assigned to incorrectly classified instances to improve poor learner. It is specifically used in supervised learning to reduce bias and variance. This algorithm works based on the principle of sequential growth of learners to make weak learners strong [19].

4. Experiments and Results

Bend-The-Truth¹⁰ is a benchmark dataset for fake news detection in Urdu language and its evaluation. This dataset consists of news articles from Sports, Social media, Education, Technology, Business, and Entertainment domains. While main stream news websites such as BBC Urdu News, CNN Urdu, Express-News, Jung News, Naway Waqat and other trustworthy news websites were referred to gather real news in Urdu language, the fake news articles were created by a group of professional writers and journalists [20]. This corpus is provided by the organizers of UrduFake 2021 shared task to develop and evaluate the working models to detect fake news in Urdu news articles [21]. Details of the corpus distributed by the topics are given in Table 1

¹⁰<https://github.com/MaazAmjad/Urdu-Fake-news-detection-FIRE2021/blob/main/>

Table 1

Distribution of Urdu news articles into topics in Bend-The-Truth dataset

Category	Business	Health	Showbiz	Sports	Technology	Total
Real	150	150	150	150	150	750
Fake	80	130	130	80	130	550

Table 2

Distribution of Urdu news articles in Train, Development and Test sets

Dataset	Articles labeled Real	Articles labeled Fake	Total
Train Set	600	438	1,038
Dev Set	150	112	262
Test Set	200	100	300

Table 3

Performance measure of the proposed models

Dataset	Fake Class			Real Class			Macro	Accuracy
	Precision	Recall	F1-score	Precision	Recall	F1-score	F1-score	
Dev. Set	0.78	0.66	0.72	0.78	0.86	0.82	0.77	0.84
Test Set	0.850	0.170	0.283	0.703	0.985	0.820	0.552	0.713

and the statistics of Train, Development (Dev) and Test sets distributed by Fake and Real labels are summarized in Table 2.

Several experiments were carried out to identify fake news by combining different features and classifiers. Three combinations of features and ensemble model with soft voting which gave good results on the Development set were used to obtain the predictions on the Test set and submitted to the organizers for final evaluation and ranking. The models submitted by the participants were evaluated on the basis on macro F1-score. Among the submitted predictions of the three models, an ensemble model trained using fastText Urdu word embeddings, word uni-grams and character n-grams performed well and obtained macro F1-score of 0.552, accuracy of 0.713 and 12th rank¹¹ in the shared task. Performances of the proposed model on Development set and Test set are given in Table 3. Using the evaluation snippet released by the organizers, precision, recall, accuracy, and F1-score for Real and Fake class and an average F1-score are calculated for the proposed model. Training set consists of only 43% fake articles and this implies that the dataset is slightly imbalanced and the same is reflecting in the results.

The performances of the individual classifiers of the ensembled model is shown in Figure 2. From the figure, it is clear that the performance of the ensemble model is better than that of the individual models as intended to be. The results also illustrates that MLP classifier which is based on NN exhibits higher macro F1-score than the individual classifiers. Further, the higher

¹¹<https://www.urdufake2021.cicling.org/results-and-rankings>

performance of all the models for Development set is due to the distribution of fake articles in the corpus. The comparison of the accuracy and macro F1-score of the top performing models with the proposed model is shown in Figure 3.

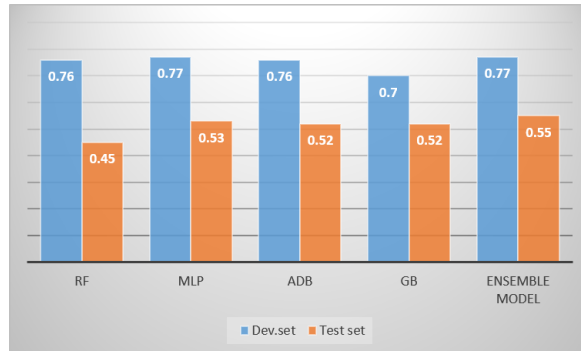


Figure 2: Comparison of macro F1-scores of the individual classifiers of the ensemble model with the proposed ensemble model

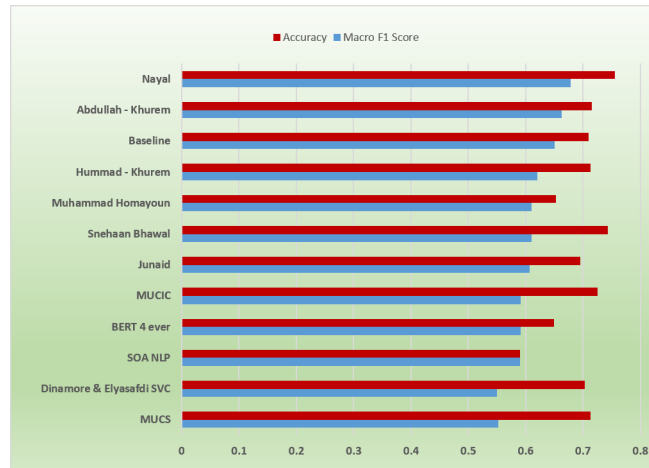


Figure 3: Comparison of macro F1-scores and accuracies of the top performing models with the proposed model

5. Conclusion and Future work

This paper presents the description of the model proposed and submitted by our team MUCS to UrduFake 2021 shared task to identify fake news articles in Urdu language. An ensemble of four ML models which was trained using a combination of fastText word vectors, word uni-grams and character n-grams has achieved a macro F1-score of 0.552 and an accuracy of 0.713. Various combinations of features and feature selection models, as well as different learning approaches in identifying fake news articles will be explored further.

References

- [1] J. P. Singh, A. Kumar, N. P. Rana, Y. K. Dwivedi, Attention-based LSTM Network for Rumor Veracity Estimation of Tweet, Springer, 2020, pp. 1–16.
- [2] B. Ghanem, P. Rosso, F. Rangel, An Emotional Analysis of False Information in Social Media and News Articles, 2020, pp. 1–18.
- [3] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news Detection on Social Media: A Data Mining Perspective, 2017, pp. 22–36.
- [4] H. Allcott, M. Gentzkow, Social Media and Fake News in the 2016 Election, 2017, pp. 211–36.
- [5] J. Tang, Y. Chang, H. Liu, Mining Social Media with Social Theories: A Survey, ACM New York, NY, USA, 2014, pp. 20–29.
- [6] M. Amjad, G. Sidorov, A. Zhila, Data Augmentation using Machine Translation for Fake News Detection in the Urdu Language, in: Proceedings of The 12th Language Resources and Evaluation Conference, 2020, pp. 2537–2542.
- [7] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, A. Gelbukh, UrduFake@ FIRE2021: Shared Track on Fake News Identification in Urdu, in: Forum for Information Retrieval Evaluation, 2021.
- [8] F. Balouchzahi, H. L. Shashirekha, Learning Models for Urdu Fake News Detection, in: FIRE (Working Notes), 2020, pp. 474–479.
- [9] A. Kumar, S. Saumya, J. P. Singh, NITP-AI-NLP@ UrduFake-FIRE2020: Multi-layer Dense Neural Network for Fake News Detection in Urdu News Articles, in: FIRE (Working Notes), 2020, pp. 458–463.
- [10] N. N. A. Balaji, B. Bharathi, SSNCSE_NLP@ Fake News Detection in The Urdu Language (UrduFake) 2020, in: FIRE (Working Notes), 2020.
- [11] N. Lina, S. Fua, S. Jianga, Fake News Detection in the Urdu Language using CharCNN-RoBERTa, in: FIRE (Working Notes), 2020.
- [12] X. Zhang, J. Zhao, Y. LeCun, Character-level Convolutional Networks for Text Classification, 2015, pp. 649–657.
- [13] S. M. Reddy, C. Suman, S. Saha, P. Bhattacharyya, A GRU-based Fake News Prediction System: Working Notes for UrduFake-FIRE 2020, in: FIRE (Working Notes), 2020, pp. 464–468.
- [14] S. Haider, Urdu Word Embeddings, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018.
- [15] M. D. Anusha, H. L. Shashirekha, An Ensemble Model for Hate Speech and Offensive Content Identification in Indo-European Languages, in: FIRE (Working Notes), 2020, pp. 253–259.
- [16] M. Z. Islam, J. Liu, J. Li, L. Liu, W. Kang, A Semantics Aware Random Forest for Text Classification, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1061–1070.
- [17] M. Bounabi, K. El Moutaouakil, K. Satori, A Probabilistic Vector Representation and Neural Network for Text Classification, in: International Conference on Big Data, Cloud and Applications, Springer, 2018, pp. 343–355.
- [18] V. Athanasiou, M. Maragoudakis, A Novel, Gradient Boosting Framework for Sentiment

Analysis in Languages where NLP Resources are not Plentiful: A Case Study for Modern Greek, Multidisciplinary Digital Publishing Institute, 2017, p. 34.

- [19] R. E. Schapire, Explaining Adaboost, in: Empirical Inference, Springer, 2013, pp. 37–52.
- [20] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, A. Gelbukh, “Bend the Truth”: Benchmark Dataset for Fake News Detection in Urdu Language and Its Evaluation, 2020, pp. 2457–2469.
- [21] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, A. Gelbukh, Overview of the Shared Task on Fake News Detection in Urdu at FIRE 2021, in: FIRE (Working Notes), 2021.