Augmentation-Based Ensemble Learning For Stance and Fake News Detection

Ilhem Salah[0000-0002-3375-3637], Khaled Jouini[0000-0001-5049-4238], and Ouajdi Korbaa[0000-0003-4462-1805]

MARS Research Lab LR17ES05, ISITCom, University of Sousse, H. Sousse, 4011, Tunisia
ilhemsalah53@gmail.com, khaled.jouini@isitc.u-sousse.tn, ouajdi.korbaa@centraliens-lille.org

Abstract. Data augmentation is an unsupervised technique used to generate additional training data by slightly modifying already existing data. Besides preventing data scarcity, one of the main interest of data augmentation is that it increases training data diversity, and hence improves models' ability to generalize to unseen data. In this work we investigate the use of text data augmentation for the task of stance and fake news detection.
In the first part of our work, we explore the effect of various text augmentation techniques on the performance of common classification algorithms. Besides identifying the best performing (classification algorithm, augmentation technique) pairs, our study reveals that the motto "*the more, the better*" is the wrong approach regarding text augmentation and that there is no *one-size-fits-all* text augmentation technique.
The second part of our work leverages the results of our study to propose a novel augmentation-based, ensemble learning approach that can be seen as a mixture between stacking and bagging. The proposed approach leverages text augmentation to enhance base learners' diversity and accuracy, ergo the predictive performance of the ensemble. Experiments conducted on two real-world datasets show that our ensemble learning approach achieves very promising predictive performances.

Keywords: Stance and Fake News Detection · Text Augmentation · Ensemble Learning · Fake News Challenge.

1 Introduction

In the era of the Internet and social media, where a myriad of information of various types is instantly available and where any point of view can find an audience, access to information is no longer an issue, and the key challenges are veracity, credibility, and authenticity. The reason for this is that any user can readily gather, consume, and break news, without verification, fact-checking, or third-party filtering. As revealed by several recent studies, fake news and misinformation are prone to spread substantially faster, wider, and deeper than genuine news and real information [21, 8].

By directly influencing public opinions, major political events, and societal debates, fake news has become the scourge of the digital era, and combating it has become a dire need. The identification of fake news is however very challenging, not only from a machine learning and Natural Language Processing (NLP) perspective, but also sometimes for the most experienced journalists [18]. That is why the scientific community approaches the task from a variety of angles and often breaks down the process into independent sub-tasks. A first practical step towards automatic fact-checking and fake news detection is to estimate the opinion or the point of view (*i.e. stance*) of different news sources regarding the same topic or claim [18]. This (sub-) task, addressed in recent research as *stance detection*, was popularized by the Fake News Challenge - Stage 1 (or FNC-1) [18], which compares article bodies to article headlines and determines if a body agrees, disagrees, discusses or is unrelated to the claim of a headline.

In this paper we propose a novel *Augmentation-based Ensemble learning* approach for stance and fake news detection. Data augmentation refers to techniques used to create new training data by slightly modifying available labelled data. Besides preventing data scarcity, one of the main interest of data augmentation is that it increases training data diversity, and hence helps to improve models' ability to generalize to unseen data [11]. Data augmentation is extensively used in Computer Vision (CV) where it is considered as one of the anchors of good predictive performance. Despite promising advances, data augmentation remains however less explored in NLP where it is still considered as the "cherry on the cake" which provides a steady but limited performance boost [23].

Ensemble learning combines the knowledge acquired by base learners to make a consensus decision which is supposed to be superior to the one attained by each base learner alone [27]. Research on ensemble learning proves that the greater are the skills and the diversity of base learners, the better are the accuracy and the generalization ability of the ensemble [27]. In this work we leverage text data augmentation to enhance both, the diversity and the skills of base learners, ergo the accuracy of the ensemble.

The main contributions of our work are therefore: (*i*) an extensive experimental study on the effect of different text data augmentation techniques on the performance of common classification algorithms in the context of stance and fake news detection. Our study provides insights for practitioners and researchers on text data augmentation and the best performing (data augmentation technique, classification algorithm) pairs; and (*ii*) a novel augmentation-based ensemble learning approach, which is a mixture of stacking and bagging.

The remainder of this paper is organized as follows. Section 2 outlines the main steps we followed to vectorize text and reduce dimensionality. Section 3 exposes the key motifs of data augmentation and the text augmentation techniques adopted in our work. Section 4 details the architecture of our novel augmentation-based ensemble learning. Section 5 briefly reviews existing work on stance and fake news detection. Section 6 presents an experimental study on two real-world fake news datasets and discusses the main results and findings. Finally, section 7 concludes the paper.

2 Text as Vectors

2.1 Pre-Processing and Feature Extraction

Machine Learning (ML) algorithms operate on numerical features, expecting input in the form of a matrix where rows represent instances and columns features. Raw news texts have therefore to be transformed into feature vectors before feeding into ML algorithms [9]. In our work, we first eliminated stop words and reduced words to their roots (*i.e.* base words) by stemming them using Snowball Stemmer from the NLTK library [16]. We next vectorized the corpus with a TF-IDF (*Term Frequency – Inverse Document Frequency*) weighting scheme and generated a term-document matrix.

TF–IDF is computed on a per-term basis, such that the relevance of a term to a text is measured by the scaled frequency of the appearance of the term in the text, normalized by the inverse of the scaled frequency of the term in the entire corpus. Despite its simplicity and its wide-spread use, the TF-IDF scheme has two severe limitations: (*i*) TF-IDF does not capture the co-occurrence of terms in the corpus and makes no use of semantic similarities between words. Accordingly, TF-IDF fails to capture some basic linguistic notions such as synonymy and homonymy; and (*ii*) The term-document matrix is high dimensional and is often noisy, redundant, and excessively sparse. The matrix is thus subject to the curse of dimensionality: as the number of features is large, poor generalization is to be expected.

2.2 Dimensionality Reduction

Latent Semantic Analysis (LSA) [3] is an unsupervised statistical topic modeling technique, overcoming some of the limitations of TF-IDF. As other topic modeling techniques, such as LDA (Latent Dirichlet Allocation [2]), LSA is based on the assumptions that: (*i*) each text consists of a mixture of topics; and (*ii*) each topic consists of a set of (weighted) terms that regularly co-occur together. Put differently, the basic assumption behind LSA is that words that are close in meaning, appear in similar contexts and form a “hidden topic”. The basic intuition behind LSA is to represent words that form a topic not as separate dimensions, but by a single dimension. LSA represents thus texts by “semantic” or “topic” vectors, based on the words that these texts contain and the set of weighted words that form each of the topics.

To uncover the latent topics that shapes the meaning of texts, LSA performs a Singular Value Decomposition (SVD) on the document-term matrix (*i.e.* decomposes it into a separate text-topic matrix and a topic-term matrix). Formally, SVD decomposes the term-document matrix $A_{t\times n}$, with t the number terms and d the number of documents, into the product of three different matrices: orthogonal column matrix, orthogonal row matrix and one singular matrix.

$$A_{t\times n} = U_{t\times n} S_{n\times n} D^T_{n\times d} \tag{1}$$

where $n = min(t, d)$ is the rank of A. By restricting the matrices T, S and D to their first $k < n$ rows, we obtain the matrices $T_{t\times k}$, $S_{k\times k}$ and $D_{d\times k}$, and hence obtain k-dimensional text vectors. From a practical perspective the key ask is to determine k, which would be reasonable for the problem (*i.e.* without major loss). In our work we used the transformer TruncatedSVD from sklearn [17]. As in [12] we set the value of k to 100D. The experimental study conducted in [12] showed that using LSA (with k set to 100D) instead of TF-IDF allows a substantial performance improvement for the tasks of stance and fake news detection.

3 Text Data Augmentation

Data augmentation aims at synthesizing new training instances that have the same ground-truth labels as the instances that they originate from [30]. Data augmentation has several well-known benefits: (*i*) preventing overfitting by improving the diversity of training data; (*ii*) preventing data scarcity by providing a relatively easy and inexpensive way to collect and label data; (*iii*) helping resolve class imbalance issues; and (*iv*) increasing the generalization ability of the obtained model.

The success of data augmentation in Computer Vision has been fueled by the ease of designing semantically invariant transformations (*i.e.* label-preserving transformations), such as rotation, flipping, etc... While recent years witnessed significant advancements in the design of transformation techniques, text augmentation remains less explored and adopted in NLP than in CV. This is mainly due to the intrinsic properties of textual data (*e.g.* polysemy), which make defining label-preserving transformations much harder [23]. In the sequel we mainly focus on off-the-shelf text augmentation techniques and less on techniques that are still in the research phase, waiting for large-scale testing and adoption. For a more exhaustive survey on text augmentation techniques, we refer the reader to [11, 1, 28].

3.1 Masked Language Models

The main idea behind Masked Language Models (MLMs), such as BERT [4], is to mask words in sentences and let the model predict the masked words. BERT, which is a pretrained multi-layer bidirectional transformer encoder, has the ability to predict masked words based on the bidirectional context (*i.e.* based on its left and right surrounding words). In contrast with other context-free models such as GLOVE and Word2Vec, BERT alleviates the problem of ambiguity since it considers the whole context of a word.

BERT is considered as a breakthrough in the use of ML for NLP and is widely used in a variety of tasks such as classification, Question/Answering, and Named Entity Recognition [22]. Inspired by the recent work of [22, 11], we use BERT as an augmentation technique. The idea is to generate new sentences by randomly masking words and replacing them by those predicted by BERT.

3.2 Back-translation (*a.k.a.* Round-trip translation)

Back-Translation is the process of translating a text into another language, then translating the new text back into the original language. Back-translation is one of the most popular means of paraphrasing and text data augmentation [15]. Google Cloud Translation API, used in our work to translate sentences to French and back, is considered as the most common tool for back-translation [11].

3.3 Synonym (*a.k.a.* Thesaurus-based augmentation)

The synonym technique, also called lexical substitution with dictionary, was until recently the most widely (and for a long time the only) augmentation technique used for textual data classification. As suggested by its name, the Synonym technique replaces randomly selected words with their respective synonyms. The types of words that are candidates for lexical substitution are: adverbs, adjectives, nouns and verbs.

The synonyms are typically taken from a lexical database (*i.e.* dictionary of synonyms). WordNet [6], used in our work for synonym replacement, is considered as the most popular open-source lexical database for the English language [11].

3.4 TF-IDF based Insertion and substitution

The intuition behind these two noising-based techniques is that uninformative words (*i.e.* having low TF-IDF scores) should have no or little impact on classification. Therefore, the insertion of words having low TF-IDF scores (at random positions) should preserve the label associated with a text, even if the semantics are not preserved. An alternate strategy is to replace randomly selected words with words having the same low TF-IDF scores (TF-IDF based substitution).

Section 6 presents an extensive study on the effect of the aforementioned augmentation techniques on the preredictive performance of ten common classification algorithms, namely, Decision Tree (DT), Support Vector Machine (SVM), Adaptive Boosting (AdaBoost), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Bagged Random Forests (Bagged RF), Light Gradient Boosting Machine (LightGBM), Gradient Boosting (GradBoost), Logistic Regression (LR), and Naive Bayes (NB). Moreover, in contrast with existing work, where text augmentation is considered as an auxiliary technique, our novel augmentation-based ensemble approach presented in next section, goes further and let augmentation shape the entire learning process.

4 Augmentation-Based Ensemble Learning

4.1 Diversity and Skillfulness in Ensemble Learning

Ensemble Learning finds its origins in the "Wisdom of Crowds" theory [26]. The "Wisdom of Crowds" theory states that the collective opinion of a group of individuals can be better than the opinion of a single expert, provided that the aggregated opinions are diverse (*i.e.* diversity of opinion) and that each individual in the group has a minimum level of competence (*e.g.* better than a random guess). Similarly, Ensemble Learning combines the knowledge acquired by a group of base learners to make a consensus decision which is supposed to be superior to the one reached by each of them separately [27]. Research on Ensemble Learning proves that the greater are the skills and the diversity of base models, the better is the generalization ability of the ensemble model [27]. Alternatively stated, to generate a good ensemble model, it is necessary to build base models that are, not only skillful, but also skillful in a different way from one another.

Bagging and stacking are among the main classes of parallel ensemble techniques. Bagging (*i.e.* Bootstrap aggregating) involves training multiple instances of the same classification algorithm, then combining the predictions of the obtained models through hard or soft voting. To promote diversity, base learners are trained on different subsets of the original training set. Each subset is typically obtained by drawing random samples with replacement from the original training set (*i.e.* bootstrap samples).

Stacking (*a.k.a.* stacked generalization) involves training a learning algorithm (*i.e.* meta-classifier) to combine the predictions of several heterogeneous learning algorithms, trained on the same training data. The most common approach to train the meta-model is via k-fold cross-validation. With the k-fold cross-validation, the whole training dataset is randomly split (without replacement) into independent equal-sized k-folds. $k-1$ folds are then used to train each of the base models and the k^{th} fold (holdout fold) is used to collect the predictions of base models on unseen data. The predictions made by base models on the holdout fold, along with the expected class labels, provide the input and the output pairs used to train the meta-model. This procedure is repeated k times. Each time a different fold acts as the holdout fold while the remaining folds are combined and used for training the base models.

4.2 Novel Augmentation Based Approach

As mentioned earlier, in conventional stacking base learners are trained on the same dataset and diversity is achieved by using heterogeneous classification algorithms. As depicted in figure 1, the classical approach for combining augmentation and stacking, is to: (*i*) apply one or several augmentation techniques to the original dataset, (*ii*) fuse the original dataset with data obtained through augmentation; and (*iii*) train base learners on the fused dataset.

In our work we adopt a different approach and train heterogeneous algorithms on different data to further promote diversity. More specifically, through an extensive experimental study (Section 6), we first identify the most accurate (augmentation technique, classification algorithm) pairs. Our meta-model is then trained on the predictions made by the most accurate pairs, using a stratified k-fold cross-validation. Figure 2 depicts the overall architecture of the proposed augmentation-based ensemble learning.

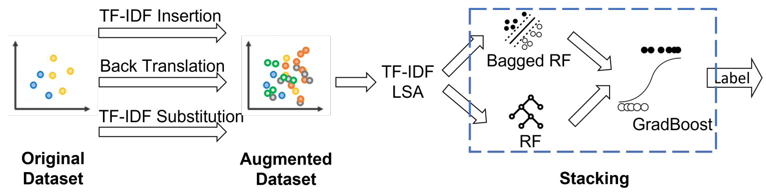

Fig. 1. Conventional approach for combining augmentation and stacking

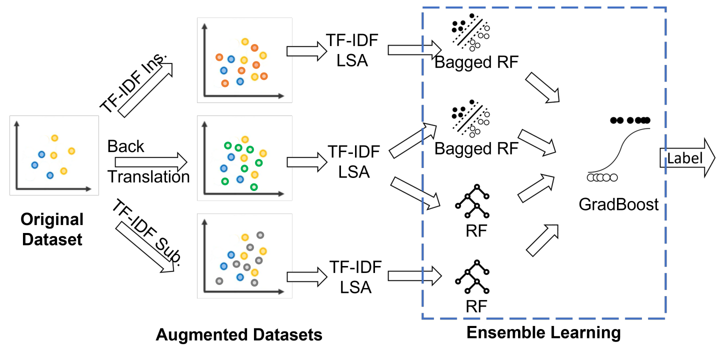

Fig. 2. Novel Augmentation-Based Ensemble Learning Approach

Our augmentation-based ensemble learning approach, can be seen as a mixture between stacking and bagging. In contrast with Bagging and like Stacking, we use an ensemble of heterogeneous learning algorithms. In contrast with stacking and like Bagging, base learners are trained on different datasets, to further promote diversity. However, unlike Bagging the considered datasets are not obtained through bootstrap sampling. Instead, they are obtained by combining the original training data with the data obtained by applying one of the text augmentation techniques. Finally, like in conventional Stacking, the meta-model is trained using a stratified K-fold cross-validation.

5 Related Work

Salient stance and fake news detection approaches adopt a wide range of different features (*e.g.*, context-based, content-based), classifiers, and learning tactics (*e.g.* stacking, bagging, etc.) [5]. Due to the lack of space, we mainly focus hereafter on ensemble approaches and on approaches that rely on content-based features. We suggest readers to refer to surveys and retrospectives on recent challenges [7, 10] for a more comprehensive overview of the current state of research.

The authors of the fake news challenge (FNC-1) [25], released a simple baseline model for the stance detection task. The proposed model achieves an F1-score of 79.53% and uses a gradient boosting (GradBoost) classifier on global co-occurrence, polarity and refutation features. The three best performing systems in the FNC-1 competition were “SOLAT in the SWEN” [20], “Team Athene”

[7] and "UCL Machine Reading" (UCLMR) [19]. "SOLAT in the SWEN" won the competition using an ensemble approach based on a 50/50 weighted average between gradient-boosted decision trees and a Convolutional Neural Network (CNN). The proposed system is based on several features: word2vec pretrained embeddings, TF-IDF, Single Value Decomposition and Word Count. The convolutional network uses pre-trained Word2Vec embeddings passed through several convolutional layers followed by three fully-connected layers and a final softmax layer for classification.

[7], the second place winner, used an ensemble composed of 5 Multi-Layer Perceptrons (MLPs), where labels are predicted through hard voting. The system of UCLMR [19], placed third, used an MLP classifier with one hidden layer of 100 units and a softmax layer for classification. In the same vein as [7], [14] uses a hard voting classifier. The ensemble is composed of three base learners, namely, MLP, Logistic Regression (LR) and X-Gradient Boosting (XGBoost). [14] experimented their approach on the dataset LIAR proposed by [29].

Recently, other published work used FNC-1 in their experiments. [5] constructed a stance detection language model by performing transfer learning on a RoBERTa deep bidirectional transformer language model. [5] leverages bidirectional cross-attention between claim-article pairs via pair encoding with self-attention. The work of [12], which is the closest to the spirit of our work, uses LSA for dimensionality reduction and a stacking-based ensemble having five base learners: GradBoost, Random Forest (RF), XGBoost, Bagging and Light Gradient Boosting Machine (Lightgbm). Besides, [12] compared LDA and LSA and found that LSA yields better accuracy. The authors in [12] experimented their approach on FNC-1 and FNN datasets.

It is worth noticing that in all the aforementioned studies, ensemble approaches yielded better results than those attained by their contributing base learners. On the other hand, despite the substantial potential improvement that text augmentation can carry out, to the best of our knowledge there exists no previous work on stance and fake news detection that compares text augmentation techniques and uses text augmentation in conjunction with ensemble learning.

6 Experimental Study

6.1 Tools & Datasets

Our system was implemented using NLTK [16] for text preprocessing, nlpaug[13] for text augmentation, SciKit-Learn (version 0.24.2) [17] for classification and Beautiful Soup for web scraping. A stratified 10-fold cross-validation was used for model fusion. The Li & al. approach was implemented as described in [12]. The experimental study was conducted without any special tuning. A large number of experiments have been performed to show the accuracy and the effectiveness of our augmentation-based ensemble learning. Due to the lack of space, only few results are presented herein.

As there are no agreed-upon benchmark datasets for stance and fake news detection [12], we used two publicly available and complementary datasets: FNC-

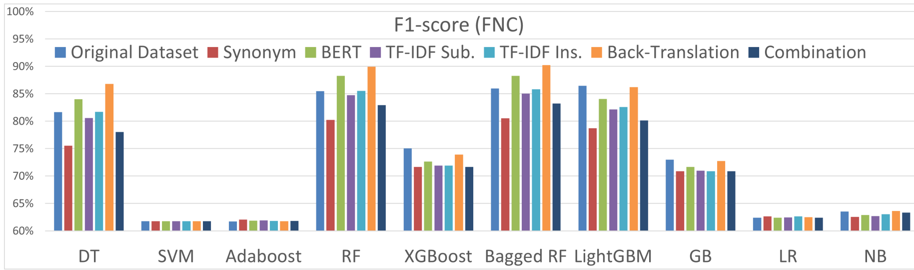

Fig. 3. F1-scores on FNC with and without text augmentation

1 [18] and FNN (*i.e.* FakeNewsNet) [24]. FNC was released to explore the task of stance detection in the context of fake news detection. Stance detection is a multinomial classification problem, where the relative stance of each headline-article pair has to be classified as either: *Agree* if the article agrees with the headline claim, *Disagree* if the article disagrees with the claim, *Discuss* if the article is related to the claim, but takes no position on the subject, and *Unrelated* if the content of the article is unrelated to the claim. The FNC-1 dataset consists of approximately 50k headline-article pairs in the training set and 25k pairs in the test set. FNN data was collected from two fact-checking websites (*i.e.* GossipCop and PolitiFact) containing news contents, along with context information. In comparison with FNN, FNC-1 provides fewer data features (4 vs. 13 features), but more data ($\approx$ 75k vs. $\approx$ 997).

6.2 Results and discussion

We ran our experiments with three objectives in mind: (*i*) identify the best performing (*Augmentation technique*, *Classifier*) pairs; (*ii*) quantify the actual performance improvement allowed by each text augmentation technique; and (*iii*) evaluate the effectiveness of our augmentation-based ensemble approach.

Best performing pairs Figure 3 (resp. 4), reports the F1-scores obtained on FNC (resp. FNN). The results presented in these charts allow to draw important conclusions regarding text augmentation:

1. *Text augmentation does not always improve predictive performance.* This can be especially observed for SVM, LightGBM, GradBoost (figure 3) and AdaBoost (figure 4), where the F1-scores on the original dataset are higher than to those obtained on the augmented datasets;
2. *There is no one-size-fits-all augmentation technique that performs well in all situations.* As depicted in figures 3 and 4, an augmentation technique may perform well when combined with a classification algorithm and poorly when combined with another. This is the case for example for the "Synonym" technique which yields the highest F1-score when combined with Adaboost and the lowest score when used with Naive Bayes (Figure 3).

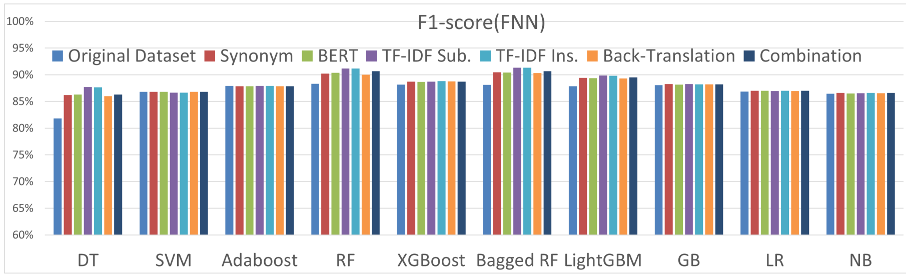

Fig. 4. F1-scores on FNN with and without text augmentation

It is worth noting that even if BERT doesn't achieve the highest F1-scores, it provides a steady performance improvement for almost all classifiers;

3. *The motto "the more, the better" is the wrong approach regarding text augmentation and targeted approaches allow often better results.* This can be observed in figures 3 and 4, where in almost all cases, combining all augmentation techniques does not yield the best F1-scores.

As shown in figure 3, the pairs (Back-translation, Bagged RF) and (Back-translation, RF) yield the highest F1-scores on FNC and increase substantially the predictive performances ($\approx +\ 4.16\%$ in comparison with the highest F1-Score that can be achieved without text augmentation). Similarly, as shown in figure 4, the pairs (Substitution TF-IDF, RF) and (Insertion TF-IDF, Bagged RF) yield the best F1-scores on the dataset FNN ($\approx +\ 5.87\%$).

Augmentation-Based Ensemble Learning As previously stated, base learners' diversity and competency are the two key success factors of any ensemble learning approach. Our ensemble approach leverages text augmentation to enhance both. Figure 2 depicts our classification model which is a mixture of stacking and bagging. In our model, we use Bagged RF and Random Forest (RF) as base classifiers and GradBoost as meta-classifier. As depicted in figure 2, each of the base classifiers is trained on a dataset composed of the original dataset and the data obtained by applying one of the augmentation techniques. The choice of the (classifier, augmentation technique) pairs was driven by the experimental study conducted in subsection 6.2. We compare our model to a more classical stacking approach, where all base classifiers are trained on the same dataset, consisting of the original dataset and the data obtained by applying one of the augmentation techniques (figure 1). We also compare our model to the approach of [12], which is one of the state-of-the-art approaches that uses LSA, stacking-based ensemble learning and K-fold cross-validation. Table 1 synthesizes the predictive performances achieved by each approach.

As reported in Table 1, the use of text augmentation allows better performances than those achieved by [12] in almost all situations. On the other hand, except for the Synonym technique over the FNC dataset, our model outperforms the classical approach in all situations. Overall, our stacking approach achieves

Table 1. F1-scores achieved by conventional stacking, [12] and the proposed approach

Model	FNC	FNN
(Insertion TF-IDF, Stacking)	85,58%	**90,92%**
(Substitution TF-IDF, Stacking)	84,57%	90,43%
(Back-Translation, Stacking)	**90,31%**	89,80%
(BERT, Stacking)	87,93%	90,26%
(Synonym, Stacking)	80,71%	90,28%
(Combination, Stacking)	83,11%	90,73%
Li & al. [12]	83,72%	88,45%
Proposed approach	**90,15%**	**91,07%**

an increase in F1-score of 7,72% (resp. 7,54%) over FNC (resp. FNN) when compared to [12].

7 Conclusion

Combating fake news on social media is a pressing need and a daunting task. Most of the existing approaches on fake news detection, focus on using various features to identify those allowing the best predictive performance. Such approaches tend to undermine the generalization ability of the obtained models.

In this work, we investigated the use of text augmentation in the context of stance and fake news detection. In the first part of our work, we studied the effect of text augmentation on the performance of various classification algorithms. Our experimental study quantified the actual contribution of data augmentation and identified the best performing (classifier, augmentation technique) pairs. Besides, our study revealed that the motto "the more, the better" is the wrong approach regarding text augmentation and that there is no one-size-fits-all augmentation technique. In the second part of our work, we proposed a novel augmentation-based ensemble learning approach. The proposed approach is a mixture of bagging and stacking and leverages text augmentation to enhance the diversity and the performance of base classifiers. We evaluated our approach using two real-world datasets. Experimental results show that it is more accurate than state-of-art methods.

As a part of our future work, we intend to explore the use of a multimodal data augmentation that involves linguistic and extra linguistic features. We also intend to explore the detection of fake news from streams under concept drifts.

References

1. Andrea Stevens Karnyoto, Chengjie Sun, B.L., Wang, X.: Augmentation and heterogeneous graph neural network for aaai2021-covid-19 fake news detection. International journal of machine learning and cybernetics p. 13 (2022). https://doi.org/10.1007/s13042-021-01503-5
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research **3**, 993–1022 (Mar 2003)

3. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science **41**(6), 391–407 (1990)
4. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR **abs/1810.04805** (2018)
5. Dulhanty, C., Deglint, J.L., Daya, I.B., Wong, A.: Taking a stance on fake news: Towards automatic disinformation assessment via deep bidirectional transformer language models for stance detection. CoRR **abs/1911.11951** (2019)
6. Fellbaum, C.: Wordnet and wordnets. In: Barber, A. (ed.) Encyclopedia of Language and Linguistics, pp. 2–665. Elsevier (2005)
7. Hanselowski, A., P.V.S., A., Schiller, B., Caspelherr, F., Chaudhuri, D., Meyer, C.M., Gurevych, I.: A retrospective analysis of the fake news challenge stance-detection task (2018)
8. Hsu, C.C., Ajorlou, A., Jadbabaie, Ali, P.: News sharing, and cascades on social networks. https://ssrn.com/abstract=3934010 or http://dx.doi.org/10.2139/ssrn.3934010 (December 2021), [Accessed: 2022-01-05]
9. Jouini, K., Maaloul, M.H., Korbaa, O.: Real-time, cnn-based assistive device for visually impaired people. In: 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). pp. 1–6 (2021)
10. Khan, J.Y., Khondaker, M.T.I., Afroz, S., Uddin, G., Iqbal, A.: A benchmark study of machine learning models for online fake news detection. Machine Learning with Applications **4**, 100032 (2021). https://doi.org/https://doi.org/10.1016/j.mlwa.2021.100032, https://www.sciencedirect.com/science/article/pii/S266682702100013X
11. Li, B., Hou, Y., Che, W.: Data augmentation approaches in natural language processing: A survey. CoRR **abs/2110.01852** (2021), https://arxiv.org/abs/2110.01852
12. Li, S., Ma, K., Niu, X., Wang, Y., Ji, K., Yu, Z., Chen, Z.: Stacking-based ensemble learning on low dimensional features for fake news detection. In: 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (2019). https://doi.org/10.1109/HPCC/SmartCity/DSS.2019.00383
13. Ma, E.: NLP Augmentation. https://github.com/makcedward/nlpaug (2019), [Accessed: 2021-05-15]
14. Mahabub, A.: A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers. SN Applied Sciences **2** (04 2020). https://doi.org/10.1007/s42452-020-2326-y
15. Marivate, V., Sefara, T.: Improving short text classification through global augmentation methods. CoRR **abs/1907.03752** (2019), http://arxiv.org/abs/1907.03752
16. NLTK.org: Natural Language Toolkit. https://github.com/nltk/nltk, [Accessed: 2021-05-15]
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
18. Pomerleau, D., Rao, D.: The fake news challenge: Exploring how artificial intelligence technologies could be leveraged to combat fake news. http://www.fakenewschallenge.org/ (2017), [Accessed: 2021-12-15]

19. Riedel, B., Augenstein, I., Spithourakis, G.P., Riedel, S.: A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. CoRR **abs/1707.03264** (2017), http://arxiv.org/abs/1707.03264
20. Sepúlveda Torres, R., Vicente, M., Saquete, E., Lloret, E., Sanz, M.: Headlinestancechecker: Exploiting summarization to detect headline disinformation. Journal of Web Semantics **71**, 100660 (09 2021). https://doi.org/10.1016/j.websem.2021.100660
21. Serrano, E., Iglesias, C.A., Garijo, M.: A survey of twitter rumor spreading simulations. In: Int. Conf. Computational Collective Intelligence (ICCCI'15). pp. 113–122. Springer International Publishing, Cham (2015)
22. Shi, L., Liu, D., Liu, G., Meng, K.: Aug-bert: An efficient data augmentation algorithm for text classification. In: Liang, Q., Wang, W., Liu, X., Na, Z., Jia, M., Zhang, B. (eds.) Communications, Signal Processing, and Systems. pp. 2191–2198. Springer Singapore, Singapore (2020)
23. Shorten, C., Khoshgoftaar, T., Furht, B.: Text data augmentation for deep learning. Journal of Big Data **8** (07 2021). https://doi.org/10.1186/s40537-021-00492-0
24. Shu, K.: FakeNewsNet (2019). https://doi.org/10.7910/DVN/UEMMHS, [Accessed: 2021-12-15]
25. Slovikovskaya, V.: Transfer learning from transformers to fake news challenge stance detection (fnc-1) task. In: Proceedings of the 12th Language Resources and Evaluation Conference. pp. 1211–1218. European Language Resources Association (2019), https://www.aclweb.org/anthology/2020.lrec-1.152
26. Surowiecki, J.: The Wisdom of Crowds. Anchor Books, 1st edn. (2005)
27. Suting, Y., Ning, Z.: Construction of structural diversity of ensemble learning based on classification coding. In: 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). vol. 9, pp. 1205–1208 (2020). https://doi.org/10.1109/ITAIC49862.2020.9338807
28. Tesfagergish Senait Gebremichael, Robertas Damaševičius, J.K.D.: Deep fake recognition in tweets using text augmentation, word embeddings and deep learning. In: Computational Science and Its Applications – ICCSA 2021: 21st International Conference. pp. 113–122. Springer Nature, Cham (2021)
29. Wang, W.Y.: "liar, liar pants on fire": A new benchmark dataset for fake news detection. CoRR **abs/1705.00648** (2017), http://arxiv.org/abs/1705.00648
30. Xie, Q., Dai, Z., Hovy, E.H., Luong, M., Le, Q.V.: Unsupervised data augmentation. CoRR **abs/1904.12848** (2019), http://arxiv.org/abs/1904.12848