Studies Towards Language Independent Fake News Detection

Soumayan Bandhu Majumder[1] and **Dipankar Das**[1]

[1]Computer Sc. & Engineering. Department, Jadavpur University
{soumayanmajumder, dipankar.dipnil2005}@gmail.com

Abstract

Fake news, one of the important topics of recent trends causes serious problems to common people and even organizations in general by spreading its threads in terms of news and social messages. The scenario becomes vulnerable while we deal with health issues like COVID19. Thus, in the present task, we have collected the tweet data on COVID19 of seven different languages. We employed two types of model, one works in a language dependent way whereas the other one aims to investigate various language independent issues. We received better results in language independent model for the languages like English, Hindi and Bengali. Results of European languages like German, Italian, French and Spanish are comparable in both language dependent and independent models.

1 Introduction

In our day-to-day world, we humans mostly generate unstructured data and mostly the textual data in a large scale. While utilizing information grounded in such textual data, our efforts become useless when we stuck in handling fake news or misinformation. We observe three main different types of news – legitimate news, fake news and satirical news and all these news differ with respect to two parameters; authenticity and intent. If authenticity of news is not verifiable or false and its intent is to mislead the readers, then it is known as fake news. On the other hand, if authenticity of the news is verifiable or true and its intent is to convey or spread the authenticity to the readers then it is known as legitimate news. Finally, if authenticity of news is not verifiable or false and its intent is oriented towards entertainment, then such type of news is known as satirical news.

In the present approach, we have considered only fake and non-fake news. Here we use linguistic features for detecting fake news. We tried to detect fake news in language dependent and independent both ways. We have also checked which features are more important than others in detecting fake news for a particular language.

Our remaining paper is divided into different sections. In Section 2 we will see some studies or previous works related to fake news detection. In Section 3 we will see our dataset, upon which we performed this research work. After that in next Section we will discuss methods or model architecture. In Section 5 we discussed about result and error analysis and lastly in Section 6 we draw a conclusion and discussed about future works.

2 Related Work

Misinformation is currently one of the balmy topics of last six to seven years. In this Fake news field particularly many researches are already been executed and many are currently also going. Researchers suggested many different ways of detecting fake news. If we generalize them, then we can come to a conclusion that we can detect a news fake or not based news content or social context. These are the only two generalize way of detecting fake news. We can call these two ways – a) news content model, and b) social context model. Again there are two divisions of news content model based upon which we can detect a fake news – a) knowledge based detection, and b) style based detection. For knowledge based method everything is depend on a knowledge base that we extracted from the news. After creating that knowledge base we have to

Proceedings of the 18th International Conference on Natural Language Processing, pages 439–446
Silchar, India. December 16 - 19, 2021.

compare it with some reliable source to check its authenticity. In style based method we mainly focus on linguistic features and based on that we predict the news. Like news content model social context model also divided in two categories, based on which we can detect a fake news – a) propagation based techniques, and b) credibility based techniques. In case of propagation based method, we have to find propagation path of the news on the social media and have to track the original source of the news. But for credibility based method, we have to find the various relationship between news article and users, publishers, posts, shares, comments etc.

Several researchers have explored the area into different ways. George et. al. uses different types of machine learning algorithms like SVM, Naïve Bayes, KNN etc. upon contextual features and linguistic features to detect fake news. In contrast, Perez-Rosas et. al. analyse seven different types of news domains and also analyse their linguistic differences in both fake and neutral news and also compare characteristics of different domains. Whereas Bedi et. al. uses knowledge based fake news detection mechanism. He creates a knowledge database first and then compare it to authorized news database to verify fake news and neutral news. Dey et. al. follows style based detection. So he first extracted features and then analyse the linguistic patterns and then apply KNN algorithm to classify news. Uppal et. al. propose discourse level analysis for deception detection of news documents.

However, these above mentioned techniques have one problem that is they can detect the fake news of a particular language and particular domain. But, one of the important issues is that whether we can detect it in a language independent fashion or not. Therefore, in the present attempt, one of our aims is to detect the fake news in language independent way. Moreover, none of the above mentioned approaches deal with less computerized and less resourced language like Bengali. Here, in the present task, we have developed dataset as well as explored the detection techniques for Bengali along with English, Hindi and other European languages.

3 Dataset Preparation

Undoubtedly, the term fake news comes into our mind while we think of social media. Thus, we aimed to collect data from a social media like twitter. We also collected newspaper data from different languages like English, Hindi and Bengali.

Already there is a popular multilingual fake news dataset present in covid19 domain, but this dataset does not contain German and Bengali languages and secondly size of our dataset is much larger than this dataset. In case of newspaper data, we crawled sentences (mostly news) from various web sources and manually labeled them. We crawled Bengali sentences from 'ABP Ananda[1]' and Hindi sentences from 'Abp News[2]' and 'Aajtak[3]', as well as 'Twitter[4]'.

In case of collecting data for European languages, we considered the data available in the CLEF shared task by participation.

In addition, we have collected COVID19 twitter data from 15th March, 2020 to 15th May 2020 of 7 different languages (German, Italian, French, Spanish, English, Hindi and Bengali). We collect our COVID19 related data from twitter using tweet-scrapper library. It has been observed in preliminary study that fake news datasets are always skewed because the frequency of real news data are much more than fake news data. Thus, we tried to maintain a similar ratio of fake and real news in each of the languages. However, we were able to collect very less number of fake tweets in Bengali and Hindi in COVID19 domain. Thus, in order to train the models for Bengali and Hindi, we had to add more data from other domain as well.

When we started collecting our dataset from twitter, we have the following tweet filtering feature options like '*tweet_id'*, *'hashtags'*, *'has_media', 'is_replied', 'is_reply_to', 'img_urls', 'links', 'likes', 'parent_tweet_id', 'replies', 'reply_to_users', 'retweets', 'screen_name', 'text', 'text_html', 'timestamp', 'timestamp_epochs', 'tweet_url', 'user_id', 'username'* and *'video_url'*.

But among these feature columns, we used only the '*text'* column and extracted textual features which we employed in our language dependent model. Some of such textual features are *'label', 'word_count', 'char_count', 'word_density', 'punctuation_count', 'title_word_count', 'upper_case_word_count',*

[1] https://bengali.abplive.com/
[2] https://www.abplive.com/
[3] https://aajtak.intoday.in/
[4] https://twitter.com/?lang=bn/hn

'noun_count', 'adj_count', 'verb_count', 'pron_count', 'adv_count', 'other_POS', 'sentiment', 'tags', 'tags_ORG', 'tags_PER', 'tags_LOC', 'tags_MISC'. For annotating our dataset, we considered the help of factcheck.com. The detail statistics of the dataset are shown in Table 1.

Language	Real # Sentence	Fake # Sentence
German	14155	302
Italian	13270	507
French	13318	300
Spanish	12113	496
English	11490	2097
Bengali	1051	449
Hindi	615	287

Table 1: Number of tweet sentences available for each language

4 Language Dependent Classification

In order to investigate the roles of language to detect fake news, here each of the models and its input features are exclusive to that particular language. We employ different types of machine learning algorithms and also ensemble them in order to achieve better results by exploring the benefits of individual machine learning classifiers.

Here we first extracted some features from the tweets. We here don't use twitter specific features because we want our experiment to be on general purpose, instead of twitter specific fake news. Therefore, we use different types of open source libraries to extract different types of features from the text. Here, we have mainly conducted experiment upon 7 different languages. Among these 7 languages we collected our COVID19 related data purely in 5 languages and for other two languages (Bengali and Hindi), we added data from other domain as the COVID19 data of these two languages were very less in number.

From each text or tweet, we extracted a couple of features for different languages using different libraries. For English and European languages, we used spacy[5] and polyglot library[6]. Like for French we have to download '*fr_core_news_sm*' (which is exclusively for French). Finally, we had to install spacytextblob library[7]. For Hindi language, we use nltk library[8] and for Bengali language, we use Bengali-NLP library[9].

We extracted 17 different features (e.g., *'word count', 'char count', 'word density', 'punctuation count', 'title word count', 'upper case word count', 'noun count', 'verb count', 'adjective count', 'pronoun count', 'adverb count', 'other POS', 'sentiment', 'tags_LOC', 'tags_MISC', 'tags_PER', 'tags_ORG'*).

Classification algorithms are of three types, such as - binary classification, multiclass classification and multi-label classification. Here, we have used binary classification algorithms, because our output is between any one of the fake and real class. We have used Logistic Regression, KNN, SVM, Random Forest, XGBOOST, Ensemble Learning and Naïve Bayes to accomplish our goals. Here we take logistic regression based model as our baseline model.

4.1 Feature Ablation Study

In order to identify the importance of different features for different languages we use random forest algorithm (for entropy) and we also use correlation matrix for checking collinearity between features. Followings are some of the hints into that direction.

English Language: For English language, we checked up to the top 15 important features. Here *sentiment* feature is the most important (more than 0.3) and *char count*, *word density* features are the next to it (close to 0.05) and others are of very less important (less than 0.05).

German Language: Here, the most important feature is *sentiment* (more than 0.6) and next to it is *title word count* feature (close to 0.1), which is very less important than sentiment and other features (less than 0.05) are of very negligible importance.

Italian Language: For Italian language, the most important feature is *sentiment* (more than 0.6) and next to it is *miscellaneous tag* feature (less than

[5] https://spacy.io/
[6] https://polyglot.readthedocs.io/en/latest/
[7] https://pypi.org/project/spacytextblob/
[8] https://www.nltk.org/
[9] https://github.com/sagorbrur/Bengali-NLP-Library

0.05) which is of very less important than the first one and other features are of very less importance.

Spanish Language: Here, we can say that *sentiment* is the most important feature (more than 0.8) and other features importance (less than 0.05) is close to 0, or we can say they are given no importance.

French Language: For French language, *sentiment* is the most important feature (more than 0.45) and then the *adjective count* feature (close to 0.05). This is also of very less important and other features (less than 0.05) are of negligible importance.

Hindi Language: For Hindi language, we can see none of the features are that important because all features have values less than 0.1. But among these, '*char count*' has the highest importance which is slightly greater than 0.07.

Bengali Language: In Bengali language, *noun count* feature has the highest importance (value greater than 0.2) and other POS feature has also some importance (value 0.1). *Char count*, *punctuation and pronoun count* have some importance but those are very less.

4.2 Results

It was noticed that our data is highly imbalanced so accuracy should not be a good metric or score to measure the performance of the models or even to compare the models. Thus, we tried different types of scores like precision, recall, f1 score for every class and macro f1 score for both the classes as a whole. We also calculated AUC (*Area Under Curve*) for each model in each language to do a better comparison among the various models.

Logistic Regression: In logistic regression model for real labelled data, the highest precision achieved is 0.99 for German and Spanish languages and the highest recall is 0.99 for German, Italian, Spanish, French and the highest F1 score is 0.99 for German, Italian and Spanish. In case of fake labelled data, our logistic regression achieved the highest precision of 0.83 for German language and the highest recall of 0.75 for Spanish and the highest F1 score of 0.79 for Spanish. Overall, in case of both real and fake data consideration, we can say Spanish language gives the best result according to both Macro F1 and AUC score.

KNN: In KNN model for real labelled data, the highest precision is 0.99 for German, Italian, Spanish, French language and the highest recall is 0.99 for German, Italian, Spanish and the highest F1 score is 0.99 for German, Italian, French and Spanish. In contrast, for fake labelled data, the highest precision is 0.89 for Spanish language and the highest recall is 0.92 for German, French and the highest F1 score is 0.88 for Italian. As a whole, by taking both real and fake data into consideration, German language gives the best result according to both Macro F1 and AUC score.

SVM: Similarly, in SVM model for real labelled data, the highest precision, recall and F1 scores are 0.99, 0.99 and 0.99 respectively for German, Italian, Spanish, French languages. In SVM model for fake labelled data, the highest precision is 0.91 for German language and the highest recall is 0.97 for Spanish and the highest F1 score is 0.92 for German. Overall, German language gives the best result according to Macro F1 and Spanish gives the best result according to AUC score.

Random Forest: In random forest model for real labelled data, the highest precision, recall and F1 scores were obtained for German, Italian, French and Spanish whereas for fake, the highest precision is 0.93 for German language and highest recall is 0.97 for Italian and the highest F1 score is 0.94 for German and Italian both. It was found that German, Italian language gives the best result according to Macro F1 score and Italian language gives the best result according to AUC score.

XGBOOST: In XGBOOST model for real labelled data, the highest precision, recall and F1 scores were obtained for German, Italian, French and Spanish. In XGBOOST model for fake labelled data, the highest precision is 0.90 for German language and the highest recall is 0.96 for Italian and the highest F1 score is 0.91 for German, Spanish. Overall, Spanish language gives the best result according to Macro F1 and Italian language gives best result according to AUC score.

Stacked Ensemble: Similarly, in stacked ensemble model for real labelled data, the highest precision recall and F1 scores were obtained of 0.99 for German, Italian and Spanish. In stacked ensemble model for fake labelled data, the highest precision is 0.92 for German language and the highest recall is 0.96 for Spanish and highest F1 score is 0.93 for German. Finally, German language gives the best result according to Macro F1 and Spanish gives the best result according to AUC score.

Naïve Bayes: In Naïve Bayes model for real labelled data, the highest precision is 0.99 for German, Italian, Spanish, French language and the highest recall is 0.99 for Spanish, Italian and the highest F1 score is 0.99 for Italian and Spanish. In Naïve Bayes model for fake labelled data highest precision is 0.82 for Italian, Spanish language and the highest recall is 0.93 for Spanish and highest F1 score is 0.87 for Spanish. Spanish language gives best result according to both Macro F1 and AUC score on both the classes.

4.3 Error Analysis

Logistic Regression: In this section first we have to know two things Type-1 error and Type-2 error. Type-1 error is false positive and Type-2 error is false negative. For rest of the paper I will use T1 error for Type-1 error and T2 error for Type-2 error.

Our main concern should be fake news because we don't want to left out any of the fake news as real news. So we need basically recall value of fake labelled data. Here Spanish have highest recall of 0.752 for fake label. So we can say for logistic regression model Spanish language has least error in detecting fake news.

KNN: For English language T1 error is 71 and T2 error is 424. In German language T1 error is 13 and T2 error is 7. For Italian language T1 error is 19 and T2 error is 14. For Spanish language T1 error is 13 and T2 error is 15. For French language T1 error is 36 and T2 error is 76. For Hindi language T1 error is 15 and T2 error is 38. For Bengali language T1 error is 22 and T2 error is 116.

Our main concern should be T2 error because we don't want to left out any of the fake news as real news. So we need basically recall of fake labelled data to calculate T2 error in relative way. Here German have highest recall of 0.92 for fake label. So we can say it has least relative Type-2 error.

SVM: For English language T1 error is 126 and T2 error is 304. In German language T1 error is 5 and T2 error is 16. For Italian language T1 error is 18 and T2 error is 6. For Spanish language T1 error is 14 and T2 error is 5. For French language T1 error is 29 and T2 error is 15. For Hindi language T1 error is 13 and T2 error is 31. For Bengali language T1 error is 15 and T2 error is 54.

Here Spanish have highest recall of 0.966 for fake label. So we can say it has least relative T2 error.

Random Forest: For English language T1 error is 74 and T2 error is 223. In German language T1 error is 7 and T2 error is 4. For Italian language T1 error is 15 and T2 error is 4. For Spanish language T1 error is 19 and T2 error is 8. For French language T1 error is 14 and T2 error is 8. For Hindi language T1 error is 15 and T2 error is 31. For Bengali language T1 error is 28 and T2 error is 98.

Here Italian language have highest recall of 0.97 for fake label. So we can say it has least relative T2 error.

XGBOOST: For English language T1 error is 116 and T2 error is 199. In German language T1 error is 9 and T2 error is 7. For Italian language T1 error is 23 and T2 error is 6. For Spanish language T1 error is 14 and T2 error is 8. For French language T1 error is 19 and T2 error is 19. For Hindi language T1 error is 14 and T2 error is 32. For Bengali language T1 error is 36 and T2 error is 89. Here Italian language have highest recall of 0.96 for fake label. So we can say it has least relative T2 error.

Stacked ensemble model: For English language T1 error is 109 and T2 error is 249. In German language T1 error is 7 and T2 error is 6. For Italian language T1 error is 17 and T2 error is 12. For Spanish language T1 error is 19 and T2 error is 5. For French language T1 error is 19 and T2 error is 20. For Hindi language T1 error is 11 and T2 error is 37. For Bengali language T1 error is 17 and T2 error is 118.

Here Italian language have highest recall of 0.96 for fake label. So we can say it has least relative T2 error.

Naïve Bayes: For English language T1 error is 973 and T2 error is 161. In German language T1 error is 96 and T2 error is 14. For Italian language T1 error is 29 and T2 error is 16. For Spanish language T1 error is 30 and T2 error is 10. For French language T1 error is 83 and T2 error is 18. For Hindi language T1 error is 38 and Ty2 error is 18. For Bengali language T1 error is 53 and T2 error is 31.

Here Spanish language have highest recall of 0.93 for fake label. So we can say it has least relative T2 error.

5 Language Independent Classification

Here, we have conducted experiments to see if we can detect fake news in language independent way or not. We here mainly focused on multilingual BERT model. This BERT model is already pre-trained on some different corpus. Therefore, we will first fine tune this multilingual BERT model with our different language corpora, then we received a fixed length embedded output through this model for each of the tweets, then we pass this output through the two layers of artificial neuron network of different unit size and at last we pass that output through the sigmoid layer.

5.1 Model Architecture

Here, we discuss about our independent language model architecture. First, in pre-processing step, we remove all urls and html tags. Then, we remove all emojis and emoticons present in the tweet. Now, we send these pre-processed tweets into BERT model as mentioned in the input format section of this thesis. After that, we choose to go with pooled output and then passed it through the 256 RELU units of artificial neural network (ANN) layer. Finally, we introduce dropout of 0.4. After that, we passed that through the 128 RELU units and lastly through the sigmoid unit which will give our ultimate output.

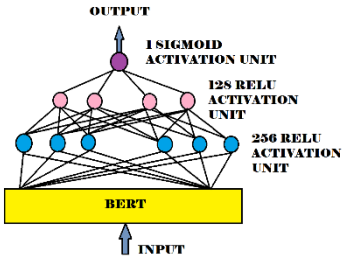

Figure1: Independent model architecture

5.2 Result

In our language independent model for real labelled data, the highest precision is 0.99 for English, Italian, Spanish, French language and the highest recall is 100% for German and the highest F1 score is 0.99 for English, German, Italian, French and Spanish.

In language independent model for fake labelled data highest precision is 100% for German language and the highest recall is 0.98 for French and the highest F1 score is 0.98 for German. Overall, by considering both the real and fake data, we can conclude that English and German language give the best result in Macro F1 score.

Langua		F1	Macro F1
English	Real	0.99	0.98
	Fake	0.97	
German	Real	0.99	0.98
	Fake	0.98	
Italian	Real	0.99	0.93
	Fake	0.88	
Spanish	Real	0.99	0.97
	Fake	0.95	
French	Real	0.99	0.94
	Fake	0.90	
Hindi	Real	0.87	0.78
	Fake	0.71	
Bengali	Real	0.86	0.68
	Fake	0.56	

Table 2: Results of Language Independent Models

5.3 Error Analysis

For English language, T1 error is 11 and T2 error is 17. In German language T1 error is 0 and T2 error is 3. For Italian language T1 error is 22 and

T2 error is 13. For Spanish language T1 error is 5 and T2 error is 7. For French language T1 error is 22 and T2 error is 2. For Hindi language T1 error is 16 and T2 error is 17. For Bengali language T1 error is 30 and T2 error is 69.

Our main concern should be T2 (Type-2) error because we don't want to left out any of the fake news as real news. So we need basically recall of fake labelled data to calculate T2 error in relative way. Here, in case of French language, the highest recall of 0.98 for fake label. So we can say it has least relative T2 error.

6 Result Comparison

After doing both language dependent and independent fake news detection, now we will compare both results.

Languages	Language Dependent		Language Independent	
	F1 score	Recall	F1 score	Recall
English	0.83	0.69	0.98	0.97
French	0.93	0.92	0.94	0.98
German	0.96	0.95	0.98	0.96
Italian	0.96	0.97	0.93	0.91
Spanish	0.95	0.97	0.97	0.95
Hindi	0.66	0.69	0.78	0.71
Bengali	0.67	0.66	0.68	0.48

Table3: result comparison between two models

Here, we compare our results based upon two parameters. Firstly, for all over performance we take F1 score as our parameter. Secondly, we take recall of fake class as our second parameter. We take this second parameter because we want to see how many of fake news are correctly predicted as fake news. Here, we give more emphasize on fake data over real or neutral data. In the above mentioned table, recall is for fake class. For English language, our independent model gives the better result in both parameters. For French, German, Italian and Spanish results of dependent and independent models are comparable in both parameters. For Hindi language, our independent model gives better results in case of both the parameters. It has also been observed that the Bengali language independent model gives better result in F1 score, but dependent model gives better result in recall value.

7 Conclusion

The present work deals with some of the modern day topics like fake news and COVID19. We first collect our data using various sources like twitter, newspaper and shared task. We analyse fake news in multilingual aspect and check how each model and language performs differently in each scenario. Though our data in Hindi and Bengali is very less but still we got some good results in some model. In future if we can get more data in Hindi and Bengali then we can build more concrete models upon these two languages. Here we also learn the basic architecture and concepts of BERT model which is one of the most popular pre-trained models of NLP. We build a language independent model using BERT multilingual which supports many languages. So with our collected data we just fine-tune BERT model with each language data. It produces some astonishing results. Though our data is less but it still gives very good results. In English, Hindi and Bengali language our language independent model outperformed most of the language dependent models. In case of European languages like German, French, Italian and Spanish both language dependent and language independent model performs very good and their results are comparable. One thing we should mention that in case of German language our language independent model predicted all real news correctly and only four fake news wrongly, which quite astonishing.

Acknowledgments

The present work is supported by the research project entitled " Claim Detection and Verification using Deep NLP: an Indian Perspective" funded by DRDO, Government of India..

References

George, J., Skariah, S., & Aleena Xavier, T. (2020). Role of Contextual Features in Fake News Detection: A Review. 2020 International Conference On Innovative Trends In Information Technology (ICITIIT). doi: 10.1109/icitiit49094.2020.9071524

P´erez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 3391–3401). Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Qi, P., Cao, J., Yang, T., Guo, J., & Li, J. (2019). Exploiting Multi-domain Visual Information for Fake News Detection. 2019 IEEE International Conference On Data Mining (ICDM). doi: 10.1109/icdm.2019.00062

Bedi, A., Pandey, N., & Khatri, S. (2019). A Framework to Identify and secure the Issues of Fake News and Rumours in Social Networking. 2019 2Nd International Conference On Power Energy, Environment And Intelligent Control (PEEIC). doi: 10.1109/peeic47157.2019.8976800

Dey, A., Rafi, R., Hasan Parash, S., Arko, S., & Chakrabarty, A. (2018). Fake News Pattern Recognition using Linguistic Analysis. 2018 Joint 7Th International Conference On Informatics, Electronics & Vision (ICIEV) And 2018 2Nd International Conference On Imaging, Vision & Pattern Recognition (Icivpr). doi: 10.1109/iciev.2018.8641018

Rajesh, K., Kumar, A., & Kadu, R. (2019). Fraudulent News Detection using Machine Learning Approaches. 2019 Global Conference For Advancement In Technology (GCAT). doi: 10.1109/gcat47503.2019.8978436

Uppal, A., Sachdeva, V., & Sharma, S. (2020). Fake news detection using discourse segment structure analysis. 2020 10Th International Conference On Cloud Computing, Data Science & Engineering (Confluence). doi: 10.1109/confluence47617.2020.9058106

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate.

Karimi, H., Roy, P., Saba-Sadiya, S., & Tang, J. (2018). Multi-Source Multi-Class Fake News Detection. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 1546–1557). Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Yichuan Li and Bohan Jiang and Kai Shu and Huan Liu (2020). MM-COVID: A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation. CoRR, abs/2011.04088.

S.B. Majumder and D. Das (2020). Detecting Fake News Spreaders on Twitter Using Universal Sentence Encoder. CLEF

Fix, E. and Hodges, J.L. (1951) Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field.

Boser, B., Guyon, I., Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh.

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32.

Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM.