Unmasking Deceptive Information using AI/ML

A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS WHICH IS ESSENTIAL FOR THE GETTING OF THE DEGREE OF

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING - DATA SCIENCE

Submitted By:-

Chirayu Baliyan (2000301540017)

Satwik Srivastava (2000301540047)

Prachi Jain (2000301540036)

Naman Sharma (2000301540034)

Under the guidance of:-

DR. TRIPTI SHARMA

(Department of CSE - Data Science)

DIVISION OF COMPUTER SCIENCE AND ENGINEERING

INDERPRASTHA ENGINEERING COLLEGE, GHAZIABAD

DR. A. P. J. ABDUL KALAM TECHNICAL UNIVERSITY,

LUCKNOW, INDIA

DEC, 2024

Department of Computer Science & Engineering - Data Science

Inderprastha Engineering College

Ghaziabad, Uttar Pradesh - 201010, India

CERTIFICATE OF ORIGINALITY

This is to certify that the project report entitled "Unmasking deceptive information using AI/ML" being submitted by Chirayu Baliyan, Satwik Shrivastava, Prachi Jain and Naman Sharma to the Department of CSE - Data Science for getting the bachelor's degree of engineering/Technology from Inderprastha Engineering College in fragmentary fulfillment of the necessity which is which is essential for the award of the degree of Bachelor of Technology, is original or legit and not copied or obtained from source without proper citation or permission. The manuscript has been thoroughly subjected to plagiarism checked by Turnitin plagiarism checker software. This work had not previously formed the basis for the award of any Degree.

Dr. Tripti Sharma

(HOD, Computer Science and Engineering - Data Science)

(Inderprastha Engineering College)

Department of Computer Science & Engineering - Data Science
Inderprastha Engineering College
Ghaziabad, Uttar Pradesh - 201010, India

CERTIFICATE OF DECLARATION

This is to certify that Project titled "Unmasking Deceptive Information using AI/ML" which is being submitted by Chirayu Baliyan (2000301540017), Prachi Jain (2000301540036), Satwik Shrivastava (2000301540047) and Naman Sharma (2000301540034) to the Department of CSE - Data Science, Inderprastha Engineering College (AKTU University) in fragmentary fulfillment of the necessity which is essential for getting the degree of Bachelor of Technology, is a record of the project work done by the students under my guidance and supervision. The content of this thesis/project works , in full or in parts, has not been submitted for another diploma or degree.

Chirayu Baliyan (2000301540017)
Satwik Shrivastava (2000301540047)
Prachi Jain (2000301540036)
Naman Sharma (2000301540034)

ACKNOWLEDGEMENT

Firstly,we would like to thank and appreciate our mentor and supervisor **Dr. Tripti Sharma (HOD, Computer Science and Engineering - Data Science)** for her constant help, support and guidance throughout the course of our project. We cannot imagine having a better mentor for our project. she has immense knowledge and an infectious dedicated attitude towards work which we hope to carry throughout our career. Without her unconditional support, constant interaction and expert guidance, this project would not have been possible. We would like to say a heartfelt thanks to her for her precious time, availability and invaluable advice regarding our project. We are also indebted to the department of Computer Science and Engineering - Data Science staff and for their support and cooperation. We also extend sincere thanks to our friends and all the people who have supported us in our entire journey.

ChirayuBaliyan(2000301540017)

Satwik Shrivastava (2000301540047)

Naman Sharma(2000301540034)

Prachi Jain (2000301540036)

Abstract

With easy access to the internet in today's world, social media is easily accessible. People share lots of content there like blogs, tweets etc. The major concern is sharing news. Shared news spreads at a very fast pace among the people, it can be fake or real. False news can create a situation of chaos. So, it is necessary to determine the news as false or real as fast as possible. To address this issue, many approaches have been proposed using LSTM, CNN, RNN but either they are unaffordable by common people in terms of computation power or their prediction accuracy is low. In this project/paper, we present a new approach to predict whether a news article is false or not on the basis of the text only. We proposed an Hybrid/ensembled learning based model to detect false news that has achieved a good accuracy. The model extracts the important features from text like name entities, part of speech tagging, sentiment, dependencies to train a deep dense model and uses Fasttext word embeddings to train another model and ensemble the output of these models to make predictions. We achieved an accuracy of 91.44% and 88.64% on training and testing data respectively.

INDEX

LIST OF FIGURES

Chapter 3

Chapter 4

CHAPTER 1:INTRODUCTION

1.1 What is False News?

The idea of false news is not developed in recent times. It was used in the past to divert people's attention to fulfill unethical targets. The recent development in technology makes the internet and web easily accessible to everyone which acts as a catalyst to it. In the past, false news didn't spread easily as there was not so great communication facility but now social media is a great communication channel which spreads the news easily. False news can be of a text, image or video. We often see people use Deepfakes to create fake videos and images and share it on social media. There will be very dangerous effects of false news. It can influence people and create chaos among them. It misleads people down the wrong path. False news can destroy images of people, corporations or a whole nation. People will start losing faith in news, media etc. and even if the news is real, they will not trust it. It raises trust issues among the people, media and government. WhatsApp is a live example of how false news is spread easily and how people get influenced by it. It is also seen that political parties also use false news to win elections in some parts of the world. Therefore, we need some kind of system to filter out the false and real news. In this paper, we bring out an ensemble classification model to classify the news as false or real.

We train and test our model on the LIAR, POLITIFACT and ISOT datasets by ensemble the deep learning models. We preprocess the data, extract the textual features from text and use the FastText word embeddings train and test our model. We have used ensemble learning for better prediction because it uses multiple models for training and it itself discards the weak classifier which is bad for our prediction.

1.2 Impact Of False News

Most of us have experienced the above said situations time and again during our lives as consumers, and we also continue to encounter this situation even today. However,

what's the common factor in all these situations is the false and misleading nature of such pieces of communications/information, which most of the time ended up being perceived as legit news items. Ironically, none of them are based upon credible/believable information and almost all such materials in circulation and move around people's devices are in abundance of misleading information.

It takes a plenty of effort on everything that the brand communicators had been toiling for to be able to ensure that their brands get the nice and best visibility, their brands' key messages are delivered, their brands' loyalty grows and their brands' share in the audience increases. Not only, it spreads wrong information, but it additionally impacts heavily on the cost of advertising and marketing communication campaigns for any brand. Moreover, many times corporations may need to adopt extensive and massive communications programmes to correct the image of their brands dealing with such situations.

Damage achieved with the aid of such misinformation doesn't stay at the level of marketing communications and related expenses and prices but cuts deep into the pinnacle-line and the lowest-line of any corporation. Perception and profits each are affected severely from the misinformation and false news about any company and its products and services. Brand interactions may be influenced by the prevalent misinformation in the audience universe to such an incredible extent that it can extensively regulate the buying behavior of consumers. Forget purchasing the products or services, even intention to experience or purchase, sharing word-of-mouth, or store visits, may be deferred or avoided, by the use of prevailing and potential consumers due to misinformation found in various diverse channels of communications.

1.3 Motivation

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more correct at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use already present data as input to

predict new output values. The substantial spread of faux news can have a significant bad impact on individuals and society. First, faux information can shatter the authenticity equilibrium of the news ecosystem as an instance.

Understanding the truth of news and messages with news detection can create a positive impact on society.Information is crucial for human's decision-making and has an impact on life behaviors. Early information exchanges happened in interactive communications through daily conversations, or came from traditional media (e.g., books, newspapers, radio and television). Such information is more truthful as it is either self-vetted or controlled by authorities. Nowadays, people are exposed to massive amounts of information through a variety of sources (e.g., web pages, blogs, posts), especially with the popularity of the Internet and social media platforms. The ease of Internet access has caused the explosive growth of all sorts of misinformation, e.g., rumor, deception, hoaxes, fake news, spam opinion, which diffuses rapidly and uncontrollably in human society. The erosion of misinformation/fake news to democracy, justice, and public trust becomes a universal and global problem, which has gained an increasing number of research interests in its detection as well as the combat toward its spread among a wide range of people and communities.

Chapter 2: Literature Survey

2.1 Introduction

The growth or increase of social media has been exponential in these times especially because of the pandemic.People had been spending maximum of their time on it and thus it is very easy to spread news through social media instead of the standard news channels.There being not any standard system in place for verifying the news on social media, sometimes some false information also receives spread and has profound effect on the users.

2.2 Existing System

In the recent past there has been many research work done in this field detecting fake news. One of the recent research projects is the work of Alvaro et al. [3], it is mainly based upon using psychological feature extraction from the complete news that allows linguistic analysis of the whole context. They used a bag of words for feature extraction and implemented it on KNN and NB with accuracy and AUC of 0.80 on average. The work in [4] has done detection of fake news for multiple languages namely Slavic, Latin, and German. They have used five different datasets for training and testing which are TwitterBR, FakeBrCorpus, FakeNewsData1, Fake-OrRealNews etc. For each dataset, they used different methods for feature extraction using custom features, Word2Vec, DCDistance, and bag-of-words. Finally, each dataset is fed to a different classification algorithm but the final accuracy achieved was not high enough to be used for detecting fake news. The work in [5] uses a neural network technique for the identification of fake news. Main idea behind this work was to consider the length of text and to apply different features. They have used two datasets namely ISOT and LIAR dataset. Their model was able to achieve 99.8% accuracy for ISOT testing dataset, 39.5% accuracy for liar testing dataset. The work in [1] has classified the fake news using ensemble techniques. They extract the NER features as well as some meta features like number of

sentences, words, character etc. in news text. They use the Random Forest, Extra Tree Classifier and Decision Tree as classification models. They get an accuracy of 100%, 44% on training and testing respectively on LIAR dataset. Arush [2] presented the work with different machine learning models and ensemble them at last. It uses the feature called speaker credibility which refers to the probability of the speaker speaking truth or fake news. [6] work has been very good on the LIAR dataset. They build a deep learning ensemble model to train and test. They did the whole work on LIAR by using all its features. They build two models one with FastText word embeddings and train it with Bi-LSTM and GRU while there is another model with rest of the features. They combine them with a voting classifier to get results. They got an accuracy of 85.9% on testing.

2.3 Need of New System

As we can see from the existing systems that the accuracy has not been enough to rely upon and a lot of possibilities are there to explore and improve. Also most of the work that has been done and has higher accuracy is based on a single dataset which restricts us. Exploring one of the ways is our model based on ensemble learning where multiple base models are combined to produce one optimal predictive model.

Chapter 3:DESIGN AND IMPLEMENTATION

3.1 Proposed System

In the proposed work, we identified textual features as name entity, pos tag, dependency, sentiment and word embeddings. We have done several steps to reach our desired output like from data preprocessing to feature extraction and model selection. The steps involved are as follows:

• LIAR,Politifact and ISOT dataset which are benchmark datasets for fake news detection were identified and explored.

• We have used text cleaning and preprocessing in order to remove the noise from the dataset.

• Extraction of features is the major step. It helps to extract the different features we need.

• Model Training and ensemble the results.

• Check the performance of our model.

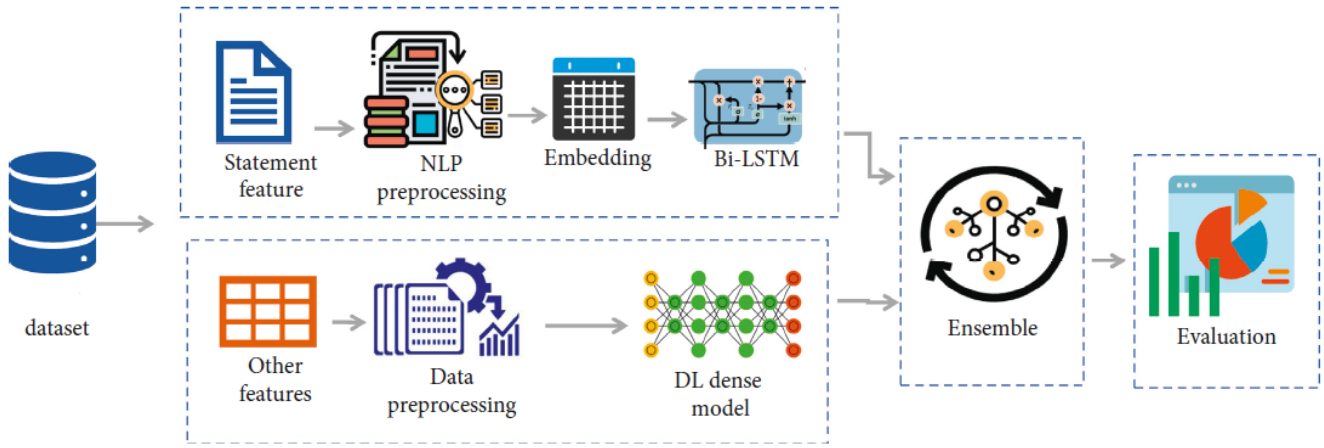


FIGURE 1: Block diagram of the proposed study methodology.

3.2 Methodology Used

3.2.1 Dataset

There are numerous datasets available on Kaggle and GitHub. We have used the LIAR, POLITIFACT and ISOT datasets which are benchmark datasets used for fake news detection.

3.2.1.1 LIAR

The Liar dataset is collected by William Wang and co, consists of six output labels i.e., mostly-true, true, half-true, barely-true, false and pant-fire but we have converted it into binary labels for our model. It is one of the largest and major datasets for fake news detection. It has been collected from various sources like interviews, radio, Speeches or Television. There is short news with various other meta information like party, state, job-title about the author, location, count of all labels counts etc. The dataset is distributed among train, test and valid. We have used the columns of news text, label. There are 11524 and 1267 news in training and testing dataset respectively.

3.2.1.2 POLITIFACT

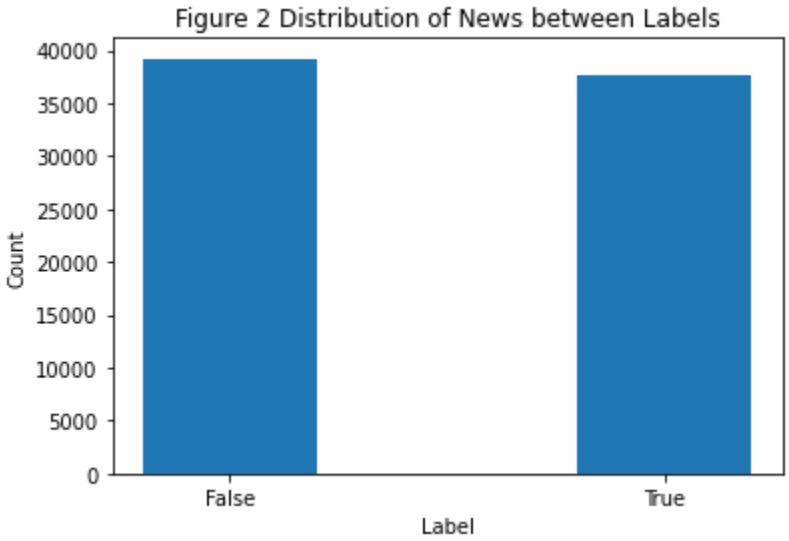
The second dataset POLITIFACT is curated from Politifact.com which has a list of fact checked data. It has around 19358 records. It has a total of 10 columns including news statement, title, information of sources, information of curators, labels. It has nine labels mostly-true, true, half-true, barely-true, false, pant-fire, half-flip, no-flip and full-flop. It is a benchmark dataset for fake news detection.

3.2.1.3 ISOT

Our next dataset is ISOT which is a combination of several existing malicious and non malicious datasets. It has five columns : title, news text, subject, date and labels. Its distribution in fake-news and real-news is 23481 and 1417 respectively.

3.2.2 Data Cleaning

It refers to removing unnecessary data that is not needed. In our work, we only concentrate on the news statement and the label. We can remove the rest of the data. We also relabelled the all labels in our dataset as 0 and 1 where 1 represents true news while 0 represents fake news. We have taken care of Nan values present in the dataset by removing it.



3.2.3 Text Pre-Processing

In this technique we generally remove all the noise from the dataset such as html tags, emojis, URL, punctuation marks, stop words etc. Stemming is also done for root word recognition. We have used the NLTK toolkit for the purpose of pre-processing. Preprocessing is done in the following ways:

● Tokenization: It is the process of breaking sentences into tokens of words. Lowering the word also helps to make all words in the same plane.

● Stop words removal: It refers to the removal of words which appears quite often like a, is etc. We also remove punctuation, URLs, hashtags etc.

● Lemmatization: It refers to reducing a word to its root. Frequency of derived words are reduced with the help of stemming.

3.3 FEATURE EXTRACTION

It refers to selection of features from the dataset. We have extracted the following features:

3.3.1 Name Entity Recognition (NER)

It basically deals with the information extraction and classifying named entities into some pre - defined categories like person, organization, percentage and so forth etc. All these features have given us a brief or quick idea whether a particular document contains the name of a particular person, or it is talking about some organization or event [13].

3.3.2 Parts of Speech tagging (POS-TAGS)

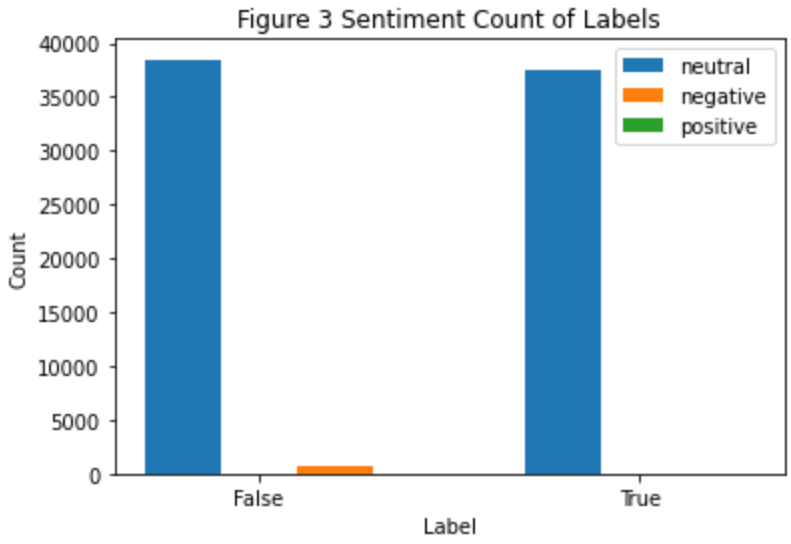
Parts of Speech are the grammatical units of language. It assigns the part of speech tag to every word like noun, verb, adjective and many more.

3.3.3 Dependency Parsing

A dependency tree is a grammatical shape added to a sentence or phrase which delineates the dependency among a word and the phrases it builds upon. Every sentence has a grammatical shape to it and with the help of dependency parsing, we will extract this grammatical structure.

3.3.4 Sentiment Feature Extraction

Sentiment feature extraction essentially deals with the study of extracting the means of subjective text. We have used the Vador sentiment library [12] to find the appropriate intention present within the text data.



3.4 Models

We have built two deep learning models and ensemble them to get better results. The first simple neural network model is trained on the features we extracted and the second Bi-LSTM model is trained on news statements. We predict the output from the two models and ensemble them to get the results. We get the probability to predict true or false from both models. We take a weighted average of both of them to predict the final result.

3.4.1 Simple Neural Network

This model has seven layers and 85 variable features. The structure of the layers is 256, 128, 128 (dropout layer), 64, 64 (dropout layer), 16 and 1 (output layer) neurons respectively. The dropout layers(0.5) were added to make the model robust. Activation functions used were a rectified linear unit “ReLU” for input and all hidden layers, while “sigmoid” for output layer. “Adam” optimizer was used while “binary_crossentropy” was used for loss of the model. “ validation accuracy” was used to evaluate the model accuracy. 25 epochs were used which had batch size = 64 with callbacks setting to monitor the “validation accuracy” and save only the best model.

3.4.2 Bi-LSTM

The main architecture of the second model is bidirectional LSTM. Bi-LSTM stands for bi-directional long short term memory. It is an extension to gated recurrent neural networks. In this neural network information can be predicted or maintained in both the directions that mean both in future and in past.

The model has 9 layers with 200 as input as per the size of the vector for each word. Embedding layer in the deep learning model is added, the size of the real-valued vector space, i.e., EMBEDDING_DIM is 300. We have used FastText[7] for word embedding. Bi-LSTM with 500 units. The outcome is passed into a neural network of 4 layers with 64, 64 (dropout layer), 16 and 1 (output layer) neurons respectively. The model was trained with 25 epochs, batch size to 64 with callbacks set to save the best model on the basis of validation accuracy.

Chapter 4:Results

We presented an ensemble learning approach to detect fake news. Our model has an accuracy of 88.64% and 91.44% on testing and training respectively. If we consider individual accuracy, the simple neural network model has an accuracy of 82.58% and 80.51% on training and testing respectively while Bi-LSTM has 90.49% and 87.98% on training and testing on the dataset contains LIAR, ISOT and POLITIFACT.

Table 1: Accuracy of Models

Models	Accuracy	
	Training	Testing
Simple Neural Network Model	82.58%	80.51%
Bi-LSTM Model	90.49%	87.98%.
Ensemble of Above	91.44%	88.64%

Chapter 5:Conclusion

5.1 Summary

We proposed a machine-learning based unmasking deceptive information detection model using an ensemble learning approach. We used the ensemble approach for training and testing purposeconsisting of a simple neural network and a Bi-LSTM model. We have used the NER, POS Tagging, Dependencies, Sentiment as our features for simple neural network and statement with embedding for Bi-LSTM. Our model achieved better results on the dataset containing LIAR, ISOT and POLITIFACT. The experimentation of the model yielded an accuracy of 88.64% and 91.44% on testing and training respectively.

5.2 Scope For Future Work

- This project may be further enhanced to provide greater flexibility and enhanced performance with certain modifications whenever necessary.
- Deep fake learning which can help to detect fake images, audio and video.
- We can tune hyperparameters to get better results.

CHAPTER 6: REFERENCES

1. Saqib Hakak, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta and Wazir Zada Khan, 2021, "An ensemble machine learning approach through effective feature extraction to classify fake news", Future Generation Computer Systems 117.

2. Arush Agarwal and Akhil Dixit, 2020, "Fake News Detection: An Ensemble Learning Approach", COE Engg. Dept. NSUT.

3. Figueira Alvaro and Oliveira Luciana, 2017, “The current state of fake news:chal-´ lenges and opportunities”, Procedia Computer Science, pg 817-825.

4. P.H.A. Faustini and T.F. Covões, 2020, “Fake news detection in multiple platforms and languages”, Expert Syst. Appl. 113503.

5. M.H. Goldani, S. Momtazi and R. Safabakhsh, 2020, “Detecting fake news with capsule neural networks”, arXiv preprint arXiv:2002.01030.

6. Nida Aslam, Irfan Ullah Khan, Farah Salem Alotaibi, Lama Abdulaziz Aldaej and Asma Khaled Aldubaikil, 2021, “Fake Detect: A Deep Learning Ensemble Model for Fake News Detection”.

7. Piotr Bojanowski, Edouard Grave, Armand Joulin and Tomas Mikolov, 2017, “EnrichingWord Vectors with Subword Information”.

8. Pooja Khurana, Deepak Kumar and Sanjeev Kumar, 2019, “Research of Fake News Spreading Through Whatsapp”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue- 6S4.

9. Hunt Allcott and Matthew Gentzkow, 2017, “Social Media and Fake News in the 2016 Election”, Journal of Economic Perspectives—Volume 31, Number 2, Pages 211–236.

10. C. Xia, C. Zhang, X. Yan, Y. Chang and P. Yu, 2018, “Zero-shot user intent detection via capsule neural networks”, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, pp. 3090–3099.

11. Y. Long, Q. Lu, R. Xiang, M. Li and C.-R. Huang, 2017, “Fake news detection through multi-perspective speaker profiles”, in: Proceedings of the Eighth International Joint Conference on Natural Language Processing, Asian Federation of Natural Language Processing, Taipei, Taiwan pp. 252–256.

12. C.J. Hutto and Eric Gilbert, 2014, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text”, Association for the Advancement of Artificial Intelli-gence.

13. Mark Neumann, Daniel King, Iz Beltagy and Waleed Ammar, 2019, “ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing”, Allen Institute for Artificial Intelligence, Seattle, WA, USA.

14. Petro Liashchynskyi and Pavlo Liashchynskyi, 2019, “Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS”, arXiv:1912.06059v1.

15. Giorgio Valentini and Francesco Masulli, 2002, “Ensembles of Learning Machines”, Conference: 13th Italian Workshop on Neural Nets, WIRN VIETRI 2002, Vietri sul Mare, Italy.

16. Kristína Machová, Miroslav Puszta, František Barčák and Peter Bednár, 2006, “A Comparison of the Bagging and the Boosting Methods Using the Decision Trees Classifiers”, article in Computer Science and Information Systems.

17. G. Biau and E. Scornet, 2016, A random forest guided tour, Test 25 (2) 197–227.

18. T.G. Dietterich, 2002, “Ensemble learning, in: The Handbook of Brain Theory and Neural Networks”, Vol. 2, MIT Press Cambridge, Massachusetts, pp. 110–125.

19. X. Dong, Z. Yu, W. Cao, Y. Shi and Q. Ma, 2020, “A survey on ensemble learning”, Front. Comput. Sci. 1–18.

20. William Yang Wang, 2017, ““Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection”, arXiv:1705.00648v1.